ESTADÍSTICA INFERENCIAL EN LA LÓGICA DE LA INVESTIGACIÓN CIENTÍFICA

EDUAR VÁSQUEZ SÁNCHEZ
GLORIA MARÍA ORTIZ BASAURI

EDITORIAL UNIVERSITARIA
UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

ESTADÍSTICA INFERENCIAL EN LA LÓGICA DE LA INVESTIGACIÓN CIENTÍFICA

EDUAR VÁSQUEZ SÁNCHEZ
GLORIA MARÍA ORTIZ BASAURI

EDITORIAL UNIVERSITARIA
UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

ESTADÍSTICA INFERENCIAL EN LA LÓGICA DE LA INVESTIGACIÓN CIENTIFÍCA

Eduar Vásquez Sánchez

Profesor principal de la Facultad de Ciencias Físicas y Matemáticas.
Universidad Nacional Pedro Ruiz Gallo.
Desarrolla las asignaturas de Metodología de la Investigación Científica y Estadística para la Investigación Científica.

Gloria María Ortiz Basauri

Profesora principal en la Facultad de Ciencias Físicas y Matemáticas.
Universidad Nacional Pedro Ruiz Gallo.
Desarrolla las asignaturas de Lógica Matemática, Análisis Matemático y Geometría Diferencial.

ESTADÍSTICA INFERENCIAL EN LA LÓGICA DE LA INVESTIGACIÓN CIENTÍFICA

Editado por:
Universidad Nacional Pedro Ruiz Gallo
Calle Juan XXIII N° 391. Lambayeque
Teléfono: 074 282081
www.unprg.edu.pe
fondoeditorial@unprg.edu.pe
Lambayeque. Perú

Hecho el depósito legal en la Biblioteca Nacional del Perú
N° 2022-09025
ISBN 978-9972-55-029-4
Primera edición digital, agosto 2022

Autores ©

Dr. Eduar Vásquez Sánchez

Dra. Gloria María Ortiz Basauri

Diseño y cubierta ©

Dr. Eduar Vásquez Sánchez

Diagramación

Lic. Mat. Luis Miguel Villegas Santamaría

Lo consignado en el presente libro es de exclusiva responsabilidad de los autores

Publicación sometida a evaluación de pares académicos.

Lambayeque - agosto 2022

“La estadística es la gramática de la ciencia”

Karl Pearson.

DEDICATORIA

Dedicamos este trabajo a nuestros hijos
Víctor Eduardo y Eduar Aníbal,
quienes nos brindan su amor
y la alegría de aprender.

A la memoria de nuestros padres
Artidoro Vásquez Ramírez y Olinda Sánchez Bustamante
Víctor Ortiz Tafur y Gosvinda Basauri Arce
por su amor infinito y sabias enseñanzas.

A nuestros hermanos
Carlos (+), Sara (+), Saúl, Castinaldo,
Juana, Gloria y Edgar;

Carmen, Nora, Elder, Maritza y Vity
por su entrañable cariño.

PRESENTACIÓN

Estadística inferencial en la lógica de la investigación científica, se elabora con el propósito de proporcionar herramientas de la estadística inferencial en el tratamiento de la información en el método científico, desde el diseño de la muestra hasta la contrastación hipótesis, de manera que el investigador realice inferencias con conocimiento de la significación estadística.

La lógica en el tratamiento estadístico de los datos en la investigación científica, subyace en la naturaleza de las variables, la estimación de sus parámetros de las variables del objeto de investigación y la elección de los modelos para la contrastación de la hipótesis respectiva; de modo que la selección de la muestra y las técnicas del muestreo cumplen un papel decisivo en el estudio de la distribución de probabilidades conociéndose de este modo el comportamiento de las variables poblacionalmente. Se enfatiza en las características de las distribuciones de probabilidad más comunes en la investigación, tales como, la distribución binomial, normal, normal estandarizada, Poisson, la t de Student y Chi cuadrada.

Para la contrastación de hipótesis se consideran los modelos de John Stuart Mill para demostrar la existencia de relaciones causales cuando se tratan factores preponderantemente socio económicos en las investigaciones y los propuestos por Campbell & Stanley para las investigaciones no experimentales, cuasi experimentales y experimentales. Una adecuada elección del diseño se corresponde con la garantía de que las inferencias estadísticas sean válidas y confiables.

Los autores

ÍNDICE

CAPÍTULO I

CAPÍTULO II

CAPÍTULO III

CAPÍTULO IV

CAPÍTULO VII

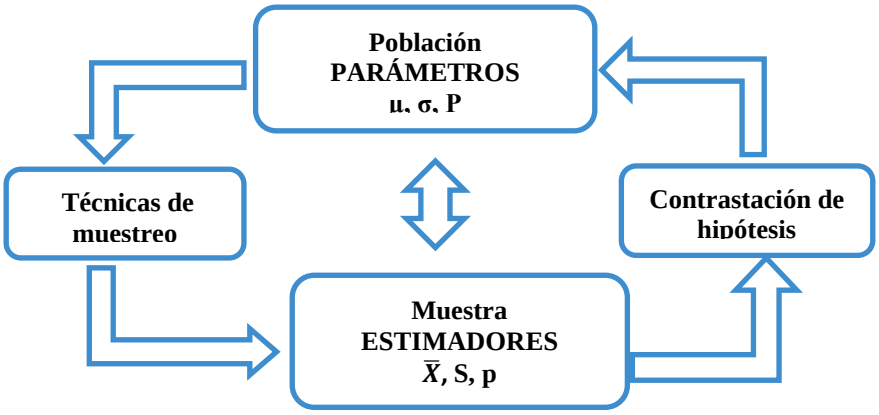
CAPÍTULO I

LA ESTADÍSTICA EN LA INVESTIGACIÓN CIENTÍFICA

"La **Estadística** como tecnología del Método Científico, proporciona instrumentos para la toma de decisiones cuando prevalecen las condiciones de incertidumbre" (Mood y Graybyll, 1972, p. 3). No es posible formular el problema científico si las variables no son susceptibles de medición, tampoco se puede plantear una hipótesis sin considerar un diseño para su contrastación, de igual manera, realizar inferencias sin conocer la distribución de las variables.

Es necesario considerar estos lineamientos en la investigación científica dado que el proceso de contrastación de la hipótesis que se plantea en la ciencia empírica está supeditado al procesamiento de información que proviene de los hechos,

sucesos o fenómenos, datos que se obtienen de muestras representativas y adecuadas aplicando muestreos de acuerdo a como se distribuyen las unidades muestrales, en las cuales las variables a investigar serán medidas utilizando instrumentos válidos y confiables, que permita la prueba lógica de contrastación, y realizar la inferencia a la población de donde se partió cerrando el circuito como se aprecia en el siguiente esquema.



1.1. EL MÉTODO CIENTÍFICO

El método científico es un proceso especial en la epistemología de la ciencia que compromete un conocimiento facto-perceptible y lógico matemático eslabonado en tres fases.

a) La **formulación del problema científico**, que surge a partir de la contemplación viva de la realidad, con la intención de conocerla a profundidad, se inicia por identificar la variable o relación entre las variables. Esta operación requiere una descripción adecuada con información válida y confiable, dentro del cuerpo de la ciencia.

b) La **formulación de la hipótesis**, es la fase donde se concreta la naturaleza de la relación de las variables, es la respuesta lógica al problema, su construcción debe ser semántica y sintácticamente correcta, tener fundamento científico y tecnológico actualizado, de modo que se pueda probar con alto grado de probabilidad.

La hipótesis direcciona la metodología de la investigación, que es responsable de la fuerza de la relación de las variables en estudio.

Las hipótesis pueden ser descriptivas o explicativas.

> Las hipótesis descriptivas son afirmaciones sobre las relaciones de covariación existente entre dos o más variables, nacen de la observación y la experiencia, son inductivas. Si las variables son cualitativas la relación es de asociación, si son cuantitativas la relación es de correlación. (Vásquez et al., 2021, p. 39).

“Las hipótesis explicativas son afirmaciones que se establecen sobre relaciones de causalidad entre dos o más variables, nacen del razonamiento y son deductivas; en este tipo de hipótesis se identifica con claridad la variable independiente y dependiente” (Vásquez et al., 2021, p. 39)

c) La **contrastación de la hipótesis**, para lo cual es necesario conocer: el tipo de investigación, la población para inferir los resultados, la muestra, el tipo del muestreo, examinar la validez y confiabilidad de los instrumentos con los que se medirán las variables, precisar los modelos paramétricos o no paramétricos para el procesamiento de los datos, además dónde y cuándo se recogerán los datos.

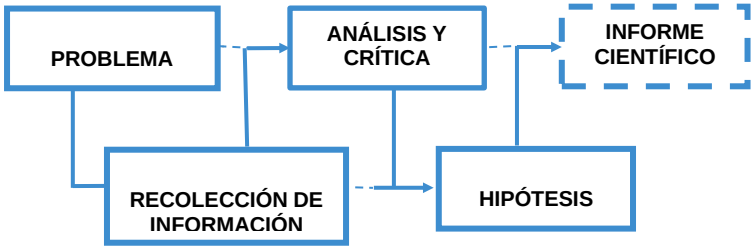
La investigación por su objeto de estudio es:

> **Pura o Básica,** incrementa el conocimiento teórico, es formal y persigue generalizaciones con vista al desarrollo de una teoría basada en principios y leyes.
>
> **Aplicada o Tecnológica,** guarda íntima relación con la investigación básica, pero busca la utilización práctica de los conocimientos. (Vásquez, 1990, p. 5)

Si se considera el tiempo en que ocurren los hechos la investigación puede ser:

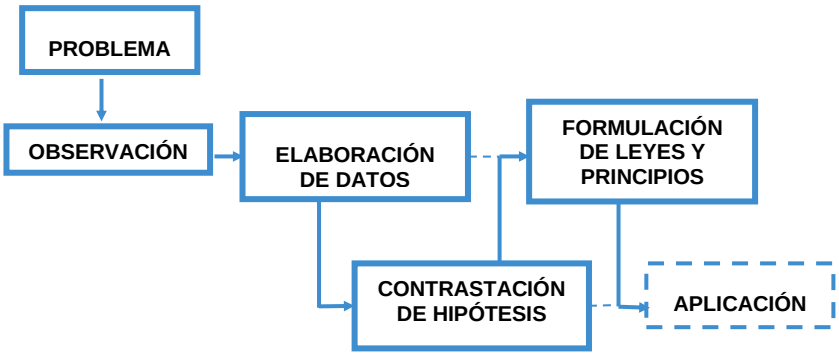
A. **Investigación Histórica,** (pasado, lo que era), comprende los registros, análisis e interpretación de los hechos pasados para comprender el presente y predecir el futuro; se investiga básicamente sobre fenómenos sociales, sobre el cambio, crecimiento o desarrollo de individuos, grupos, procesos, etc.

Esquema de la investigación histórica



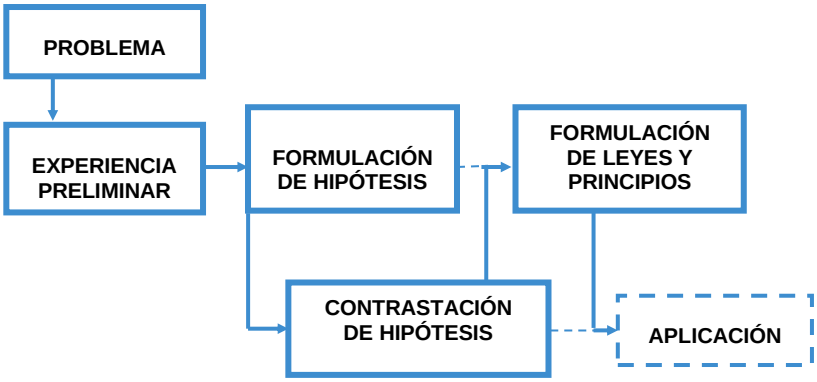
B. **Investigación Descriptiva,** (presente, lo que es), abarca descripciones, registros, análisis e interpretación de una situación problemática, procesos o fenómenos que se realizan sobre aspectos dominantes en el presente, muchas veces lleva consigo algún tipo de comparación y responde al **qué, quién** y **dónde.**

Esquema de la investigación descriptiva



C. **Investigación Experimental,** (futuro, lo que será), el objetivo es establecer las relaciones de causa-efecto entre las variables que se tratan, de manera que puedan explicarse el **cómo** y el **porqué.**

Esquema de la investigación experimental



1.2. PLANTEAMIENTO DE UNA INVESTIGACIÓN CIENTÍFICA

Al iniciar una investigación debe necesariamente responderse las preguntas siguientes:

1. ¿Qué se va a investigar?
2. ¿Qué variables intervienen en la investigación?
3. ¿Cuál es la variable dependiente?
4. ¿Qué factores pueden considerarse influyentes en la variable dependiente?
5. ¿Cómo se medirán estos factores?
6. ¿Se trabajará con todos los elementos de la población?
7. ¿Cuán grande será el tamaño de la muestra?
8. ¿Cómo se recolectará la información?
9. ¿Dónde se va a investigar?
10. ¿Se realizó esta o similar experiencia anteriormente?
11. ¿Qué conclusiones se obtuvieron y qué técnicas se usaron en investigaciones anteriores?
12. ¿Con qué error de muestreo se trabajaron en las experiencias anteriores?
13. ¿Cómo se procesará la información?

14. ¿Con qué técnicas se analizará la información para probar la hipótesis?
15. ¿Con qué medida de error se realizará la investigación?
16. ¿En qué medida serán válidos nuestros resultados?

Generalmente, la **población** a estudiar no se puede conocer completamente, por esta razón es necesario investigarla en una **muestra** representativa y adecuada, manejable con metodologías confiables, para definirla globalmente considerando su validez científica.

1.3. ENFOQUE ESTADÍSTICO DE UNA INVESTIGACIÓN CIENTÍFICA

Para entender el enfoque estadístico de la investigación consideremos el siguiente problema.

Un ginecólogo del Hospital Regional de Chiclayo, sospecha que, a más partos en la madre, el tiempo de dilatación va aumentando.

Las preguntas que se plantean serán:

- ¿Qué se va a investigar y qué variables intervienen?

- El tiempo de dilatación antes del parto (variable continua) en función del número de partos (variable discreta).
- ¿Cuál es la variable dependiente?

  El tiempo de dilatación denotado por Y será la variable dependiente dada en horas desde el inicio de la dilatación hasta el momento del parto.
- ¿Qué factores pueden considerarse influyentes en la variable dependiente y cómo se medirán?

  Los factores de posible influencia en el tiempo de dilatación las denotamos por X_i

  X_1: Edad de la madre, entre los 15 y 45 años: variable continua.

  X_2: Peso de la madre en kilogramos, normal de acuerdo a su estatura: variable continua.

  X_3: Número de partos de la madre (1 primípara, 2, 3 o más multípara): variable discreta.

  X_4: Actitud psicológica hacia el parto, (1: favorable sin problemas. 2: desfavorable con problemas): variable cualitativa.

  Para la investigación se considera toda madre con actitud favorable al parto.

X_5: Preparación para el parto (1.con preparación, 2.sin preparación): variable cualitativa.

Para la investigación se considera toda madre sin preparación para el parto.

X_6: Nivel socio-económico de la madre (1. acude a clínica, 2. acude a maternidad común): variable cualitativa.

Para la investigación, se toma toda madre que acude a una maternidad común de Chiclayo.

¿Serán suficientes estos factores?

¿Es posible considerar constante otro factor?

¿Es correcto los niveles establecidos para los factores X_3 y X_6?

¿Cuál es la unidad de muestreo?

La unidad de muestreo es:

La madre parturienta con una edad entre 15 y 45 años que proceden de Chiclayo; que acude a la maternidad común para el parto, con peso normal de acuerdo a su estatura, con actitud favorable hacia el parto y sin preparación para el parto.

- ¿Cuán grande será la muestra?

  La muestra será representativa y adecuada, dependerá del grado de confiabilidad y de las características de la población. De no conocerse dichas características, se

sugiere una **muestra piloto**, para evaluar el error que se pudiere cometer en el cálculo de la muestra.

- ¿Cómo se recolectará la información?

  Dependerá de la técnica de muestreo, tomando en cuenta los factores de posible influencia. El instrumento para la toma de datos para esta investigación será la ficha clínica.

- ¿Dónde se va a realizar la investigación?

  En el Hospital Regional de Chiclayo.

- ¿Se realizó esta o similar experiencia antes y en qué lugar, a qué conclusiones se llegó?

  Debe buscarse antecedentes de investigaciones parecidas.

- ¿Con qué error de muestreo se trabajaron las investigaciones anteriores?

  El error de muestreo está referido a la comparación del estimador con el parámetro.

- ¿Qué técnicas estadísticas se usaron en los antecedentes?

  Se refiere al uso de modelos paramétricos para variables cuantitativas o no paramétricos para variables cualitativas.

- ¿Con qué medida de error realizaremos nuestra investigación?

  Se refiere al error de muestreo permitido por el investigador sobre la base del comportamiento de la variable dependiente.

Formulación de la Hipótesis

El número de partos influye significativamente en el tiempo de dilatación del parto.

1.4. MODELO ESTADÍSTICO QUE DESCRIBE EL EXPERIMENTO

En esta fase se determina la forma abstracta y simplificada del fenómeno real, en términos de los factores y las restricciones para dicho modelo. Una vez que se lleva a cabo la experiencia se formula las siguientes preguntas:

- ¿Cómo procesaremos la información?
- ¿Con qué técnicas estadísticas analizaremos la información para contrastar la hipótesis?

- ¿Con qué medida de error se realizará la investigación?
- ¿Qué validez tendrán los resultados?

Para responder a estas preguntas, la estadística proporciona técnicas confiables para la **evaluación de parámetros y toma decisiones con mínimo error**. Para contrastar la hipótesis, se considera el tiempo de dilatación de las madres primíparas y se compara con el tiempo de dilatación de las madres multíparas, el modelo sería: **Comparación de medias poblacionales**; para la segunda parte se aplica el **análisis de regresión** donde la variable tiempo de dilatación está en función del número de partos.

AUTOEVALUACIÓN I

1.Elija la secuencia lógica de las preguntas para investigar

1) ¿Qué se va a investigar?
2) ¿En quién o en qué?
3) ¿Dónde?
4) ¿Cuándo?

a) 1,4,2,3

b) 2,2,3,4

c) 1,2,4,3

2. Experimentalmente se Investiga:

a) Lo que es

b) Lo que será

c) Lo que fue

3. El 10% de instituciones Educativas investigadas son de educación primaria de cuarto y quinto año.

Esta conclusión es de carácter.

a) Descriptiva

b) Explicativa

c) Inferencial.

4. Se quiere investigar el (CI) de los estudiantes del quinto año de primaria de una universidad:

i. ¿Qué se va investigar?

.....

ii. ¿Qué variables intervienen en la investigación?

.....

iii. ¿Qué factores pueden influir en la variable en estudio?

.....

iv. ¿Cómo se medirán estos factores?

.....

v. ¿Se trabajará con todos los elementos de la población?

.....

vi. ¿Cómo se determinará el tamaño de la muestra?

.....

vii. ¿Cómo se recolectará la información?

.....

viii. ¿Dónde se va a investigar??

.....

ix. ¿Existen antecedentes de la investigación?

.....

x. ¿Cuáles son las hipótesis del problema?

.....

xi. ¿Con qué técnicas estadísticas podríamos analizar nuestra información para probar las hipótesis formuladas?

.....

5. Proponga un esquema diferente al que se presenta en el texto para la investigación: histórica, descriptiva y experimental.

CAPÍTULO II

LA INFORMACIÓN ESTADÍSTICA

La información para la investigación científica, debe partir de una muestra representativa y adecuada, tomando en cuenta las fuentes generadoras, tales como: naturaleza de la población y el modo de obtenerla, concordantes con la hipótesis formulada y el objetivo de la investigación; siendo los instrumentos de medición de las variables, ya sean físicos o documentales y el personal capacitado, medios fundamentales en la obtención de los datos.

2.1. SELECCIÓN DE UNA MUESTRA

Fijado el objetivo de la investigación, delimitar la **población** sobre cual recaerá la investigación es tarea de vital importancia; por ejemplo: si se investiga sobre la demanda de un determinado bien o servicio, el estudio puede llevarse a cabo, bien por el lado de los vendedores o de los compradores; el universo de los vendedores es más pequeño que de los compradores; el interrogatorio a los vendedores es más fácil, pero éstos no podrán darnos algunas características personales de los compradores.

En la práctica encontramos tres fases ligadas entre sí para la delimitación del universo muestral o población.

- Definir la unidad estadística o de observación.
- La delimitación espacial.
- La delimitación temporal.

1. La **unidad de observación**

   La unidad de observación puede ser una persona, un animal, un objeto, un fenómeno, un grupo de personas, una vivienda, una explotación agrícola, etc. de tal

manera que un elemento del universo empírico, pertenezca al universo o población de la investigación.

2. La **delimitación espacial.**

   La delimitación espacial **e**n una investigación debe indicarse el lugar donde ocurren los hechos, sucesos o fenómenos; así, si se investiga **niños con primer grado de desnutrición**, esta información no es suficiente para iniciar la investigación pues es necesario precisar, el lugar de dónde son los niños; de ser el caso, distrito, provincia, departamento o región.

3. La **delimitación en el tiempo**

   La delimitación en el tiempo de las unidades de observación completa la información para la aplicación de los instrumentos de medición de las variables del problema, permitiendo la contextualización del estudio. De esta manera, las observaciones **sobre desnutrición de los niños** pueden ser del año 2018 al 2021, solo del 2020 o del primer semestre de 2021 etc.

Figura 2.1 Delimitaciones de la población en la investigación científica.

2.2. DETERMINACIÓN DEL TAMAÑO DE MUESTRA

Para realizar **inferencia estadística,** la muestra debe ser representativa y adecuada a la población de la que procede.

La **representatividad de la muestra** está referida a la elección de sus elementos sin sesgo, sin tendencia, de modo que, todos tengan la misma probabilidad de ser elegidos, para esto, se usa la técnica de la aleatoriedad en los muestreos.

La **adecuacidad de la muestra** está dada por las unidades de muestreo que garanticen las mismas características de la

población; esto es, si la muestra es adecuada, las inferencias o generalizaciones serán válidas.

La muestra se diseña al inicio de la investigación,
sí es pequeña, los resultados pueden ser no válidos,
y si es muy grande, implica pérdida de recursos.

La muestra está condicionada por:

1. Las variables del problema. Si están muy relacionadas, se requiere, muestras pequeñas, si están poco correlacionadas se necesitan muestras grandes.

2. La homogeneidad o heterogeneidad de las poblaciones, poblaciones homogéneas, muestras pequeñas y poblaciones heterogéneas, muestras grandes.

3. El nivel de confiabilidad para estimar el promedio o proporción verdadera de la población. Para el 95% de confiabilidad el valor de Z en la normal estandarizada es de 1.96.
4. La exactitud de los resultados que el investigador desea obtener, se llama también error permitido o grado de error entre el estimador y el parámetro de la variable que se investiga, para resultados más exactos, muestras más grandes.

A partir de investigaciones anteriores y bases teóricas se pueden obtener parámetros iniciales μ y σ (promedio y desviación estándar) en el caso de variables cuantitativas o P y Q (proporción de elementos con o sin la característica en estudio) para variables cualitativas, los que servirán para determinar el tamaño de muestra. Estos parámetros se usan para estimar el error permisible o error de muestreo h con las fórmulas $h = |\bar{X} - \mu|$ en el caso de variables cuantitativas o $h = |p - P|$ en el caso de variables cualitativas, evaluando de esta forma que tan distante de los parámetros reales de la población objetivo se encuentren los estimadores $\bar{X}$, S y p que se calculan con la muestra a determinarse. De no conocerse los parámetros iniciales por corresponder a teorías o estándares aún no investigados, se hace uso de muestreos pilotos.

2.2.1. Tamaño de muestra para estimar μ

A) Cuando la varianza σ^2 es conocida:

$$n = \frac{Z^2\sigma^2}{h^2} \qquad n = \frac{NZ^2\sigma^2}{h^2N + Z^2\sigma^2}$$

Población infinita o desconocida. Población finita o conocida.

B) Cuando la varianza σ^2 es desconocida:

$$n = \frac{Z^2 S^2}{h^2} \qquad\qquad n = \frac{N Z^2 S^2}{h^2 N + Z^2 S^2}$$

Población N desconocida

Población N conocida

sabiendo, para ambos casos, que:

Z Factor de confiabilidad que se obtiene de la Tabla Z.

σ^2 Varianza poblacional de la variable problema.

h Error permitido, diferencia entre el estimador y el parámetro, valor experimental y real de la variable en estudio. $h = |\bar{X} - \mu|$

S^2 Varianza muestral, estimador de σ^2.

N Tamaño de la población.

n Tamaño de la muestra definitiva.

Para determinar la muestra definitiva, generalmente se realiza un muestreo piloto, en donde el investigador toma referencia del promedio y la desviación estándar de la variable en estudio, que le ayuda a evaluar h si no está precisado.

Ejemplo 1

Se desea estudiar el peso en una población de niños de grupo etario determinado, ¿cuál es el tamaño de muestra para esta investigación?

Solución:

La población de niños no es conocida, tampoco se conoce como varía el peso. Aleatoriamente se elige una muestra piloto de 20 niños donde los pesos en kilogramos son:

14.8	14.2	15.4	15.3	15.0	15.2	15.3	14.2	15.2	15.5
14.5	15.5	15.2	15.0	14.3	15.2	14.0	14.0	14.2	15.3

De esta muestra preliminar se obtiene el promedio, la desviación estándar y la varianza.

$\bar{X} = 14.865kg.$ $S = 0.5344kg.$ $S^2 = 0.2856kg^2$

Con esta información sobre las características de la variable en estudio ya se puede determinar la muestra definitiva admitiendo un error de 0.15 kg, y estableciendo un 95% de confiabilidad. (Z= 1.96, ver en la tabla Z del apéndice)

$$n = \frac{Z^2 S^2}{h^2} = \frac{(1.96)^2(0.2856)}{(0.15)^2} = 48.76 \cong 49$$

Luego, la investigación se debe realizar en 49 niños.

Ejemplo 2

El cociente intelectual (C.I.) para niños tiene las siguientes características: $\mu = 100$ y $\sigma = 15$. Calcular una muestra para investigar sobre el C.I. en I.E.S. Santa Rosa, con 2 unidades de error y 95% de confiabilidad.

Sustituyendo los valores se tiene:

$$n = \frac{Z^2\sigma^2}{h^2} = \frac{(1.96)^2(15)^2}{(2)^2} = 216$$

Luego la muestra estará constituida por 216 niños.

2.2.2. Tamaño de muestra para estimar P

Ya sea para estimar μ o P se debe especificar, que tan cerca deseamos que esté nuestra estimación del valor real que se está calculando, indicar qué nivel de confianza se desea y finalmente hacer una estimación de la variabilidad de la población.

Las estimaciones de variabilidad de la población se pueden calcular a partir de otros estudios o mediante una muestra piloto.

A) **Cuando la proporción P de la población que tiene la característica de interés es conocida**.

$$n = \frac{Z^2 PQ}{h^2} \qquad n = \frac{NZ^2 PQ}{h^2 N + Z^2 PQ}$$

Población infinita o desconocida. Población finita o conocida.

B) **Cuando la proporción P de la población que tiene la característica de interés es desconocida**.

$$n = \frac{Z^2 pq}{h^2} \qquad n = \frac{NZ^2 pq}{h^2 N + Z^2 pq}$$

Población desconocida Población conocida

sabiendo, para ambos casos, que:

Z Valor de la distribución normal estandarizada correspondiente al nivel de confianza escogido.

P Proporción con la característica de interés y Q=1-P su complemento.

p Estimador de P y $\mathrm{h} = 1 - P$ su complemento.

h Error permitido, diferencia entre el estimador *p* y el parámetro *P*, $\mathrm{h} = p - P$.

Cuando no se dispone de una estimación de P, lo práctico es buscar la proporción de elementos con la variable en estudio a

partir de una muestra piloto. Se obtiene el máximo valor de n haciendo $P = 0.50$.

Desconocida P, proporción con característica en estudio, se asigna 50% máximo, además h (error permisible) no debe ser mayor de 6%.

Ejemplo 1

Tomado de Vásquez:

> Una asistenta de un seguro social de salud desea saber el tamaño de muestra para hacer una estimación de la proporción de ciudadanos que están de acuerdo con la afirmación: Son demasiadas las personas que deberían estar trabajando y que están recibiendo pensiones de bienestar social.
>
> La investigadora desea que la estimación de p, esté a 2.5% del valor verdadero con 95% de confiabilidad.
>
> Investigadores que llevaron a cabo un estudio similar encontraron que el 60% de la gente entrevistada, estaba de acuerdo con la afirmación; por lo tanto, la estimación de la muestra se calculará de la manera siguiente, considerando *p=0.6*.

$$n = \frac{Z^2 pq}{h^2} = \frac{1.96^2(0.6)(0.4)}{(0.025)^2} = 1{,}475.17 \cong 1{,}476$$

luego, la asistenta deberá entrevistar a 1476 personas.

Si no hubiera conocido de la estimación de P , se utilizaría el valor de P = 0.5, que hace máximo el tamaño de la muestra. (1990, p.30)

$$n = \frac{Z^2 PQ}{h^2} = \frac{1.96^2(0.5)(0.5)}{(0.025)^2} = 1{,}536.64 \cong 1{,}537$$

así, la asistenta debería considerar 1 537 personas.

Ejemplo 2

> Se desea conocer la frecuencia de cáncer uterino en los pueblos jóvenes de Chiclayo. Se han hecho estudios en Lima y se han encontrado una incidencia de 6 por mil. Cuál debe ser el tamaño de muestra adecuado para la investigación si se espera que el valor de la estimación no difiera más del 2 por mil del verdadero, con 95% de seguridad. (Vásquez, 1990, p. 30)

Se debe usar la primera fórmula, pues no se conoce la población.

$$n = \frac{1.96^2(0.006)(0.994)}{0.002^2} = 5{,}727.8 \cong 5{,}728.$$

Ejemplo 3

¿Cuál es la proporción de trabajadores que manifiestan conformidad con la política laboral de su empresa, si de 150,

120 responden estar satisfechos? Los datos que contamos son: N=1500 empleados, h = 3.2 % y Z=1.96.

$$n = \frac{NZ^2pq}{h^2N + Z^2pq} = 428.69 \cong 429$$

Con este tamaño de muestra ya se puede determinar la proporción de empleados que muestren conformidad.

Ejemplo 4.

Un instituto de educación tiene 200 profesores. En un estudio piloto el 73% están de acuerdo en cambiar el plan de estudios. Determine el tamaño de muestra para investigar sobre la aceptación del nuevo plan al 95% de seguridad y el 5% de error.

$$n = \frac{NZ^2pq}{h^2N + Z^2pq} = \frac{(200)(1.96)^2(0.73)(0.27)}{(0.05)^2(200) + (1.96)^2(0.73)(0.27)} \cong 121$$

donde:

Z = 1.96

n = Muestra

p = Proporción con característica del evento, 0.73 o 73%

q = Proporción sin la característica.

q = 1-p = 1- 0.73 = 0.27 o 27%.

h = 0.05 o 5%

N = Población: 200 profesores.

Por lo tanto, 121 profesores están de acuerdo en cambiar el plan de estudios.

2.3. DISTRIBUCIÓN DE LA POBLACIÓN Y LA MUESTRA

2.3.1. Distribución de frecuencias de una población

Ejemplo 1

La distribución que sigue corresponde a los niveles del cociente intelectual (C.I.) de una población de 987 estudiantes de pregrado de una universidad peruana.

Tabla 2.1 Distribución del cociente intelectual (C.I) de la población de 987 estudiantes de pregrado de una universidad peruana.

Niveles de C.I	Promedio de clase (X_i)	Frecuencia n_i	Frecuencia relativa (h_i)	Frecuencia acumulada (N_i)
60 - 70	65	26	0.026	26
70 - 80	75	56	0.057	82
80 - 90	85	145	0.147	227
90 -100	95	230	0.233	457
100-110	105	235	0.238	692
110-120	115	182	0.185	874
120-130	125	82	0.083	956
130-140	135	31	0.031	987
TOTAL	---	987	1.000	---

Fuente: Departamento de evaluación pedagógica de la universidad.

Para esta distribución los parámetros del cociente intelectual de la población son:

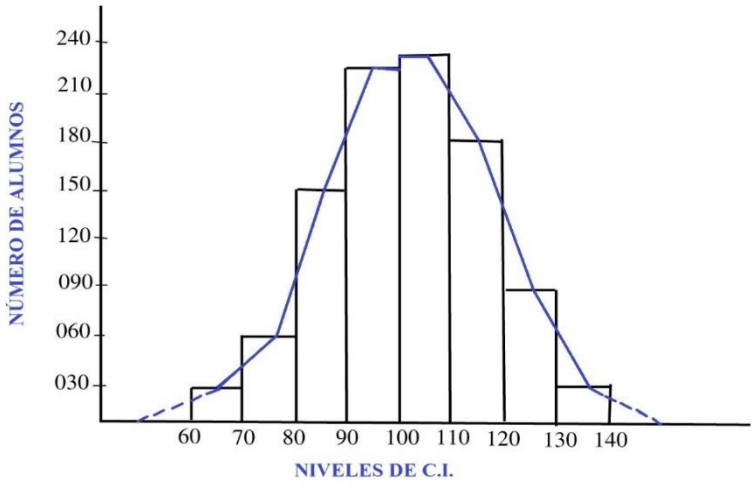
$$\text{Promedio } = \mu = \frac{\sum X_i n_i}{N} = \frac{100105}{987} = 101.45$$

$$\text{Varianza } = \sigma^2 = \frac{987(10392275) - (100105)^2}{(987)^2} = 242.43$$

$$\text{Desviación Estándar } = \sigma = \sqrt{\sigma^2} = 15.57$$

El polígono de frecuencia de h_i representa a la distribución de probabilidades por intervalos del cociente intelectual (C.I.); con N_i se construye el polígono acumulativo u ojiva de la distribución.

Figura 2.2 Cociente intelectual (C.I.) de la población de 987 estudiantes de pregrado de una universidad peruana.



2.3.2. Distribución de frecuencias de una muestra.

Ejemplo 2

Considere los datos de la siguiente tabla.

Tabla 2.2 Distribución del cociente intelectual (C.I.) de una muestra de 275 estudiantes de pregrado de una universidad peruana.

Niveles de C.I	Promedio de clase (Xi)	Frecuencia ni	Frecuencia relativa (hi)	Frecuencia acumulada (Ni)
60-70	65	2	0.007	2
70-80	75	4	0.015	6
80-90	85	43	0.156	49
90-100	95	88	0.320	137
100-110	105	91	0.331	228
110-120	115	35	0.127	263
120-130	125	7	0.026	270
130-140	135	5	0.018	275
TOTAL	---	275	1.000	---

Para esta distribución, los estimadores del cociente intelectual de la muestra son:

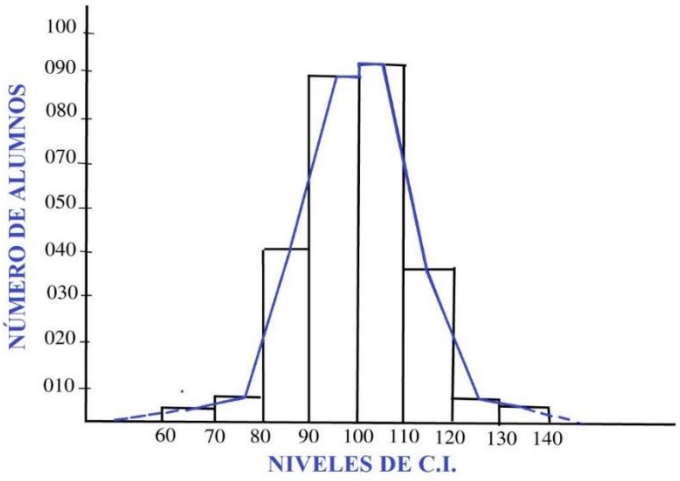
$$\text{Promedio } = \bar{X} = \frac{\sum X_i n_i}{n} = \frac{27575}{275} = 100.27$$

$$\text{Varianza} = S^2 = \frac{\sum (X_i - \bar{X})^2 n_i}{n-1} = \frac{n\sum {X_i}^2 n_i - (\sum X_i n_i)^2}{n(n-1)} =$$

$$S^2 = \frac{275(2802475) - (27575)^2}{275(274)} = 136.69$$

Desviación estándar $= \mathrm{S} = \sqrt{S^2} = 11.69$

Figura 2.3 Distribución del cociente intelectual (C.I.) de una muestra de 275 estudiantes de pregrado de una universidad peruana.



Como se puede apreciar, tanto la población como la muestra se distribuyen análogamente. La diferencia entre estimador y el parámetro del promedio ($|\bar{X} - \mu| = h$) se denomina error de muestreo o error permisible en la investigación; disminuye al aumentar el tamaño de la muestra.

El promedio μ del C.I. en la población es 101.42 con $\sigma = 15.57$, en la muestra $\bar{X}$=100.27 con una desviación estándar S de 11.69.

Las características más importantes para describir la población y la muestra respectivamente son μ, σ y $\bar{X}$, S. Con estas características se puede obtener porcentajes de información contenidos en intervalos determinados por el número de desviaciones estándar:

$$\mu \pm \sigma = 68.26\% = \bar{X} \pm S$$
$$\mu \pm 2\sigma = 95.44\% = \bar{X} \pm 2S$$
$$\mu \pm 3\sigma = 99.74\% = \bar{X} \pm 3S$$

Los resultados a cerca de la población o la muestra se pueden presentar en gráficos, los cuales permiten visualizar mejor las características más resaltantes en estudio.

Los gráficos más frecuentes son los histogramas y polígonos de frecuencias absolutas, relativas y acumuladas.

Los gráficos polares, sectoriales, de series de tiempo, gráficos de Gantt etc. son también utilizados.

2.4. DISTRIBUCIÓN BIDIMENSIONAL

Cuando se estudian dos características en forma simultánea de la población (o muestra) la información se presenta en tablas bidimensionales, de doble entrada o de correlación.

Ejemplo 1

Se requiere estudiar la forma como se distribuye el cociente intelectual (C.I.) de acuerdo a la edad en los estudiantes de pregrado de una universidad peruana. Para esto se hará uso de una muestra de 275 alumnos.

Tabla 2.3 Distribución del cociente intelectual según edad de 275 estudiantes de pregrado de una universidad peruana.

C.I.(x)	Edad y / Y_i / X_i	16-18	18-20	20 -22	22-24	24-26	Total
		17	19	21	23	25	
60 - 70	65	1	0	1	0	0	2
70 - 80	75	1	1	2	0	0	4
80 - 90	85	4	8	14	10	7	43
90-100	95	11	27	23	18	9	88
100-110	105	9	29	27	19	7	91
110-120	115	5	7	9	11	3	35
120-130	125	1	3	2	1	0	7
130-140	135	1	2	2	0	0	5
TOTAL	---	33	77	80	59	26	275

Fuente: Departamento de evaluación pedagógica de la universidad.

Con esta tabla de doble entrada se puede estimar los parámetros de cada una de las variables, generalmente se usan para determinar si existe alguna **relación funcional** que explique el comportamiento de estas variables (regresión) o el **grado de relación** entre ellas (correlación).

El parámetro ρ indica correlación entre variables, se expresa por:

$$\rho = \frac{\text{Cov}\,(X,Y)}{\sigma_x \sigma_y} = \frac{\sum\sum(X_i - \mu_x)\left(Y_i - \mu_y\right)}{N\sigma_x \sigma_y}$$

el estimador de ρ es r,

$$r = \frac{Cov\,(X,Y)}{S_x S_y} = \frac{\sum\sum(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_x S_y}$$

$$\text{Cov}\,(X,Y) = \frac{\sum x_i y_j n_{ij}}{n-1} - \bar{X}\,\bar{Y}$$

Para el caso de la covarianza de la distribución bidimensional en el ejemplo se iniciaría con:

$$\text{Cov}(X,Y)$$

$$= \frac{(65 \times 17 \times 1 + 75 \times 17 \times 1 + \cdots + 125 \times 25 \times 0 + 135 \times 25 \times 0)}{274}$$

$$- (100.27)(20.76)$$

$$\text{Cov}\,(X,Y) = 6.84$$

$$r = \frac{\text{Cov}\,(X,Y)}{S_x S_y} = \frac{6.84}{(11.69)(2.32)} = 0.252$$

Este resultado indica una correlación muy baja entre los niveles de cociente intelectual C.I. y la edad.

Los estimadores de la variable edad:

$$\bar{Y} = \frac{\sum Y_i n_i}{n} = \frac{5711}{275} = 20.76 \text{ años de edad.}$$

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2 n_i}{n-1} = \frac{n\sum {Y_i}^2 n_i - (\sum Y_i n_i)^2}{n(n-1)} =$$

$$S^2 = \frac{275(120075) - (5711)^2}{(275)(274)} = 5.3763$$

$$S = 2.32 \text{ años de edad.}$$

Análogamente, se pueden calcular estos estimadores para la variable *X*, cociente intelectual.

Tabla 2.4 Cálculo de $\overline{X}$ y S del cociente intelectual de la tabla bidimensional de los 275 estudiantes.

Marca de clase del CI(X_i)	Frecuencia marginal del CI n	$X_i n_i$	X_i^2	$X_i^2 n_i$
65	2	130	4225	8450
75	4	300	5625	22500
85	43	3655	7225	310675
95	88	8360	9025	794200
105	91	9555	11025	1003275
115	35	4025	13225	462875
125	7	875	15625	109375
135	5	675	18225	91125
Total	**275**	**27575**	**84200**	**2802475**

$$\bar{X} = \frac{\sum X_i n_i}{n} = \frac{27575}{275} = 100.27$$ de cociente intelectual.

$$S_X^2 = \frac{n\sum X_i^2 n_i - (\sum X_i n_i)^2}{n(n-1)} = 136.69$$

S_X = 11.69 unidades de CI.

2.5. EL MUESTREO

Una investigación científica se hace generalmente con una parte de la población por algunas de las siguientes razones:

- Es menos costoso observar una parte en lugar de toda la población**.**
- Una muestra se puede observar en menos tiempo.
- Los elementos de la población se deterioran; como en el caso de pruebas industriales (pruebas de elasticidad, tracción, resistencia al calor, etc.)
- Limitaciones de recursos humanos con entrenamiento.
- Por homogeneidad de la población.

2.5.1. Técnicas de muestreo

Tomado de Vásquez et al.:

> El proceso para seleccionar un conjunto de elementos, personas o cosas de una población de la que se consideren representativos, con el fin de estudiarlos y caracterizar esta población se denomina muestreo; el muestreo está condicionado al tipo de investigación y fundamentalmente al diseño para la contrastación de las hipótesis, dado que cada muestreo tiene diferente metodología, los muestreos pueden ser probabilístico y no probabilístico. (2021, p. 92)

A. Muestreo probabilístico

De acuerdo a Vásquez et al: “el muestreo es probabilístico si todos los elementos de la población tienen las características que se van a investigar y tienen la misma probabilidad de salir elegidos para la muestra, de acuerdo a como se presenta la población pueden ser”: (2021, p. 92)

A1. Muestreo aleatorio simple

> El Muestreo aleatorio simple **(**M.A.S), se aplica si la población es finita y sus elementos son homogéneos respecto a las características en estudio. El número de elementos en la muestra es n, para métodos de selección de igual probabilidad, f es la fracción de muestreo uniforme

general para los elementos, siendo la fracción de muestreo $f = \frac{n}{N}$ y $f = \frac{n}{N} \times 100$ es el porcentaje de la población que representa la muestra. Para seleccionar una muestra (n) se someten a todos los elementos de la población (N) a sorteo o a un ordenador para generar números aleatorios, de tal manera que cada elemento tiene la misma probabilidad de salir elegido en la muestra. Este muestreo se utiliza en los experimentos científicos propiamente dichos. (Vásquez et al., 2021, p. 92)

A2. Muestreo aleatorio sistemático

Es un muestreo aleatorio simple, si se tiene una población ordenada se puede dividir en intervalos iguales. Del primer intervalo se selecciona al azar un elemento, y con los que ocupan el mismo lugar en el resto de intervalos se forma la muestra. “Es decir, para extraer una muestra de tamaño n, se elige un número $k = \frac{N}{n}$, luego al azar se selecciona otro número “a” tal que $1 \leq a \leq k$ llamado número de arranque. La muestra queda determinada con los siguientes elementos:” (Vásquez et al., 2021, p. 92).

$a,\ a + k,\ a + 2k,\ \ldots,\ a + (n - 1)k.$

A3. Muestreo aleatorio estratificado

En un muestreo estratificado, la población de N unidades es primero dividida en subpoblaciones de N_1, $N_{2,} \dots, N_L$ unidades, respectivamente. Estas subpoblaciones no se sobreponen y juntas forman la totalidad de la población, por lo que $N_1 + N_2 + \cdots + N_L = N$. Las subpoblaciones son llamadas estratos. Para obtener un beneficio completo de la estratificación, deben ser conocidos los valores de N_h. Una vez que han sido determinado los estratos se saca una muestra de cada uno, la obtención se realiza independientemente en estratos diferentes. Los tamaños de muestra dentro de los estratos son representados por n_1, $n_{2,} \dots, n_L$ respectivamente. (Neumann, 2014)

Debe existir proporcionalidad entre el tamaño de la muestra al tamaño del estrato; la muestra será más grande si el estrato es más variable (asignación óptima). Al estratificar una muestra se busca la máxima homogeneidad de la variable de estudio en cada estrato y la máxima heterogeneidad entre los estratos.

Si la muestra se reparte igual en los diferentes estratos, se llama igual afijación, si el reparto de la muestra es

proporcional a los estratos se llama con afijación proporcional y si la muestra se reparte tomando en cuenta la dispersión de los estratos se llama afijación óptima o de Neyman.

A4. Muestreo aleatorio por conglomerados o áreas

Adecuado cuando la población es grande y los elementos de interés que presentan características similares se distribuyen en toda la población, por eso no se puede estratificar, o sus elementos presentan dificultades para sortearse. Para obtener una muestra, primero se divide a la población en conglomerados o subpoblaciones, luego cada conglomerado se somete a muestreo aleatorio simple, según los tamaños de muestras asignadas a cada uno de ellos.

Según el tipo de investigación se utiliza muestreos que minimicen los errores, el muestreo por etapas sucesivas reduce los errores.

B. Muestreo no probabilístico

Si los muestreos no son probabilísticos no se pueden hacer generalizaciones a la población, ya que la muestra no es

representativa, esto es, los elementos de la población no han sido elegidos aleatoriamente.

Los más usados en la investigación son:

B1. Muestreo por cuotas

> “Se usa generalmente cuando se tiene buen conocimiento de los estratos de la población o de los elementos más representativos o adecuados para los fines de la investigación” (Vásquez et al., 2021, 95).
>
> Se fijan unas cuotas que consisten en un número de individuos que reúnen determinadas condiciones por ejemplo 20 personas de 30 a 50 años de sexo femenino residentes en San Eduardo; una vez determinada la cuota se eligen las primeras que se encuentren que cumplen esas características. Se utiliza principalmente en encuestas de opinión.

B2. Bola de Nieve

> En este muestreo se localiza algunos individuos, los cuales conducen a otros, y estos a otros y así hasta conseguir en las preguntas un punto de saturación y

obtener una muestra suficiente. Se usa cuando se hacen estudios con poblaciones marginales, sobre delincuentes, sectas, determinados tipos de enfermos etc. (Vásquez et al., 2021, 96)

B3. Muestreo Criterial

Llamado también a juicio de expertos, crítico o discrecional, intencional u opinático; en este muestreo el investigador selecciona directa o indirectamente los individuos para su estudio, la muestra es representativa, si se incluyen elementos o grupos supuestamente típicos. Se usan para estudiar grupos humanos con características especiales, como en sondeos preelectorales. También pueden escogerse elementos a los que se tiene fácil acceso, como el caso de los profesores que usan con frecuencia sus propios alumnos. (Vásquez et al., 2021, 97)

B4. Muestreo por Conveniencia

"Llamado también deliberado, errático, accidental o por conveniencia. Los elementos de la muestra son elegidos a criterio del investigador, sobre lo que él cree que puedan aportar a la investigación" (Vásquez et al., 2021,

97-98). Existen también los muestreos por itinerarios en donde se proporciona el perfil por ejemplo de las personas que tiene que entrevistar o los muestreos intencionales que elige el investigador por las características que quiere investigar o los que se ofrecen en ser investigados.

2.5.2. Selección de muestras con números aleatorios

Supongamos que la población tiene **50 observaciones** y que sus elementos están enumerados de 1 a N (N = 50, tamaño de una población).

1- 131	11 - 92	21 - 87	31 - 95	41 - 68
2 - 79	12 - 64	22 - 90	32 - 260	42 - 295
3- 305	13 - 80	23 - 239	33 - 240	43 - 280
4- 419	14 - 90	24 - 137	34 - 121	44 - 120
5 -180	15 - 110	25 - 170	35 - 126	45 - 148
6 - 75	16 - 78	26 - 98	36 - 85	46 - 91
7 - 85	17 - 74	27 - 275	37 - 299	47 - 310
8 - 295	18 - 270	28 - 235	38 - 225	48 - 250
9 - 118	19 - 125	29 - 145	39 - 138	49 - 155
10-120	20 - 145	30 - 115	40 - 139	50 – 122

Supongamos una muestra de tamaño 24 (n = 24).

1. Como la población tiene 2 dígitos (N = 50) se toma en la tabla de dígitos al azar, una columna que contenga dos cifras.

2. Según indique la columna de dígitos elegida, se van separando todos aquellos menores o iguales a N, como elementos para la muestra tomándose en cuenta solo una vez.
3. Si se agota la columna de dígitos antes de completar la muestra, se utiliza otra columna y así sucesivamente hasta completar "n".

Para el ejemplo se usa la primera columna de dígitos que inicia con **51772** y la sub - columna que comienza con 51. (Tabla de Números Aleatorios, Anexos). Según esto las observaciones integrantes de la muestra son:

24 – 137	21 – 87	23 – 239	4 – 419
45 – 148	50 – 122	2 – 79	42 – 295
30 – 115	7 – 85	17 – 74	29 – 145
3 – 305	27 – 275	25 – 170	10 – 120
15 – 110	8 – 295	40 – 139	36 – 85
9 - 118	39 – 138	34 – 121	48 – 250

Números aleatorios

51772	74640	42331	29044	46621	62898	93582
24033	23491	83587	06568	21960	21387	76105
45939	60173	52078	25424	11645	55870	56974
30586	02133	75797	45406	31041	86707	12973
03585	79353	81938	82322	96799	85659	36081

Para la **población** de 50 datos, los parámetros μ y σ^2 son:

$$\mu = \frac{\sum X_i}{N} = 161.86$$

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{N} = \frac{N\sum X_i^2 - (\sum X_i)^2}{N^2}$$

$$\sigma^2 = \frac{50(1670521) - (8093)^2}{2500} = 7211.7604$$

$$\sigma = 84.92$$

Y para la **muestra** de 24 datos, los estimadores $\bar{X}$ y S^2 son:

$$\bar{X} = \frac{\sum X_i}{n} = \frac{3894}{24} = 162.25$$

$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n(n-1)} = \frac{n\sum X_i^2 - (\sum X_i)^2}{n(n-1)}$$

$$S^2 = \frac{24(810654) - (3894)^2}{24(23)} = 7776.1956$$

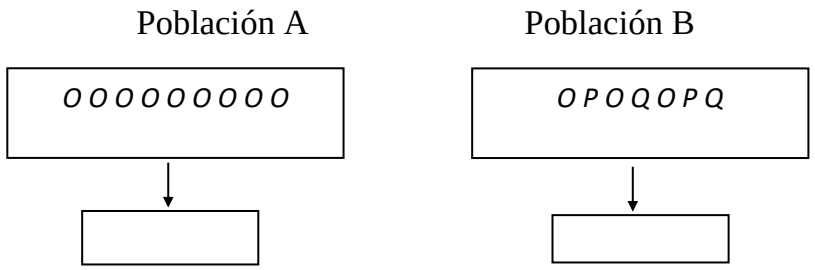
$$S = 88.18$$

μ, σ^2, P (Promedio, varianza, proporción), son características de la población generalmente **no conocidas**; como la **investigación científica** necesita de estos parámetros para contrastar sus hipótesis y poder formular sus principios y modelos, resulta imprescindible **estimarlos.** La muestra, para calcular $\bar{X}$, S^2, y p estimadores de los parámetros (promedio, varianza, proporción), debe ser representativa y adecuada asegurando así la validez y confiabilidad de las generalizaciones.

AUTOEVALUACIÓN II

1) La población, objeto de estudio, se divide en subpoblaciones lo más heterogéneo posible:
   A) ¿En el muestreo sistemático?
   B) ¿En el muestreo estratificado?
   C) ¿En el muestreo por conglomerados?

2) Selecciona una muestra de cada población:



3) Dadas las variables:

- Número de personas por familia.
- Color de ojos de los cursillistas.
- Cantidad de H_2S (ácido sulfhídrico) producidos en una
- fermentación anaeróbica (sin oxígeno).
- Temperatura de los pacientes en una clínica.
- Grupos sanguíneos de las personas.
- Contenido de creatinina en la sangre.
- Vías de administración de vacunas.

- Edad de los estudiantes de enfermería.
- Color y pétalos de una rosa.

Identifíquelas según sean: Cualitativas, discretas, continuas.

4) Haciendo uso de la tabla de números al azar de la población de 50 datos que se considera en el texto, seleccione 2 muestras de 7 y 12 elementos respectivamente.

1……	5……
2……	6……
3……	7……
4……	

1……	5……	9 ……
2……	6……	10……
3……	7……	11……
4……	8……	12……

5) Considere, como 2 poblaciones diferentes, las muestras antes seleccionadas y calcule sus parámetros:

$\mu_1 =$ _____ $\sigma_1^2 =$ _____ $\sigma_1 =$ _____

$\mu_2 =$ _____ $\sigma_2^2 =$ _____ $\sigma_2 =$ _____

6) De las 2 poblaciones consideradas en la pregunta 6 saque una muestra de 5 y 10 respectivamente y calcule sus estimadores:

$\overline{X}_1 =$ _____ $S_1^2 =$ _____ $S_1 =$ _____ $CV_1 = \frac{S_1}{\overline{X}_1} \times 100$

$\overline{X}_2 =$ _____ $S_2^2 =$ _____ $S_2 =$ _____ $CV_2 = \frac{S_2}{\overline{X}_2} \times 100$

7) Edina, de 11 años y 5 meses de edad y alumna del sexto grado, respondió correctamente 85 preguntas de una prueba normalizada de lectura, consistente en 110 preguntas. Rosa, de 8 años de edad que está en tercer grado, respondió bien a 42 preguntas, en la misma prueba. Suponga que la norma del sexto grado es de 98 y la de tercer grado 31 (en ambas normas se combinan los resultados de uno y otro sexo).

a) Cuál es la población de referencia de Edina?

…….

b) ¿Cuál es la población de referencia de Rosa?

…….

c) Suponiendo que la prueba sea una buena medida de capacidad de lectura, ¿quién lee mejor en este momento?

…….

d) ¿Cuál de ellas tiene mayor probabilidad de que su maestra la considere una buena lectora? ¿Por qué?

…….

8) Usted supone que existe relación entre el sodio en el plasma sanguíneo y el sodio en los glóbulos rojos. Si se ha propuesto investigar este problema, está obligado a diseñar una prueba estadística en base:

A) ¿A la delimitación del universo objetivo?

B) ¿Al tipo de muestreo?

C) ¿Al objetivo de la investigación?

9) Los anestésicos A, B y C antes de administrarse en personas se prueban en conejos con los siguientes resultados:

112 machos y 88 hembras responden positivamente a la combinada AB. 104 machos probados con A, 126 con B y 80 con C responden positivamente. 88 machos y 112 hembras responden negativamente la combinada AC. 96 hembras probadas con A, 74 con B y 120 con C responden negativamente. 80 machos y 120 hembras responden positivamente a la combinada BC.

Presente una tabla analítica que proporcione: Totales por anestésicos, sexo y respuesta positiva o negativa. Y muestre una gráfica que permita interpretar objetivamente los resultados.

10) Una variable en estudio tiene 100 de varianza, se permite 4u de error en el muestreo y 95% de seguridad en los resultados. El tamaño de muestra es:

A) 25

B) 96

C) 49

11) El muestreo indicado para estudiar la influencia de la televisión en la cultura infantil es:

A) ¿Sistemático?

B) ¿Estratificado?

C) ¿Por conglomerados?

12) Las variables X, Y son las variables del problema con: $S_x = 0.2$, $S_y = 4$. Elija una desviación típica para calcular la muestra y justifique su respuesta.

CAPÍTULO III

DISTRIBUCIONES DE PROBABILIDAD

Una **distribución de probabilidad** de datos corresponde, a una distribución de frecuencias relativas de la población. Son distribuciones **teóricas** que se emplean como patrones de referencia para analizar distribuciones **empíricas** reales, interpretándose, cómo se distribuye la población en cada punto $P(X = x)$ o en un conjunto de puntos $P(X \leq x)$.

Una distribución de probabilidades ayuda a describir un experimento o predecir futuros resultados cuando se efectúa un número aceptable de repeticiones.

3.1. DISTRIBUCIÓN BINOMIAL $B(X, n\,p, n\,p\,q)$

La probabilidad de obtener x triunfos en n pruebas independientes está dado por:

$$P(x) = C_x^n\ p^x\ q^{n-x}$$

donde:

x = 0, 1, 2, …, n.

p = probabilidad de éxito en cada prueba.

q = 1-p = probabilidad de fracaso.

$\mathbf{C}_x^n = \frac{n!}{(n-x)!x!}$

n! = n (n-1) (n-2) … (2)(1)

0! = 1

Características

Media = esperanza matemática = $\mu = np$

Varianza = $\sigma^2 = npq$

Desviación estándar = $\sigma = \sqrt{npq}$

Ejemplo 1

Si la prueba es el lanzamiento de una moneda se tiene:

P(cara) = $\frac{1}{2}$ = p P(sello) = $\frac{1}{2}$ = q

La probabilidad de obtener 2 caras en 4 lanzamientos será:

x = 2, n = 4, $p = \frac{1}{2}$, $q = \frac{1}{2}$

$P(x = 2) = C_2^4 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = 6\left(\frac{1}{2}\right)^4 = 0.375.$

Ejemplo 2

Un tratamiento de cierta enfermedad tiene una probabilidad de 0.6 de aliviar todos los signos y síntomas. Determine la probabilidad de que, de 6 pacientes tratados, se curen exactamente 2.

$$P(x = 2) = C_2^6 (0.6)^2 (0.4)^4 = 15(0.009216) = 0.138$$

La probabilidad de que a lo más 2 pacientes se mejoren está dado por:

$$P(x \leq 2) = P(x = 0) + P(x = 1) + P(x = 2)$$

$$= C_0^6 (0.6)^0 (0.4)^6 + C_1^6 (0.6)^1 (0.4)^5 + C_2^6 (0.6)^2 (0.4)^4$$

$$= C_0^6 (0.4)^6 + 6(0.6)(0.4)^5 + 15(0.6)^2 (0.4)^4$$

$$P(X \leq 2) = 0.1792$$

Para un total de 100 pacientes cabe esperar que mejoren:

$\mu = n\,p = 100(0.6) = 60$ pacientes

$\sigma = \sqrt{npq} = 100(0.6)(0.4) = 4.89 \approx 5$ pacientes.

3.2. DISTRIBUCIÓN DE POISSON $P(X, \lambda, \lambda)$

Si n es grande y p es pequeña, las probabilidades binomiales se suelen aproximar mediante la fórmula:

$$P(x) = \frac{e^{-(np)} np^x}{x!} \qquad x = 0, 1, 2\ldots n, \qquad e = 2.71821$$

Ejemplo 1

Una compañía de seguros contra incendios, asegura 10,000 casas dispersas por todo Lima. Suponiendo la probabilidad de que una casa cualquiera se incendie sea p = 1/5000, ¿cuál es la probabilidad de que el fuego destruya a 0, 1, 2 de las 10,000 casas durante un año determinado?

Si n = 10,000, $p = \frac{1}{5000}$ y $n\,p = 2$.

Luego:

$$P(x = 0) = \frac{e^{-2}2^0}{0!} = 0.1353$$

$$P(x = 1) = \frac{e^{-2}2^1}{1!} = 0.2706$$

$$P(x = 2) = \frac{e^{-2}2^2}{2!} = 0.2706$$

Ejemplo 2

La probabilidad de que una persona se contagie con un virus es 0.001. Determine la probabilidad de que, de 2000 personas,

a) Únicamente 2 personas se contagien.

b) Más de 2 personas se contagien.

Sí n = 2000, p = 0.001 y np = 2

a) P (x = 2) = $\frac{e^{-2}2^2}{2!}$ = 0.2706

b) $P(x > 2) = 1 - [P(x = 0) + P(x = 1) + P(x = 2)]$

$$= 1 - \left[\frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!}\right]$$

P(x > 2) = 1 - 0.6765 = 0.3235.

En caso, que la distribución de Poisson no se relacione con la binomial, n p = λ , luego la probabilidad de obtener X éxitos sería:

$$P(x) = \frac{e^{-\lambda}\lambda^x}{X!} \text{ con X = 0, 1, 2, 3…n, …}$$

y sus características son:

Media = esperanza matemática = $\mu = \lambda$ (promedio de éxitos)

Varianza =$\sigma^2 = \lambda$

Desviación estándar = $\sqrt{\lambda}$

Ejemplo 3

Una empresa recibe un promedio de 3 cheques sin fondos por día, la probabilidad de que reciba exactamente uno, un día determinado es:

Como λ = 3(promedio) se tiene:

$$\mathrm{P}(\mathrm{x} = 1) = \frac{e^{-3}3^1}{1!} = 0.1493.$$

Ejemplo 4

Si el número de accidentes por semana, variable aleatoria, en una intersección de mucho tráfico sigue una distribución de Poisson con λ = 4.8, la probabilidad que haya 3 accidentes exactamente en una semana dada es:

P (x = 3) = $\frac{e^{-4.8}4.8^3}{3!}$ = 0.1516

Con un promedio = $\mu = \lambda$ = 4.8

y su desviación estándar es $\sigma = \sqrt{\lambda}$ = 2.19

La probabilidad que haya al menos un accidente será:

P (x $\geq$ 1) = 1- $\frac{e^{-4.8}4.8^0}{0!}$ = 0.9917.

La distribución de **Poisson** es utilizada para el estudio de la probabilidad de ocurrencia en 2 tipos de eventos generales:

1) Al considerarse una gran **cantidad de un medio** tal como el agua del mar, el aire, la sangre o las pastas de hornear; dentro del cual se encuentra un número total de entidades pequeñas discretas, tales como plancton, bacterias, granulocitos o pasas; entonces si se examina una pequeña cantidad del medio, la probabilidad de que se encuentre x de las entidades en la porción examinada es de Poisson.

2) Lo referente a **sucesos en el tiempo**, éstos pueden ser: emisiones de partículas radioactivas, llamadas que entran a una central telefónica o pacientes que entran a una línea de espera; así la probabilidad de que ocurran x sucesos en un intervalo de tiempo prefijado es de Poisson.

3.3. DISTRIBUCIÓN NORMAL $N(X, \mu, \sigma^2)$

La información que sigue corresponde a la distribución poblacional de 188 estudiantes de maestría de una universidad peruana según sus notas en el ciclo de verano 2021.

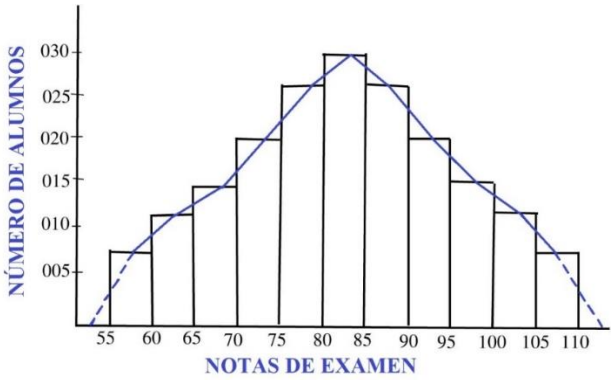
Tabla 3.1 Distribución poblacional de 188 estudiantes de maestría de una universidad peruana según sus notas en el ciclo de verano 2021.

Notas	Promedio Interválico (X_i)	Frecuencia n_i	Frecuencia relativa (h_i)
55-60	57.5	7	0.037
60-65	62.5	11	0.059
65-70	67.5	15	0.080
70-75	72.5	20	0.106
75-80	77.5	26	0.138
80-85	82.5	30	0.160
85-90	87.5	26	0.138
90-95	92.5	20	0.106
95-100	97.5	15	0.080
100-105	102.5	11	0.059
105-110	107.5	7	0.037
TOTAL	---	188	1.000

Con esta distribución se puede determinar la probabilidad de que, por ejemplo, un estudiante tenga un puntaje entre 60 y 65, lo que sería h_2 = 0.059. La probabilidad de que un estudiante tenga una nota promedio de 87.5 es 0.138 etc. El histograma y

polígonos de frecuencia hacen que pueda apreciarse más objetivamente la forma de esta distribución.

Figura 3.1. Distribución poblacional de 188 estudiantes de maestría de una universidad peruana según sus notas en el ciclo de verano 2021.



Gran número de variables tienden a distribuirse siguiendo la forma de este polígono (curva) de allí su nombre distribución normal; como las probabilidades son las frecuencias relativas, se puede asociar cada frecuencia relativa con el área de cada rectángulo; de modo que, si el número de observaciones es grande y la amplitud de los intervalos se reduce cada vez más, el polígono del histograma se transforma en una curva que cubre un área aproximada a la unidad.

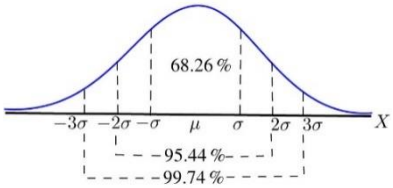
La distribución normal queda definida por:

i) Valores para la variable x en estudio: $-\infty < x < +\infty$.

ii) La función de probabilidad: $P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, tiene

las siguientes características:

Media	= Esperanza matemática = μ
Varianza	= σ^2
Desviación estándar	= σ,



Los intervalos μ±σ, μ±2σ y μ±3σ contienen el 68.26%, 95.44% y 99.74% de unidades de la población respectivamente.

Para determinar la probabilidad con la distribución normal debe estandarizarse cada valor experimental de la variable.

3.4. DISTRIBUCIÓN NORMAL ESTANDARIZADA N(Z,0,1)

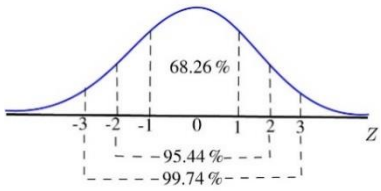
Todas las variables con distribución normal $N(X, \mu, \sigma^2)$ pueden transformarse en otra distribución normal más simple con $\mu=0$ y $\sigma^2=1$, esto es: $N(Z, 0, 1)$, a esta distribución se le llama **normal estandarizada** o normal reducida:

La distribución normal estandarizada queda definida en el recorrido de -3.9 < Z < +3.9 valores de la variable estandarizada, con una probabilidad $P(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, sabiendo que $Z = \frac{X - \mu}{\sigma}$, y tiene las siguientes características:

Media = $\mu = 0$

Varianza = $\sigma^2 = 1$

D. estándar = $\sigma = 1$

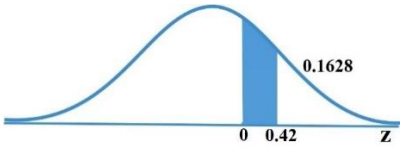


Los intervalos 0 ± 1, 0 ± 2 y 0 ± 3 contienen el 68.26%, 95.44% y 99.74% unidades de la población respectivamente. Para el cálculo de probabilidad (áreas) se usa la Tabla Z, la cual se presenta en anexos, su manejo se ilustra en los siguientes ejemplos.

Tabla Z

Z	0	1	2	3	4
0.0	0.0000	0.0040	0.0080	0.0120	0.0160
0.1	0.0398	0.0438	0.0478	0.0517	0.0557
0.2	0.0793	0.0832	0.0871	0.0910	0.0948
0.3	0.1179	0.1217	0.1255	0.1293	0.1331
0.4	0.1554	0.1591	0.1628	0.1664	0.1700

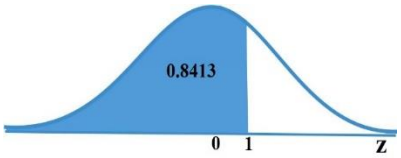
1) Probabilidad cuando Z = 0.42



P (Z = 0.42) = 0.1628

P (Z = -0.42) = P (Z = 0.42) = 0.1628, por ser simétricos.

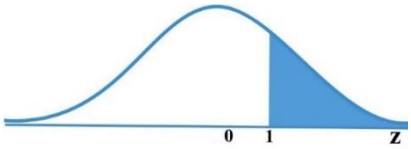
2) Probabilidad cuando Z ≤ 1



P (Z≤1) = 0.50 +P (Z = 1) = 0.5 + 0.3413 = 0.8413

P (Z≤1) = P (Z ≥ -1)

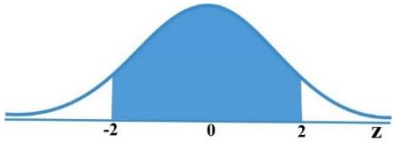
3) P (Z ≥ 1) = 0.5 – P(Z = 1) = 0.5 - 0.3413 = 0.1587



$P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - 0.8413$

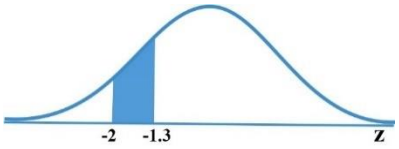
$P(Z \geq 1) = 0.1587$

4) $P(-2 \leq Z \leq 2) = 2P(Z = 2) = 2(0.4772) = 0.9544$



Propiedad: el área entre $\mu \pm 2\sigma$, ó $0 \pm 2 = 95.44\%$

5) $P(-2 \leq Z < -1.3) = P(1.3 < Z \leq 2)$



$P(Z = 2) - P(Z = 1.3) = 0.4772 - 0.4032 = 0.074$

Si los datos provienen de la medición directa, para calcular probabilidades es necesario estandarizarlos tal como apreciamos en el ejemplo siguiente.

Las notas de examen de los 188 estudiantes de maestría de una universidad peruana se distribuyen en forma normal con μ = 82.5 y σ = 12.5, Tabla 3.1.

$$\mu = \frac{\sum X_i n_i}{\mathrm{N}} = \frac{15510}{188} = 82.5$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2 n_i}{N} = \frac{N\sum X_i^2 n_i - (\sum X_i n_i)^2}{N^2}$$

$$\sigma^2 = \frac{188(1309175) - (15510)^2}{(188)^2} = 157.4468$$

σ = 12.55

Determinar la probabilidad de que:

a) Un estudiante elegido al azar tenga una nota entre 85 y 90.
b) Un estudiante tenga una nota mayor que 100.
c) Un estudiante tenga exactamente la nota de 70.
d) ¿Cuántos estudiantes tienen una nota entre 85 y 90?
e) ¿Cuántos tienen una nota mayor de 100?
f) La nota por debajo de la cual está el 75 % de la población.

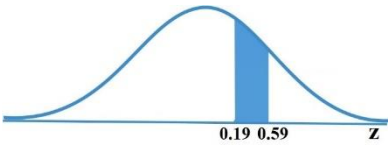
Solución

a) P (85≤ X ≤90) = probabilidad de que la nota del estudiante esté entre 85 y 90

Estandarizando las notas entre 85 y 90:

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{85 - 82.5}{12.55} = 0.19$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{90 - 82.5}{12.55} = 0.59$$

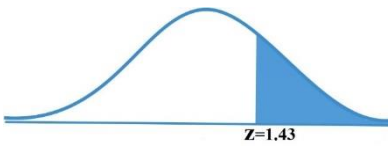


$$P(0.19 \leq Z \leq 0.59) = P(Z = 0.59) - P(Z=0.19)$$
$$= 0.2224 - 0.075$$
$$= 0.147$$

b) P (X>100) Los estudiantes con notas mayores que 100 deben tener al menos una nota de 100.5

Estandarizando 100.5:

$$Z = \frac{X - \mu}{\sigma} = \frac{100.5 - 82.5}{12.55} = 1.43$$



$$P(Z > 1.43) = 0.5 - P(Z = 1.43)$$
$$= 0.5 - 0.4236 = 0.0764$$

c) P (X = 70). Los alumnos con nota de exactamente 70, realmente deben tener un calificativo entre 69.5 y 70.5

P ($69.5 \leq X \leq 70.5$):

Estandarizando:

$$Z_1 = \frac{69.5 - 82.5}{12.55} = -1.03$$

$$Z_2 = \frac{70.5 - 82.5}{12.55} = -0.95$$

$$P(-1.03 \leq Z \leq -0.95) = P(Z = -1.03) - P(Z = -0.95)$$
$$= 0.3485 - 0.3289$$
$$= 0.0196$$

d) Conociendo la probabilidad P y la población N, los estudiantes con nota entre 85 y 90 está dado por:

NP = 188 (0.147) - 27.6 = 28 estudiantes.

e) Los estudiantes con nota mayor de 100 serán:

$$Nº = NP = 188\ (0.0764) = 14.3 = 15$$

f) El 75% = es área (probabilidad) con Z = 1.15 porque 0.75/2 = 0.375 = 0.3749 luego, la nota por debajo de la cual está el 75% de la población será:

$$1.15 = \frac{X - 82.5}{12.55}$$

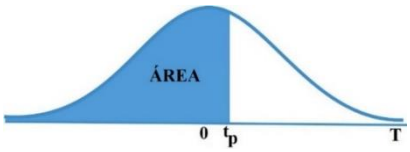
X = (12.55) (1.15) + 82.5

X = 96.93 puntos.

3.5. DISTRIBUCIÓN t student

Es una distribución de probabilidad de variable continua, así como la distribución normal $N(X, \mu, \sigma^2)$ o la distribución normal estandarizada $N(Z, 0, 1)$.

Se representa por $t_{n-1} = \frac{x-\mu}{\frac{s}{\sqrt{n}}}$ con *n-1* grados de libertad.



siendo sus características:

Media $= \mu = 0$

Varianza $= \sigma^2 = \frac{n}{n-2}$

Desviación estándar $= \sigma = \sqrt{\frac{n}{n-2}}$

Cuando n es grande ($n \geq 30$) la distribución t se aproxima a la distribución normal.

Para el cálculo de valores de t, debe conocerse el nivel de probabilidad y los grados de libertad (gl) que están dados por el tamaño de la muestra menos uno: $\nu = n - 1$.

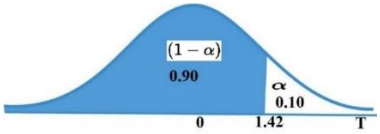
$$P\ (T_{n-1} \leq t_p)\ =\ 1-\alpha$$

1-α = nivel de probabilidad

Ejemplo 1

El valor de t, dado el 90% de probabilidad y 7 gl será:

$t_{0.90(7)}$ =1.42, entonces P ($T_7 \leq 1.42$) = 0.90



P ($T_7 \geq$-1.42) = 0.90

Por ser la curva simétrica.

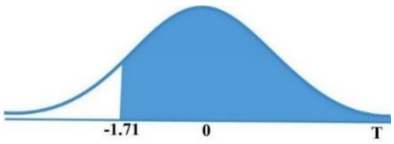
Tabla t

ν	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	t_{095}	$t_{0,90}$
1	63.66	31.82	12.71	06.31	03.08
2	09.92	06.96	04.30	02.92	01.89
3	05.84	04.54	03.18	02.35	01.64
4	04.60	03.75	02.78	02.13	01.53
5	04.03	03.36	02.57	02.02	01.48
6	03.71	03.14	02.45	01.94	01.44
7	03.50	03.00	02.36	01.90	01.42
8	03.36	02.90	02.31	01.86	01.40

ν = gl = grados de libertad.

Ejemplo 2

El valor de t con 95% de probabilidad y 25 gl, se obtendría de una lectura directa; gráfica y analíticamente se tendría del modo siguiente:



P ($T_{25} \geq -t_p$) = 0.95

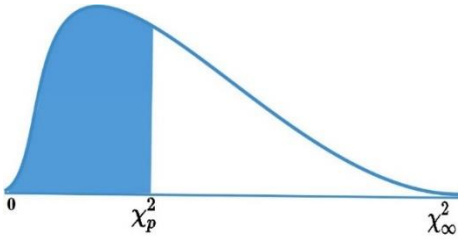
$t_{0.95(25)} = 1.71$

($T_{25} \geq$ -1.71) = 0.95

3.6. DISTRIBUCIÓN χ^2

Es una distribución de variable continua, los valores de χ^2 chi-cuadrado se determinan conociendo la probabilidad y los grados de libertad, se representa por la expresión:

$$\chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma^2}$$



siendo sus características:

Media = μ = n (grados de libertad)

Varianza = $\sigma^2 = 2n$

D. estándar = $\sigma\sqrt{2n}$

Valores de χ^2 se calculan conociendo la probabilidad y los grados de libertad.

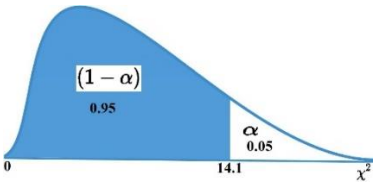
Ejemplo 1.

1-α = 0.95 = probabilidad (α = 0.05)

n = 8 = número de ítems experimentales, muestra.

n - 1 = 7 grados de libertad.

luego el valor de $\chi^2_{0.95(7)}$ =14.1



$P(\chi^2_7 \leq \chi^2_{\mathrm{p}}) = 0.95$ $P(\chi^2_7 \leq 14.1) = 0.95$

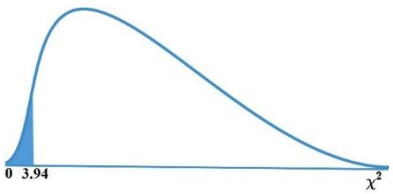
Tabla χ^2

ν	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$
1	07.88	06.63	05.20	03.84	02.71
2	10.60	09.21	07.38	05.99	04.61
3	12.80	11.30	09.35	07.81	06.25
4	14.90	13.30	11.10	09.49	07.78
5	16.70	15.10	12.80	11.10	09.24
6	18.50	16.80	14.40	12.60	10.60
7	20.30	18.50	16.00	14.10	12.00
8	22.00	20.10	17.50	15.50	13.40

$\nu = n - 1$ = gl = grados de libertad.

Ejemplo 2.

Muestra = 11, probabilidad = 0.05



$P(\chi^2_{10} \leq \chi^2_p) = 0.05$ $\chi^2_{0.05(10)} = 3.94$

$P(\chi^2_{10} \leq 3.94) = 0.05$, por complementación

$P(\chi^2_{10} \geq 3.94) = 0.95$

Esperanza = promedio =μ= n =10

Varianza = σ^2 = 2n =20

D. estándar = $\sigma\sqrt{20} = 4.47$

3.7. DISTRIBUCIÓN F

Es una distribución de variable continua, asimétrica, sesgada a la derecha, se usa para estudiar la razón de las varianzas de dos poblaciones distribuidas normalmente a partir de sus muestras y también en el análisis de varianza de un factor (ANOVA).

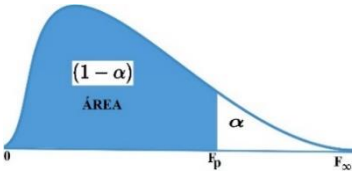
Para los valores de F es necesario la confiabilidad y los grados de libertad del numerador y denominador. Una muestra de tamaño *n* tendrá *n-1* grados de libertad.

La distribución F se representa por la expresión:

$$F_{n_1-1,n_2-1} = \frac{\frac{n_1 S_1^2}{n_1 - 1}}{\frac{n_2 S_2^2}{n_2 - 1}}$$

donde n_1 y n_2 son los tamaños de muestras.

Observe que $F_{n_1-1,n_2-1} \neq F_{n_2-1,n_1-1}$, además se tiene que la probabilidad $P\left(F_{n_1-1,n_2-1} \leq F_p\right) = 1 - \alpha$, siendo $1-\alpha$ el grado de confiabilidad, su gráfica es:



Ejemplo 1

Considere $n_1 = 5,\ n_2 = 6,\ 1 - \alpha = 0.95$

$F_p = F_{0.95,\,4,\,5} = 5.19$

$P\,(F_{4,\,5} \leq 5.19) = 0.95$

Ejemplo 2

Si $n_1 = 6$ y $n_2 = 5$,

$1 - \alpha = 0.95$

$F_{0.95,\,5,\,4} = 6.26$

$P\,(F_{5,4} \leq 6.26) = 0.95$.

Valores de $F_{0.01}$

n_1-1 / n_2-1	numerador						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1	40.52	50.00	54.03	56.25	57.64	58.59	59.28
2	98.50	99.00	99.20	99.20	99.30	99.30	99.40
3	34.10	30.80	29.50	28.70	28.20	27.90	27.70
4	21.20	18.00	16.70	16.00	15.50	15.20	15.00
5	16.30	13.30	12.10	11.40	11.00	10.70	10.50
6	10.90	10.90	09.78	09.15	08.75	08.47	08.26
7	12.20	09.55	08.45	07.85	07.48	07.19	06.99
8	11.30	08.65	07.59	07.01	06.63	06.37	06.18

Para probabilidades pequeñas se usa la siguiente fórmula:

$$F_{\alpha,n_1-1,n_2-1} = \frac{1}{F_{1-\alpha,n_2-1,n_1-1}}$$

Ejemplo 3

Si se tiene:

$P\left(F_{4,5} \leq F_p\right) = 0.05$

entonces $F_{0.05,4,5} = \frac{1}{F_{0.95,5,4}} = \frac{1}{6.26} = 0.16$

3.8. DISTRIBUCIÓN DE LOS ESTIMADORES $\bar{X}, S^2, p$

Dada una población, se pueden extraer muchas muestras y para cada una se puede calcular su promedio y desviación estándar respectiva Según el **teorema del límite central**, todas las medias muestrales se distribuyen alrededor de la media de la población así **la media de todas las medias es la media poblacional**.

Si las muestras tienen más de 30 datos, las medias se distribuyen alrededor de la media poblacional en forma gaussiana, el promedio de los promedios es el promedio de la población y la desviación estándar de los promedios es el error estándar de la media.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

3.8.1 Distribución de Medias Muestrales $\bar{X}$

Para ilustrar la distribución de las medias muestrales se usará una población pequeña.

Consideremos el número de horas de estudio de una población de 5 alumnos de psiquiatría.

Alumnos	Nº de horas
A	2
B	3
C	4
D	5
E	6
TOTAL	**20**

Para la población de 5 datos sus parámetros serán:

$$\mu = \frac{\sum X_i}{N} = \frac{20}{5} = 4$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$$

$$\sigma^2 = \frac{N\sum X_i^{\,2} - \left(\sum X_i\right)^2}{N^2}$$

$$\sigma^2 = \frac{5(90) - (20)^2}{5^2} = 2$$

$$\sigma = 1.41$$

Tomamos todas las muestras de tamaño 2 (sin repetición), el número de muestras será $= \binom{n}{x} = \binom{5}{2} = \frac{5!}{(5-2)!5!} = 10$ y las muestras son: (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6).

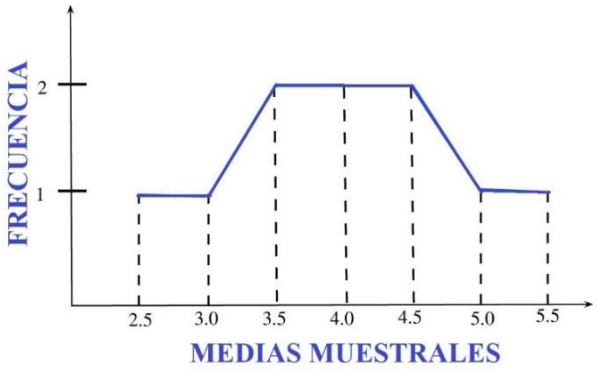
El promedio de cada muestra es:

2.5, 3.0, 3.5, 4.0, 3.5, 4.0, 4.5, 4.5, 5.0, 5.5.

Tabla 3.2. Distribución de las medias muestrales

Medias Muestrales (X_i)	Frecuencia n_i	Probabilidad (h_i)
2.5	1	0.1
3.0	1	0.1
3.5	2	0.2
4.0	2	0.2
4.5	2	0.2
5.0	1	0.1
5.5	1	0.1
Total	10	1.0

Figura 3.2 Distribución de 10 medias muestrales



Los estimadores son $\bar{\bar{X}}$:

$$\bar{\bar{X}} = \frac{\sum \bar{X}_i n_i}{n} = \frac{40}{10} = 4 \qquad E(\bar{x}) = \mu = 4$$

$$\sigma^2_{\bar{X}} = \frac{\sum(\bar{X}_i - \bar{\bar{X}})n_i}{n} = \frac{n\sum \bar{X}_i^2 n_i - (\sum \bar{X}_i n_i)^2}{n^2}$$

$$\sigma^2_{\bar{X}} = \frac{10(167.5) - (40)^2}{100} = 0.75$$

$$\sigma^2_{\bar{X}} = 0.866$$

En la práctica:

$$\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{X}} = \sqrt{\frac{2}{2}} \cdot \sqrt{\frac{5-2}{5-1}} = \sqrt{\frac{3}{4}} = 0.866$$

$\sqrt{\frac{N-n}{N-1}}$ factor de corrección para poblaciones finitas, acá N es 5, la muestra de tamaño 2.

Se considera que una población es infinita si es mayor o igual a 100,000; se aplica entonces $\sqrt{\frac{N-n}{N-1}}$, solo si N< 100,000 y n ≥ al 5% de N, este factor se omite si N es infinita, o las muestras se extraen con repetición; en tales casos la media muestral se distribuye como

$$N(\bar{x}, \mu, \frac{\sigma}{\sqrt{n}}).$$

Obsérvese la forma como se distribuye las medias muestrales, el polígono se suaviza, aproximándose a la distribución normal cuando crece N.

3.8.2 Distribución de la Varianza Muestral S^2

Las muestras:

(2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6)

Las medias:

2.5, 3.0, 3.5, 4.0, 3.5, 4.0, 4.5, 4.5, 5.0, 5.5.

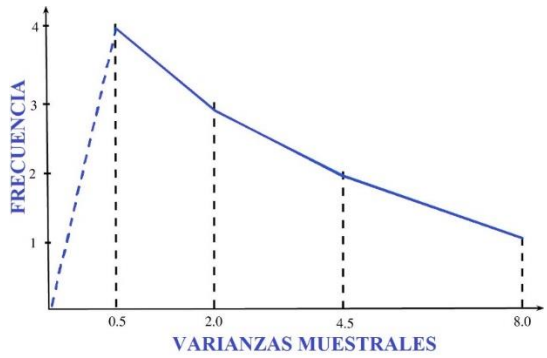
Las varianzas:

0.5, 2.0, 4.5, 8.0, 0.5, 2.0, 4.5, 0.5, 2.0, 0.5.

Tabla 3.3 Distribución de las varianzas muestrales

Varianzas Muestrales (S^2)	Frecuencia n_i	Probabilidad (h_i)
0.5	4	0.4
2.0	3	0.3
4.5	2	0.2
8.0	1	0.1
Total	10	1.0

Figura 3.3. Distribución de 10 varianzas muestrales



3.8.3 Distribución de la proporción muestral p

Con frecuencia se desea conocer, por ejemplo, la proporción de aprobados en una universidad, la proporción de votantes en un referéndum, etc. La relación entre la proporción poblacional y la distribución de la proporción muestral se ilustra pedagógicamente con un ejemplo muy pequeño.

En un curso taller de Bioestadística hay 2 alumnos fumadores A, B, y 2 no fumadores C y D.

Consideramos los 4 alumnos como una población y representemos con 1 al fumador y con 0 al no fumador.

Tabla 3.4 Datos y cálculos auxiliares para obtener la Proporción y desviación estándar poblaciones

Alumnos	X	**(X – P)**	**$(X - P)^2$**
A	1	0.5	0.25
B	1	0.5	0.25
C	0	-0.5	0.25
D	0	-0.5	0.25
4	2		1.00

Parámetros

Proporción de fundadores = P = $\frac{\sum X}{N} = \frac{2}{4} = 0.5 = 50\%$

Proporción de no fumadores = Q = 1 – P = 0.5 = 50%

$$\sigma_p^2 = \frac{\sum(X-P)}{N} = \frac{N\sum X^2 - (\sum X)^2}{N^2} = \frac{4(2) - (2)^2}{4^2} = 0.25$$

$$\sigma_p = \sqrt{0.25} = 0.5$$

equivalente $\mu_P = P = 0.5$; $\sigma_P = \sqrt{PQ} = 0.5$

Consideremos todas las muestras de tamaño 2 (sin repetición).

nº de muestras = $\binom{n}{x} = \binom{4}{2} = 6$

AB, AC, AD, BC, BD, CD

La proporción de fumadores en cada muestra

$$\frac{2}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{0}{2}$$

Tabla 3.5 Distribución de las proporciones muestrales

Proporciones muestrales (P_i)	Frecuencia n_i	Probabilidad (h_i)
$\frac{0}{2} = 0.0$	1	0.167
$\frac{1}{2} = 0.5$	4	0.667
$\frac{2}{2} = 1.0$	1	0.167
Total	6	0.998

Estimadores.

$$\bar{p} = \frac{\sum p_i n_i}{n} = \frac{3}{6} = \frac{1}{2} = 0.5$$

Media proporcional muestral con $E(\bar{p}) = P$

$$\sigma_p^2 = \frac{\sum(p_i - \bar{p})^2 n_i}{n} = \frac{n\sum p_i^2 n_i - (\sum p_i n_i)^2}{n^2}$$

$$\sigma_p^2 = \frac{6(2) - (3)^2}{(6)^2} = 0.0833$$

$$\sigma_p = 0.28$$

En la práctica se utiliza:

$$\sigma_p = \sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_p = \sqrt{\frac{(0.5)(0.5)}{2}}\sqrt{\frac{4-2}{4-1}} = 0.28$$

$$\sigma_p = \sqrt{\frac{PQ}{n}}$$

Si la población es infinita o muy grande

La variable estandarizada es $=Z = \frac{(\mathrm{p}-\mathrm{P})}{\sigma p} \rightarrow N(Z, 0{,}1)$ se utiliza como aproximación de la distribución de la proporción muestral.

Las distribuciones muestrales son indispensables para el análisis de los parámetros; su principal aplicación está en la contrastación de Hipótesis.

AUTOEVALUACIÓN III

1) Calcule las medidas de centralización y dispersión que se indican para la siguiente distribución de frecuencias.

X	0	1	2	3	4
n	4	3	2	1	0

$\bar{x}$ = …… Md =..... S =…..... C.V. =…….

2) 20 observaciones se distribuyen simétricamente como se observa.

Intervalos	**Frecuencia.**
8 - 10	2
10 – 12	5
12 – 14	6
14 – 16	5
16 - 18	2

Calcule el promedio, la mediana y la moda y grafique el polígono de frecuencias absolutas y acumuladas.

3) ¿Suponga que el curso se dicta en dos grupos A y B.-En base a las estadísticas que se presentan, ¿cuál es el grupo más eficiente?, por qué?:

$\bar{X}_A = 13$ $\bar{X}_B = 12.5$
$S_A = 4$ $S_B = 3.5$
$M_d = 13$ $M_d = 12$

4) Un tratamiento de cierta enfermedad tiene 0.7 de probabilidad de aliviar todos los signos y síntomas, ¿cuál es la probabilidad de que 10 pacientes tratados mejore más de uno?

   A) 0.9998564
   B) 0.0001436
   C) 0.0000059

5) La presión sistólica en reposo de cierta población pediátrica tiene:

   μ=115 m Hg y σ^2 = 225 (mm Hg)^2, ¿cuál es la probabilidad que un niño tenga una presión sistólica menor a 140 mm Hg?

   A) 0.0516
   B) 0.9525
   C) 0.4484

6) De un total de 1 000 familias con dos hijos, ¿en cuántos cabe esperar que tenga al menos un niño?

   A) 750
   B) 250

C) 125

7) La proporción de complicaciones operatorias en cierto procedimiento de reconstrucción vascular es del 20%. En 50 casos semejantes, ¿cuál es la probabilidad de que haya a lo más 5 pacientes con complicaciones operatorias?

A) 0.0295

B) 0.0559

C) 0.0854

8) Las pruebas estandarizadas para medir C.I. tienen $\mu = 100$ y $\sigma = 16$. Si consideramos 1 000 alumnos en nuestra universidad, ¿cuántas tendrán $\mu+2\sigma$?

A) 95.45%

B) 478

C) 955

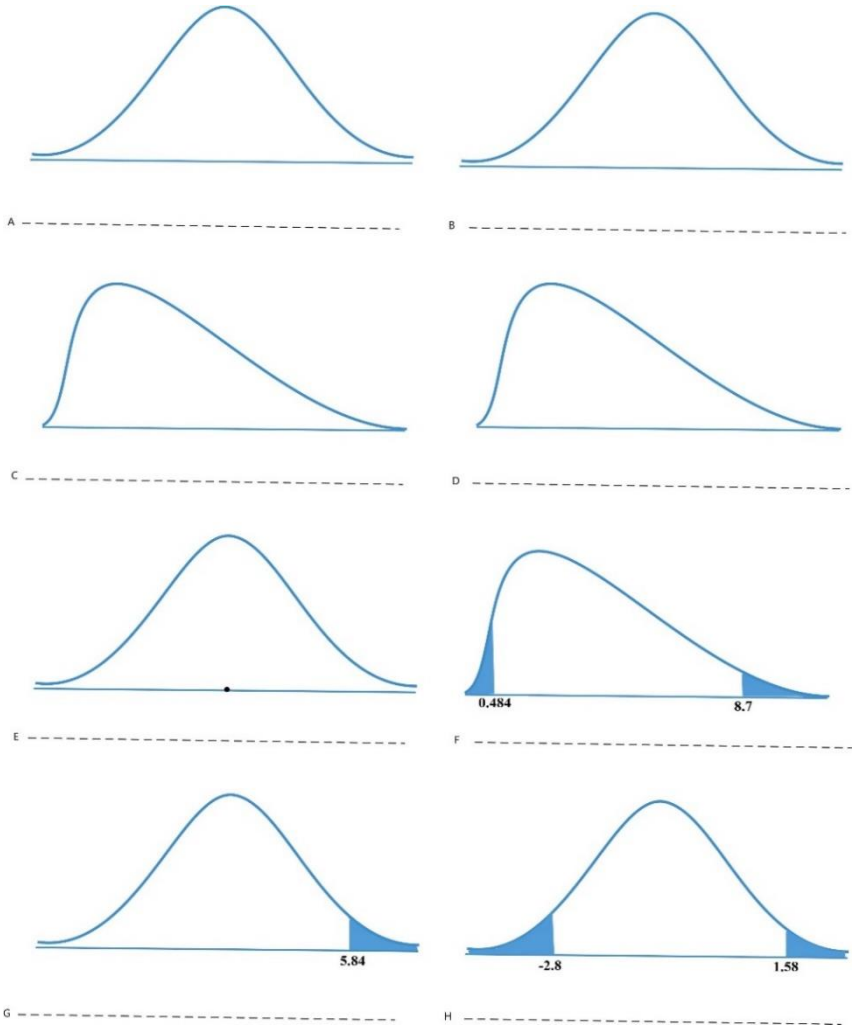
9) Con los datos que cuenta determine los valores que faltan según las tablas correspondientes para los siguientes casos:

Tabla z	
$Z_1 = 1.2,$	$Z_2 = -1.2$
$P (Z_1 + Z_2)$	……
P = 0.9932	Z =…….
P = -2.85	P =…….
$P (-1 \leq Z \leq 2)$	P =……
$P (Z > -3)$	= ……
Tabla t	
P = 0.975 , 23 gl	t =……..
P = 0.995 , 20 gl	t =……..
t = 1.76 , 14 gl	P =……..
t =1.81 , 10 gl	P =……..
Tabla χ^2	
P = 0.99 , 27 gl	X^2 =……..
P = 0.975 , 8 gl	X^2 = ……..
X^2 = 16.7 , 5 gl	P = ……...
X^2 = 43 , P = 0.99	gl ……...

Tabla F ($\alpha = 0.05$)		
gl numerador 24	gl denominador 17	F = ……...
F = 4.84	gl numerador …	gl. denominador …

Tabla F ($\alpha = 0.01$)		
F = 4.16	gl numerador ……	gl denominador ……
gl numerador ….	gl denominador = 12	F……

10) Nombre las distribuciones:



11) Si los elementos A, B, C, D, E, constituyen una población, ¿cuántos y cuáles son las muestras sin sustitución de tamaño 3?.

……..

12) Para población finita $\sigma_{\bar{X}}$ = …….

Para población infinita $\sigma_{\bar{X}}$ = …….

13) De una población infinita se extrae una muestra de tamaño 5 con P = 0.6, calcule:

σ_p =……E (p) =…….

14) En una distribución donde n = 100, $\sigma_p^2 = 0.0025$, la probabilidad es: ……

15) Para la distribución poblacional

Alumnos	Nº de Horas
A	3
B	4
C	5
D	6
E	7
F	8

μ = ……. σ = …….

Tome todas las muestras de tamaño 2 y calcule la distribución de las medias muestrales.

$\bar{\bar{X}}$= …… $\sigma_{\bar{X}}$ = ……..

Grafique la distribución de la población, de las medias muestrales y las varianzas.

CAPÍTULO IV

ESTIMACIÓN INTERVÁLICA DE PARÁMETROS

La estimación interválica consiste en determinar el rango de valores donde es probable que se encuentre el parámetro deseado.

Si se conocen la media y la varianza de la distribución de una variable se puede determinar las **probabilidades** de varios sucesos; pero como los *parámetros* de las variables son atributos de la población original, generalmente son desconocidos y el problema es entonces averiguar los valores más cercanos acerca de los mismos, partiendo de una muestra aleatoria, es decir, hay que **estimar** su valor

numérico dentro del cual se espera que esté el parámetro. (Vásquez, 1990, p. 83)

4.1. ESTIMACIÓN PUNTUAL

Para la estimación puntual utilizamos

> la información de una sola muestra seleccionada aleatoriamente, se trata de determinar un solo valor numérico que sea buen indicador del verdadero valor del parámetro; así si se desea la estatura media de los estudiantes de la Universidad Nacional "Pedro Ruiz Gallo", el promedio de una muestra representativa y adecuada será un buen **estimador.** (Vásquez, 1990, p. 83)

Ejemplo 1

El tamaño en centímetros de 6 yemas de espárragos de una parcela experimental es como sigue: 1.85, 2.71, 2.91, 2.70, 2.64, 3.25

Estimación puntual del promedio:

$$\bar{X} = \frac{\sum X_i}{n} = 2.68\text{cm.} \qquad (\ \bar{X}_{\bar{X}} = \mu\ )$$

Estimación puntual de la desviación estándar:

$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} = 0.46\text{cm.} \qquad (S_{\bar{X}} = \frac{\sigma}{\sqrt{n}})$$

4.2. ESTIMACIÓN INTERVÁLICA

La inferencia estadística es probabilística,

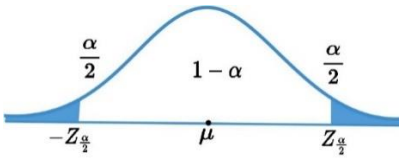
> partiendo de una muestra es necesario admitir alguna noción de variabilidad, lo que está sujeto al muestreo y al grado de confiabilidad para el intervalo en el cual se espera se encuentre el parámetro de la población que se investiga. Esto es lo que se conoce como estimación interválica. (Vásquez, 1990, p. 84)

4.2.1. Intervalo de confianza para μ, varianza conocida

Se representa por:

$$P(\bar{X} - Z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z\frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Siendo Z el valor del grado de confiabilidad $(1-\alpha)$ establecido por distribución normal estandarizada. (Tabla Z).



Ejemplo 1

El promedio del cociente intelectual (C.I.) es de 89 en una muestra de 81 alumnos de una institución educativa. Se desea saber la media de C.I. en el centro educativo con 95% y 99% de probabilidad. Se conoce que el C.I. se distribuye normalmente con σ =16.

Solución:

n = 81 $\quad$ 1 – α = 95% = 0.95

Z, para el 95% = 1.96

$\overline{X}$ = 89 $\quad$ σ = 16

Sustituyendo en el intervalo confidencial

$$P\left(\bar{X} - Z\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(89 - 1.96\frac{16}{\sqrt{81}} < \mu < 89 + 1.96\frac{16}{\sqrt{81}}\right) = 0.95$$

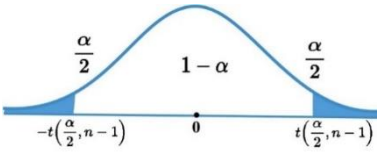
$$P(85.25 < \mu < 92.48) = 0.95$$

$$P\left(89 - 2.58\frac{16}{\sqrt{81}} < \mu < 89 + 2.58\frac{16}{\sqrt{81}}\right) = 0.99$$

La media del C.I. en el centro educativo con 95% de probabilidad se encuentra entre 92.48 y 85.52. así mismo, al 99% de seguridad el intervalo sería [84.41, 93.59] (Z = 2.58).

4.2.2 Intervalo de confianza para μ, varianza desconocida.

$$P\left(\bar{X} - t\frac{S}{\sqrt{n}} < \mu < \bar{X} + t\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$



Los valores de t dependen de los grados de libertad (n-1 elementos de la muestra) y confiabilidad (1 – α). (Tabla t).

Ejemplo 1

Se ha medido el pH en 10 recién nacidos, de mujeres normales, obteniéndose $\bar{X}$=7.35 y S = 0.21. Halle un intervalo de confianza del 95% para el pH medio poblacional.

Solución:

n = 10 $\qquad\qquad$ 1 – α = 95% = 0.95

$\bar{X}$ =7.35

S = 0.21 $t_{0.975(9)} = 2.26$

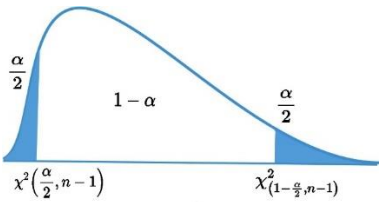
$$P\left(\bar{X} - t\frac{S}{\sqrt{n}} < \mu < \bar{X} + t\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$
$$P\left(7.35 - 2.26\left(\frac{0.21}{\sqrt{10}}\right) < \mu < 7.35 + 2.26\left(\frac{0.21}{\sqrt{10}}\right)\right) = 0.95$$
$$P[7.2 < \mu < 7.5] = 0.95$$

Con el 95% de seguridad la media poblacional del pH de los recién nacidos estará entre 7.2 y 7.5.

4.2.3 Intervalo de confianza para σ, distribución normal

Se representa por

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{(1-\frac{\alpha}{2},n-1)}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{(\frac{\alpha}{2},n-1)}}}\right) = 1 - \alpha$$



Los valores de χ^2 dependen de los grados de libertad y nivel de probabilidad (Tabla χ^2).

Ejemplo 1

Para estudiar el contenido del calcio de una población fósil del gasterópodo sphaeronassa mutabilis **se** toma muestra de 10 especímenes:

36.39, 36.19, 34.20, 35.15, 35.47, 35.22, 36.11, 35.63, 36.63, 35.59.

donde:

n = 10

n - 1 = 9 gl

$S^2 = 0.5094$

S = 0.7137

Para 95% de probabilidad
$\chi^2_{(\frac{\alpha}{2},n-1)} = \chi^2_{(0.025,9)} = 2.70$
$\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)} = \chi^2_{(0.975,9)} = 19.02$

Calcular el intervalo confidencial para la desviación estándar al 95%.

Solución

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi^2\left(1-\frac{\alpha}{2},n-1\right)}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2\left(\frac{\alpha}{2},n-1\right)}}\right) = 1-\alpha$$

$$P\left(\sqrt{\frac{(9)(0.5094)}{19.02}} < \sigma < \sqrt{\frac{(9)(0.5094)}{2.70}}\right) = 0.95$$

$$P(0.491 < \sigma < 1.303) = 0.95$$

Con un 95% de seguridad se puede afirmar que la desviación estándar del contenido porcentual de calcio de la población fósil estará entre 0.49 y 1.30.

4.2.4. Intervalo de confianza para P, parámetro de la binomial

Para tamaños de muestras razonables (n ≤ 30) existe un intervalo basado en la aproximación a la normal.

Se representa por:

$$P\left(p - Z\sqrt{\frac{pq}{n}} < P < p + Z\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

La frecuencia relativa de casos favorables en n pruebas independientes es $\frac{x}{n}$ que se toma p como estimador de P. Los valores de Z son de la normal estandarizada.

Ejemplo 1

Se inocula a una nidada experimental de 200 pollos con un organismo sospechoso de causar un conjunto de signos clínicos. Después de 14 días de inoculación, 137 de las aves manifiesta los signos característicos; suponiendo que no haya problemas de infección, hallar el intervalo confidencial de P al 95% y 99%.

Solución

$n = 200$ $\qquad p = \frac{x}{n} = \frac{137}{200} = 0.685$

(0.685 = probabilidad de que un pollo inoculado muestre los signos).

Z = 1.96 para el 95% y para el 99%, Z = 2.58.

Sustituyendo en:

$$P\left(p - Z\sqrt{\frac{pq}{n}} < P < p + Z\sqrt{\frac{pq}{n}}\right) = 1 - \alpha$$

$$P\left(0.685 - 1.96\sqrt{\frac{(0.685)(0.315)}{200}} < P < 0.685 + 1.96\sqrt{\frac{(0.685)(0.315)}{200}}\right) = 0.95$$

$$P(0.621 < P < 0.749) = 0.95$$

Con 95% de seguridad la verdadera proporción estará entre el 62.1% y 74.9%; para el 99% de confiabilidad la proporción estará entre 60% y 77%.

4.2.5. Intervalo de confianza para μ_1-μ_2 de dos distribuciones normales, de varianzas conocidas.

Está dado por:

$$P\left((\bar{x}_1 - \bar{x}_2) - Z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

con n_1 y n_2, muestras independientes y sus respectivos promedios $\bar{x}_1$ y $\bar{x}_2$; μ_1 y μ_2 promedios de las poblaciones con varianzas σ_1^2 y σ_2^2 respectivamente.

Ejemplo 1

Supóngase que se puede admitir que los niveles de hemoglobina en varones blancos de más de 11 años tengan distribución normal con varianza 1.642 y los de 11 años o menos con 0.867. Si en una muestra de 10 varones mayores de 11 años y en 20 menores de 11 años se tiene como medias 14.47 mg/100ml. y 12.64 mg/100ml. Respectivamente. Calcule el intervalo para $\mu_1 - \mu_2$ al 95% de confiabilidad.

Solución

$n_1 = 10$ $\quad n_2 = 10$ $\quad 1-\alpha = 95\%$ $\quad Z = 1.96$

$\bar{x}_1 = 14.47$ mg/100ml. $\quad \sigma_1^2 = 1.642(\text{mg/100ml.})^2$

$\bar{x}_2 = 12.64$ mg/100ml. $\quad \sigma_2^2 = 0.867(\text{mg/100ml.})^2$

Sustituyendo:

$$P\left((14.47-12.64)-1.96\sqrt{\frac{1.462}{10}+\frac{0.867}{20}} < \mu_1-\mu_2 < (14.47-12.64)+1.96\sqrt{\frac{1.462}{10}+\frac{0.867}{20}}\right) = 0.95$$

$$P(0.98 < \mu_1 - \mu_2 < 2.68) = 0.95$$

Cero no pertenece al intervalo; entonces, las poblaciones difieren probablemente en nivel medio de hemoglobina y difieren como mínimo en 0.98 mg/100ml. o como máximo en 2.68 mg/100ml.

4.2.6. Intervalo de confianza para μ_1-μ_2 con varianzas desconocidas pero iguales.

Se representa por:

$$P\left((\bar{x}_1 - \bar{x}_2) - tS_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + tS_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1 - \alpha$$

dada la confiabilidad $1-\alpha$, el valor de t se determina con $n_1 + n_2 - 2$ grados de libertad. S_p^2 es una varianza ponderada.

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}}$$

$$\left(\sigma_1^2 = \sigma_2^2 = \sigma_p^2 = S_p\right)$$

Ejemplo 1

Las notas de 13 estudiantes de pesquería y 12 de microbiología en bioestadística son como sigue:

Pesquería:	10, 15, 16, 08, 13, 18, 12, 14, 12, 11, 14, 13, 12,
Microbiología:	12, 11, 14, 08, 10, 16, 12, 07, 10, 15, 11, 10.

Calcule un intervalo al 95% para la $\mu_1 - \mu_2$.

Solución:

$n_1 = 13$ $\quad \bar{x}_1 = 12.92$ $\quad S_1^2 = 6.74$

$n_2 = 12$ $\quad \bar{x}_2 = 11.33$ $\quad S_2^2 = 7.15$

$t_{\frac{\alpha}{2}} = t_{1-\frac{\alpha}{2}}$ $\quad t_{(1-\frac{\alpha}{2}, n_1+n_2-2)} = t_{(0.975,22)} = 2.07$

$S_p = 2.63$

Sustituyendo en la fórmula, se tiene:

$$P\left(12.92 - 11.33 - (2.07)2.63\sqrt{\frac{1}{13}+\frac{1}{12}} < \mu_1 - \mu_2 < 12.92 - 11.33 + (2.07)2.63\sqrt{\frac{1}{13}+\frac{1}{12}}\right) = 0.95$$

$$P(-0.59 < \mu_1 - \mu_2 < 3.77) = 0.95$$

Como el intervalo contiene el "cero", entonces podría darse que: $\mu_1 = \mu_2$.

4.2.7. Intervalo de confianza para μ_1-μ_2 con varianzas desiguales y desconocidas

Se representa por:

$$P\left((\bar{x}_1 - \bar{x}_2) - t\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = 1 - \alpha$$

El valor de t se toma con "f" grados de libertad

$$\boldsymbol{f} = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 + 1} + \frac{\left[\frac{S_2^2}{n_2}\right]}{n_2 + 1}} - \mathbf{2}$$

Ejemplo 1

La diuresis, en lt/día, de 17 personas sometidas a un fármaco activo, al cabo de dos semanas, fue:

1.7, 1.7, 1.8, 1.7, 1.9, 1.9, 1.7, 1.6, 1.6, 1.9, 1.8, 1.5, 1.8, 2.0,1.7,2.0 y 1.6.

La diuresis en otras 17 personas sometidas a un placebo fue:

1.5, 1.4, 1.5, 1.3, 1.3, 1.7, 1.3, 1.5, 1.6, 1.7, 1.5, 1.2, 1.2, 1.5, 1.4, 1.3, 1.3.

Construya un intervalo del 95% para la $\mu_1 - \mu_2$.

Solución

$$n_1 = 17 \qquad \bar{x}_1 = 1.76 \qquad S_1^2 = 0.0213$$

$$n_2 = 17 \qquad \bar{x}_2 = 1.42 \qquad S_2^2 = 0.0244$$

$$\boldsymbol{f} = \frac{\left[\frac{0.0213}{17} + \frac{0.0244}{17}\right]^2}{\frac{\left[\frac{0.0213}{17}\right]^2}{18} + \frac{\left[\frac{0.0244}{17}\right]^2}{18}} - \mathbf{2} = \mathbf{34}\ \boldsymbol{gl}.$$

Como en la tabla no hay 34 gl, se interpola por promedios.

(0.975, 28) = 2.05

(0.975, 40) = 2.02

(0.975, 34) = 2.035

Sustituyendo:

$$P\left((1.76 - 1.42) - 2.035\sqrt{\frac{0.0213}{17} + \frac{0.0244}{17}} < \mu_1 - \mu_2 < (1.76 - 1.42) + 2.035\sqrt{\frac{0.0213}{17} + \frac{0.0244}{17}}\right) = 0.95$$

Se obtiene: $P[0.23 < \mu_1 - \mu_2 < 0.45] = 0.95$

4.2.8. Intervalo de confianza para $P_1 - P_2$, parámetros de dos distribuciones binominales

Se representa por:

$$P\left((p_1 - p_2) - Z\sqrt{\frac{p_1 n_1}{n_1} + \frac{p_2 n_2}{n_2}} < P_1 - P_2 < (p_1 - p_2) + Z\sqrt{\frac{p_1 n_1}{n_1} + \frac{p_2 n_2}{n_2}}\right) = 1 - \alpha$$

Ejemplo 1

Dos muestras de 200 y 150 pollos son inoculadas con organismos aislados de diferente enfermedad, después de 14 días de inoculación 137 de la primera y 98 de la segunda nidada, se encuentran clínicamente enfermas.

Hallar un intervalo del 95% para $\boldsymbol{P_1 - P_2}$, proporciones de aves enfermas dentro de los 14 días de inoculación por las dos soluciones de organismos.

Solución:

$$n_1 = 200 p_1 = \frac{137}{200} = 0.685 \quad q_1 = \frac{63}{200} = 0.315$$

$$n_2 = 150 p_2 = \frac{98}{150} = 0.653 \qquad q_2 = \frac{52}{150} = 0.3$$

Sustituyendo

$$P\left((0.685 - 0.653) - 1.96\sqrt{\frac{(0.685)(0.315)}{200} + \frac{(0.653)(0.347)}{150}} < P_1 - P_2 < (0.685 - 0.653) - 1.96\sqrt{\frac{(0.685)(0.315)}{200} + \frac{(0.653)(0.347)}{150}}\right) = 0.95$$

Se obtiene $P = (-0.068 < P_1 - P_2 < 0.132) = 0.95$

Como el cero queda en el intervalo, no se puede concluir que existe diferencia entre las proporciones de aves enfermas, tal como lo mide la incidencia de signos clínicos en 14 días.

AUTOEVALUACIÓN IV

1. Analice e interprete el intervalo del parámetro μ

$$P\left(\bar{x} - \frac{Z\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{Z\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

……

……

2. El consumo en gramos de un aditivo en la ración alimenticia de 15 palomas es como sigue:

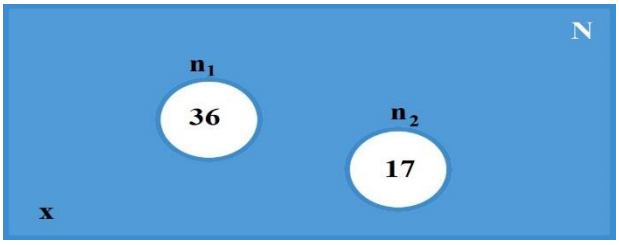
24.0, 24.1, 24.5, 23.2, 21.2, 24.0, 24.1, 22.2, 21.1, 23.7, 23.2, 26.0, 22.8, 24.4, 22.9.

a) Determine el intervalo confidencial del 95% para el consumo del aditivo.

b) ¿Cuál es el intervalo del 99% para σ?

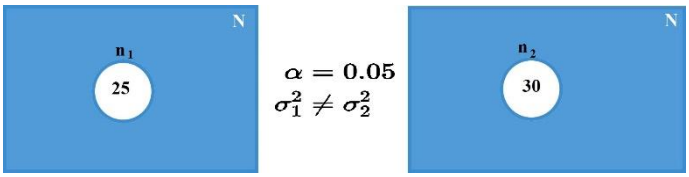
3. Con los elementos que tiene la figura, elabore intervalos para los parámetros.

a)



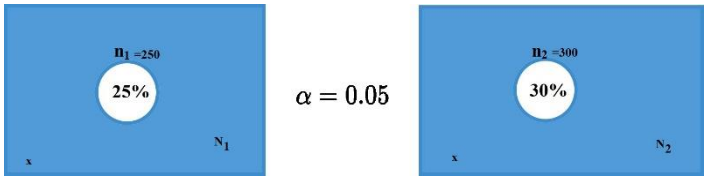
……

b)



……

c)



……

4. Un agrónomo afirma que el abono preparado con cáscara de plátano influye en el crecimiento del maíz en los primeros 15 días de cultivo, para tal investigación se preparan dos parcelas. En la parcela A se aplica el abono, la parcela B sirve de testigo; la altura de 15 plantaciones de maíz es la siguiente:

   A: 37,47,35,40,38,39,45,40,45,43,37,42,41,43,47

   B: 30,40,30,35,35,33,38,35,35,28,22,31,33,27,37

   a) Estime e interprete un intervalo del 95% para la μ_1-μ_2.

   ……

   b) Explique, que indica si el intervalo calculado contiene el cero.

CAPÍTULO V

DISEÑOS PARA CONTRASTACIÓN DE HIPÓTESIS

Formulada la hipótesis científica es necesario:

> “diseñar la estrategia para su contrastación. Los diseños se eligen antes de la recogida de datos de acuerdo a la naturaleza de la hipótesis y a los objetivos porque indican si la investigación es teórica o aplicada” (Vásquez et al., 2021, p. 68). Del diseño usado en la contrastación depende la validez de los resultados del estudio.

Kerlinger (1982, p. 217) recomienda tomar en cuenta el principio MAXMINCON consistente en:

1. Maximizar el efecto de la variable independiente o varianza del tratamiento.
2. Minimizar el error dado por los instrumentos y las muestras.
3. Controlar las variables extrañas, porque aportan varianza fuera del tratamiento.

5.1 MODELOS DE JOHN STUART MILL

John S. Mill siguió los principios de Francis Bacon, para demostrar la existencia de relaciones causales y propuso cuatro modelos lógicos: semejanza, diferencia, variaciones concomitantes y residuos.

5.1.1. Modelo lógico de semejanza

Para la hipótesis de que la aplicación del **cálculo económico** aumenta la **eficiencia productiva** de la empresa, se selecciona dos empresas diferentes una perteneciente a la industria alimenticia y la otra a la industria básica.

Si **E** es el aumento de eficiencia productiva y **X** el cálculo económico, las características propias de cada empresa son: materia prima, producción, equipos, fuerza de trabajo calificada y no calificada, etc. y las diferencias entre ellas se denotan con A, B y C en la primera empresa y en la segunda D, E y F el esquema resulta:

Si: A + B + C + X --------- E

y: D + E + F + X ---------- E

entonces: X ---------- E

Si en las dos empresas se ha elevado la eficiencia productiva y solo tienen en común la aplicación del Cálculo económico (X), entonces X es la causa de E.

El modelo lógico de semejanza para la demostración de hipótesis se puede enunciar de la siguiente manera:

"*Si dos o más casos del fenómeno que se investiga tienen solo una circunstancia, ésta es la causa del fenómeno en cuestión*"

5.1.2. Modelo lógico de diferencia

Dos empresas iguales (pertenecientes a la misma rama, con líneas de producción idénticas, etc.), salvo en el hecho de que funciona bajo el sistema de cálculo económico (X) y la otra aún

no lo ha aplicado. Si en la primera empresa se constata un aumento de la eficiencia productiva (E) y en la segunda se observa el incumplimiento de diversos indicadores productivos, se puede concluir que la aplicación del cálculo económico es la causa o parte de la causa del mejoramiento de la eficiencia. El esquema lógico será:

Si: A + B + C + X --------- E
y: A + B + C ---------- no E
entonces: X---------- E.

El modelo lógico de diferencia establece que:

"*Sí un caso en el cual aparece el fenómeno que se investiga y otro caso en el cual no aparece y tienen todas las circunstancias comunes excepto una, presentándose ésta en el primero, la circunstancia única en la cual difieren ambos casos es la causa, o una parte indispensable de la causa de dicho fenómeno*".

5.1.3. Modelo lógico de las variaciones concomitantes

"Un fenómeno que varía de cualquier manera, siempre que otro fenómeno varía de una manera particular, es una causa o un

efecto de este fenómeno, o está conectado con él por algún hecho de causalidad”

La eficiencia productiva de una empresa va aumentando a medida que se introduce paulatinamente el cálculo económico, puede concluir que existe cierta relación causal entre ambos procesos.

Si: A + B + C + X ---------- E

y: A + B + C + X^1 --------- E

A + B + C + X^2 --------- E

A + B + C + X^3 --------- E

entonces: X ---------- E

5.1.4. Modelo Lógico de residuo

“Si de un fenómeno complejo extraemos lo que es efecto de una parte de las circunstancias, el resto o residuo de dicho fenómeno ha de ser efecto de las circunstancias restantes”. (Rodríguez et al, 1984, p. 102)

Si se sabe que una parte del aumento de la eficiencia económica de una empresa agrícola en la que se aplica el cálculo económico se debe a la calidad de la tierra, a la atención cultural y técnica que se le brinda a los cultivos y a las buenas vías de

comunicación que la circundan, entonces la otra parte del aumento de la eficiencia económica de la empresa se debe o guarda relación con la aplicación del cálculo económico. Su esquema sería:

Si: $A + B + C + X$ ---------- $E^1 + E^2 + E^3 + E^4$

y: $A + B + C + X^1$ --------- $E^1 + E^2 + E^3$

entonces: X---------- E^4

Sí se conoce con antelación que A es causa de E^1;

y B es causa de E^3,

Entonces X es causa de E^4.

5.2 MODELOS DE CAMPBELL & STANLEY

En este trabajo se considera entre otros, los diseños no experimentales, cuasi experimentales y experimentales, propuestos por Campbell y Stanley.

5.2.1. Diseños No Experimentales

"Se usan para describir las características de un fenómeno existente, relacionar sucesos pasados con sucesos actuales, examinar las relaciones entre variables. Se llama así porque no es posible formar grupos de estudio al azar, ni manejar variables independientes". (Vásquez et al, 2021, p. 68).

Los más comunes son: "Antes y después", "Sólo Después" o "Ex Post-Facto" y "Sólo Después con grupo de comparación".

i. Diseño solo después o ex Posfacto

> "Se utiliza un solo grupo con el objeto de obtener un informe aislado de lo que existe en el momento del estudio, sirve para una investigación previa, para resolver problemas de identificación" (Vásquez et al., 2021, p. 68-69)

G_1	O_x
Grupo o muestra	**Mediciones de la variable en estudio**

ii. Diseño antes y después

Se trabaja con un solo grupo, que funciona como experimental y testigo del mismo, en línea.

G_1	O_{1x}	O_{2x}
Grupo o muestra	**Medición de la variable antes**	**Medición de la variable después**

iii. Diseño solo después con un grupo de comparación

"En este diseño se comparan dos grupos, uno que ha experimentado un estímulo y el otro que no, se deberá conocer el efecto de dicho estímulo. Se desconoce la situación de antes de ambos grupos y se hace una observación a cada uno de ellos sólo después, para determinar la diferencia $\mathbf{O_1 - O_2}$ es significativa" (Vásquez et al., 2021, p. 69). Su diagrama es:

G_1	**X**	**O_1** **Observaciones del grupo 1**
G_2	**-**	**O2** **Observaciones del grupo 2**

iv. Diseños correlacionales

"Los diseños correlacionales se utilizan cuando se trata de buscar la relación entre dos o más variables, sí las variables son cualitativas, la relación es de asociación, y si son cuantitativas, la relación es de correlación **propiamente** dicha" (Vásquez et al., 2021, p. 69).

G	**O_x** **Observación de la variable X**
	R (Relación)
	O_y **Observación de la variable Y**

5.2.2. Diseños cuasi experimentales

"Los diseños Cuasi-Experimentales, se usan cuando existen relaciones causales sin tener pleno control sobre las variables. Se llaman así por la dificultad de asignar aleatoriamente las unidades de análisis a la muestra y la de formar grupos de control" (Vásquez et al., 2021, p. 70).

Los más usados son: "Con grupo control no equivalente" y el de "Series de tiempo".

i. Diseño pretest y postest con un solo grupo

Un solo grupo sirve como experimental y testigo de sí mismo, en línea.

G_1	O_1	X	O_2
	Observaciones antes		**Observaciones después**

ii. Diseño con dos grupos "después"

Este diseño es parecido al clásico, se aplica el instrumento a los dos grupos, después del estímulo solo al primer grupo, sin conocerse la condición de "antes" de los grupos, luego la prueba se aplica a: **O_1- O_2.**

G_1	X	O_1 **Observación de G1, después del tratamiento**
G_2	-	O_2 **Observación del G2**

iii. Diseño de grupo control no equivalente

G_1	O_1 Observación antes	X tratamiento	O_2 Observación después
G_2	O_3 Observación antes	-	O_4 Observación después

iv. Diseños de series temporales.

G_1	0_1	0_2	0_3	0_4	X	0_5	0_6	0_7	0_8
	Observaciones				Trat.	Observaciones			

v. Diseños de muestras temporales equivalentes

G_1	X	0_1	-	O_2	X	O_3	-	O_4
	Trat.	Obs.		Obs.	Trat.	Obs.		Obs.

vi. Diseño de materiales equivalentes

G_1	MA	0_1	X_0	O_2
G_2	MB	0_3	X_1	0_4

vii. Diseño con muestras diferentes

"Se utilizan dos grupos diferentes bajo la influencia del mismo estímulo [X], en tiempos diferentes (Falcón et al., 1978, p. 41), luego se busca la diferencia de O_2-O_1" (Vásquez et al., 2021, p. 73).

"Estos diseños sirven para evaluar los efectos a mediano y largo plazo porque tiene bases para suponer que la influencia de la variable independiente sobre la dependiente tarda en manifestarse. Su esquema es el siguiente:" (Vásquez et al., 2021, p. 73).

Grupos	**Tiempo 1**		**Tiempo 2**	
G_1	**X**	**O_1**		
G_2	**-**		**X**	**O_2**

vii. Diseños compensados

Grupos	**Tiempo**							
	1		**2**		**3**		**4**	
G_1	**X_1**	**O**	**X_2**	**O**	**X_3**	**O**	**X_4**	**O**
G_2	**X_2**	**O**	**X_4**	**O**	**X_1**	**O**	**X_3**	**O**
G_3	**X_3**	**O**	**X_1**	**O**	**X_4**	**O**	**X_2**	**O**
G_4	**X_4**	**O**	**X_3**	**O**	**X_2**	**O**	**X_1**	**O**

5.2.3. Diseños experimentales

Son usados para estudiar las relaciones de tipo causal, este diseño permite valuar los tratamientos y la respuesta y controlar las variables extrañas. El diseño del experimento busca minimizar el error. En estos diseños se hace la selección aleatoria de los elementos de la muestra y se manipula la variable independiente, los más frecuentes son, diseño con grupo control pre y post test o clásico, diseño con cuatro grupos de Solomon, diseño con grupo control sólo después para una variable independiente, y para dos o más variables independientes los diseños factoriales 2 x 2, 3 x 2 etc. (Vásquez et al, 2021, p. 74)

i. Diseño de un grupo control con postest únicamente

GR_1	**X**	**O_1**
GR_2	**-**	**O_2**

ii. Diseño de una muestra separada pre y postest

GR_1	**0_1**	**X**	
GR_2		**X**	**0_2**

iii.Diseño con grupo control pre y post test o clásico

GR_1:	**O_1**	**X**	**O_2**
GR_2:	**O_3**	**-**	**O_4**

iv. Diseños con cuatro grupos de Solomon

GR_1	**0_1**	**X**	**0_2**
GR_2	**0_3**	**-**	**0_4**
GR_3		**X**	**0_5**
GR_4		**-**	**0_6**

v. Diseño factorial de 2x2, A dos niveles y B dos niveles

Estos diseños permiten calcular la influencia de las diferentes combinaciones de tratamiento, dos del factor A y dos del factor B, tal como se muestra en el gráfico.

A_1B_1	**A_1B_2**
A_2B_1	**A_2B_2**

vi. Diseño con estímulo creciente

Se toman grupos parecidos donde uno no recibe estímulo y es considerado testigo, a los demás (experimentales) se aplica el

“estímulo” en diferentes magnitudes, tal como se muestra al considerar cuatro grupos.

A	Sin estímulo	A_1
B	Intensidad X	B_1
C	Intensidad 2X	C_1
D	Intensidad 3X	D_1

5.3 MODELOS ANALÍTICOS

Los diseños analíticos se aplican en investigaciones médicas, permiten estudiar las relaciones de **causa a efecto** comparando un grupo experimental con uno de grupo control, los grupos se forman por presencia (+) o ausencia (-) del factor de riesgo. (Watanabe, T., De Córdova, M., 1990). Entre estos diseños tenemos:

i. Diseños de corte transversal retrospectivos

De causa a efecto o estudios de estimación de riesgo, se utiliza para estudiar morbi1idad o prevalencia, se representa en una tabla de 2 x 2.

Riesgo	**Daño o efecto**		Total
	si (+)	**no (-)**	
si (+)	a	b	a + b
no (-)	c	d	c + d
Total	a + c	b + d	a+ b+ c+ d

El riesgo relativo (RR) está relacionado con la razón de prevalencia. Si RR=1 no hay relación entre el factor de riesgo y el evento, en cambio si RR>1 existe relación.

$$Razón\ de\ prevalencia = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

ii. Diseños de casos y controles

Son estudios retrospectivos de efecto a causa. Las enfermedades son de largo período de latencia.

La **comparación** entre casos y controles se establece utilizando tablas de 2 x 2. **Se ingresa por el daño o efecto**.

Factor	Daño o efecto		total
	sí (+) casos	**no (-) controles**	
si (+)	verdaderos positivos a	falsos positivos b	a + b
no (-)	falsos negativos c	verdaderos negativos d	c + d
Total	a + c	b + d	n

Para el cálculo del riesgo relativo se usa la razón de los productos cruzados el Odds Ratio (OR).

$$O.R. = \frac{ad}{bc}$$

OR<1 existe poca probabilidad que ocurra el evento, OR=1 no hay relación entre variables, OR>1 a mayor valor, mayor.

iii. Diseño de cohortes

Son longitudinales de causa a efecto. Se observan individuos expuestos a un determinado factor de riesgo, **la comparación** entre individuos expuestos y no expuestos a un determinado factor se establece usando una tabla de 2 x 2. **Se ingresa por el factor de riesgo o exposición.**

Factor de riesgo	Daño o efecto		Total
	si (+)	no (-)	
si (+)	a	b	a + b
no (-)	c	d	c + d
	-	-	n

La tasa de incidencia del daño se calcula con el riesgo relativo

$$R.R. = \frac{riesgo\ de\ daño\ en\ expuestos}{riesgo\ de\ daño\ en\ NO\ expuestos} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

iv. Diseño para la sensibilidad, especificidad y valor predictivo de la prueba

Resultado del diagnóstico	Patología		Total
	Si (+)	No (-)	
Positivo (+)	Verdaderos Positivos (a)	Falsos Positivos (b)	a+b
Negativo (-)	Falsos Negativos (c)	Verdaderos Negativos(d)	c+d
total	a+c	b+d	n

Para analizar la sensibilidad y especificidad se ingresa en la tabla por la patología.

$$\textbf{Sensibilidad} = \frac{\text{VP}}{(\text{VP} + \text{FN})}$$

$$= \frac{\text{a}}{\text{a} + \text{c}} = \frac{\text{verdaderos positivos}}{\text{total de enfermos}}$$

$$\textbf{Especificidad} = \frac{\text{VN}}{(\text{VN} + \text{FP})}$$

$$= \frac{\text{d}}{\text{d} + \text{b}} = \frac{\text{verdaderos negativos}}{\text{total de no enfermos}}$$

Para el análisis del valor predictivo de la prueba se ingresa en la tabla por el resultado del procedimiento diagnóstico.

$$\textbf{Valor predictivo positivo} = \frac{\text{VPP}}{(\text{VP} + \text{FP})}$$

$$= \frac{a}{a + b} = \frac{\text{verdaderos positivos}}{\text{total de pruebas positivas}}$$

$$\textbf{Valor predictivo Negativo} = \frac{\text{VPN}}{(\text{VN} + \text{FN})}$$

$$= \frac{\text{d}}{\text{d} + \text{c}} = \frac{\text{verdaderos negativos}}{\text{total de pruebas negativas}}$$

V. Diseño de los ciegos

Estos diseños se usan para determinar la eficacia de los tratamientos en salud, permiten eliminar los factores intencionalmente asociados con el tratamiento. A los pacientes del grupo control generalmente se administra placebos.

En el diseño del ciego simple, el paciente no sabe si ha sido asignado al grupo experimental o al control.

En el diseño del doble ciego tanto el paciente como el investigador desconocen quienes tienen el factor experimental.

vi. Diseños de tipo cualitativo

Se tipifican como diseños cualitativos los estudios

1. Etnográficos.
2. Fenomenológicos
3. Investigación-acción.
4. Narrativo-biográfica.

5.4 VALIDEZ DE LOS DISEÑOS PARA LA CONTRASTACIÓN DE HIPÓTESIS

Los diseños masa adecuados para la contrastación de hipótesis garantizar validez interna y externa. La validez interna se ve afectada cuando en el experimento intervienen variables extrañas que generan efectos que pueden confundirse con los de la variable independiente. Por su parte, la validez externa se ve afectada cuando en el experimento están ejerciendo influencia ciertos factores que evitan que los resultados se puedan inferir a toda la población. (Vásquez et al., 2021, p. 84)

Los factores en Vásquez et al., que afectan la validez interna son:

Historia: Está referido a acontecimientos específicos ocurridos entre la primera y la última medición de la variable en estudio, éstos suelen manifestarse con mayor frecuencia en experimentos a largo plazo.

Maduración: Son procesos internos que los participantes desarrollan debido al paso del tiempo, los más comunes son los procesos biológicos o psicológicos.

Administración de test: Es la influencia que ejercen tests tomados previamente sobre los que se administran posteriormente.

Instrumentación: Son cambios en los instrumentos de medición o en los observadores o calificadores que pueden producir variaciones en las mediciones.

Regresión estadística: Se produce cuando en el procesamiento de los datos se incluyen valores atípicos o extremos, muy altos o muy bajos, que regresionan los datos en el promedio produciendo variaciones más allá de los efectos de la variable experimental.

Selección de la unidad de estudio: La selección sesgada de participantes para los grupos de comparación hace perder la acción del tratamiento experimental, de allí que la

selección aleatoria garantiza la confiabilidad en los resultados.

Mortalidad experimental: Trata de la pérdida de participantes en el experimento, se da mayormente en experimentos de periodos prolongados. Toda investigación debe garantizar no perder unidades experimentales.

Interacción de selección de la unidad de estudio y maduración: Se refiere a la inadecuada selección de la unidad de estudio, pues no se ha tomado en cuenta que éstas pueden madurar durante el transcurso del experimento, variando los resultados esperados. (2021, pp. 84-85)

Los factores en Vásquez et al., que afectan la validez externa son:

Interacción del test y la variable independiente: Ocurre cuando no se toma en cuenta la sensibilidad a la variable independiente producida por las mediciones en la variable de estudio.

Interacción de selección de la unidad de estudio y variable independiente: Ocurre cuando no se toma en cuenta las características de las unidades de estudio en el muestreo y los sesgos por la falta de randomización para la selección de los grupos, afectando la inferencia de los resultados a la población.

Efectos reactivos de los dispositivos experimentales: Ocurre cuando los dispositivos experimentales modifican la actitud de las unidades de estudio, pudiendo ocasionar predisposición o rechazo a participar en el experimento, modificando los resultados.

Interferencia de múltiples variables independientes: Se da cuando se apliquen tratamientos múltiples a los mismos participantes, sin esperar el tiempo necesario que asegure que un tratamiento no influye en el otro. (2021, pp. 85-86)

En la investigación tecnológica, el diseño se convierte en una estrategia metodológica que permite la formulación y resolución de problemas con procesos de razonamiento lógico, inductivos y deductivos; es una etapa visualizada creativamente, para la que se necesita conocimientos, habilidades, destrezas, manejo de leyes, principios, teorías y competencias en el uso de las TIC. (Vásquez et al., 2021, p. 86)

CAPÍTULO VI

CONTRASTACIÓN DE HIPÓTESIS

Corresponde a la tercera fase del método científico, en la cual se hace uso de métodos y técnicas para elegir la estrategia más adecuada para usar la información empírica o base teórica para la contrastación de la hipótesis que daría solución al problema científico planteado.

6.1. HIPÓTESIS ESTADÍSTICAS

La única forma de tener una certeza absoluta de la veracidad o falsedad de la hipótesis estadística es examinando toda la población, pero esto con frecuencia es impracticable o imposible y entonces nos vemos obligados a sacar una muestra de la población. La muestra sacada de la población se usa para tomar una **decisión** como si la hipótesis fuera verdadera o falsa. El proceso de usar la muestra para contrastar si la hipótesis es verdadera (o falsa) se llama prueba estadística de la verdad (o falsedad) de la hipótesis o **contrastación de hipótesis**.

De hecho, no hay certeza de que no se cometerá una equivocación, existiendo dos tipos de error, cualquiera de los cuales se puede cometer.

Error tipo I, (α). Llamado también nivel de significación, es la máxima probabilidad de rechazar la hipótesis siendo cierta o si la hipótesis establecida es verdadera y la consideramos como
falsa. Si la hipótesis establecida es falsa y se considera como verdadera se comete un **error tipo** II, (β).
Nuestra decisión de aceptar o rechazar una hipótesis se basará en la información que obtengamos haciendo observaciones y en el riesgo que estamos dispuestos a correr de que nuestra decisión puede ser errónea.

Hipótesis nula Ho. Es la Hipótesis que se debe probar, se llama también hipótesis de ninguna diferencia, de allí el nombre de **nula;** esta conjetura de la realidad (universo o población) puede ser cierta o falsa.

Hipótesis alternativa H_1. Llamada también **hipótesis de investigación,** es la hipótesis contra la cual se va a contrastar Ho. Se busca evidencia en la muestra, que nos induzca a rechazar Ho a favor de H_1.

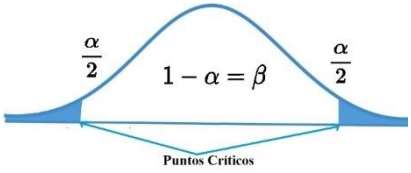
Función pivotal o estadístico de prueba; es toda variable aleatoria cuya **distribución muestral** está relacionada con H_1 de cuyo valor depende la decisión. Si este valor cae en la zona de rechazo, la decisión es **rechazar** Ho; en caso contrario no hay evidencia para rechazarlo.

Región crítica o zona de rechazo. Una vez establecido el nivel de significación, quedan determinadas dos regiones para la prueba: zona **de rechazo** = α y **zona de aceptación** = 1 - α =β.

El punto que separa estas zonas se llama **punto crítico** y toma valor de la distribución teórica con la que mantiene concordancia la muestra.

Se llama **prueba de una cola**; si la región crítica se ubica a la derecha o izquierda del punto crítico; son sugeridas por la hipótesis de investigación, **mayor que o menor que** respectivamente.

Si la región de rechazo se reparte en las dos direcciones se tiene la prueba de dos colas, es sugerida también por la hipótesis de la modalidad diferente que.



El valor p mide la participación del azar en los resultados de los experimentos.

Si p resulta alto, el azar interviene mucho, la relación entre las variables dependientes e independientes es aparente, no es de causa a efecto, si p resulta bajo, el azar interviene poco, la diferencia es real.

Generalmente, en la investigación científica se tiene:

si p< 0.05 los resultados del experimento son ***significativos****,*

si p <0.01 los resultados son ***altamente significativos****.*

Figura 6.1 Toma de decisión al contrastar una hipótesis en base a los resultados de la muestra.

Decisión basada en la muestra	Conjetura de la realidad	
	H_0: Es Verdadera	H_0: Es Falsa
No rechazar H_0:	Decisión correcta	**Error tipo II**
Rechazar H_0:	**Error tipo I**	Decisión correcta

Proceso en la contrastación de hipótesis

1. Planteamiento de las hipótesis nula Ho y alternativa H_1.
2. Elegir el nivel de significación α
3. Descripción de la **población** para inferencia de los resultados: **distribución teórica.**
4. Elección del modelo estadístico de prueba, **distribución empírica**: **muestra**
5. Decisión**:** Comparar el valor experimental, con el valor de la distribución teórica. (tablas estadísticas del texto)
6. Conclusión

6.2. PRUEBA DE HIPÓTESIS DE μ, σ, P.

6.2.1. Contrastación de $\bar{X}$ con μ conocido σ^2

Si la media μ de una población tiene un valor determinado μ_0 y la varianza un valor conocido σ^2, entonces la hipótesis nula se plantea como:

$$H_0 : \mu = \mu_0$$

Y el estadístico de prueba será:

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

por ser la distribución de la muestra. *Z* es de uso general cuando n≥30.

Ejemplo 1

El perímetro toráxico de niños recién nacidos, se distribuye normalmente con μ = 13 pulgadas y σ = 0.7. Se estudia un grupo de 25 niños recién nacidos, procedente de una población

que vive en una región muy apartada de la que se piensa constituye una isla genética y se obtiene $\bar{X}$ = 12.6 pulgadas de perímetro toráxico. ¿Se puede afirmar con 95% de seguridad que los niños son de una población con parámetros diferentes a μ = 13 y σ = 0.7?

Solución:

Hipótesis nula:

Ho: $\mu = \mu_0$

Hipótesis Alternativa:

$H_{1:}\ \mu < \mu_0$, porque el promedio de la muestra es menor:

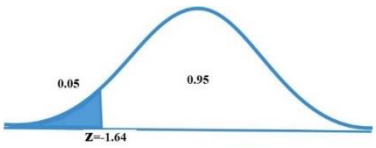
$\bar{X} = 12.6$

Según el proceso de contrastación:

1. Las hipótesis

$$H_0 : \mu = 13$$
$$H_1 : \mu < 13$$

2. $\alpha = 0.05$ (área crítica).
3. Z= 1.64, según la tabla de N(Z,0,1)



4. El estadístico de prueba será:

$$Z_e = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Sustituyendo los valores poblacionales y muestrales se tiene:

$$Z_e = \frac{12.6 - 13}{\frac{0.7}{\sqrt{25}}} = -2.85$$

5. Decisión:

Se rechaza H_0, porque el valor de Z_e cae en la zona crítica.

6. Conclusión:

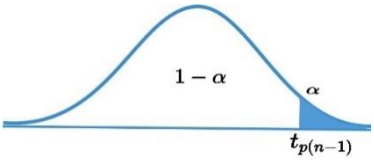
Como p < 0.05, los niños estudiados no se consideran, muestra de la población donde μ = 13 y σ = 0.7, muestran diferencia significativa.

6.2.2. Contrastación de $\bar{X}$ con μ desconocido σ^2

H_o: μ = μ_o

H_1 : μ > μ_o

α: : O. O5



$$t_{(n-1)} = \frac{\bar{X} - \mu_o}{\frac{S}{\sqrt{n}}}$$

Una distribución normal tiene el inconveniente de que no siempre se conoce σ, entonces es natural reemplazar σ por S y usar:

$$t_{(n-1)} = \frac{\bar{X} - \mu_o}{\frac{S}{\sqrt{n}}}$$

En vez de Z, aunque por término medio $S^2 = \sigma^2$ siendo en la mayoría de las veces $S^2 < \sigma^2$, si n es muy grande, la distribución t se diferencia poco de Z.

t es usada generalmente cuando n < 30

Ejemplo 1

En una muestra de 25 niños en edad escolar, se determinó que $\bar{X}$ =94 con S=18 . ¿Se puede afirmar con 95% de seguridad que los niveles de C.I. de la población de donde se sacó la muestra están por debajo de la normal?

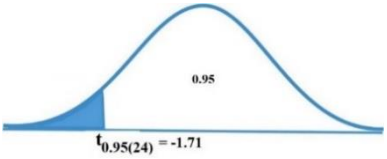
Los niveles normales de C.I. son: (μ = 100; σ = 16).

Solución:

H_o: μ = 100

H_1: μ < 100

α: 0.05



$$t_e = \frac{\bar{X} - \mu_o}{\frac{S}{\sqrt{n}}} = \frac{94 - 100}{\frac{18}{\sqrt{25}}} = -1.66$$

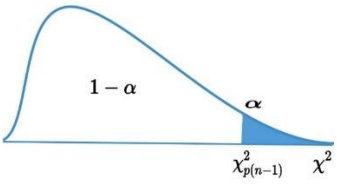
t_e cae en la zona de aceptación.

No se puede afirmar que los niveles de C.I. encontrados en la muestra estén por debajo de los normales; no se tienen argumentos para rechazar H_o. La diferencia encontrada se debe a causas aleatorias.

6.2.3. Prueba de hipótesis para la varianza σ^2

$H_0: \sigma^2 = \sigma_0^2$

$\chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma_0^2}$



La varianza σ^2 de una población normal tiene algún valor fijo $\sigma_0{}^2$. Obtenida una muestra aleatoria de la población y calculada S^2 si es marcadamente más pequeña o más grande que σ_0^2. se rechazará ese valor razonable para σ^2. La distribución muestral:

$$\chi^2_{(n-1)} = \frac{(n-1)S^2}{\sigma_0^2}$$

es utilizada para contrastar las hipótesis $\sigma^2 = \sigma_0{}^2$

Ejemplo 1

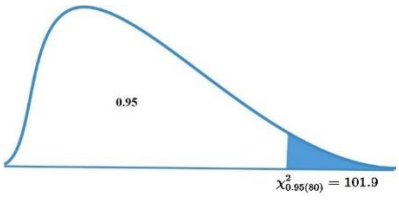
El Gerente de producción de **Bayer** desea verificar que la varianza de llenado de empaques de cierto producto no sea menor de 196 g. Toda vez que examinando 81 paquetes de la producción dio 400 g de S^2.

Solución:

H_o: $\sigma^2 = 196$ g

H_1: $\sigma^2 > 196$ g

α: 0.05



$$\chi_e^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(80)(400)}{196} = 163.26$$

χ_e^2 cae en la zona de rechazo. Con lo que se puede afirmar con 95% de seguridad que efectivamente el llenado de empaques se está excediendo a 14 de desviación estándar.

Ejemplo 2:

Se ha puesto varios años un examen **tipo,** con $\mu = 70$ y $\sigma^2 = 9$. Una escuela que utilizaba por primera vez este examen se lo puso a un grupo de 25 estudiantes y obtuvieron una media de $\bar{X} = 71$ y una varianza de $S^2 = 12$.

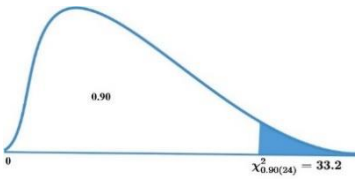
¿Hay una razón para dudar de que las calificaciones de todos los estudiantes de la escuela tuvieran una varianza de 9?

Solución:

H_o: $\sigma^2 = 9$

H_1: $\sigma^2 > 9$

α: 0.10

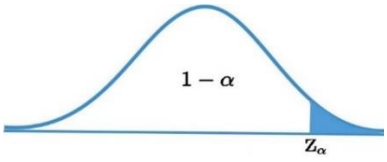


$$\chi_e^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(24)12}{9} = 32$$

χ_e^2 se ubica en la zona de aceptación de H_0, por tanto, no tenemos razones suficientes para dudar (con 90% de seguridad) que la varianza sea igual a 9.

6.2.4. Contrastación de p con la proporción poblacional P

H_0: $P = P_0$



$$Z = \frac{p - P_0}{\sqrt{\frac{p(1-p)}{n}}}$$

P_0 es el supuesto valor (hipotético) que se va a contrastar. El muestreo es binomial, se tiene una sucesión de observaciones independientes cada una de las cuales resulta **favorable o desfavorable** de modo que la probabilidad de favorable en una prueba es p y desfavorable (1 - p) = q, (p + q = 1).

Algunos autores prefieren utilizar:

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$

como estadístico de prueba, arguyendo que $\frac{P_0(1-P_0)}{n}$ es la varianza de p, se prefiere aquí $\frac{p(1-p)}{n}$ como la varianza aproximada de p.

Ejemplo 1

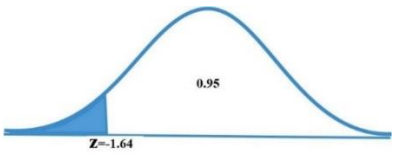
Se afirma que en la Universidad "Pedro Ruiz Gallo", en la Facultad de Ciencias Económicas, Administrativas y contables no más del 30% de alumnos trabajan y estudian. Para verificar la sospecha se toma una muestra aleatoria de 300 alumnos donde se encontró que 75 trabajan además de estudiar. ¿Se puede seguir manteniendo la primera afirmación con 95% de seguridad?

Solución:

H_0: P = 0.30

H_1: P < 0.30

α: 0.05



$$z_e = \frac{p - P_o}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.25 - 0.30}{\sqrt{\frac{(0.25)(0.75)}{300}}} = -2$$

$Z_e < Z_T$ se rechaza H_0. En consecuencia, si es fundada la afirmación de que los alumnos que trabajan y estudian están por debajo del 30%.

6.2.5. Contrastación de $\mu_1 - \mu_2$, σ_1^2 y σ_2^2 conocidas

H_0: $\mu_1 = \mu_2$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

El interés más frecuente es saber si las 2 poblaciones tienen la misma media por lo que se plantea $\mu_1 - \mu_2 = 0$, se toma 2 muestras aleatorias independientes de tamaños n_1 y n_2 de 2 poblaciones con distribución normal.

La función pivotal es la normal típica donde la zona crítica termina: $Z_{(1-\frac{\alpha}{2})}$ o $Z_{(\frac{\alpha}{2})}$ que es equivalente; y la zona crítica unilateral lo determina $Z_{1-\alpha}$ o Z_{α}.

La distribución muestral de $\overline{X_1} - \overline{X_2}$ es para muestras grandes aproximadamente normal con media $\mu_1 - \mu_2$ y con varianza

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Si $\sigma_1^2 = \sigma_2^2 = \sigma^2$ la función pivotal será:

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Ejemplo 1

Se desea saber si el período de gestación determina una menor concentración de magnesio sérico en la sangre. Se toma muestras de sangre de 10 universitarias en estado de gestación y a 8 no gestantes en condiciones análogas al grupo experimental, obteniendo los siguientes datos:

Gestantes:	1.16, 2.25, 2.36, 2.04, 1.24, 1.96, 1.68, 1.55, 1.38, 1.42
No Gestantes:	2.21, 2.10, 2.65, 2.30, 1.94, 2.72, 1.80, 1.56

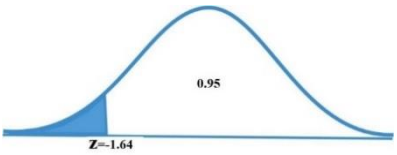
La varianza poblacional en gestantes es de 0.25 y en no gestantes 0.16; se asume además la distribución normal de las poblaciones, (use α= 0.05).

Solución:

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 < \mu_2$

α: 0.05



Datos de muestra: $n_1 = 10$ $\overline{X_1} = 1.704$, $n_2 = 8$ $\overline{X_2} = 2.16$

$$Z_e = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1.704 - 2.16}{\sqrt{\frac{0.25}{10} + \frac{0.16}{8}}} = -2.15$$

Como $Z_e < Z_T$ se decide rechazar H_0. El estado de gestación hace disminuir la concentración promedio de magnesio sérico en la sangre, afirmación confiable hasta el 95%.

6.2.6. Contrastación de $\mu_1 - \mu_2$, de dos poblaciones σ_1^2 y σ_2^2 desconocidas y $\sigma_1^2 \neq \sigma_2^2$

H_0: $\mu_1 = \mu_2$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ con f gl.}$$

Si σ_1^2 y σ_2^2 son desconocidos, pero se tiene la evidencia de que cada población se distribuye normalmente se estimara σ_1^2 y σ_2^2 por S_1^2 y S_2^2 respectivamente obteniendo el estadístico t con f gl.

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$f = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 + 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 + 1}} - 2$$

El valor de f no siempre es entero, comúnmente el valor más próximo de la tabla es suficiente.

Ejemplo 1

Los datos que siguen representan excreciones totales de histidina en mg. De muestras de orina en 24 horas, de 5 hombres y 10 mujeres a dieta sin restricciones.

Hombres:	229, 236, 435, 172, 432
Mujeres:	197, 224, 115, 74, 138, 135, 107, 204, 200, 138

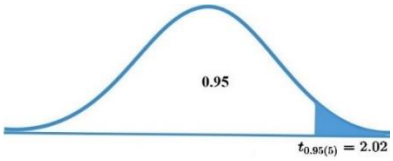
¿La concentración de histidina en los hombres es mayor que en las mujeres?

Solución:

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$

α: 0.05



$n_1 = 5 \quad \bar{X}_1 = 300.8 \quad S_1^2 = 15291.70$

$n_2 = 10 \quad \bar{X}_2 = 153.2 \quad S_2^2 = 2484.62$

$$f = \frac{\left[\frac{15291.7}{5} + \frac{2484.62}{10}\right]^2}{\frac{\left[\frac{15291.7}{5}\right]^2}{5+1} + \frac{\left[\frac{2484.62}{10}\right]^2}{10+1}} - 2$$

f = 4.9 $\cong$ 5 grados de libertad

$$t_e = \frac{300.8 - 153.2}{\sqrt{\frac{15291.7}{5} + \frac{2484.62}{10}}} = 2.56$$

El valor experimental cae en la zona de rechazo, $t_e > t_T$ lo que demuestra con 95% de seguridad que las concentraciones de histidina en los hombres son mayores que en las mujeres (al 99% se acepta H_0; $t_T = 3.36$).

6.2.7. Prueba de hipótesis de $\mu_1 - \mu_2$, de dos poblaciones $\sigma_1^2 = \sigma_2^2 = \sigma^2$ desconocidas

$$H_0 : \mu_1 = \mu_2$$

$$t = \frac{(\bar{X}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ con } n_1 + n_2 - 2 \text{ gl}$$

Dado que las poblaciones son normales con $\sigma_1^2 = \sigma_2^2 = \sigma^2$ no conocida, σ^2 se estima con S_p^2 y la distribución de $\bar{x}_1 - \bar{x}_2$ es:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$ con $(n_1 + n_2 - 2)$ grados de libertad.

Donde $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$

Ejemplo 1

Se desea comprobar con 95% de seguridad si el rendimiento en Bioestadística es mayor en la Facultad de Enfermería que en la Facultad de Medicina Humana de la Universidad "Pedro Ruiz Gallo". Para esto el investigador toma una muestra de notas de cada Facultad en forma aleatoria.

Enfermería:	10,15,16,08,13,18,12,14,11,14,13,12
M. Humana:	12,11,14,08,10,16,12,07,10,15,11,10

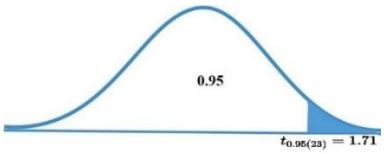
Solución:

Las poblaciones no tienen varianzas conocidas, pero si es razonable pensar que sean iguales.

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 > \mu_2$

α: 0.05



13 +12 – 2 = 23 gl

Reemplazando estadísticas de las muestras:

$n_1 = 13$, $\bar{X}_1 = 12.92$, $S_1^2 = 6.74$, $n_2 = 12$ $\bar{X}_2 = 11.33$ $S_2^2 = 7.15$

$$S_p = \sqrt{\frac{(13-1)6.74 + (12-1)7.15}{13+12-2}} =$$

$$S_p = 2.63$$

$$t_e = \frac{12.92 - 11.33}{2.63\sqrt{\frac{1}{13} + \frac{1}{12}}} = 1.51$$

Decisión: t_e < tabular, se acepta Ho, no se puede decir que los rendimientos en Enfermería en el curso de Bioestadística sean mayores que en Medicina, la diferencia encontrada en las muestras se debe a causas aleatorias.

6.2.8. Prueba de hipótesis para $\mu_1 - \mu_2$ con $\sigma_1^2 \neq \sigma_2^2$ desconocidas y $n_1, n_2 \geq 30$.

Ho: $\mu_1 - \mu_2$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Como σ_1^2 y σ_2^2 son desconocidas y diferentes debería para el contraste usarse t con f gl, pero en razón de que las muestras son grandes $n \geq 30$ se utiliza Z como función pivotal para la prueba.

Ejemplo 1

Para evaluar los efectos clínicos de un nuevo esteroide en el tratamiento de personas falta de peso crónico, se necesitó 40 personas quienes recibieron dosis de 25 mg. de la droga por un período de 12 semanas y otro grupo de 40 de características análogas recibió una dosis de 50 mg. por el mismo periodo.

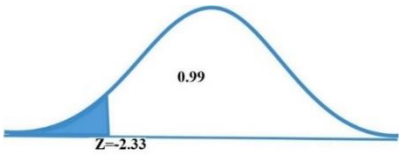
Después de 6 meses se observó que las personas del primer grupo habían ganado en promedio 9 lb con una S=7 lb, mientras que los segundos ganaron 11 libras con 6 de desviación estándar. Existe diferencia real confiable hasta 99%?

Solución:

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

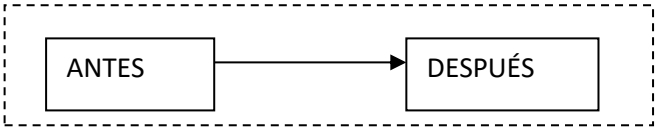
$\alpha = 0.01$



$$Z_e = \frac{9 - 11}{\sqrt{\frac{7^2}{40} + \frac{6^2}{40}}} = -1.37$$

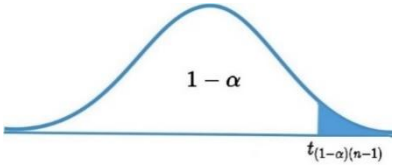
Decisión: No se rechaza Ho, porque Z_e cae en la zona de aceptación, luego no existe diferencia real en el peso promedio ganado, la diferencia de peso promedio en muestra no es debido a la dosis administrada del esteroide, se atribuye a causas aleatorias.

6.3. PRUEBA DE HIPÓTESIS PARA MUESTRAS PAREADAS



$H_0: \mu_A = \mu_D$

$$t = \frac{\bar{d} - (\mu_A - \mu_D)}{\frac{S_d}{\sqrt{n}}}$$



La población es hipotética y suponemos que el grupo disponible para la experimentación es muestra aleatoria de esa

población y que la población de **antes** menos **después** está normalmente distribuida.

Como la población de diferencias está normalmente distribuida, consideramos que la distribución muestral de $\bar{d} = \frac{\sum d_i}{n}$ es normal con $\mu = 0$ y $\sigma_d^2 = \frac{S^2}{n}$.

Como no se conoce σ_d y n es pequeña (n < 30) el estadístico de prueba es:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$ con n-1 grados de libertad.

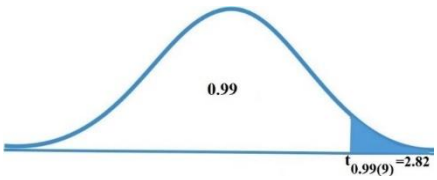
Ejemplo 1

Para estudiar la eficacia de una dieta de adelgazamiento se obtuvieron las siguientes cifras de peso en libras, **antes y después** en una muestra de 10 mujeres casadas de 30 y 40 años de edad.

Paciente:	1	2	3	4	5	6	7	8	9	10
Antes	134	147	165	152	139	122	138	147	153	178
Después	131	140	164	153	133	122	135	148	147	165
Diferencia	3	7	1	-1	6	0	3	-1	6	13

Se puede afirmar que la dieta es eficaz hasta con 99%.

Solución



$H_0: \mu_A = \mu_D$
$H_1: \mu_A > \mu_D$
$\alpha: 0.01$

$\bar{d} = 3.7$
$S_d = 4.39$
$n = 10$

$$t_e = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{3.7}{\frac{4.39}{\sqrt{10}}} = 2.66$$

No se rechaza H_0, $t_e <$ t teórica, se acepta H_0, la dieta no es eficaz porque el peso real después de la dieta es igual al peso antes de la dieta.

Ejemplo 2

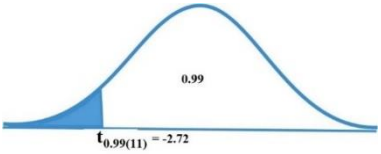
Se desea investigar si cierta cantidad de marihuana altera el número de pulsaciones en jóvenes que fuman por primera vez. Para esto a 12 jóvenes que se ofrecían para la experiencia se les proporcionó una cantidad específica, controlando el número de pulsaciones por minuto **antes y después** de haberla fumado.

Sujeto	1	2	3	4	5	6	7	8	9	10	11	12
Antes	70	59	79	68	80	64	75	58	68	62	61	75
Después	80	70	88	77	90	75	87	65	80	74	70	82
Diferencia	-10	-11	-9	-9	-10	-11	-12	-7	-12	-12	-9	-7

¿Se confirma la hipótesis con 99% de seguridad?

Solución:

$H_0: \mu_A = \mu_D$
$H_1: \mu_A > \mu_D$
$\alpha: 0.01$



$\bar{d} = -9.91$
$S_d = 1.78$
$n = 12$

$$t_e = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{-9.91}{\frac{1.78}{\sqrt{12}}} = -19.28$$

H_0 se rechaza, concluyendo que: En los jóvenes que fuman marihuana por primera vez, el número de pulsaciones aumenta, para esta afirmación se cuenta con el 99% de seguridad.

6.3.1. Contrastación para la diferencia $P_1 - P_2$

H_0: $P_1 = P_2$

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

La distribución muestral de la diferencia, $P_1 - P_2$ es aproximadamente normal si n_1 y n_2 son lo suficientemente grandes, con media P_1-P_2 y una desviación típica:

$$\sigma_{(p_1-p_2)} = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

Como siempre P_1 y P_2 se desconocen,$\sigma_{(p_1-p_2)}$ se estima mediante:

$$S_{(p_1-p_2)} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

donde p_1 y p_2 son las proporciones muestras aleatorias independientes de las poblaciones N_1 y N_2 respectivamente.

Si $P_1 = P_2 = P$, se combina las 2 muestras y se obtiene una estimación de P mediante:

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

donde X_1 elementos con característica de interés en la muestra de N_1 y X_2 el número de elementos que presentan la característica de interés en la muestra extraída de N_2 obteniendo el siguiente estadístico de prueba:

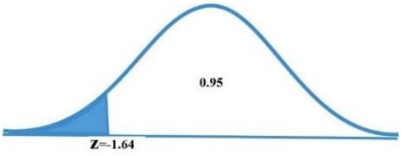
$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}}$$

Ejemplo 1

Un botánico en una universidad de California investiga dos métodos para conseguir resistencia a una enfermedad de una determinada planta. De 200 plantas tratadas con el método A, 50 resultaron resistentes y de 250 tratadas con el método B, 88 resultaron resistentes; ¿puede garantizar el botánico el método B con 95% de seguridad?

Solución:

$H_0: P_A = P_B$
$H_1: P_A < P_B$
$\alpha = 0.05$



$$\boldsymbol{p_A = \frac{50}{200} = 0.25}$$
$$\boldsymbol{q_A = \frac{150}{200} = 0.75}$$
$$\boldsymbol{p_B = \frac{88}{250} = 0.352}$$
$$\boldsymbol{q_B = \frac{162}{250} = 0.648}$$
$$\boldsymbol{q_A = 1 - p_A, q_B = 1 - p_B}$$

$$\boldsymbol{Z = \frac{(p_A - p_B) - (P_A - P_B)}{\sqrt{\frac{p_A q_A}{n_A} + \frac{p_B q_B}{n_B}}} = \frac{0.25 - 0.352}{\sqrt{\frac{(0.25)(0.75)}{200} + \frac{(0.352)(0.648)}{250}}} = -2.37}$$

Se rechaza H_0, el botánico si puede garantizar el método B incluso hasta con 99% (Z_T = -2.33).

Ejemplo 2

Un antropólogo sospecha que los individuos de 2 poblaciones, con doble bucle de cabello en la región occipital, es la misma. Con el objetivo de probar esta hipótesis, toma muestras aleatorias independientes de cada población y determina la proporción de individuos en cada muestra con esta característica:

Población	Muestra	Individuos con la característica
N_1	100	23
N_2	120	32

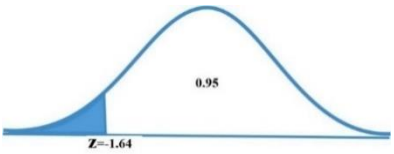
¿Puede confirmar su hipótesis al 95% de seguridad?

Solución

$H_0 : P_1 = P_2$

$H_1 : P_1 < P_2$

$\alpha : 0.05$



$$p_1 = 0.23 \quad p_2 = 0.27$$

$$\bar{p} = \frac{23 + 32}{100 + 120} = 0.25$$

se usa $\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$ como estimador de P porque se considera $P_1 = P_2 = P$.

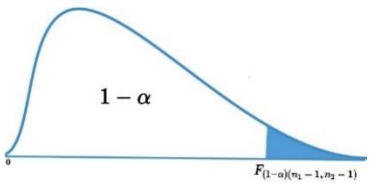
$$Z_e = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} = \frac{0.23 - 0.27}{\sqrt{\frac{(0.25)(0.75)}{100} + \frac{(0.25)(0.75)}{120}}} = -0.68$$

como Z_e < Z tabular. El antropólogo no tendría razones para no aceptar que $P_1 = P_2$

6.3.2. Contrastación de varianzas poblacionales normalmente distribuidas

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$F = \frac{\frac{n_1 S_1^2}{n_{1-1}}}{\frac{n_2 S_2^2}{n_{2-1}}}$$



Las varianzas muestrales S_1^2 y S_2^2 se derivan de 2 muestras independientes n_1 y n_2 extraídas de 2 poblaciones normalmente distribuidas con σ_1^2 y σ_2^2 respectivamente.

La prueba se denomina **razón entre dos varianzas** y se distribuye como: F con $n_1 - 1$ y $n_2 - 1$ gl donde $n_1 - 1$ son los grados de libertad del numerador y $n_2 - 1$grados de libertad del denominador.

Si la prueba es bilateral la región crítica derecha lo determina el valor $F_{\left(1-\frac{\alpha}{2}\right)(n_1-1,n_2-1)}$ y la región crítica izquierda $F_{\frac{\alpha}{2}(n_1-1,n_2-1)}$ cuando no se cuenta con tablas que den valores de $F_{\frac{\alpha}{2}}$ se utiliza:

$$F_{\frac{\alpha}{2}(n_1-1,n_2-1)} = \frac{1}{F_{\left(1-\frac{\alpha}{2}\right)(n_1-1,n_2-1)}}$$

$\frac{\sigma_1^2}{\sigma_2^2}$ = R_0 se propone también como H_0

Ejemplo 1

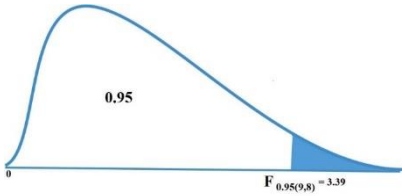
El departamento de producción de laboratorios **Bayer** ha notado que el contenido en el frasco de un insecticida de 3.2 onzas parece más variable que el de sus competidores y quiere utilizar en las máquinas de llenado un adaptador que según su fabricante aumenta la precisión en el llenado. Para ensayo se instaló un adaptador en una línea de llenado y otra línea en forma usual, se tomó como control. ¿En base a la información que se presenta, tiene razón el fabricante del adaptador?

Sin adaptador: 3.20,3.21,3.19,3.21,3.20,3.18,3.19, 3.20, 3.19, 3.20
Con adaptador: 3.20,3.19,3.20,3.20,3.19,3.20,3.19,3.19,3.20

Los datos se tomaron durante el proceso de llenado en forma aleatoria.

Solución:

$H_0: \sigma_1^2 = \sigma_2^2$
$H_1: \sigma_1^2 > \sigma_2^2$
$\alpha: 0.05$



$n_1 = 10 \quad \bar{X}_1 = 3.197 \quad S_1^2 = 0.00009$

$n_2 = 9 \quad \bar{X}_2 = 3.195 \quad S_2^2 = 0.00002$

$$F_e = \frac{\frac{n_1 S_1^2}{n_1 - 1}}{\frac{n_2 S_2^2}{n_2 - 1}} = \frac{\frac{10(0.00009)}{9}}{\frac{9(0.00002)}{8}} = 4.44$$

Se rechaza H_0, el adaptador satisface las expectativas del departamento de producción, el fabricante tiene razón.

Ejemplo 2

Las presiones sistólicas de pie y en supino en promedio son iguales, pero no hay seguridad de sus varianzas. Para

determinar si sus varianzas son iguales se toma la presión a 13 personas y se encuentra:

Varianza en pie = $S_p^2 = 90.06$

Varianza en supino= $S_s^2 = 83.17$

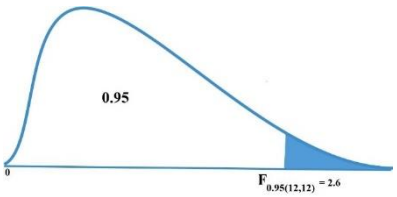
¿Hay diferencia significativa de varianzas?

Solución:

$H_0: \sigma_p^2 = \sigma_s^2;$ $\quad \frac{\sigma_p^2}{\sigma_s^2} = 1$

$H_1: \sigma_p^2 > \sigma_s^2;$ $\quad \frac{\sigma_p^2}{\sigma_s^2} > 1$

$\alpha: 0.05$



$$F_e = \frac{\frac{n_1 S_1^2}{n_1 - 1}}{\frac{n_2 S_2^2}{n_2 - 1}} = \frac{\frac{13(90.06)}{12}}{\frac{13(83.17)}{12}} = 1.08$$

Se acepta H_0, lo que quiere decir que también las varianzas de las presiones sistólicas de pie y en supino son iguales. Obsérvese la razón, sensiblemente mayor que 1.

6.4. PRUEBA DE LA BONDAD DE AJUSTE

H_0: Está referida a distribuciones teóricas para describir frecuencias observadas.

$$\chi^2 = \sum^{K} \frac{(f_o - f_e)^2}{f_e} \text{ con K} - 1 \text{ gl.}$$

En las pruebas de bondad de ajuste existe una sola variable y lo que se prueba es el **patrón de frecuencias hipotético**, o distribución de la variable. El patrón esperado (binomial - Poisson) generalmente es la distribución normal.

Las frecuencias observadas pueden colocarse en una sola fila o columna, cada frecuencia observada corresponderá a una clase K o nivel de variable de modo que la Ji-cuadrado tendrá K-1 grados de libertad, porque se utiliza un solo total de frecuencias, n para calcular la frecuencia teórica o esperada.

Para evitar algún grado de distorsión de la prueba se debe buscar que en cada clase del factor que se estudia, la frecuencia esperada no sea menor de 5, la clase (o casilla) que no se ajuste a este criterio deben combinarse con las categorías adyacentes.

Ejemplo 1

La cátedra de psiquiatría de la Facultad de Medicina de la Universidad de San Marcos ha clasificado a 145 estudiantes en

5 grupos: Excelentes, muy buenos, satisfactorios, regulares y malos. ¿Las aptitudes de los estudiantes siguen una distribución normal?

Tabla 6.1 Las aptitudes de los estudiantes psiquiatría de la Facultad de Medicina de la universidad de San Marcos

Clase	Nº de Estudiantes
Excelentes	27
M. Buenos	29
Satisfactorios	41
Regulares	25
Malos	23
Total	**145**

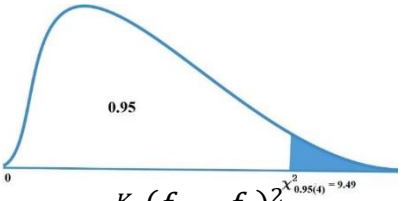
En base a los datos que se muestran, pruebe la hipótesis con un nivel del 5 %

Solución

H_0: La aptitud se distribuye normalmente.

H_1: La aptitud no se distribuye normalmente.

$\alpha = 0.05$



$$\chi^2_{\exp} = \sum^{K} \frac{(f_o - f_e)^2}{f_e} = 174.34$$

No se puede aceptar la hipótesis de normalidad.

Según el modelo normal estandarizado el 99.74% de las frecuencias esperadas (teóricas) deben estar en $\pm$ 3, en este rango se distribuyen las clases en que está dividida la variable en estudio.

$\frac{R}{K} = \frac{6}{5} = 1.2$ = amplitud estandarizada para las nuevas clases.

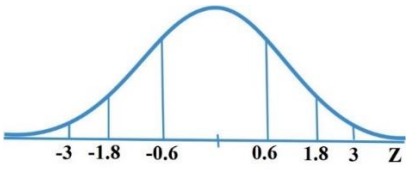


Tabla 6.2 Cálculos para la bondad de ajuste.

Intervalos en Z		Área (probabilidad)	np	Frecuencia Esperada	Frecuencia Observada
-3	-1.8	0.035	5.075	05	27
-1.8	-0.6	0.240	34.800	35	29
-0.6	0.6	0.450	65.250	65	41
0.6	1.8	0.240	34.800	35	25
1.8	3	0.035	5.075	05	23
Total		1.000		145	145

Conocida la probabilidad p por clase y n la muestra, np es la frecuencia esperada así:

145 (0.035) = 5.075, 145 (0.240) = 34.800 que siendo datos de recuento se toma al entero más próximo 5 y 35 respectivamente.

$$\chi_e^2 = \frac{(27-5)^2}{5} + \frac{(29-35)^2}{35} + \frac{(41-65)^2}{65} + \frac{(25-35)^2}{65} + \frac{(23-5)^2}{5} = 174.34$$

Mientras mayor es la discrepancia entre las frecuencias observadas y teóricas, mayor es la diferencia entre los valores de la Ji-cuadrado.

6.5. PRUEBA DE INDEPENDENCIA DE CRITERIOS

H_0: Suposición de que los criterios o variables son estadísticamente independientes.

$$\chi^2 = \sum_i^r \sum_j^s \frac{(f_{oij} - f_{eij})}{f_{eij}} \qquad \text{con } (r-1)(s-1) \text{ gl}$$

Las pruebas de independencia incluyen dos variables; el investigador quiere saber respecto a una población, si dos

criterios o variables de clasificación están relacionados: H_1, (si no están relacionados, son independientes: H_0).

Como intervienen 2 variables, las frecuencias observadas se distribuyen en una tabla de 2 vías o **tablas de contingencia** cuyas dimensiones son $r \; x \; s$; r = clases o niveles del criterio fila y s = clases o niveles del criterio columna. Los grados de libertad para la Ji-cuadrado será (r-1) (s-1).

Dada las hipótesis de independencia de las dos variables, la frecuencia esperada relacionada con cada casilla de una tabla de contingencia debería ser proporcional al total de frecuencias observadas incluidas en la fila y columna en la que está ubicada la casilla, en relación a la muestra.

$$f_e = \frac{\sum r \sum s}{n}$$

Ejemplo 1

Se selecciona una muestra de 2000 naranjas de C**uculí** de una cosecha, cada naranja se clasifica de acuerdo a su color y dulzor. Pruebe la hipótesis de que el dulzor y color son independientes con 95% de seguridad.

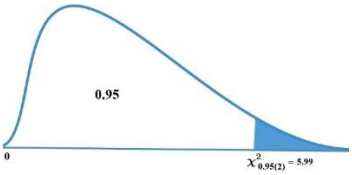
Dulzor	**Color**			
	Claro	**Mediano**	**Oscuro**	**Total**
Dulce	1000	340	140	1480
No dulce	200	130	190	520
Total	1200	470	330	2000

Solución:

H_0: El dulzor y color en las naranjas de Cuculí son independientes

H_1: El dulzor y color no son independientes

α: 0.05



Los grados de libertad = (r-1) (s-1) = (2-1) (3-1) = 2

Las frecuencias esperadas: $f_e = \frac{\sum r \sum s}{n}$

$$f_{e11} = \frac{1480x1200}{2000} = 888$$

$$f_{e12} = \frac{1480x470}{2000} = 347.8 \approx 348$$

$$f_{e13} = \frac{1480 \times 330}{2000} = 244.2 \approx 244$$

$$f_{e21} = \frac{520x1200}{2000} = 312$$

$$f_{e22} = \frac{520x470}{2000} = 122.2 \approx 122$$

$$f_{e23} = \frac{520x330}{2000} = 85.8 \approx 86$$

Dulzor	Color			
	Claro	Mediano	Oscuro	Total
Dulce	888	348	244	1480
No dulce	312	122	86	520
Total	1200	470	330	2000

$$\chi_e^2 = \frac{(1000-888)^2}{888} + \frac{(340-348)^2}{348} + \cdots + \frac{(190-86)^2}{86} = 225.13$$

Decisión: Se rechaza H_0. Luego el color y dulzor en las naranjas de cuculí no son independientes, existe algún grado de interacción entre los criterios de clasificación.

6.6. PRUEBA DE HIPÓTESIS EN TABLAS 2×2

Una tabla de contingencia con dos filas y dos columnas es un caso particular de independencia de criterios, se presenta cuando los sujetos se categorizan según 2 factores o criterios, cada uno de los cuales ocurre en dos niveles. Como gl=1, la aproximación de la distribución de χ^2 puede mejorarse mucho aplicando la corrección siguiente:

$$\chi^2 = \frac{(|ad - bc| - 0.5n)^2 n}{(a+b)(a+c)(b+d)(c+d)}$$

Criterio B	**Criterio A**		
	A 1	**A 2**	**Total**
B 1	a	b	(a + b)
B 2	c	d	(c + d)
Total	(a + c)	(b + d)	a + b + c + d = n

La **frecuencia teórica** mínima para la tabla de 2x2 no debe ser menor de 5 y no debe emplearse χ^2 cuando n es menor que 40. La corrección por continuidad puede omitirse si n > 50 y usar como función pivotal.

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(f_{oij} - f_{eij}\right)^2}{f_{eij}} \quad \text{con 1 gl}$$

La hipótesis de contraste es de independencia y concordancia de criterios respectivamente.

Ejemplo 1

A 60 individuos de una población se les ha hecho un estudio de mercado clasificándolos por sexo y según prefiera una marca A o B de cierto producto, obteniéndose los siguientes resultados:

Sexo	Marca		
	A	B	Total
Hombre	25	9	34
Mujer	7	19	26
Total	**32**	**28**	**60**

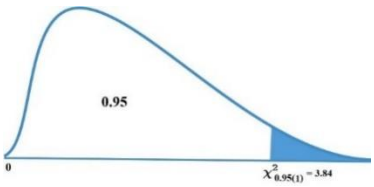
Se desea saber si hay una influencia de sexo frente a la preferencia de una determinada marca del producto.

Solución:

H_0: La preferencia por el tipo de producto es independiente del sexo.

H_1**:** La preferencia por el tipo de producto depende del sexo.

$\alpha = 0.05$



Sexo	Marca		
	A	B	Total
H	a 25	b 9	34 (a+b)
M	c 7	d 19	26 (c+d)
Total	32 (a+c)	28 (b+d)	25+9+7+19 a + b + c + d = 60

$$\chi_e^2 = \frac{(|ad - bc| - 0.5n)^2 n}{(a+b)(a+c)(b+d)(c+d)} = \frac{(|25 \times 19 - 9 \times 7| - 0.5 \times 60)^2 60}{34 \times 32 \times 28 \times 26} = 11.05$$

χ_e^2 cae en la zona de rechazo, luego con seguridad del 95% existe influencia del sexo en la preferencia del producto.

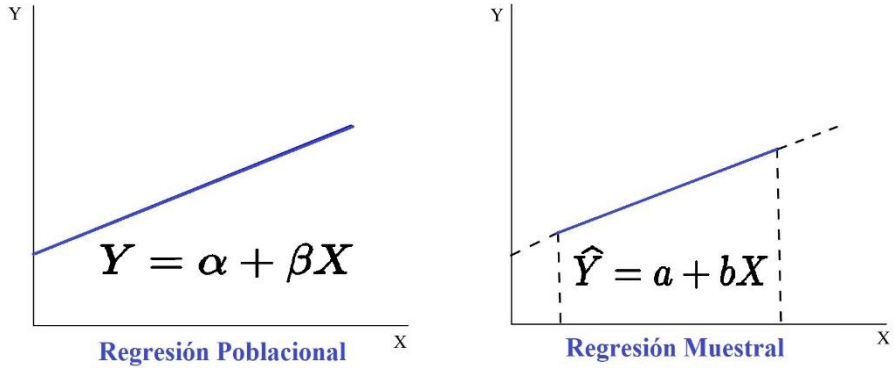
Como n > 50 se puede utilizar conservadoramente:

$$\chi_e^2 = \frac{(f_0 - f_e)^2}{f_e} = \frac{\left[25 - \frac{34 \times 32}{60}\right]^2}{\frac{34 \times 32}{60}} + \frac{\left[9 - \frac{34 \times 28}{60}\right]^2}{\frac{34 \times 28}{60}} + \frac{\left[7 - \frac{26 \times 32}{60}\right]^2}{\frac{26 \times 32}{60}} + \frac{\left[19 - \frac{26 \times 28}{60}\right]^2}{\frac{26 \times 28}{60}}$$

$\chi_e^2 = 12.85$ valor que también rechaza H_0.

6.7. PRUEBA DE HIPÓTESIS PARA UNA REGRESIÓN LINEAL

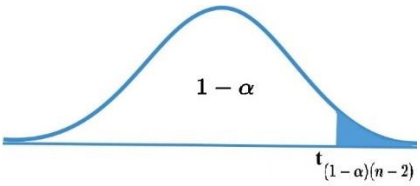
Figura 6.2 Representación de la regresión lineal de la muestra y la población.



H_o: β = 0, Y es independiente de X

H_1: β > 0, Y es dependiente de X

α = 0.05



$$t = \frac{(b - \beta_0)S_x\sqrt{n-1}}{S_{y/x}}$$

La regresión debe proporcionar un **ajuste** tan bueno para los datos como sea posible. El método de **mínimos cuadrados es** que mejor ajusta los datos a una recta. La regresión poblacional se estima tomando una muestra.

La variable X es el factor causa mientras que Y el efecto.
El modelo Y = a + bx predice, que valor asume probablemente

$\hat{Y}$ es la predicción cuando la variable X toma cualquier valor .

La regresión $Y = \alpha + \beta X$ se estima por $\hat{Y} = a + bX$ donde:

$$a = \bar{Y} - b\bar{X}, \qquad \bar{Y} = \frac{\sum Y}{n} \qquad \bar{X} = \frac{\sum X}{n}$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

La función pivotal para probar la hipótesis del coeficiente de regresión β, como es una t con n-2 grados de libertad necesita:

Desviación típica de X,

$$S_X = \sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n-1)}}$$

Error típico de estimación,

$$S_{Y/X} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n-2}}$$

cada Y obtendrá de reemplazar el valor experimental de X en

$$\hat{Y} = a + bx$$

Ejemplo 1

Halle la regresión para predecir la albúmina circulante del volumen del plasma y pruebe la hipótesis H_0: β=0 en base a la información que se obtiene de 5 pacientes.

Paciente	Volumen del Plasma X cm^3	Albúmina total circulante Y g
A	2.5	119
B	2.8	133
C	2.4	121
D	2.5	129
E	3.2	146

Solución

El diagrama de dispersión descubre una relación lineal

$$Y = a + bx:$$

Tabla 6.3 Tabla de cálculos para la regresión

X	Y	X Y	X^2	Y^2	$\hat{Y}$	$(Y - \hat{Y})^2$
02.5	119	0297.5	06.25	14161.	123.98	24.8004
02.8	133	0372.4	07.84	17689	133.35	00.1225
02.4	121	0290.4	05.76	14641	120.86	00.0196
02.5	129	0322.5	06.25	16641	123.98	25.2004
03.2	146	0467.2	10.24	21316	145.83	00.1718
13.4	648	1750.0	36.34	84448	648.00	50.1718

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{5(1750) - (13.4)(648)}{5(36.34) - (13.4)^2}$$

$$= \quad 31.2149$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{648}{5} = 129.6$$

$$\bar{X} = \frac{\sum X}{n} = \frac{13.4}{5} = 2.68$$

$$\text{a} = \bar{Y} - \text{b}\bar{X} = 129.6 - 31.2149(2.68) = 45.944$$

El modelo de estimación será: $\hat{Y} = 45.944 + 31.2149\ X$

El modelo explica que, por cada cm^3 de plasma, la albúmina se incrementa en 31.2149 g.

$$S_X = \sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n-1)}} = \sqrt{\frac{5(36.34) - (13.4)^2}{5(5-1)}} = 0.327\text{cm}^3$$

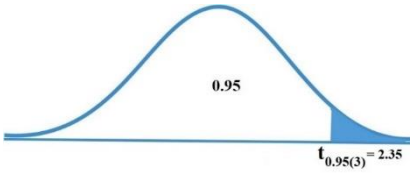
$$S_{Y/X} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{50.1718}{5-2}} = 4.089g.$$

La prueba de hipótesis:

$H_0 : \beta = 0$

$H_1 : \beta > 0$

α : 0.05



$$t_e = \frac{(b - \beta_0)S_x\sqrt{n-1}}{S_{y/x}} = \frac{(31.2149 - 0)(0.327)\sqrt{5-1}}{4.089} = 4.99$$

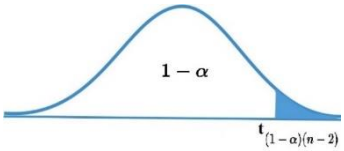
t_e, cae en la zona de rechazo de H_0, luego la albúmina circulante está influenciada determinantemente por el volumen del plasma.

6.8. PRUEBA DE HIPÓTESIS PARA CORRELACION ρ

H_0 : ρ = 0 , X y Y no están correlacionados

H_1 : ρ > 0, X y Y están correlacionados

α: 0.05



$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

En el análisis de correlación no es posible precisar la variable dependiente o independiente como se hace en el análisis de regresión, pudiendo intercambiar las posiciones.

Normalmente resulta imposible examinar una población bivariante, donde su correlación es:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{Cov}\,(X,Y)}{\sigma_X \sigma_Y}$$

y para estudiar el grado de correlación de las variables se toma una muestra donde el estimador de ρ es:

$$r = \frac{n\sum XY - \sum X \sum Y}{\{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]\}^{1/2}} \qquad \text{o} \qquad r = b\frac{S_X}{S_Y}$$

r^2 se denomina **coeficiente de determinación** y explica la cantidad de variación en la variable dependiente que es debida a la variable independiente se expresa adecuadamente por $100r^2$.

El coeficiente de correlación r tiene el mismo signo que el coeficiente de regresión b.

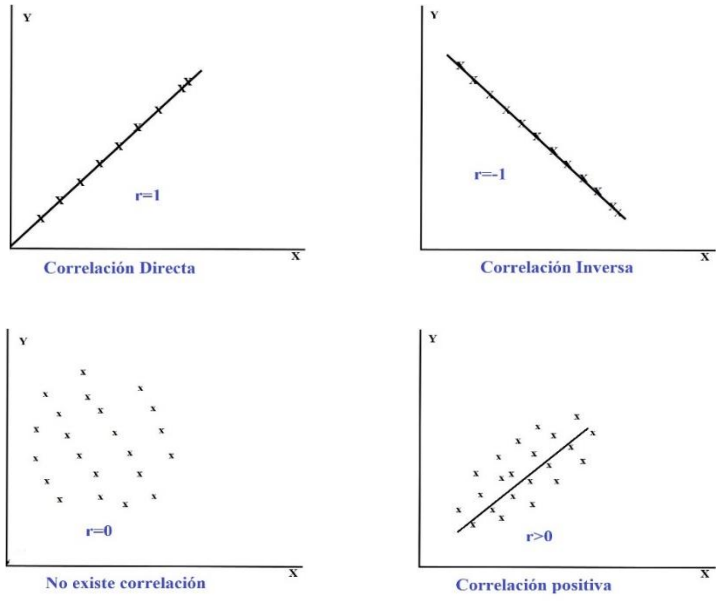
Si r = 1, las variables están correlacionadas perfectamente en forma creciente.

Si r =-1, la correlación es perfecta en forma decreciente – (mientras una variable crece la otra decrece).

Si r = 0, (lo que nos proponemos contrastar) las variables son independientes, no están correlacionadas.

El coeficiente r varía entre : $-1 \leq r \leq 1$

Figura 6.3 representaciones del grado de correlación entre variables.



Interdependencia establecida por “r”

Coeficiente r	Grado de interrelación
0.0 ± 0.2	Nula
± 0.2 ± 0.4	Poco significativa
± 0.4 ± 0.7	Significativa
± 0.7 ± 0.9	Bastante significativa
± 0.9 ± 1.0	Muy significativa

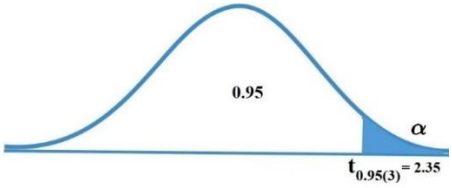
Ejemplo 1

Para mayor comprensión y secuenciación lógica trataremos el problema que ilustra la regresión.

H_0: $\rho = 0$ La albúmina circulante no está correlacionada con el volumen del plasma.

H_1: $\rho > 0$ La Albúmina circulante y el volumen del plasma están correlacionados positivamente.

α: 0.05



$$t_e = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.9448\sqrt{5-2}}{\sqrt{1-0.8926}} = 4.99$$

t_e, cae en la zona crítica: se rechaza H_0. La Albúmina circulante y el volumen del plasma están correlacionados.

El cálculo de r:

$$= \frac{5(1750) - (13.4)(648)}{\{[5(36.34) - (13.4)^2][5(84448) - (648)^2]\}^{1/2}} = 0.9448$$

$$\text{ó } r = b\frac{S_x}{S_y} = 31.2149x\frac{0.3271085}{10.807405} = 0.9448$$

$100r^2 = 100(0.9448)^2 = 89.26\%$ lo que indica que el 89.26% de las variaciones en la albúmina circulante es debido al volumen del plasma.

AUTOEVALUACIÓN VI

1. Los problemas más adecuados para hacer inferencias científicas son:

   A) Descriptivos ( ) B) Explicativos ( )

2. El valor P confiere significación estadística a los resultados:

   A) Cuando es mayor que α ( )

   B) Cuando es menor que α ( )

3. P<0.05 indica:

   A) Alta participación del azar ( )

   B) Baja participación del azar ( )

4. El valor P se usa en investigaciones:

   A) De tipo experimental ( )

   B) De tipo prospectivo ( )

5. A) El valor P depende del punto crítico ( )

   B) El punto crítico depende del valor P ( )

6. La hipótesis de investigación es:

   A) H_0 ( ) B) H_1 ( )

7. La máxima probabilidad de cometer un error tipo I es:

A) α ( ) B) 1-α ( )

8. H_0 se rechaza si el valor de la función experimental:

A) Cae en la zona de rechazo ( )

B) Cae en la zona de aceptación ( )

9. En la función pivotal se reemplaza:

A) Las estadísticas de la muestra ( )

B) Los parámetros de la población ( )

10. La función pivotal se elige de acuerdo:

A) Al tamaño de la muestra. ( )

B) Al nivel de significación α ( )

11. Para investigar se ha establecido, muestras grandes y muestras chicas. Según el tipo de prueba asigne el tamaño de muestra, $\geq$30, <30.

Z-Promedio (_____) t-Proporción (_____)

Z-Proporción (_____) χ^2 varianza (_____)

t-Promedio (_____) F-Varianzas (_____)

12. Marque con F o V en caso sea falso o verdadero.

    ( ) El investigador busca que la situación real del experimento coincida con el modelo estadístico.

    ( ) El investigador que obtiene un P< 0.05 da validez a sus hallazgos.

13. Se han efectuado dos series de contajes de colonias de bacterias en 9 placas Petri, cada una de dichas series realizada por un método diferente:

Método A	221	141	63	249	292	79	161	357	118
Método B	323	202	80	198	323	97	181	416	139

¿Existe diferencia significativa entre los métodos?

14. 20 personas con asma crónica de intensidad análoga se dividen en dos grupos de 11 y 9 respectivamente para probar la eficacia de dos fármacos. Los resultados se dan en una escala objetiva de 0 a 100:

Fármaco A	23	28	32	18	32	41	43	48	18	47	40
Fármaco B	52	58	53	55	64	58	67	72	48		

¿Existe diferencia significativa en cuanto a la eficacia de los fármacos?

15. La prueba χ^2 (chi cuadrado) no paramétrica es para:
    A) ¿Contrastar la media?
    B) ¿Contrastar la varianza?
    C) ¿La independencia de criterios?

16. En las pruebas de bondad de ajuste participan:
    A) ¿Un criterio?
    B) ¿Dos criterios?
    C) ¿Más de dos criterios?
17. Una hipótesis alternativa en la prueba de bondad de ajuste se enuncia como:

    ……

    Y una hipótesis nula en una tabla de contingencias se plantea como:

    ……

    La frecuencia esperada, f_e, en la bondad de ajuste se define por: ..... y en la independencia de criterios por

    ……
18. a) Un fabricante afirma que de cada 100 familias 25 consumen su producto, se seleccionó 200 familias y solo 50 de ellas eran consumidores; ¿es cierto lo afirmado por el fabricante? Justifique su respuesta.

    ……

b) Según la siguiente tabla ¿existe diferencia significativa entre afectados y no afectados?

Respuesta	Vacuna	
	Inoculados	**No inoculados**
Afectados	9	21
No afectados	24	12

19. En una muestra de 335 casos policiales de personas que han cometido alguna falta, se clasificó cada caso según la clase social del delincuente y su condena o amonestación, como muestra la tabla:

Condena	Clase Social			
	Alta	**Media**	**Baja**	**Total**
Cárcel	10	45	143	198
Multa	68	60	79	157
Total	78	105	172	355

En base a los datos, podríamos concluir que; ¿la clase social y la condena están relacionados?

……

20. Establecer la diferencia entre el análisis de regresión y correlación entre dos variables

……

21. Nombre cada uno de los elementos de las fórmulas y la utilización de las mismas.

$$t = \frac{(b-\beta_0)S_x\sqrt{n-1}}{S_{y/x}} \qquad t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

…….

…….

22. El diagrama de dispersión es importante porque:

……

El coeficiente de correlación indica:

……

El coeficiente de determinación indica:

……

Si $0.2 \leq r \leq 0.4$ el grado de asociación es:

…….

.

23. ¿Pueden 2 variables estar correlacionadas y no seguir un modelo lineal? Si ( ) No ( ).

¿Por qué?

……..……

24. Proponga 2 ejemplos donde se aplique el modelo de regresión:

1……

2……

…

25. Los datos que siguen son de 10 regiones de un pais en vías desarrollo:

Incidencia de bocio (Y) en la población (%)	60 75 50 55 45 33 50 52 35 30
Cantidad de Yodo (X) en el agua (g/litro)	02 02 03 05 07 08 08 08 10 12

a) Presente el análisis de dispersión

b) Grafique el modelo de estimación.

c) Explique $\hat{Y}$, en caso de que X = 1.5

d) ¿Cuál es el grado de interrelación?

e) ¿Puede afirmarse que la incidencia del bocio en la población se debe a la cantidad de yodo en el agua?

f) Pruebe las hipótesis $H_0 : \beta = 0$, $H_0 : \rho = 0$ con $\alpha = 0.05$.

g) Comente el coeficiente de determinación.

……

……

CAPÍTULO VII

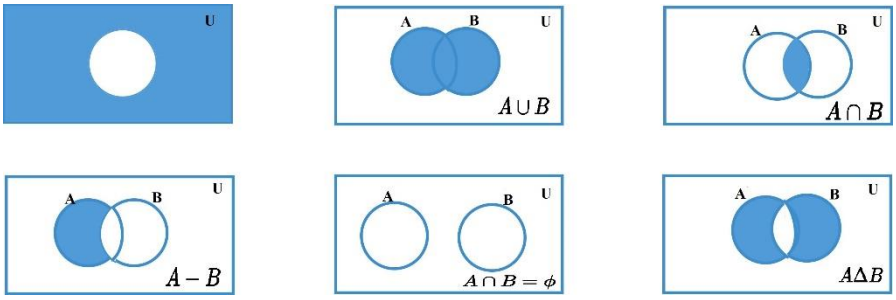
PROBABILIDADES

El estudio del cálculo de probabilidades se origina cuando se piensa en una experiencia aleatoria cuyos resultados son inciertos, esta experiencia real o conceptual precisa de una "medida de esa incertidumbre" llamada **probabilidad.**

Como la teoría de probabilidades es semejante a la teoría de conjuntos se puede establecer la siguiente analogía.

Teoría de Probabilidades	**Teoría de Conjuntos**
Espacio Muestral “Ω”	Conjunto Universal “U”
Suceso elemental	Elemento del conjunto
Evento o suceso	Subconjunto
Eventos incompatibles	Conjuntos disjuntos
Unión de eventos	Unión de conjuntos
Evento imposible	Conjunto vacío
Evento contrario	Conjunto complementario
Intersección de eventos	Intersección de conjuntos

Estas relaciones permiten utilizar la representación convencional de **Venn** muy familiar en la teoría de conjuntos, al cálculo de probabilidades.



En estos diagramas el cuadrilátero representa el espacio muestral “Ω”, los círculos representan los sucesos o eventos y el sombreado el resultado del experimento.

7.1. ANÁLISIS COMBINATORIO

Para calcular el número de resultados posibles de evento o el número de elementos de un conjunto se usa técnicas conocidas como **análisis combinatorio** basado en la factorial de un número**.**

Notación factorial del número n: n!

El producto de los enteros positivos de 1 a n inclusive, se denota por n! y se lee “n factorial”.

n! = n (n - 1) (n – 2) ... 3.2.1

Si **n = 1** tenemos: 1! = 1 (1 - 1)!

0! =1

5! = 5 (4) (3) (2) (1) = 120

Permutación nPr o P(n, r)

Es cualquier r-arreglo de un número n de objetos diferentes en un orden establecido, así: el número de permutaciones de r-arreglos de n objetos diferentes, sin sustitución está dado por:

$$P(n,r) = \frac{n!}{(n-r)!}$$

Ejemplo 1

¿De cuántas maneras se pueden depositar cuatro cartas en siete buzones, no depositando más de una carta en cada buzón?

Solución:

Aquí n = 7, r = 4, luego:

$$P\ (7,4) = \frac{7!}{(7-4)!} = \frac{7!}{3!} = 7\times6\times5\times4 = 840$$

Respuesta: de 840 maneras.

El número de permutaciones de n objetos tomados todos a la vez es n!, esto es: P(n, n) = n!

Ejemplo 2

¿De cuántas maneras 3 jóvenes y 2 señoritas pueden sentarse en una fila?

¿De cuántas maneras pueden sentarse si los jóvenes se sientan junto a las señoritas?

Solución

i) ¡Las cinco personas se pueden sentar en una fila de 5! maneras, es decir:

¡P (5,5) = 5! = 5 x 4 x 3 x 2 x 1 = 120

de 120 maneras.

ii) Existe dos formas para distribuirlos según el sexo

HHHMM o MMHHH, en cada caso los jóvenes pueden sentarse de 3! ¡maneras y las señoritas de 2!, luego en total hay: 2 x 3! x 2! = 2 x 6 x 2 = 24 maneras.

Combinación ${}^{n}C_{r}$ o $\left(\frac{n}{r}\right)$

Es la colección de un número de objetos sin considerar algún orden. Luego el número de combinaciones de un conjunto de n objetos diferentes tomando r al mismo tiempo está dado por:

$$ {}^{n}C_{r} = \frac{n!}{(n-r)!r!} $$

Ejemplo 1

Un alumno debe responder necesariamente 8 de 10 preguntas en su evaluación.

i) ¿Cuántas formas de elegir tiene?

ii) ¿Cuántas, si las 3 primeras preguntas son obligatorias?

Solución:

i) Las 8 preguntas pueden seleccionarse de:

$$^{10}C_8 = \frac{10!}{(10-8)!8!} = \frac{10!}{2!8!} = \frac{10\times 9\times 8}{2!} = 45$$

ii) Si contesta las 3 primeras preguntas, entonces puede escoger las otras 5 de las 7 preguntas que quedan de:

$$^{7}C_5 = \frac{7!}{(7-5)!5!} = \frac{7\times 6\times 5!}{2!5!} = \frac{7\times 6}{2!} = 21 \quad \textit{maneras}$$

* Para todo r ≤ n se define:

$$\binom{n}{r} = \frac{n(n-1)(n-2)\ldots(n-r+1)}{r(r-1)\ldots 3\times 2\times 1} = \frac{n!}{(n-r)!\,r!}$$

Estos números son llamados coeficientes del binomio:

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r}b^r$$

Así

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{1-2}a^{n-2}b^2 + \cdots + nab^{n-1} + b^n$$

Para aclarar, combinaciones y permutaciones; determinemos el número de combinaciones de las 3 letras a, b, c tomadas de dos en dos. Vemos que cada combinación que contenga dos letras produce 2 permutaciones, así:

Combinaciones	Permutaciones
ab	ab, ba
ac	ac, ca
bc	bc, cd

Luego:

El número de combinaciones multiplicando por 2, equivale al número de permutaciones totales.

$$ {}^3C_2 \times 2 = {}^3P_2 \text{ ó } {}^3C_2 = \frac{{}^3P_2}{2} $$

Entonces

$$ {}^3P_2 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = 3! = 6 \text{ permutaciones} $$

$$ {}^3C_2 = \frac{3!}{(3-2)!2!} = \frac{3!}{1!2!} = 3 \text{ combinaciones.} $$

Puesto que cada combinación de n objetos tomados “r” a la vez determina a r! permutaciones de los objetos, se deduce que:

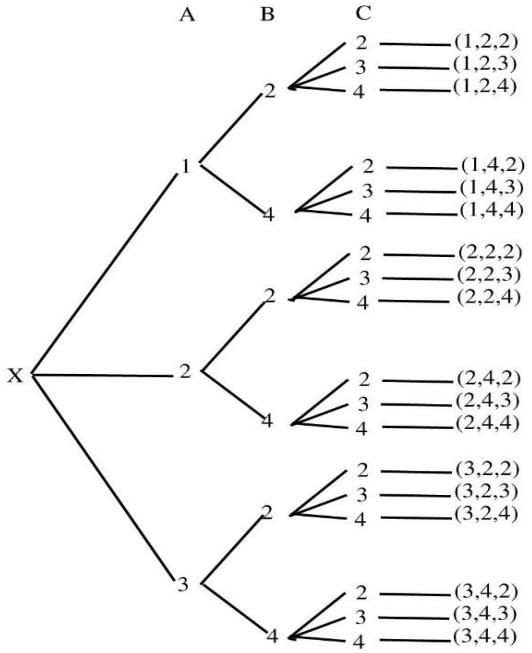
$$ {}^nP_r = {}^nC_r \,.\, r! \quad , \qquad {}^nC_r = \frac{{}^nP_r}{r!} = \frac{n!}{(n-r)!r!} \quad . $$

7.2. DIAGRAMA DE ÁRBOL

Los diagramas de árbol sirven para ilustrar el principio de **continuidad** o la forma como se relacionan los elementos de los conjuntos de un acontecimiento, estos diagramas fundamentalmente permiten conseguir resultados para **contar** el número de arreglos diferentes.

Ejemplo 1

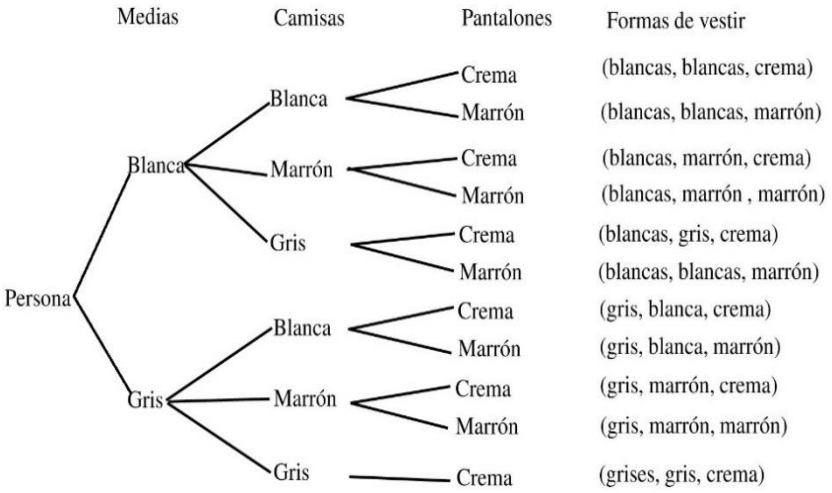
Sean los conjuntos A = {1,2,3}, B = {2,4}, C = {2,3,4} el conjunto producto A x B x C tendrá 18 elementos, como se muestra en la parte terminal del diagrama del árbol.



Ejemplo 2

Una persona tiene dos pares de medias, blancas y grises; tres camisas, blanca, marrón y gris y dos pantalones, crema y marrón.

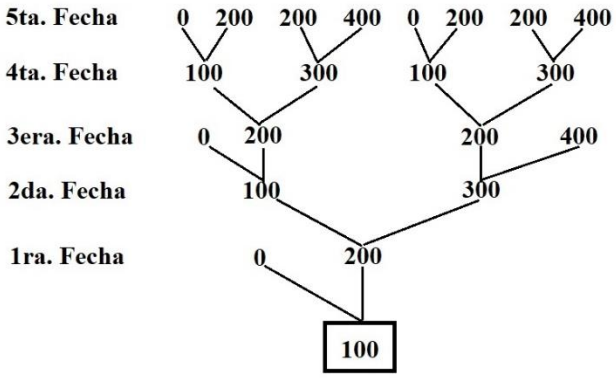
Queremos saber cuántas maneras diferentes puede vestirse esta persona.



Luego la persona puede vestirse de doce maneras diferentes, tal como se observa en el terminal del diagrama.

Ejemplo 3

Un hincha de fútbol puede apostar hasta 5 fechas cuando juega su equipo; en cada partido gana o pierde 100 soles. El hincha inicia con 100 soles y dejará de apostar si antes del quinto partido pierde todo su dinero o si gana 300 soles (o tiene 400 soles). ¿En cuantos partidos va la apuesta?



Nuestro diagrama describe el camino en que la apuesta puede ocurrir, cada número del diagrama denota la cantidad que el hincha tiene en ese punto; observamos que la apuesta puede suceder de 10 maneras diferentes (cada terminal del árbol). Como se ve suspenderá la apuesta antes que llegue a la quinta fecha sólo en tres casos.

7.3. ESPACIO MUESTRAL DISCRETO "Ω"

Es el conjunto de resultados del experimento, tal que:

i) cada elemento del conjunto representa un resultado de experimento.

ii) A cualquier resultado del experimento le corresponde uno y sólo uno de los elementos del espacio muestral.

Suceso "W" (evento). Es un subconjunto de un espacio muestral.

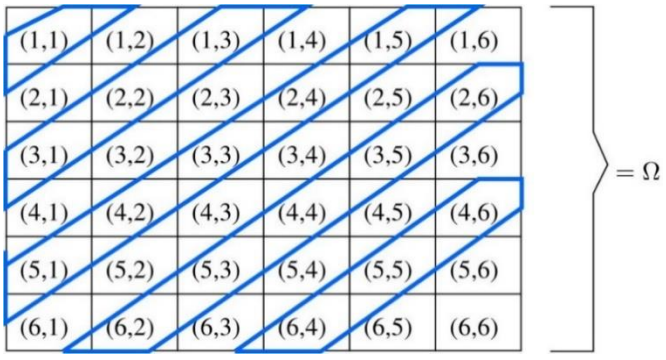
Suceso elemental. (Punto muestral). Suceso que no puede ser descompuesto.

Número de elementos del espacio muestral. n(Ω)

La determinación del número de elementos de un espacio muestral es de suma importancia, se denota por n(Ω) y para un suceso cualquiera tal como A, por n(A) y no viene a ser sino la totalidad de sucesos elementales o puntos muestra del suceso; de modo que si tenemos el suceso imposible correspondiente el conjunto vacío Ø, n(Ø) = 0

Explicación:

Si el experimento es "Lanzamiento de un par de dados corrientes" uno rojo y el otro blanco, sus resultados pueden ser.



En un experimento se puede interesar en:

Conocer el número de resultados posibles del experimento.

Ω = espacio muestral

Ω = {(1,1), (1,2), ..., (6,6)}

n (Ω) = 36

Que la suma de los puntos sea 10.

A = {suceso donde la suma de puntos sea 10}

A = { (4,6), (5,5), (6,4) }

n(A) = 3

La suma de puntos sea 3.

B = {(1.2), (2,1)}

n(B) = 2

Que el dado rojo muestre 6 y el blanco 2.

C = {(6,2)}, n(C)= 1

Que en el dado rojo aparezca 3 y 4 en el dado blanco.

D = { (3,3) , (3, 4) , (4,3) , (4,4) }

n(D) = 4.

Esta determinación de subespacios dentro de Ω se conoce como *muestreos de **Euler***.

El reconocimiento apropiado del espacio muestral es básico para la determinación de la probabilidad de ocurrencia o no ocurrencia de un suceso.

Consideremos el siguiente experimento:

Lanzar 2 monedas al aire, si c = cara y s = sello; el espacio muestral o número total de casos posible será:

$$\Omega = [CC, CS, SC, SS]$$

Al arrojar las 2 monedas a la vez o una por una se tiene un sub conjunto de Ω; por ejemplo, A = {CC}, suceso elemental donde las 2 monedas caen cara, de igual modo el resultado {SS} le corresponde al suceso elemental de obtener 2 sellos.

Si deseamos observar el suceso "al menos una cara" entonces el sub conjunto correspondiente será:

B = {CS, SC, CC}

n(B) = 3

Este subconjunto o subespacio muestral puede ser particionado.

Ejemplo 1

Suceso en el cual la segunda moneda caiga cara.

$$B_1 = \{SC, CC\}$$

$$n(B_1) = 2$$

El espacio muestral para una familia con 3 hijos será:

$$\Omega = \{HHH, HHM, HMH, MHH, HMM, MHM, MMH, MMM\}$$

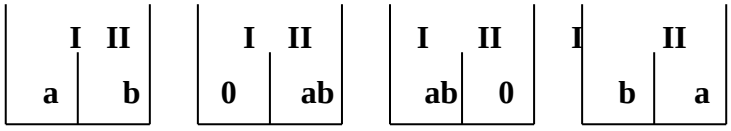
n (Ω) = 8

Para el suceso que; la familia tenga al menos 2 hijos hombres, el sub espacio será:

$$W_1 = \{HHH, HHM, HMH, MHH\} \text{ con } n(W_1) = 4$$

Para el suceso que; la familia no tenga hijos hombres será:
$W_2 = \{MMM\}$ con $n(W_2) = 1$.

Si 2 objetos pueden diferenciarse; en 2 celdas numerada pueden ubicarse del modo siguiente:



y el espacio muestral quedará:

$$W = \{(a, b), (0, ab), (ab, 0), (b, a)\}$$

7.4. PROBABILIDAD DE UN SUCESO

Si A es un SUCESO entonces su probabilidad de ocurrencia está dada por:

$$P(A) = \frac{n(A)}{n(\Omega)}$$

n(A) = Número de resultado a favor

n(Ω) = Número de resultados posibles

P(A) = Probabilidad de A

Observe que P(A) cumple con los axiomas:

. $P(A) \geq 0$

. $P(\Omega) = 1$

. $P(A \cup B) = P(A) + P(B)$ si A y B son sucesos mutuamente excluyentes (disjuntos).

El primer axioma nos dice que la probabilidad de un suceso debe ser número positivo o cero.

El segundo axioma establece que la probabilidad de un suceso seguro es 1.

El tercer axioma permite determinar la probabilidad de cualquier suceso (evento) siempre que se conozcan las probabilidades de los sucesos elementales.

De estos axiomas se pueden establecer que:

$$0 \leq P(A) \leq 1$$

de manera que si P(A) se acerca a 1, la ocurrencia será muy probable y menos probable si se acerca a cero.

Si los sucesos son mutuamente excluyentes su intersección es vacía y consecuentemente P(Ø) = 0.

La probabilidad de la no ocurrencia del suceso A está dado por

$$P(A') = 1 - P(A)$$

puesto que

$$P(\Omega) = 1$$

Ejemplos

¿Cuál es la probabilidad de que una familia con 2 hijos, no tenga hijos hombres?

Solución

Ω = {HH, HM, MH, MM} (espacio muestral)

A = {ningún hijo hombre} (suceso)

A = {MM}

Luego $P(A) = \frac{n(A)}{n(\Omega)} = \frac{1}{4}$.

¿Entre familia con 3 hijos?

Solución:

Ω = {HHH, HHM, HMH, MHH, HMM, MHM, MMH, MMM}

A = {MMM}

Entonces $P(B) = \frac{n(B)}{n(\Omega)} = \frac{1}{8}$.

¿Cuál es la probabilidad tener 10 como suma de puntos en el lanzamiento de un par de dados?

Solución:

Ω = {(1,1), (1,2), ... , (6,6)}

n (Ω) = 36

A = {(6,4), (5,5), (4,6)}

n (A) = 3

Luego la probabilidad pedida $P(A) = \frac{n(A)}{n(\Omega)} = \frac{1}{36} = \frac{1}{12}$.

Seis jóvenes, A, B, C, D, E, F, están seleccionados para formar un comité de 3 personas, ¿Cuál es la probabilidad de que se encuentren A y B conformando el comité?

Solución

Ω = {ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF, DEF,}

n (Ω) = 20

S = {ABC, ABD, ABE, ABF}

n (S) = 4

Entonces

$$P(S) = \frac{n(S)}{n(\Omega)} = \frac{4}{20} = \frac{1}{5}$$

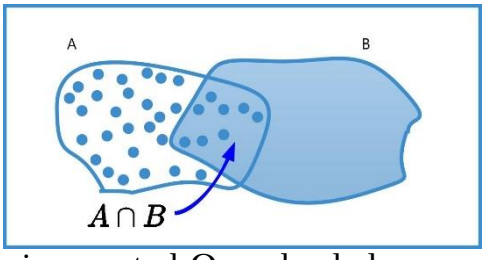
.

La probabilidad de que no se encuentren será:

$$P(S)' = 1 - P(S) = 1 - \frac{1}{5} = \frac{4}{5}$$

7.5. PROBABILIDAD DE LA UNIÓN DE SUCESOS

P(A) y P(B) son probabilidades de ocurrencia para los sucesos A y B entonces P(A∪B) representa la probabilidad de que ocurra A o B.



Dado un espacio muestral Ω en donde los eventos A y B son subconjuntos que tienen una porción común llamada intersección y denotada por A ∩ B.

Como un punto está asociado una probabilidad, la suma de probabilidades para los puntos en A corresponde a P(A) y la suma de probabilidad para los puntos en B corresponde a P(B); entonces:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

donde $P(A \cap B)$ representa la probabilidad de que ocurran simultáneamente A y B. En la suma de las probabilidades P(A)+P(B) se incluyen dos veces los puntos de $A \cap B$; por lo tanto, $P(A \cap B)$ debe restarse de la suma P (A) + P (B).

Para demostrar este teorema definimos: $A \cup B$ y B como sucesos mutualmente excluyentes de manera que en:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (I) *

$A \cup B = A \cup (A' \cap B)$

$B = (A \cap B) \cup (A' \cap B)$

Aplicando probabilidades

$P(A \cup B) = P(A \cup (A' \cap B))$

$P(A \cup B) = P(A) + P(A' \cap B)$ (I)

y

$$P(B) = P((A \cap B) \cup (A' \cap B))$$

$$P(B) = P(A \cap B) + P(A' \cap B) \qquad \text{(II)}$$

de esta relación tenemos que:

$$P(B) - P(A \cap B) = P(A' \cap B)$$

y sustituyendo $P(A' \cap B)$ en (I) se tiene:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

que es la relación (I)*

Si A y B son disjuntos su intersección es vacía entonces el teorema anterior se reduce a una forma más simple del *Teorema de la Suma* (Teorema de la probabilidad total)

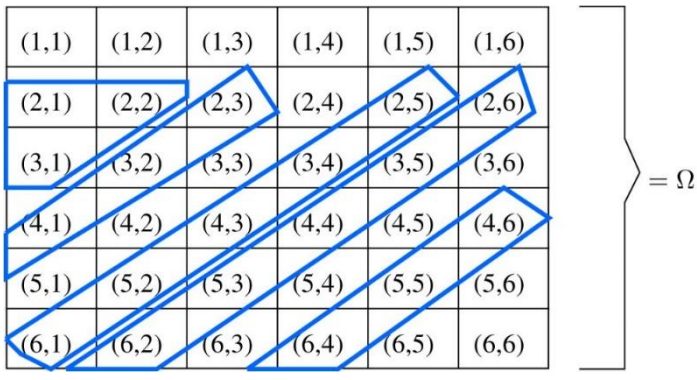
$$P(A \cup B) = P(A) + P(B)$$

Ejemplo 1

¿Cuál es la probabilidad de que en el lanzamiento de dos dados obtengamos, una suma par o una menor que cuatro?

Solución

Para observar con claridad graficaremos el espacio muestral.



Ω = { (1,1), (1,2), ... , (6,6)} conjunto de pares ordenados (x, y), n (Ω) = 36

A = {(x, y) / x + y = un número par}

A = {(1,1), (2,2), ... , (6,6)}

n (A) = 18

B = {(x, y) / x + y < 4}

B = {(1,1), (2,1), (1,2)}

n (B) = 3

Los resultados que cumplan las dos condiciones correspondan a la parte en común que la denotamos por A $\cap$ B.

A $\cap$ B = {(x, y) / x + y es par menor que 4}

A $\cap$ B = {(1, 1)}

n (A $\cap$ B) = 1

Como a cada punto del espacio muestral se asocia una probabilidad de 1/36 entonces la probabilidad del evento será:

P (A $\cup$ B) = P (A) + P (B) – P (A $\cap$ B)

$$= \frac{n(A)}{n(\Omega)} + \frac{n(B)}{n(\Omega)} - \frac{n(A \cap B)}{n(\Omega)}$$

$$= \frac{18}{36} + \frac{3}{36} - \frac{1}{36}$$

$$P(A \cup B) = \frac{20}{36} = \frac{5}{9}.$$

Para el mismo experimento ¿Cuál será la probabilidad de que por lo menos uno de ellos muestra un número mayor que 4?

Solución:

A = {(x, y) / x sea mayor que 4}

A = { (5, 1) , (5, 2) , (5, 3) , (5, 4) , (5, 5) , (5, 6) , (6, 1) , (6, 2) , (6, 3) , (6, 4), (6, 5) , (6, 6) };

n (A) = 12

B = {(x, y) / y sea mayor que 4}

B = { (1, 5) , (2, 5) , (3, 5) , (4, 5) , (5, 5) , (6, 5) , (1, 6) , (2, 6) , (3, 6) , (4, 6) , (5, 6) , (6, 6) } ;

n (B) = 12

$A \cap B$ = {(x, y) / x y y sean mayores que 4}

$A \cap B$ = {(5, 5), (5, 6), (6, 5), (6, 6)}

N ($A \cap B$) = 4, de modo que

P ($A \cup B$) = P (A) + P (B) – P ($A \cap B$)

$$= \frac{12}{36} + \frac{12}{36} - \frac{4}{36} = \frac{20}{36} = \frac{5}{9}$$

En el mismo experimento ¿Cuál es la probabilidad de obtener como suma 6 u 11?

Solución:

Ω = {(1,1), (1,2), ... , (6,6)}

n (Ω) = 36

A = {(x, y) / x + y = 6}

n (A) = 5

B = {(x, y) / x + y = 11}

B = {(6,5), (5,6)}

n (B) = 2

Como los eventos son mutuamente excluyentes su intersección es vacía entonces su probabilidad será:

$$P(A \cup B) = P(A) + P(B)$$

$$= \frac{n(A)}{n(\Omega)} + \frac{n(B)}{n(\Omega)} = \frac{5}{36} + \frac{2}{36} = \frac{7}{36}$$

7.6. PROBABILIDAD CONDICIONAL

Se trata de encontrar la probabilidad de suceso teniendo cierta información acerca de el, o simplemente se desea conocer la probabilidad de ocurrencia de un suceso dado la ocurrencia de otro.

Luego la probabilidad de A dado B se denota por P(A/B) y queda definida por:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Aquí A, B y A ∩ B son sucesos de un espacio muestral Ω.

El suceso B establecido como **condición** se convierte en un **espacio muestral reducido** asegurándonos de este modo que P(B) = 1 por ser un suceso ya ocurrido y entonces P(A Ω B) es tomada en relación a B.

P (A / B) cumple con:

- P (A / B) ≥ 0
- P (Ω / B) = 1

- P [(A ∪ B) / C] = P (A / C) + P (B / C)

El primer axioma se satisface inmediatamente, pues P(A∩B) ≥ 0 y si dividimos por P(B) que es mayor que cero se tiene:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \geq 0$$

Para el segundo axioma debe observarse que Ω ∩ B = B luego:

$$P(\Omega/B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

En el tercer axioma, como A∩B = ∅ por ser disjuntos tenemos A∩C y B∩C son mutuamente excluyentes y luego

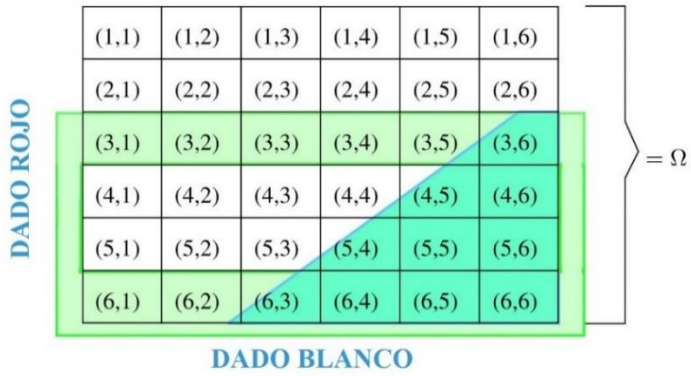
$$\begin{aligned} P[A \cup B)/C] &= \frac{P \quad [(A \cup B) \cap C]}{P(C)} \\ &= \frac{P \quad [(A \cap C) \cup (B \cap C)]}{P(C)} \\ &= \frac{P(A \cap C) + P(B \cap C)}{P(C)} \\ &= \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C)}{P(C)} \end{aligned}$$

P[(A∪B)/C] = P(A/C) + P(B/C), lo que deseábamos verificar.

Ejemplo 1

¿Cuál es la probabilidad de que, en el lanzamiento de un dado rojo y un blanco, la suma de puntos sea mayor que 8; si se sabe que el dado rojo muestra un número divisible por 3?

Solución:



(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Ω = {(1, 1), (1, 2), ... , (6, 6)}, sabiendo que x son puntos del dado rojo y son puntos del dado blanco.

A = {(x, y) / x+y sea mayor que 8}

A = {(3,6), (4,5), (4,6), (5,4), (5,5), (5,6), (6,3), (6,4), (6,5), (6,6)}

n (A) = 10

B = {(x, y) / x sea un número divisible por 3}

B = {(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

n (B) = 12

A ∩ B = {(x, y) / x+y sea mayor que 8 y x sea un número divisible por 3}

$A \cap B = \{(3, 6), (6, 3), (6, 4), (6, 5), (6, 6)\}$

$n(A \cap B) = 5$

Ahora sí, la probabilidad de que la suma de puntos sea mayor que 8 si el dado rojo muestra un número divisible por 3, será:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

P (B) = 1 (suceso que ya ocurrió) y

$$P(A \cap B) = \frac{n(A \cap B)}{n(B)} = \frac{5}{12}$$ (relativo a B)

se tiene: $P(A/B) = \frac{5}{12}$.

Ejemplo 2

En cierta sección, 25% de los estudiantes desaprobaron matemáticas, 15% desaprobaron química y 10% desaprobaron matemática y química. Seleccionamos un estudiante al azar.

i) Si desaprobó química. ¿Cuál es la probabilidad de que desaprobó matemáticas?

ii) Si desaprobó matemáticas, ¿Cuál es la probabilidad de que desaprobó química?

iii) ¿Cuál es la probabilidad de que desaprobó matemática y química?

Solución:

A = {Estudiantes que desaprobaron matemáticas}

P (A) = 0.25

B = {Estudiantes que desaprobaron química}

P (B) = 0.15

A∩B={Estudiantes que desaprobaron matemáticas y química}

P (A∩B) = 0.10

Entonces:

i) La probabilidad de que el estudiante desapruebe matemáticas, dado que haya desaprobado química es

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.10}{0.15} = \frac{2}{3}$$

ii) La probabilidad de que el estudiante desapruebe química, dado que haya desaprobado matemática es

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{0.10}{0.25} = \frac{2}{5}$$

iii) La probabilidad de que desaprobó matemáticas o química es

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = 0.25 + 0.15 - 0.10 = \frac{3}{10}$$

7.7 TEOREMA DE BAYES

Sea el conjunto de eventos $\{E_1, E_2, \ldots E_n\}$ una partición del espacio muestral Ω, donde $P(E_i) \neq 0$, $i = 1, 2, \ldots, n$, entonces, para cualquier evento E para el que

$P(E) \neq 0$ y para $1 \leq k \leq n$

$$P(E_k / E) = \frac{P(E_k)\, P(E / E_k)}{\sum_{i=1}^{n} P(E_i)\, P(E / E_i)}$$

Para la demostración de la probabilidad condicional se tiene:

$$P(E_K / E) = \frac{P(E_K \cap E)}{P(E)}$$

sustituyendo

$$P(E_K \cap E) = P(E_K) P(E / E_K)$$

$$P(E) = \sum_{i=1}^{n} P(E_i) P(E / E_i)$$

llegamos a:

$$P(E_K / E) = \frac{n(E_K) P(E / E_K)}{\sum_{i=1}^{n} P(E_i) P(E / E_i)}$$

$$P(E_i / E) = \frac{n(E_i) P(E / E_i)}{\sum_{i=1}^{n} P(E_i) P(E / E_i)}$$

Ejemplo 1

En una fábrica; una máquina A produce el 30%, una máquina B el 25% y la máquina C el 45%; el 1% de la producción total de la máquina A es defectuosa, el 1.2% de la producción total de la máquina B es también defectuosa y de C el 2%.

En un día las 3 máquinas producen 10,000 artículos. Un artículo es seleccionado al azar de la producción total de un día resulta ser defectuoso. ¿Cuál es la probabilidad de que el artículo sea producido por la máquina A?, por B?, ¿por C?

Solución:

E= {El artículo seleccionado es defectuoso}

E_1={producción de la máquina A}

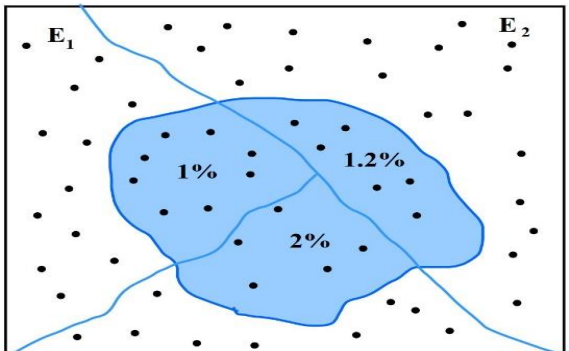
P (E_1) = 30/100

E_2 = {producción de la máquina B}

P(E_2) =25/100

E_3 = {producción de la máquina C}

P (E_3) = 45/100



entonces la probabilidad de que el artículo defectuoso lo haya producido la máquina A está dada por:

$$P(E_2/E)=\frac{P(E_1)P(E/E_1)}{P(E_1)P(E/E_1)+P(E_2)P(E/E_2)+P(E_3)P(E/E_3)}$$

que sustituyendo

$$P(E/E_1)=1/100 P(E/E_2)=\frac{1.2}{100} \qquad \text{y} \qquad P(E/E_3)=\frac{2}{100}$$

resulta:

$$P(E_1/E)=\frac{\left(\frac{30}{100}\right)\left(\frac{1}{100}\right)}{\left(\frac{30}{100}\right)\left(\frac{1}{100}\right)+\left(\frac{25}{100}\right)\left(\frac{1.2}{100}\right)+\left(\frac{45}{100}\right)\left(\frac{2}{100}\right)}$$

$$P(E_1/E)=\frac{1}{5}$$

La probabilidad de que el artículo defectuoso lo haya producido la maquina B es:

$$P(E_2/E)=\frac{P(E_2)P\left(E/E_2\right)}{P(E_1)(E/E_1)+P(E/E_2)+P(E_3)P(E/E_3)}$$

$$\mathrm{P(E_2/E)} = \frac{(2.5/100)(1.2/100)}{(30/100)(1/100) + (25/100)(1.2/100) + (45/100)(2/100)}$$

$$P(E_2 / E) = \frac{1}{5}$$

Para C su probabilidad es:

$$P(E_3 / E) = \frac{3}{5}$$

La suma de las 3 probabilidades necesariamente tiene que ser 1.

Ejemplo 2

Carmen, Rosa y Elena, lavan la vajilla familiar. Carmen emplea el 40% del tiempo, Rosa y Elena comparten igualmente el 60% restante. La probabilidad de que una vajilla sea quebrada cuando Carmen lava es 0.02, para Rosa es 0.03 y para Elena 0.02. Los padres no saben quién está lavando la vajilla y escuchan quebrarse una.

¿Cuál es la probabilidad de que Carmen la esté lavando? ¿Rosa? ¿Elena?

Solución

Considerando los sucesos

E = {una vajilla se quiebra}

E_1 = {Carmen está lavando}

E_2 = {Rosa está lavando}

E_3 = {Elena está lavando}

Luego la probabilidad de que Carmen estuvo lavando será:

$$P(E_1 / E) = \frac{P(E_1)P(E / E_1)}{P(E)}$$

$$P(E) = \sum_{i=1}^{3} P(E_i)P(E / E_i)$$

$$= P(E_1)P(E / E_1) + P(E_2)P(E / E_2) + P(E_3)P(E / E_3)$$

$$= (0.40)(0.02) + (0.30)(0.03) + (0.30)(0.02)$$

$$P(E) = 0.023$$

$$P(E_1 / E) = \frac{(0.40)(0.02)}{0.023} = \frac{0.008}{0.023} = \frac{8}{23}$$

Para Rosa:

$$P(E_2 / E) = \frac{P(E_2)P(E / E_2)}{P(E)}$$

$$P(E_2 / E) = \frac{(0.30)(0.03)}{0.023} = \frac{0.009}{0.023} = \frac{9}{23}$$

Para Elena:

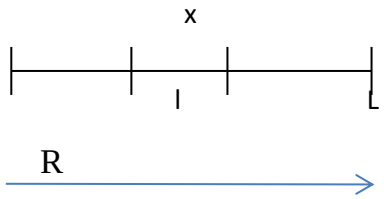
$$P(E_3 / E) = \frac{6}{23}$$

7.8. ESPACIO MUESTRAL CONTÍNUO

Si la variable experimental es de tipo **continua** tendremos un espacio muestral continuo. Los espacios muestrales que consideramos, son aquellos de medida **geométrica** finita tales como **longitud, área o volumen** y en los cuales un punto se selecciona al azar.

Si A es un evento, su probabilidad de ocurrencia está dado por:

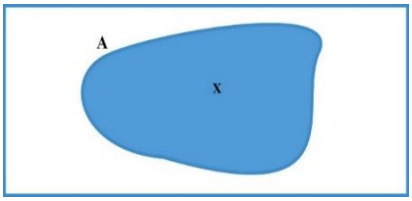
Si es en "R", la probabilidad de que el punto seleccionado pertenezca a 1 es:



$$P(A) = \frac{\text{longitud } \mathbf{l}}{\text{longitud } \mathbf{L}}$$

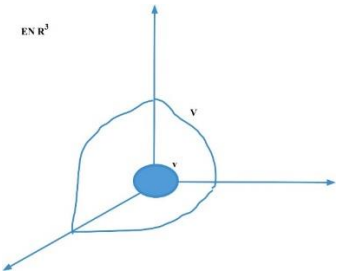
Si es en "R^2", la probabilidad de que el punto elegido pertenezca a la región A será:

$$P(A) = \frac{superficie\ de\ \boldsymbol{A}}{superficie\ de\ \boldsymbol{S}}$$



Si es en "R^3", la probabilidad de que el punto elegido pertenezca al volumen $\boldsymbol{V}$ es:

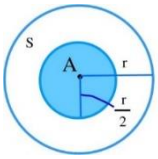
$$P(A) = \frac{volumen\, \boldsymbol{v}}{volumen\, \boldsymbol{V}}$$



Ejemplo 1

En un círculo de radio r se traza un círculo de radio r/2. ¿Cuál es la probabilidad de que al marcar un punto al azar pertenezca al círculo menor?

Solución:



A = {graficar un punto al azar

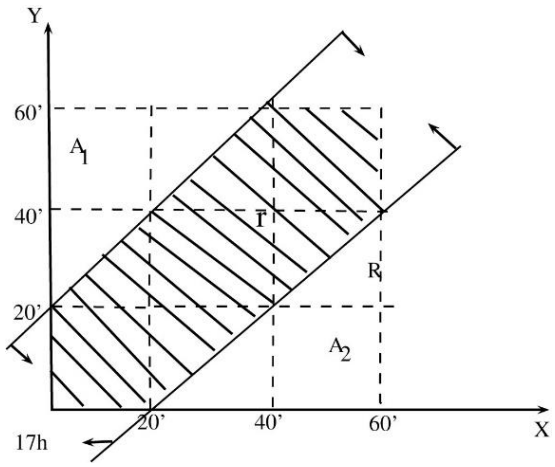
$$P(A) = \frac{superficie\ de\ \boldsymbol{A}}{superficie\ de\ \boldsymbol{S}} \qquad P(A) = \frac{\pi(\frac{r}{2})^2}{\pi r^2} = \frac{1}{4}$$

Ejemplo 2

Las personas A y B han acordado encontrarse en una esquina determinada entre las 17 y 18 horas, con la condición

de que si uno de ellos llega primero, espera a la otra durante 20 minutos.

¿Cuál es la probabilidad de que se encuentren?



X = tiempo en que llega A: 17 horas ≤ x ≤ 18 horas

Y = tiempo en que llega B : 17 horas ≤ y ≤ 18 horas

El tiempo de espera: $|X - Y| \leq 20'$

r = región favorable a la cita

r = { $(x, y) / |x - y| \leq 20'$ }

R = Región probable a la cita

R = { (x, y)/17 horas ≤ x ≤ 18 horas, 17 horas ≤ y ≤ 18 horas}

Si E = {suceso en que A y B se encuentran}

$$P(E) = \frac{superficie\ de\ \boldsymbol{r}}{superficie\ de\ \boldsymbol{R}}$$

sup R = 60´ x 60´ = 3600´2

$\sup r = \sup R - \sup (A_1 + A_2)$

$$= \sup R - \sup\left[2\left(\frac{b \times h}{2}\right)\right]$$

$$\sup r = 3600^{\prime 2} - 2\left(\frac{40^{\prime} x 40^{\prime}}{2}\right) = 2000^{\prime 2}$$

$$P(E) = \frac{2000^{\prime 2}}{3600^{\prime 2}} = \frac{5}{9}$$

La probabilidad de que no se encuentran $P(E^{\prime}) = 1 - P(E) = \frac{4}{9}$.

AUTOEVALUACIÓN VII

1. Muestre todas las señales que se pueden obtener con seis banderas, dos de las cuales son rojas, tres son verdes y una morada.

2. Con Pedro Paco y Genaro, ¿cuántos grupos de dos personas se pueden formar.

3. En una tienda se venden 8 sillas de color verde y 8 de color rojo, también hay 10 de madera y seis de plástico y se sabe que de las seis de plástico hay cinco que, son verdes, ¿qué tan probable es que si se escoge una silla cualquiera sea verde o de plástico?

4. Una fábrica de celulares dispone de la máquina A y B que elaboran el 60 % y el 40% de la producción; el % de celulares defectuosos que produce cada máquina es de 5% y 10% respectivamente. ¿Calcular la probabilidad de que el celular haya sido producido por la máquina A sabiendo que es defectuoso?

AUTOEVALUACION FINAL

1. Se ha obtenido un tamaño de nuestra definitiva con información de muestreo piloto; si la variable en estudio es continua con distribución normal, el estadístico de prueba es:

   A) $\frac{x-u_0}{\frac{\sigma}{\sqrt{n}}}$

   B) $\frac{p-P_0}{\sqrt{\frac{pq}{n}}}$

   C) $\frac{x-u_0}{\frac{s}{\sqrt{n}}}$

2. Los factores que condicionan el uso de las pruebas estadísticas son:

   A) Modelo estadístico - tipo de población - tipo de medición - potencia de la prueba-– Tamaño de muestra.

   B) Realidad problemática – objetivo de la investigación - hipótesis planteada – grado de confiabilidad.

3. El modo correcto de seleccionar la prueba la contrastación científica es:

   C) Hacer el proyecto de investigación → ejecutar la investigación → buscar la prueba.

   D) Hacer el proyecto de investigación → buscar la prueba → ejecutar la investigación.

4. Asigne la función pivotal para la prueba:

   σ^2 conocido (promedio):

   ……

   σ^2 desconocido (promedio):

   ……

   $\sigma^2_1 \neq \sigma^2_2$ conocido (diferencia, de promedios):

   …….

   .

   $\sigma^2_1 = \sigma^2_2 = {}_\sigma{}^2$ desconocido (diferencia de promedios):

   ……

Discuta los resultados:

5. Para comprobar si el perejil efectos teratógenos, se experimentó en 100 ratas de laboratorio dando P = 0.01%.

……

6. La hipótesis de que el laboratorio produce cáncer pulmonar es realmente verdadera. Un estudio realizado es fumadores da $P > 0.05$.

……

7. La hipótesis. La TV. determina delictividad humana es realmente falsa, un estudio da $P<0.01$.

……

8. En las hipótesis que se siguen, identifique H_0 y H_1:
   ( ) La altura influye en la presión arterial de los pobladores.
   ( ) La presión arterial del poblador de las ciudades sobre los 3500 metros es 20 milímetros mayor que la del poblador costeño.
   ( ) La miseria es causa de la mayor parte de delitos comunes.
   ( ) La miseria es causa del 70% de delitos.
   ( ) El estado de gestación altera la concentración del magnesio sérico.
   ( ) La concentración media del magnesio sérico es igual en gestantes y no gestantes.
9. Para los ejemplos tratados en el ejercicio anterior, identifica el Diseño, el modelo, la hipótesis de la investigación y el estadístico de prueba.

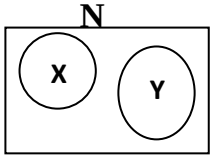
10. Todas las pruebas que se han tratado se derivan de los modelos:(señalados en el texto).

    a. Comparación de un estadístico con un parámetro poblacional
    b. Comparación de dos estadísticos.
    c. Comparación de frecuencias observadas y teóricas (esperadas).
    d. Relación entre variables.

Analice estadísticamente los diagramas:

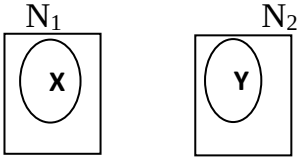
a) 

……

b) 

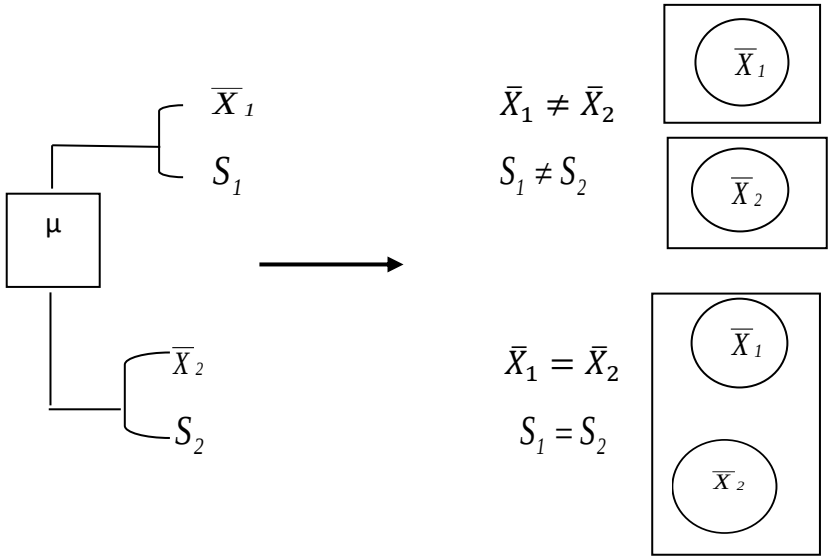
$S_X = S_Y$

……..

c) 

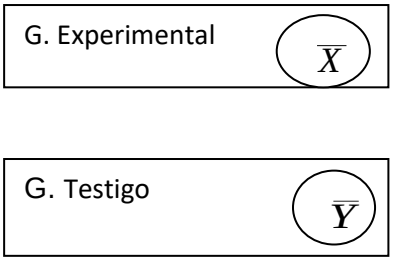
$S_X \neq S_Y$

d)



……

e)



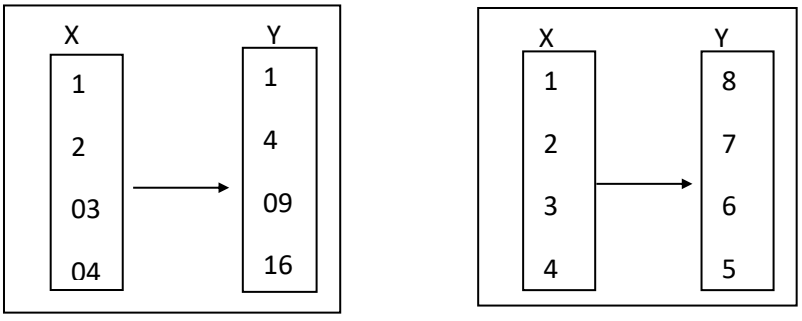
p< 0.05

……

p>0.05

……

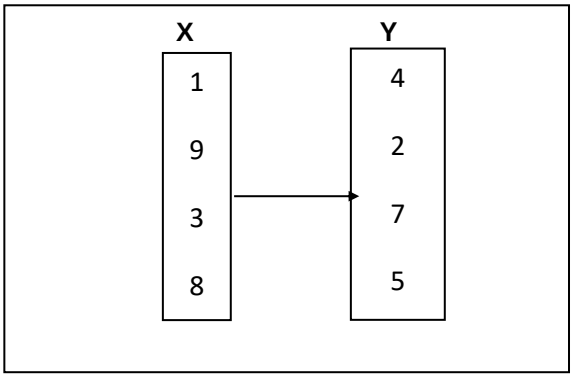
f)



r = ……

r =……..

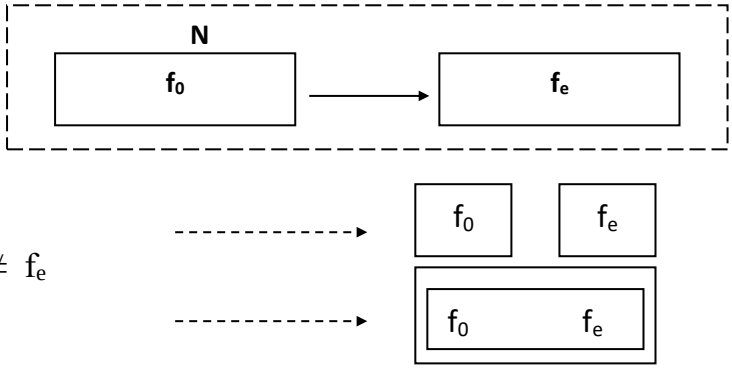
g)



r =

……

h)



$f_0 \neq f\, f_0 \neq f_e$

REFERENCIAS BIBLIOGRÁFICAS

Arnau Gras, J. (1978). *Psicología experimental: Un enfoque metodológico.* Editorial Trillas.

Ary, D., Jacobs, L. C. y Razavieh, A. (1982). *Introducción a la investigación Pedagógica* (Trad. J. Salazar y J. Pecina). Nueva editorial interamericana. (Trabajo original publicado en 1979).

Azorín Poch F. (1972). *Curso de Muestreo y Aplicación*. Editorial Aguilar.

Bancroft H. (1974). *Introducción a la Bioestadística* (Trad. N. Mittelman). Editorial Universitaria de Buenos Aires. (Trabajo original publicado en 1959).

Bernal C. (2000). *Metodología de la Investigación para Administración y Economía.* Editorial Prentice/ Hall Internacional.

Calzada J. (1966). *Estadística General con Énfasis en Muestreo.* Editorial Jurídica S.A.

Campbell, D., & Stanley, J. (1979). *Delineamentos experimentais e quase-experimentais de pesquisa* (R. A. D. Dio, Trans.): Editora da Universidade de Sao Paulo.

Cazau, P. (2006). *Introducción a la investigación en ciencias sociales*. Buenos aires.

Dión, Carlos. (1995). *Curso de Lógica*. Editorial Mc. Graw Hill.

Dixon W. y Massey F. (1973). *Introducción al Análisis Estadístico.* Editorial Mc. Graw Hill.

Exebio, C y Zavaleta V. (1979). *Estadística Aplicada a la Investigación Científica.* Universidad Nacional de Trujillo.

Falcón, P., Zavaleta, V., Guerrero, J., Bocanegra, R., Guerra, A. y Morales, F. (1978). *Metodología de la Investigación Científica*. Centro de Promoción Educativa Universidad Nacional de Trujillo.

Freund J. y Williams F. (1973). *Elementos Modernos de Estadística Empresarial* (Trad. J. Castaño Ceballos). Editorial Prentice/Hall Internacional. (Trabajo original publicado en 1972).

Hernández R., Fernández C. y Baptista P. (1998). *Metodología de la investigación.* Editorial Prentice/Hall Internacional.

Kerlinger, F. (1984). Investigación del Comportamiento (Interamerica Ed. Segunda ed.). México.

Levin, R. I. (1988). *Estadística para administradores* (Trad. M. Efrén Alatorre). Prentice-Hall Hispanoamericana. (Trabajo original publicado en 1987).

Mode E., (1970). *Elementos de Probabilidades y Estadística.* Editorial Reverté. México.

Mood A. y Graybill F. (1972). *Introducción a la teoría de la estadística* (Trad. F. Azorín). Editorial Aguilar. (Trabajo original publicado en 1963).

Neuman, M. (2014). *Seminarios sobre evaluación de tecnologías y sistemas de tratamiento de RSU*. Estrucplan. https://estrucplan.com.ar/ seminarios-sobre-evaluacion-de-tecnologias-y-sistemas-de-tratamiento-de-rsu/

Rodríguez F., Barrios I. y Fuentes M. (1984). *Introducción a la Metodología de las Investigaciones Sociales.* Editora Política. La Habana. 1984.

Vásquez, E. (1990). *Estadística para la Investigación Científica.* Publicación auspiciada por Concytec.

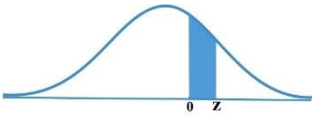
Vásquez E., Rodríguez, N., Ortiz G. y Vásquez-Ortiz E. (2021). *El Proyecto de Investigación*. Editorial Universitaria. UNPRG. Chiclayo-Perú.

Watanabe T., De Córdova M. (1990). *Elaboración de Proyectos de Investigación en Salud.* UNMSM. Lima. Perú.

NÚMEROS ALEATORIOS

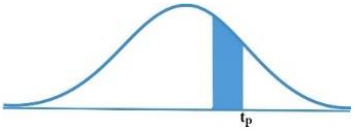
51772	74640	42331	29044	46621	62898	93582
24033	23491	83587	6568	21960	21387	76105
45939	60173	52078	25424	11645	55870	56974
30586	2133	75797	45406	31041	86707	12973
3585	79353	81938	82322	96799	85659	36081
64937	3355	95863	20790	65304	55189	745
15630	64759	51135	98527	62586	41889	25439
9448	56301	57683	30277	94623	85418	68829
21531	91157	77331	60710	52290	16835	48653
91097	17480	29414	6829	87843	28195	27279
50532	25496	95652	42457	73547	76552	50020
7136	40876	79971	54195	25708	51817	36732
27989	64728	10744	8396	56242	90985	28868
85184	73949	36601	46253	477	25234	9908
54398	21154	97810	36764	32869	11785	55261
65544	34371	9591	7839	58892	92843	72828
8263	65952	85762	64236	39238	18776	84303
39817	67906	48236	16057	81812	15815	63700
62257	4077	79443	95203	2479	30763	92486
53298	90276	62545	21944	16530	3878	7516
4186	19640	87056	73533	92520	51759	32783
10863	97453	90581	52850	13932	38045	58107
37428	93507	94271	34654	61261	51658	70502
17169	88116	42187	62230	64164	74266	93657
50884	14070	74950	98839	12848	14316	89167
65253	11822	15804	41783	60065	8274	38953
88036	24034	67283	63127	47593	97941	65191
6652	41982	49159	56605	13466	65194	28614
71590	16159	14676	32646	48427	38276	43042
47152	35683	47280	48625	61842	40851	22811
24819	52984	76168	25174	5647	7041	48805
72484	94923	75936	9517	78124	91188	69171
99431	50995	20507	25990	17246	70539	82473
36574	72139	70185	93088	18509	35479	36801
59009	38714	38723	35259	95078	4961	98140
91341	84821	63886	34878	62890	41462	8471
99247	46149	3229	6225	1709	13377	75116
85915	19219	45943	32946	77203	38975	10365
54083	23631	5825	98529	72586	53115	65749
95715	2526	33537	40366	9702	39856	55330

AREAS BAJO LA CURVA NORMAL TIPIFICADA DE 0 a z



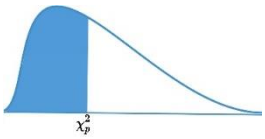
z	0	1	2	3	4	5	6	7	8	9
0	0,0000	0,0040	0,0080	0,0120	0,016	0,0199	0,0239	0,0279	0,0319	0,0359
0.1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0754
0.2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0.3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0.4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0.5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0.6	0,2258	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2518	0,2549
0.7	0,2580	0,2512	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0.8	0,2381	0,2910	0,2939	0,2967	0,2996	0,3023	0,3051	0,3078	0,3106	0,3133
0.9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1.1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1.2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1.3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1.4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1.5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1.6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1.7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1.8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1.9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2.1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2.2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2.3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2.4	0,4918	0,4920	0,4922	0,4925	0.4927	0,4929	0,4931	0,4932	0,4934	0,4936
2.5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2.6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2.7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2.8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2.9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3.1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3.2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3.3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3.4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3.5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3.6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3.7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3.8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3.9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

DISTRIBUCIÓN t STUDENT CON v GRADOS DE LIBERTAD (ÁREA SOMBREADA=p)



v	$t_{0,995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$	$t_{0,80}$	$t_{0,75}$	$t_{0,70}$	$t_{0,60}$	$t_{0,65}$
1	63,66	31,82	12,71	6,31	3,08	1,376	1,000	0,727	0,325	0,158
2	9,92	6,96	4,30	2,92	1,89	1,061	0,816	0,617	0,289	0,142
3	5,84	4,54	3,18	2,35	1,64	0,978	0,765	0,584	0,277	0,137
4	4,60	3,75	2,78	2,13	1,53	0,941	0,741	0,569	0,271	0,134
5	4,03	3,36	2,57	2,02	1,48	0,920	0,727	0,559	0,267	0,132
6	3,71	3,14	2,45	1,94	1,44	0,906	0,718	0,553	0,265	0,131
7	3,50	3,00	2,36	1,90	1,42	0,896	0,711	0,549	0,263	0,130
8	3,36	2,90	2,31	1,86	1,40	0,889	0,706	0,546	0,262	0,130
9	3,25	2,82	2,26	1,83	1,38	0,883	0,703	0,543	0,261	0,129
10	3,17	2,76	2,23	1,81	1,37	0,879	0,700	0,542	0,260	0,129
11	3,11	2,72	2,20	1,80	1,36	0,876	0,697	0,540	0,260	0,129
12	3,06	2,68	2,18	1,78	1,36	0,873	0,695	0,539	0,259	0,128
13	3,01	2,65	2,16	1,77	1,35	0,870	0,694	0,538	0,259	0,128
14	2,98	2,62	2,14	1,76	1,34	0,868	0,692	0,537	0,258	0,128
15	2,95	2,60	2,13	1,75	1,34	0,866	0,691	0,536	0,258	0,128
16	2,92	2,58	2,12	1,75	1,34	0,865	0,690	0,535	0,258	0,128
17	2,90	2,57	2,11	1,74	1,33	0,863	0,689	0,534	0,257	0,128
18	2,88	2,55	2,10	1,73	1,33	0,862	0,688	0,534	0,257	0,127
19	2,86	2,54	2,09	1,73	1,33	0,861	0,688	0,533	0,257	0,127
20	2,84	2,53	2,09	1,72	1,32	0,860	0,687	0,533	0,257	0,127
21	2,83	2,52	2,08	1,72	1,32	0,859	0,686	0,532	0,257	0,127
22	2,82	2,51	2,07	1,72	1,32	0,858	0,686	0,532	0,256	0,127
23	2,81	2,50	2,07	1,71	1,32	0,858	0,685	0,532	0,256	0,127
24	2,80	2,49	2,06	1,71	1,32	0,857	0,685	0,531	0,256	0,127
25	2,79	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
26	2,78	2,48	2,06	1,71	1,32	0,856	0,684	0,531	0,256	0,127
27	2,77	2,47	2,05	1,70	1,31	0,855	0,684	0,531	0,256	0,127
28	2,76	2,47	2,05	1,70	1,31	0,855	0,683	0,530	0,256	0,127
29	2,76	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
30	2,75	2,46	2,04	1,70	1,31	0,854	0,683	0,530	0,256	0,127
40	2,70	2,42	2,02	1,68	1,30	0,851	0,681	0,529	0,255	0,126
60	2,66	2,39	2,00	1,67	1,30	0,848	0,679	0,527	0,254	0,126
100	2,62	2,36	1,98	1,66	1,29	0,845	0,677	0,526	0,254	0,126
∞	2,58	2,33	1,96	1,645	1,28	0,842	0,674	0,524	0,253	0,126

DISTRIBUCIÓN χ^2_p
CON v GRADOS DE LIBERTAD
(AREA SOMBREADA = p)



v	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,975}$	$\chi^2_{0,95}$	$\chi^2_{0,90}$	$\chi^2_{0,75}$	$\chi^2_{0,50}$	$\chi^2_{0,25}$	$\chi^2_{0,10}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	7,88	6,63	5,02	3,94	2,71	1,32	0,455	0,102	0,0158	0,0009	0,0010	0,0002	0,0000
2	10,6	9,21	7,38	5,99	4,61	2,71	1,39	0,575	0,211	0,103	0,0506	0,0201	0,0100
3	12,8	11,3	9,35	7,81	6,25	4,11	2,37	1,21	0,584	0,352	0,216	0,115	0,072
4	14,9	13,3	11,1	9,49	7,78	5,39	3,36	1,92	1,06	0,711	0,484	0,297	0,207
5	16,7	15,1	12,8	11,1	9,24	6,33	4,35	2,67	1,61	1,15	0,831	0,554	0,412
6	18,5	16,8	14,4	12,6	10,6	7,84	5,35	3,45	2,20	1,64	1,24	0,872	0,676
7	20,3	18,5	16,0	14,1	12,0	9,04	6,35	4,25	2,83	2,17	1,69	1,24	0,969
8	22,0	20,1	17,5	15,5	13,4	10,2	7,34	5,07	3,49	2,73	2,18	1,65	1,34
9	23,6	21,7	19,0	16,9	14,7	11,4	8,34	5,90	4,17	3,33	2,70	2,09	1,73
10	25,2	23,2	20,5	18,3	16,0	12,5	9,34	6,74	4,87	3,94	3,25	2,56	2,16
11	26,8	24,7	21,9	19,7	17,3	13,7	10,3	7,58	5,58	4,57	3,82	3,05	2,60
12	28.3	26,2	23,3	21,0	18,5	14,8	11,3	8,44	6,30	5,23	4,40	3,57	3,07
13	29,8	27,7	24,7	22,4	19,8	16,0	12,3	9,30	7,04	5,89	5,01	4,11	3,57
14	31,3	29,1	26,1	23,7	21,1	17,1	13,3	10,2	7,79	6,57	5,63	4,66	4,07
15	32,8	30,6	27,5	25,0	22,3	18,2	14,3	11,0	8,55	7,26	6,26	5,23	4,60
16	34, 3	32,0	28,8	26,3	23,5	19,4	15,3	11,9	9,31	7,96	6,91	5,81	5,14
17	35,7	33,4	30,2	27,6	24,8	20,5	16,3	12,8	10,1	8,67	7,56	6,41	5,70
18	37,2	34,8	31,5	28,9	26,0	21,6	17,3	13,7	10,9	9,39	8,23	7,01	6,26
19	38,6	36,2	32,9	30,1	27,2	22,7	18,3	14,6	11,7	10,1	8,91	7,63	6,84
20	40,0	37,6	34,2	31,4	28,4	23,8	19,3	15,5	12,4	10,9	9,59	8,26	7,43
21	41,4	38,9	35,5	32,7	29,6	24,9	20,3	16,3	13,2	11,6	10,3	8,90	8,03
22	42,8	40,3	36,8	33,9	39,8	26,0	21,3	17,2	14,0	12,3	11,0	9,54	8,64
23	44,2	41,6	38,1	35,2	32,0	27,1	22,3	18,1	14,8	13,1	11,7	10,2	9,26
24	45,6	43,0	39,4	36,4	33,2	28,2	23,3	19,0	15,7	13,8	12,4	10,9	9,89
25	46,9	44,3	40,6	37,7	34,4	29,3	24,3	19,9	16,5	14,6	13,1	11,5	10,5
26	48,3	45,6	41,9	38,9	35,6	30,4	25,3	20,8	17,3	15,4	13,8	12,2	11,2
27	49,6	47,0	43,2	40,1	36,7	31,5	26,3	21,7	18,1	16,2	14,6	12,9	11,8
28	51,0	48,3	44,5	41,3	37,9	32,6	27,3	22,7	18,9	16,9	15,3	13,6	12,5
29	52,3	49,6	45,7	42,6	39,1	33,7	28,3	33,6	19,8	17,7	16,0	14,3	13,1
30	53,7	50,9	47,0	43,8	40,3	34,8	29,3	24,5	20,6	18,5	16,8	15,0	13,8
40	66,8	63,7	59,3	55,8	51,8	45,6	39,3	33,7	29,1	26,5	24,4	22,2	20,7
50	79,5	76,2	71,4	67,5	63,2	56,3	49,3	42,9	37,7	34,8	32,4	29,7	28,0
60	92,0	88,4	83,3	79,1	74,4	67,0	59,3	52,3	46,5	43,2	40,5	37,5	35,5
70	104,2	100,4	95,0	90,5	85,5	77,6	69,3	61,7	55,3	51,7	48,8	45,4	43,3
80	166,3	112,3	106,6	101,9	96,6	88,1	79,3	71,1	64,3	60,4	57,2	53,5	51,2
90	128,3	124,1	118,1	113,1	107,6	98,6	89,3	80,6	73,3	69,1	65,6	61,8	59,2
100	140,2	135,8	129,6	124,3	118,5	109,1	99,3	90,1	82,4	77,9	74,2	70,1	67,3

Valores de $F_{0.01}$

Grados de libertad del denominador																			
Grados de libertad del denominador	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5928	5982	6023	6056	6106	6157	6208	6235	6261	6287	6313	6339	6386
2	98,5	99,0	99,2	99,2	98,3	99,3	99,4	99,4	99,4	99,4	99,4	9,4	99,4	99,5	99,5	99,5	995	99,5	99,5
3	34,1	30,8	29,5	287	28,2	27,9	27,7	27,5	27,3	27,2	27,1	2,4,9	26,7	26,6	26,5	24,4	263	26,2	26,1
4	21,2	18,0	16,7	160	15,5	15,2	15,0	14,8	14,7	14,5	14,4	1,2	14,0	13,9	13,8	13,7	12,7	12,6	13,5
5	16,3	13,3	12,1	11,4	11,0	197	10,5	10,3	10,2	10,1	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,05	6,97	6,88
7	12,2	9,55	8,45	7,85	7,46	7,19	3,99	6,84	6,72	6,62	6,47	6,31	6,16	6,97	5,99	5,91	5,22	5,74	5,65
8	11,3	8,65	7,59	7,01	4,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,23	5,20	5,12	5,03	4,95	4,86
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,43	4,40	431
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,08	4,94	4,85	4,71	4,56	4,41	4,33	1,25	4,17	4,03	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,24	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,85	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	470	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	327	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	456	4,32	4,14	4,00	3,69	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,06	2,96	2,87
16	8,53	6,23	5,29	4,77	1,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,9	2,54	2,75
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,7	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,34	275	2,68	2,57
19	8,19	5,93	5,01	4,50	4,17	3,94	3,7	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	276	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,73	2,69	2,61	252	2,2
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,48	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,20	2,62	2,54	2,45	235	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,69	2,74	2,66	2,58	2,48	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,36	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,53	2,45	2,26	2,27	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	230	2,24	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	220	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,41	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1, 59	1,47	1,32	1,00

Valores de $F_{0.05}$

Grados de libertad del denominador																				
Grados de libertad del denominador		1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
	1	161	200	216	225	230	234	237	239	241	242	244	246	2,48	249	250	251	252	253	254
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,86	5,91	5,96	5,80	5,77	5,75	5,72	5,69	5,66	5,63
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
	8	5,32	4,46	4,07	3,84	3,69	5,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	2,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	2,96	1,91	1,86	1,81	1,76
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
	60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
	120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
	∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

El hombre
encuentra a Dios,
detrás de cada puerta
que la Ciencia logra abrir.

ALBERT EINSTEIN

ISBN: 978-9972-55-029-4