NOTAS DE CLASE ESTADÍSTICA I: ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

Elvis Lacruz Calderón
Walter Andrés Ortiz Vargas

Índice general

1. **5**

2. Introducción a la inferencia estadística **7**

2.1. Funciones de distribución 7

2.1.1. Ejercicios resueltos 10

2.2. Diseño de experimentos 12

2.3. Distribuciones de probabilidad 13

2.3.1. Variables aleatorias discretas 14

2.3.2. Variables aleatorias Continuas 16

2.4. Muestreo aleatorio 20

2.4.1. Muestreo aleatorio simple 21

2.4.2. Muestreo sistemático 21

2.4.3. Muestreo aleatorio estratificado 21

2.4.4. Muestreo por conglomerados 22

2.5. Estimación 22

2.5.1. Estimación puntual 22

2.5.2. Estimación por máxima versimilitud 23

2.5.3. Estimación Intervalo de confianza 24

2.5.4. Estimación Intervalo de confianza Media poblacional Caso 2: 25

2.6. Pruebas de hipótesis 27

2.6.1. Hipótesis estadística 27

2.6.2. Nivel de significación 28

2.6.3. Prueba de hipótesis- Comparación de medias I 31

2.6.4. Prueba de hipótesis- Comparación de medias II 31

2.6.5. Prueba de hipótesis- Comparación de medias III 32

2.7. Ejercicios resueltos 33

Capítulo 1

Capítulo 2

Capítulo 3

Introducción a la inferencia estadística

En este capítulo se estudiará un conjunto de elementos de pruebas con el objetivo de generar datos que, al ser analizados estadísticamente, proporcionen evidencias objetivas que permitan responder las interrogantes planteadas por el experimentador sobre determinada situación. Como parte inicial se estudiarán las respectivas propiedades que compete una función matemática de distribución, así como las propiedades de los estimadores, y funciones generadoras de momentos y las respectivas técnicas de estimación de parámetros (Puntual, intervalo de confianza y máxima verosimilitud). La inferencia estadística se ocupa de estudiar los métodos necesarios para extraer, o inferir, conclusiones válidas e información sobre una población a partir del estudio experimental de una muestra de dicha población. Los métodos utilizados en la inferencia estadística dependen de la información previa que se tenga de la población a estudiar. Por otro lado, a través del contraste de hipótesis, la estadística proporciona procedimientos óptimos para decidir la aceptación o el rechazo de afirmaciones o hipótesis acerca de la población en estudio. Las hipótesis se contrastan comparando sus predicciones con los datos experimentales. Si coinciden dentro de un margen de error, la hipótesis se mantiene. En caso contrario se rechaza y hay que buscar hipótesis o modelos alternativos que expliquen la realidad. Para seguir la secuencia es muy recomendable y de forma obligatoria hacer una lectura exhaustiva sobre las distribuciones de probabilidad, ya que en este caso serán de suma importancia, teniendo en cuenta el tipo de información que se tenga en los respectivos problemas.
Es de aclarar que se necesitarán distribuciones de probabilidad en las cuales se hará enfásis en el curso correspondiente a la asignatura de Probabilidad.

3.1. Funciones de distribución

Definición 3.1.1: Función de distribución

Una función de distribución de una variable aleatoria es una aplicación matemática:

$$f : \mathbb{R} \to [0, 1]$$

dicha función cumple: **Variable aleatoria discreta:**

- $f(x) \geq 0$
- $\sum_{i=1}^{n} f_i(x) = 1$

Variable aleatoria continua:

- $f(x) \geq 0$
- $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$

Ejemplo 3.1.1

Verificar si las siguientes, son funciones de distribución:

x	-1	2	3	4
$f(x)$	0.30	0.25	0.10	0.35

x	-2	4	6	8
$f(x)$	0.36	0.29	0.14	0.35

x	0	1	2	3	4
$f(x)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

- $f(x) = \begin{cases} \frac{8}{7x^2} & 1 \leq x \leq 8 \\ 0 & o.c. \end{cases}$
- $f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2 - x & 1 < x \leq 2 \\ 0 & O.C. \end{cases}$
- *Gran número de fenómenos aeronáuticos tienen asociada una variable aleatoria con función de distribución* $f(x) = \begin{cases} ke^{-kx} & x > 0, k > 0 \\ 0 & O.C. \end{cases}$
  ¿Puede tomar k cualquier valor? Si $k = 0{,}1$ *es una función de distribución?*

Definición 3.1.2: Esperanza Matemática

Sea X una variable aleatoria con función de distribución $f(x)$. La media o el valor esperado está dado por:
Variable aleatoria discreta:

$$E(X) = \sum_{i=1}^{n} x_i f_i(x)$$

Variable aleatoria continua:

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx$$

Proposición 3.1.1: Propiedades de la esperanza

Sea X una variable aleatoria y $a, b \in \mathbb{R}$, entonces satisface las siguientes propiedades:

- $E(a) = a$
- $E(X + a) = \mu + a$
- $E(aX) = aE(X)$
- $E(aX + b) = a\mu + b$

Definición 3.1.3: Varianza de una función de distribución

Sea X una variable aleatoria con función de distribución $f(x)$. La media o el valor esperado está dado por:

Variable aleatoria discreta:

$$V(X) = \sum_{i=1}^{n} x_i^2 f_i(x) - (E(X))^2$$

Variable aleatoria continua:

$$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = E(X^2) - (E(X))^2$$

Proposición 3.1.2: Propiedades de la varianza

Sea X una variable aleatoria y $a, b \in \mathbb{R}$, entonces satisface las siguientes propiedades:

- $V(a) = 0$
- $V(X + a) = V(x)$
- $V(aX) = a^2 V(X)$
- $V(aX + b) = a^2 V(x)$
- $Cov(X, Y) = E(XY) - E(X)E(Y)$

Ejemplo 3.1.2

Una empresa industrial compra varios procesadores de textos nuevos al final de año, el número de procesadores exacto depende de la frecuencia de reparación del año anterior. Suponga que el número de procesadores de texto X, tiene la siguiente función de distribución.

x	*0*	*1*	*2*	*3*
$f(x)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{5}$	$\frac{1}{5}$

Cuál es el valor esperado de procesadores que debe comprar la empresa? ¿Cuál es la varianza y desviación típica?

Ejemplo 3.1.3

La distribución del tiempo que transcurre antes de que una lavadora requiera una reparación mayor fue dada por:

$$f(y) = \begin{cases} \frac{1}{4}e^{-y/4} & y \geq 0 \\ 0 & O.C \end{cases}$$

Cuál es el tiempo medio que transcurre antes de requerir reparación? ¿Cuál es la varianza y desviación típica?

3.1.1. Ejercicios resueltos

Ejercicio resuelto Variable discreta 1

Sea X una variable aleatoria discreta cuya función de distribución está dada por:

x	-2	-1	0	1	2	3
$P(x)$	a	0.2	b	0.3	0.01	0.1

1. Encontrar el valor de a y b si el valor espado es $0{,}38$
   Solución:
   Para econtrar los valores de a y b se procede a usar que la función debe cumplir que $\sum_{i=1}^{5} p(x_i) = 1 \rightarrow a + 0{,}2 + b + 0{,}3 + 0{,}01 + 0{,}1 = 1 \rightarrow a + b + 0{,}61 = 1 \rightarrow a + b = 1 - 0{,}61 \rightarrow a + b = 0{,}39$
   Por otro lado,
   $$\begin{aligned} E(x) = 0{,}38 \rightarrow \sum_{i=1}^{5} x_i p(x_i) = 0{,}38 \\ &\rightarrow (-2)(a) + (-1)(0{,}2) + (0)(b) + (1)(0{,}3) + (2)(0{,}01) + (3)(0{,}1) = 0{,}38 \\ &\rightarrow -2a + 0{,}42 = 0{,}38 \\ &\rightarrow -2a = 0{,}38 - 0{,}42 \\ &\rightarrow a = \frac{-0{,}04}{-2} = 0{,}02 \end{aligned}$$
   Se sabe que $a + b = 0{,}39 \rightarrow b = 0{,}39 - a \rightarrow b = 0{,}39 - 0{,}02 = 0{,}37$

2. Calcular la varianza y la desviación típica
   Solución:
   La varianza se calcula usando la expresión $V(x) = E(x^2) - (E(x))^2$ de modo que
   $$\begin{aligned} E(x^2) &= \sum_{i=1}^{5} x_i^2 p(x_i) = (-2)^2(0{,}02) + (-1)^2(0{,}2) + (0)^2(0{,}37) + (1)^2(0{,}3) + (2)^2(0{,}01) + (3)^2(0{,}1) \\ &= (4)(0{,}02) + (1)(0{,}2) + (0)(0{,}37) + (1)(0{,}3) + (4)(0{,}01) + (9)(0{,}1) = 1{,}52 \end{aligned}$$

y Así, $V(x) = E(x^2) - (E(x))^2 = 1{,}52 - (0{,}38)^2 = 1{,}3756$
Por otro lado la desviación típica está dada por: $DT(x) = \sqrt{V(x)} = \sqrt{1{,}3756} = 1{,}1728$

Ejercicio resuelto variable discreta 2

Comprobar en el siguiente caso si es función de distribución y en caso afirmativo calcular la esperanza, varianza y desviación estándar.

X	-1	2	3	4
f(X)	0.35	0.40	0.20	0.05

Se puede observar que todas las imagenes son $f(x) \geq 0$ con lo cual se cumple la primera propiedad, con respecto a la segunda $\sum_{i=1}^{4} f_i(x) = 0{,}35 + 0{,}40 + 0{,}20 + 0{,}05 = 1$ luego es una función de distribución .

$E(x) = \sum_{i=1}^{n} x_i f_i(x) = (-1)(0{,}35) + (2)(0{,}40) + (3)(0{,}20) + (4)(0{,}05) = 1{,}25$

$E(x^2) = \sum_{i=1}^{n} x_i f_i(x) = (-1)^2(0{,}35) + (2)^2(0{,}40) + (3)^2(0{,}20) + (4)^2(0{,}05) = 4{,}55$

$V(x) = E(x^2) - (E(x))^2 = 4{,}55 - (1{,}25)^2 = 2{,}9875$; $DT(x) = \sqrt{2{,}9875} = 1{,}728439$

Ejercicio resuelto variable continua 1

Comprobar en el siguiente caso si es función de distribución y en caso afirmativo calcular la esperanza, varianza y desviación estándar.

$f(x) = 0{,}25e^{-0{,}25x} x \geq 0$ se tiene que $f(x) \geq 0$

$\int_{-\infty}^{\infty} f(x)dx = 1 \rightarrow \int_0^{\infty} 0{,}25e^{-0{,}25x}dx = 0{,}25\int_0^{\infty} e^{-0{,}25x}dx = \frac{-0{,}25}{0{,}25}e^{-0{,}25x}|_0^{\infty} = (-1)(e^{-\infty} - e^0) = (-1)(0-1) = 1.$

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx \rightarrow 0{,}25\int_0^{\infty} xe^{-0{,}25x}dx \rightarrow \begin{cases} u = x \rightarrow du = dx \\ dv = e^{-0{,}25x}dx \rightarrow v = -\frac{1}{0{,}25}e^{-0{,}25x} \end{cases}$$

$$= 0{,}25\left(xe^{-0{,}25x}|_0^{\infty} - \int_0^{\infty} e^{-0{,}25x}dx\right) = 0{,}25\left(xe^{-0{,}25x}|_0^{\infty} + \frac{1}{0{,}25}e^{-0{,}25x}|_0^{\infty}\right) = 4$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)dx \rightarrow 0{,}25\int_0^{\infty} x^2 e^{-0{,}25x}dx \rightarrow \begin{cases} u = x^2 \rightarrow du = 2xdx \\ dv = e^{-0{,}25x}dx \rightarrow v = -\frac{1}{0{,}25}e^{-0{,}25x} \end{cases}$$

$$= 0{,}25\left(x^2e^{-0{,}25x}|_0^{\infty} - \int_0^{\infty} xe^{-0{,}25x}dx\right) = 0{,}25\left(x^2e^{-0{,}25x}|_0^{\infty} + \frac{2}{0{,}25}(xe^{-0{,}25x}|_0^{\infty} + \frac{1}{0{,}25}e^{-0{,}25x}|_0^{\infty})\right) = 32$$

$v(x) = E(x^2) - (E(x))^2 = 32 - (4)^2 = 16$; $DT(x) = \sqrt{16} = 4$

3.2. Diseño de experimentos

Definición 3.2.1: Diseño de experimentos

Consiste en planear y realizar un conjunto de pruebas con el objetivo de generar datos que, al ser analizados estadísticamente, proporcionen evidencias objetivas que permitan responder las interrogantes planteadas por el experimentador sobre determinada situación.

Definición 3.2.2: Experimento

Es un cambio en las condiciones de operación de un sistema o proceso, que se hace con el objetivo de medir el efecto del cambio en una o varias propiedades del producto o resultado.

Definición 3.2.3: Unidad experimental

Pieza(s) o muestra(s) que se utiliza para generar un valor que sea representativo del resultado de la prueba.

Definición 3.2.4: Variable de respuesta

A través de esta(s) variable(s) se conoce el efecto o los resultados de cada prueba experimental.

Definición 3.2.5: Error aleatorio

Es la variabilidad observada que no se puede explicar por los factores estudiados; resulta del pequeño efecto de los factores no estudiados y del error experimental.

Definición 3.2.6: Error experimental

Componente del error aleatorio que refleja los errores del experimentador en la planeación y ejecución del experimento.

Definición 3.2.7: Población finita

Es aquella en la que se pueden medir todos los individuos para tener un conocimiento exacto de sus características.

Definición 3.2.8: Parámetros

Características que, mediante su valor numérico, describen a un conjunto de elementos o individuos.

Ejemplo 3.2.1

Los parámetros poblacionales estudiados en este curso son:
media :μ
varianza: σ^2
Desviación típica/estándar σ.
Proporción: P *o* π

Definición 3.2.9: Inferencia estadística

Son las afirmaciones válidas acerca de la población o proceso basadas en la información contenida en la muestra.

Definición 3.2.10: Estadístico

Cualquier función de los datos muestrales que no contiene parámetros desconocidos.

Ejemplo 3.2.2

Los parámetros poblacionales estudiados en este curso son:
media :$\bar{X}$
varianza: s^2
Desviación típica/estándar s.
Proporción: $\hat{P}$ *o* $\hat{\pi}$.

3.3. Distribuciones de probabilidad

En esta sección se presentan las distribuciones más cononcidas en inferencia y sus características.

Definición 3.3.1: Distribución de probabilidad de X

Relaciona el conjunto de valores de X con la probabilidad asociada con cada uno de estos valores.

3.3.1. Variables aleatorias discretas

Distribución Uniforme discreta

$X \sim U(x_1, \ldots, x_n)$

$$f_X(x) = \begin{cases} \frac{1}{N}, & si \;\; x = 1, 2, \ldots \\ 0 & O.C. \end{cases}$$

- $E(X) = \frac{N+1}{2}$
- $V(X) = \frac{N^2-1}{12}$
- $m(t) = \sum_{k=1}^{N} \frac{1}{N} e^{kt}$

Distribución Binomial

$X \sim Bin(n, p)$

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & si \;\; x = 1, 2, \ldots \\ 0 & O.C. \end{cases}$$

- $E(X) = np$
- $V(X) = np(1-p)$
- $m(t) = (pe^t + (1-p))^n$

Distribución Bernoulli

$X \sim Ber(p)$

$$f_X(x) = \begin{cases} (1-p), & si \;\; x = 0 \\ p & si \;\; x = 1 \\ 0 & O.C. \end{cases}$$

- $E(X) = p$
- $V(X) = p(1-p)$
- $m(t) = (pe^t + (1-p))$

Distribución Geométrica

$X \sim G(p)$

$$f_X(x) = \begin{cases} p(1-p)^x, & si \;\; x = 1, 2, \ldots \\ 0 & O.C. \end{cases}$$

- $E(X) = \dfrac{1-p}{p}$
- $V(X) = \dfrac{1-p}{p^2}$
- $m(t) = \dfrac{p}{1-(1-p)e^t}$

Distribución Binomial Negativa

$X \sim B_N(n, p)$

$$f_X(x) = \begin{cases} \dbinom{n+x-1}{x} p^n (1-p)^x, & si \;\; x = 1, 2, \ldots \\ 0 & O.C. \end{cases}$$

- $E(X) = \dfrac{n(1-p)}{p}$
- $V(X) = \dfrac{(1-p)n}{p^2}$
- $m(t) = \left(\dfrac{p}{1-(1-p)e^t} \right)^n$

Distribución Hipergeométrica

$X \sim H(n, R, N)$

$$f_X(x) = \begin{cases} \dfrac{\dbinom{R}{x}\dbinom{N-R}{n-x}}{\dbinom{N}{n}}, & si \;\; x = 1, 2, \ldots \\ 0 & O.C. \end{cases}$$

- $E(X) = \dfrac{nR}{N}$
- $V(X) = \dfrac{nR(N-R)(N-n)}{N^2(N-1)}$

Distribución Poisson

$X \sim Poiss(\lambda)$

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & si \;\; x = 1, 2, \ldots \\ 0 & O.C. \end{cases}$$

- $E(X) = \lambda$
- $V(X) = \lambda$
- $m(t) = exp(\lambda(e^t - 1))$

3.3.2. Variables aleatorias Continuas

Distribución Uniforme

$X \sim U[a, b]$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & si \;\; a \leq x \leq b, \\ 0 & O.C. \end{cases}$$

- $E(X) = \frac{a+b}{2}$
- $V(X) = \frac{(b-a)^2}{12}$
- $m(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$

Distribución Exponencial

$X \sim exp(\lambda)$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & si \;\; x > 0, \\ 0 & O.C. \end{cases}$$

- $E(X) = \frac{1}{\lambda}$
- $V(X) = \frac{1}{\lambda^2}$
- $m(t) = \frac{\lambda}{\lambda - t}$

Distribución Normal

$X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- $E(X) = \mu$
- $V(X) = \sigma^2$
- $m(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$

Distribución Gamma

$X \sim \Gamma(n, \lambda)$

$$f_X(x) = \begin{cases} \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda e^{-\lambda x}, & si \;\; x > 0, \\ 0 & O.C. \end{cases}$$

$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt$

- $E(X) = \frac{n}{\lambda}$
- $V(X) = \frac{n}{\lambda^2}$
- $m(t) = \left(\frac{\lambda}{\lambda - t} \right)^n$

Distribución Beta

$X \sim \beta(a, b)$ donde $a >, b > 0$

$$f_X(x) = \begin{cases} \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, & si \;\; 0 < x < 1, \\ 0 & O.C. \end{cases}$$

$\beta(a, b) = \int_0^\infty x^{a-1}(1-x)^{b-1} dx$

$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

- $E(X) = \frac{a}{a+b}$
- $V(X) = \frac{ab}{(a+b+1)(a+b)^2}$

- $m(t) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(a+b)}{\Gamma(a+b+k)\Gamma(b)} \frac{t^k}{k!}$

Distribución χ^2

Si $X \sim N(0,1)$ $X \sim \chi^2_n$

$$f_X(x) = \frac{2^{n/2} x^{n/2-1} e^{-x/2}}{\Gamma(n/2)}$$

- $E(X) = n$
- $V(X) = 2n$
- $m(t) = (1-2t)^{-n/2}$

Distribución t-student

$X \sim T_n$

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

- $E(X) = 0$
- $V(X) = \frac{n}{n-2}$
- $m(t)$ no existe

Distribución F

$X \sim F_{n,m}$

$$f_X(x) = \frac{\Gamma\left(\frac{n+m}{2}\right) n^{n/2} m^{m/2}}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} x^{n/2-1} (m+nx)^{-(m+n)/2}$$

- $E(X) = \frac{m}{m-2}$
- $V(X) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$
- $m(t)$ no existe

Distribución Weibull

$X \sim W(a,b)$

$$f_X(x) = \begin{cases} abx^{b-1}e^{-ax^b}, & si \;\; x > 0, \\ 0 & O.C. \end{cases}$$

- $E(X) = \left(\frac{1}{a}\right)^{\frac{1}{b}} \Gamma(1+\frac{1}{b})$
- $V(X) = \left(\frac{1}{a}\right)^{\frac{2}{b}} \left[\Gamma(1+\frac{2}{b}) - \Gamma^2(1+\frac{2}{b})\right]$

Distribución Lognormal

$X \sim LN(\mu, \sigma^2)$

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}x} e^{-\frac{(ln(x)-\mu)^2}{2\sigma^2}}, & si \;\; x > 0, \\ 0 & O.C. \end{cases}$$

- $E(X) = exp\left(\mu + \frac{\sigma^2}{2}\right)$
- $V(X) = exp\,(2\mu + \sigma^2)\,(e^{\sigma^2} - 1)$

Distribución Cauchy

$X \sim C(\theta, b)$

$$f_X(x) = \frac{1}{\pi b} \frac{1}{1 + \left(\frac{x-\theta}{b}\right)^2}$$

- $E(X)$ no existe

Distribución Doble exponencial (Laplace)

$X \sim L(\mu, b)$

$$f_X(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

- $E(X) = \mu$

- $V(X) = 2b^2$
- $m(t) = \dfrac{exp(\mu t)}{1 - t^2 b^2}$

Distribución Potencia

$X \sim POT(\mu, b, \gamma)$

$$f_X(x) = \frac{1}{2b\Gamma(1+\frac{1}{\gamma})} exp\left(-\left|\frac{x-\mu}{b}\right|^{\gamma}\right)$$

Distribución Logística

$X \sim LG(\mu, b)$

$$f_X(x) = \frac{1}{b} \frac{exp\left(-\left(\frac{x-\mu}{b}\right)\right)}{\left(1 + exp\left(-\left(\frac{x-\mu}{b}\right)\right)\right)^2}$$

- $E(X) = \mu$
- $V(X) = \dfrac{\pi^2}{3} b^2$
- $m(t) = e^{\mu t} \beta(1 - bt, 1 + st)$

3.4. Muestreo aleatorio

El proceso que consiste en inferir resultados a la población a partir de la muestra se denomina inferencia estadística. La confiabilidad de las conclusiones extraídas concernientes a una población dependen de si la muestra se ha escogido apropiadamente de manera que represente bien a la población. Una técnica para obtener muestras representativas de la población es el muestreo aleatorio.

Definición 3.4.1

Se denomina muestreo aleatorio a todo proceso que asegure en cualquier momento del mismo igual probabilidad de ser incluidos en la muestra a todos los elementos que pertenezcan a la población en dicho momento.

3.4.1. Muestreo aleatorio simple

Consiste en que los elementos se escogen del total de la población en forma individual con una oportunidad igual e independiente .Por lo general se utiliza una tabla de números aleatorios. Si la población es infinita el muestreo aleatorio ocurre cuando la extracción de los elementos de la muestra se hace con o sin reemplazo. Si la población es finita de tamaño N, el muestreo aleatorio ocurre también si la extracción es con o sin reemplazo. Con reemplazo, la probabilidad de cada elemento de ser extraído es $1/N$. Si es, sin reemplazo, la probabilidad de cada elemento de ser elegido es $1/N$ en la primera extracción, es de $1/(N-1)$ en la segunda extracción, es $I/(N-2)$ en la tercera extracción, etc.

Ejemplo 3.4.1

Seleccionar una muestra al azar simple es similar a la que se realiza en la extracción aleatoria de números en una lotería.

3.4.2. Muestreo sistemático

Los elementos de la población son elegidos por medio de intervalos uniformes a partir de un listado ordenado. El $k-$ésimo elemento de la muestra es $k = N/n$, donde n es el tamaño de la muestra y N el tamaño de la población.

Ejemplo 3.4.2

Elegir una muestra sistemática de 100 alumnos de VIU que tiene 3000 alumnos, $k = 3000/100 = 30$. *El primero se elige en forma aleatoria de los 30 primeros de la lista y los demás sistemáticamente cada 30 alumnos de la lista.*

3.4.3. Muestreo aleatorio estratificado

Con una o más características importantes (estratos). Después se obtiene por separado una muestra aleatoria simple o sistemática en cada estrato. El tamaño de cada submuestra debe ser proporcional al tamaño del estrato para asegurar representatividad.

Ejemplo 3.4.3

Para obtener una muestra aleatoria de 600 electores de una población de 600,000 electores de los cuales 300,0000 son de clase baja, 200,000 de clase media y 100,000 de clase alta. Se deben elegir al azar 300 de clase baja, 200 de clase media y 100 de clase alta.

3.4.4. Muestreo por conglomerados

Los elementos de la población se dividen en forma natural en subgrupos. Luego se eligen al azar los subgrupos que forman la muestra.

Ejemplo 3.4.4

Al estudiar las mensualidades que se pagan en los colegios concertados donde no es posible tener una lista de todas las mensualidades, pero puede obtenerse una lista de los colegios concertados (grupos). Entonces, con esta lista puede obtener una muestra aleatoria de colegios y así obtener las mensualidades que se pagan en estos colegios.

3.5. Estimación

Uno de los objetivos en la Estadística es hacer inferencias con respecto a una población, teniendo en cuenta la información contenida en una muestra. En este apartado se estudian técnicas que permiten cómo utilizar los valores de una muestra para obtener información sobre el valor de un parámetro poblacional.

3.5.1. Estimación puntual

Definición 3.5.1: Estimador puntual

Estadístico que estima un valor específico de un parámetro.

Ejemplo 3.5.1

La media poblacional es un estimador de la media muestral $\hat{\mu} = \bar{X}$

Se pueden proponer tantos estimadores para los parámetros como se quiera.
En lo que sigue se denotará cómo parámetro θ y estimador del parámetro $\hat{\theta}$.

Propiedades de los estimadores

- Un estimador es insesgado, si verifica $E[\hat{\theta}] = \theta$.
- En caso de no ser insesgado, es sesgado y este está dado por $E[\hat{\theta}] - \theta$
- Eficiencia :Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de θ. Diremos que $\hat{\theta}_1$ es mas eficiente que $\hat{\theta}_2$ si verifica que $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.
- Error cuadrático medio: $ECM(\hat{\theta}) = V(\hat{\theta}) + sesgo(\hat{\theta})^2$

Ejemplo 3.5.2

Verificar si los siguientes estimadores son insesgados. Si tiene por media poblacional, μ y desviación típica σ. Cuál es más eficiente?

- $\hat{\theta} = 0{,}5x_1 + 0{,}15x_2 + 0{,}25x_3 + 0{,}1x_4$
- $\hat{\theta} = \dfrac{x_1 + 2x_2 + 3x_3}{6}$
- $\hat{\theta} = \frac{x_3 - 4x_2}{-3}$

Ejemplo 3.5.3

Para la siguiente variable aleatoria, con función $f(x) = \theta x^{\theta-1}, \theta > 0, 0 \leq x \leq 1$. se ha tomado una muestra de 3 observaciones Calcular los sesgos para los siguientes estimadores y las estimaciones puntuales para $(0{,}7, 0{,}1, 0{,}3)$.

- $\hat{\theta} = \dfrac{x_1^2 + 2x_2^2 + 3x_3}{6}$
- $\hat{\theta} = \bar{X}$
- $\hat{\theta} = \dfrac{x_3 - 2x_1 + 4x_2}{6}$

3.5.2. Estimación por máxima versimilitud

La estimación por máxima verosimilitud es un método de optimización que supone que la distribución de probabilidad de las observaciones es conocida

Definición 3.5.2

Sea $x_1, \ldots, x_n$ una muestra aleatoria(no necesariamente simple)de una población X con función de probabilidad P_θ (o con función de densidad f_θ).
El estimador de máxima verosimilitud (probabilidad conjunta) de θ es el formado por los valores $\hat{\theta_1}, \ldots, \hat{\theta_n}$ que maximizan lafunción de verosimilitud de la muestra $x_1, \ldots, x_n$ obtenida:

$$L(\theta) = L(X;\theta) = L(x_1, \ldots, x_n;\theta) = \begin{cases} P_\theta(x_1)\ldots P_\theta(x_n) & \textit{Caso discreto} \\ f_\theta(x_1)\ldots f_\theta(x_n) & \textit{Caso continuo} \end{cases}$$

Por practicidad se recomienda encontrar el estimador de máxima verosimilitud es considerar la función soporte o Log-verosimilitud $log(L(X;\theta))$, en lugar de la función deverosimilitud $L(X;\theta)$,ya presenta los mismos máximos y mínimos.

Pasos para encontrar el estimador por máxima versimilitud

- Escribir la función de verosimilitud $L(X;\theta)$

- Escribir el logaritmo de la verosimilitud $logL(X;\theta)$
- Obtener θ_i según corresponda mediante $(log(L(X;\theta)))' = \frac{\partial}{\partial\theta_i}(log(L(X;\theta))) = 0$ y denotarlo por $\hat{\theta}_i$.
- Comprobar que es un máximo, con el criterio de la segunda derivada, tal que: $(log(L(X;\theta)))''(\hat{\theta}_i) = \frac{\partial^2}{\partial\theta_i^2}(log(L(X;\theta))(\hat{\theta}_i) < 0$.

3.5.3. Estimación Intervalo de confianza

Definición 3.5.3: Estimación Intervalo de confianza:

Rango donde se estima que está el valor de un parámetro poblacional.
En otras palabras, se tiene que

$$P(L \leq \theta \leq M) = (1-\alpha)\,\%$$

La cantidad $(1-\alpha)\,\%$ se le llama el nivel de confianza.

Estimación Intervalo de confianza Media poblacional Caso 1:

Para dicha estimación debe ser conocido el valor de la varianza o desviación estándar o típica poblacional y el estadístico (media muestral); es decir $\bar{x}$ y σ^2 donde.

$$\mu \in \bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

N.C.	$99{,}73\,\%$	$99\,\%$	$98\,\%$	$96\,\%$	$95{,}45\,\%$	$95\,\%$	$90\,\%$	$80\,\%$
$Z_{\alpha/2}$	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28

Cuadro 3.1: Tabla distribución normal

Ejemplo 3.5.4

Para evaluar la resistencia media de los envases se toma una muestra aleatoria de $n = 20$ *piezas. De los resultados se obtiene que* $\bar{X} = 55{,}2$ *y* $\sigma^2 = 9$. *Estime con una confianza de* $95\,\%$*, ¿cuál es la resistencia promedio de los envases?*

3.5.4. Estimación Intervalo de confianza Media poblacional Caso 2:

Para dicha estimación debe ser conocido valor de los estadísticos (media , varianza muestral); es decir $\bar{x}$ s^2. La varianza poblacional(desviación típica) desconocida σ^2 donde.

$$\mu \in \bar{x} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

Ejemplo 3.5.5

En la fabricación de discos compactos una variable de interés es la densidad mínima (grosor) de la capa de metal, la cual no debe ser menor de 1.5 micras. Se sabe por experiencia que la densidad mínima del metal casi siempre ocurre en los radios 24 y 57, aunque en el método actual también se miden los radios 32, 40 y 48. Se hacen siete lecturas en cada radio dando un total de 35 lecturas, de las cuales sólo se usa la mínima. A continuación se presenta una muestra histórica de 18 densidades mínimas: 1.81, 1.97, 1.93, 1.97, 1.85, 1.99, 1.95, 1.93, 1.85, 1.87, 1.98, 1.93, 1.96, 2.02, 2.07, 1.92, 1.99, 1.93. Encuentre un intervalo de confianza de $99\,\%$ *para la media de la densidad mínima.*

Estimación Intervalo de confianza proporción:

Para dicha estimación se debe tener una proporción estimada, $\hat{p} = \frac{n}{N}$ donde:

$$P \in \hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Ejemplo 3.5.6

En la producción de una planta se está evaluando un tratamiento para hacer que germine cierta semilla. De un total de 60 semillas se observó que 37 de ellas germinaron. Estimar con una confianza de $90\,\%$*, la proporción de germinación que se logrará con tal tratamiento.*

Estimación Intervalo de confianza Media de dos poblaciones :

Para dicha estimación deben ser conocido valor de los estadísticos (media , varianza muestral); es decir $\bar{x_1}, \bar{x_2}$ $s_1^2.s_2^2$ Las varianzas poblacionales (desviación típica) son desconocidas σ_1^2, σ_2^2 donde.

$$\mu_1 - \mu_2 \in \bar{x_1} - \bar{x_2} \pm t_{\alpha/2,n_1+n_2-1} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

donde $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

Ejemplo 3.5.7

En un laboratorio bajo condiciones controladas, se evaluó, para 10 hombres y 9 mujeres, la temperatura que cada persona encontró más confortable. Los resultados en grados Fahrenheit fueron los siguientes:

M	75	77	78	79	77	73	78	79	78	
H	74	72	77	76	76	73	75	73	74	75

Encontrar un intervalo de confianza para la diferencia de medias entre hombres y mujeres al $98\,\%$

Estimación Intervalo de confianza dos proporciones:

Para dicha estimación se debe tener una proporción estimada, $\hat{p}_i = \frac{n_i}{N}$ $i = 1, 2$ donde:

$$P_1 - P_2 \in \hat{p_1} - \hat{p_2} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$$

Ejemplo 3.5.8

Si el total de personas evaluadas del problema anterior es de 100. Teniendo en cuenta que se incluyen niños.
Encontrar un intervalo de confianza para la diferencia de proporciones entre hombres y mujeres al $96\,\%$

Estimación Intervalo de confianza para la varianza

Para dicha estimación debe ser conocido el valor de la varianza o desviación estándar o típica (el estadístico); es decir s^2 donde.

$$\sigma^2 \in \left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \right)$$

$\chi^2_{\alpha/2,n-1}$ (Chi cuadrado)

Tamaño de la muestra:

Se considera el error, como

$$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2.$$

3.6. Pruebas de hipótesis

Para tomar decisiones estadísticas se debe partir de afirmaciones o conjeturas con respecto a la población en el que estamos interesados. Tales suposiciones, putde.i ser verdaderas o no. Una conjetura hecha sobre una población o sobre sus parámetros deberá ser sometida a comprobación experimental con el propósito de saber si los resultados de una muestra aleatoria extraída de esa población, contradicen o no tal conjetura.

3.6.1. Hipótesis estadística

Definición 3.6.1: Hipótesis estadística

Es una afirmación sobre los valores de los parámetros de una población o proceso, que puede probarse a partir de la información contenida en una muestra.

Definición 3.6.2: Hipótesis nula: H_0

Es considerada como la que hace referencia al valor del parámetro que se quiere probar como verdadero.

Definición 3.6.3: Hipótesis alternativa: H_1 o H_a

Corresponde a la falsedad o estableciendo que el parámetro puede ser mayor, menor o igual, de acuerdo con la propuesta hecha en la hipótesis nula.

Ejemplo 3.6.1

Sea θ_0 el valor del parámetro desconocido θ de una población cuya distribución se supone conocida, entonces son hipótesis nulas y alternativas respectivamente las siguientes afirmaciones:

- $H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$
- $H_0 : \theta \geq \theta_0$ $H_1 : \theta < \theta_0$
- $H_0 : \theta \leq \theta_0$ $H_1 : \theta > \theta_0$

Pruebas de hipótesis

La prueba de una hipótesis estadística es un proceso que nos conduce a tomar la decisión de aceptar o rechazar la hipótesis nula H_0 , en contraposición de la hipótesis alternativa H_1 y en base a los resultados de una muestra aleatoria seleccionada de la población en estudio. La hipótesis nula H_0 es la primera hipótesis que se plantea, y debe ser establecida de manera que especifique un valor θ_0 del parámetro θ en estudio. Cabe destacar que algunos autores plantean la hipótesis nula como $H_0 : \theta = \theta_0$ $H_1 : \theta < \theta_0$ y $H_0 : \theta = \theta_0$ $H_1 : \theta > \theta_0$

Nota 3.6.1

La aceptación de una hipótesis significa que los datos de la muestra no proporcionan evidencia suficiente para refutarla. El rechazo significa que los datos de la muestra lo refutan.

Tipos de prueba

El tipo de prueba depende básicamente de la hipótesis alternativa H_1. Se denomina prueba de una cola a toda prueba de hipótesis donde la alternativa H_1 es unilateral. Si la alternativa es bilateral, la prueba se denomina prueba de dos colas.

- Prueba bilateral (dos colas)
  $H_0 : \theta = \theta_0$ $H_1 : \theta \neq \theta_0$
- Prueba unilateral (Una cola, izquierda)
  $H_0 : \theta = \theta_0$ $H_1 : \theta < \theta_0$
- Prueba unilateral (una cola, derecha).
  $H_0 : \theta = \theta_0$ $H_1 : \theta > \theta_0$

3.6.2. Nivel de significación

α es el recicproco de la confianza, el cual debe ser fijado antes de escoger la muestra. Al tomar la decisión de aceptar o rechazar la hipótesis nula$H_0 : \theta = \theta_0$ O en base a los resultados obtenidos de una muestra aleatoria seleccionada de la población con estudio; hay cuatro posibles situaciones que determinan si la decisión tomada es correcta o incorrecta, como se muestra a continuación:

Decisión	H_0 verdadera	H_0 falsa
Rechazar H_0	Error tipo I Probab: α	Decisión correcta Probab: $1-\beta$
Aceptar H_0	Decisión correcta Probab: $1-\alpha$	Error tipo II Probab: β

Cuadro 3.2: Nivel significación

Definición 3.6.4: Error tipo I

Se denomina error tipo I, al error que se comete al rechazar una hipótesis nula H_0 cuando esta realmente es verdadera

Definición 3.6.5: Error tipo II

Se denomina error tipo II, al error que se comete al aceptar una hipótesis nula H_0 cuando en realidad es falsa.

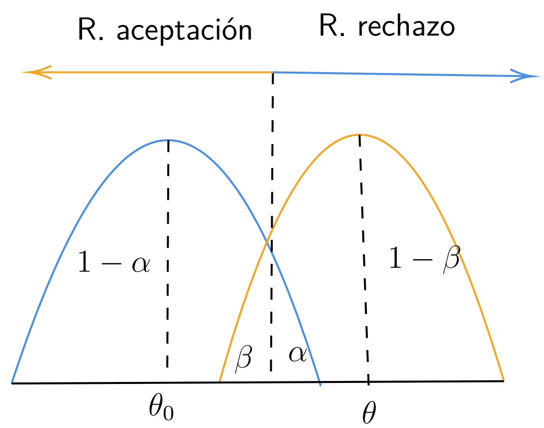


Figura 3.1: Errores

Definición 3.6.6: Potencia

La potencia de una prueba es la probabilidad de tomar la decisión acertada de, rechazar H_0 cuando ésta es falsa o de aceptar H_1 cuando ésta es verdadera

La potencia de una prueba es calculada por $1-\beta$. El nivel de significación se fija previamente por lo general en $\alpha = 0{,}05$ o $\alpha = 0{,}01$. Si para un valor dado de α , se rechaza la hipótesis H_0 , entonces se dice que los resultados muéstrales obtenidos, no sólo son diferentes por efectos del azar, si no que son realmente significativamente diferentes al nivel $\alpha \times 100\,\%$, es decir; se espera que de 100 resultados muéstrales en $\alpha \times 100\,\%$ de las veces se rechazará la hipótesis nula H_0 cuando realmente es verdadera.

Estadístico de prueba o contraste

Número calculado a partir de los datos y de H_0, cuya magnitud permite discernir si se rechaza o no la hipótesis nula. Es de aclarar que según corresponda el estadístico de contraste se dennominará EC_0 el calculado y EC_α el teórico (prueba unilateral) $EC_{\alpha/2}$ (prueba bilateral) . Por otro lado se utilizarán como teóricos usando las distribuciones normal, t-student, chi cuadrado, F-fisher según corresponda. (Los valores se encontrarán en los textos guía de la bibliografía o la guía de estudio, estos son conocidas como tabla de distribuciones.)

Región de rechazo

Es el conjunto de posibles valores del estadístico de prueba que llevan a rechazar la hipótesis nula.

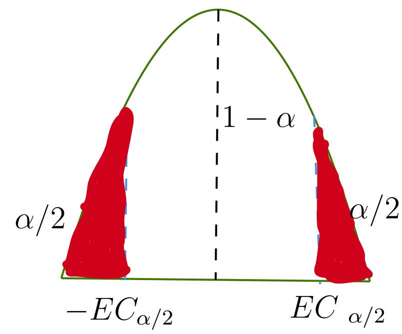


Figura 3.2: Region de rechazo bilateral

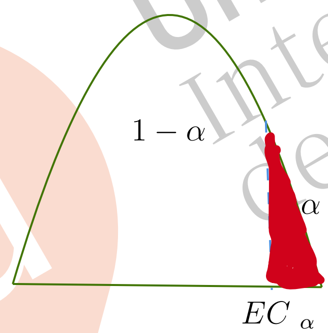


Figura 3.3: Region de rechazo unilateral derecha

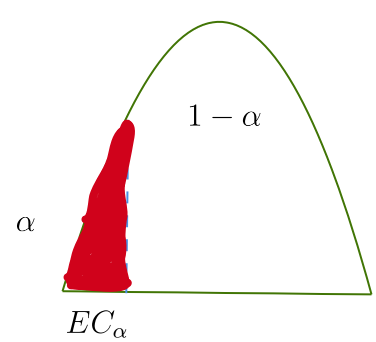


Figura 3.4: Region de rechazo unilateral izquierda

3.6.3. Prueba de hipótesis- Comparación de medias I

- $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0 (\mu \leq \mu_0 \ \ \mu \geq \mu_0)$
- $\alpha = 0{,}05, 0{,}01, 0{,}1$
- Estadistico de prueba o contraste:
  varianza poblacional conocida

$$Z_0 = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

- Región de rechazo $|Z_0| > Z_{\alpha/2}$, $Z_0 > Z_\alpha$, $Z_0 < -Z_\alpha$

Ejemplo 3.6.2

La viscosidad de un detergente liquido debe promediar 800 *centistokes a* 25žC*. Se colecta una muestra aleatoria de 16 lotes del detergente y la viscocidad promedio es de* 812 *Suponga que la desviación estándar de la viscosidad es* $\sigma = 25$ *centistokes. Enunciar las hipótesis que deberán probarse. Probar las hipótesis utilizando* $\alpha = 0{,}05$ *¿A qué conclusiones llega?*

3.6.4. Prueba de hipótesis- Comparación de medias II

- $H_0 : \mu_1 = \mu_2$
  $H_1 : \mu_1 \neq \mu_2 (\mu_1 \leq \mu_2 \ \ \mu_1 \geq \mu_2)$
- $\alpha = 0{,}05, 0{,}01, 0{,}1$
- Estadistico de prueba o contraste:
  varianzas poblacionales conocidas

$$Z_0 = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Región de rechazo $|Z_0| > Z_{\alpha/2}$, $Z_0 > Z_\alpha$, $Z_0 < -Z_\alpha$

Ejemplo 3.6.3

Un fabricante de calculadoras electrónicas puede usar dos tipos de plástico. La resistencia a la ruptura de este plástico es importante. Se sabe que $\sigma_1 = \sigma_2 = 1$ *de muestras aleatorias* $n_1 = 10$ *y* $n_2 = 12$ *se obtiene que* $\bar{y_1} = 162{,}5$ $\bar{y_2} = 155$*. La compañía no empleará el plástico 1a menos que su resistencia a la ruptura le exceda a la del plástico 2 por almenos 10. Con base en la información muestral ¿Deberá usarse el plástico 1? para responder esta pregunta* *se deben establecer y probar las hipótesis apropiadas usando* $\alpha = 0{,}05$.

3.6.5. Prueba de hipótesis- Comparación de medias III

- $H_0 : \mu = \mu_0$
  $H_1 : \mu \neq \mu_0 (\mu \leq \mu_0 \ \ \mu \geq \mu_0)$
- $\alpha = 0{,}05, 0{,}01, 0{,}1$
- Estadistico de prueba o contraste:
  varianza poblacional desconocida
  $$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$
- Región de rechazo $|t_0| > t_{\alpha/2,n-1}$ $t_0 > t_{\alpha,n-1}$, $t_0 < -t_{\alpha,n-1}$

Ejemplo 3.6.4

La vida de anaquel de una bebida carbonatada es motivo de interés. Se seleccionan 10 botellas al azar y se prueban obteniéndose los siguientes resultados (dias):

$$108, 138, 124, 163, 124, 159, 106, 134, 115, 139$$

Es posible que la vida media de anaquel exceda los 120 dias? Probar las hipótesis usando $\alpha = 0{,}01$.

Prueba de hipótesis- Comparación de medias IV

- $H_0 : \mu_1 = \mu_2$
  $H_1 : \mu_1 \neq \mu_2 (\mu_1 \leq \mu_2 \ \ \mu_1 \geq \mu_2)$
- $\alpha = 0{,}05, 0{,}01, 0{,}1$
- Estadistico de prueba o contraste:
  varianzas poblacionales conocidas
  $$t_0 = \frac{\bar{x_1} - \bar{x_2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
  donde $s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
- Región de rechazo $|t_0| > t_{\alpha/2,n_1+n_2-2}$, $t_0 > t_{\alpha,n_1+n_2-2}$ $t_0 < -t_{\alpha,n_1+n_2-2}$

Ejemplo 3.6.5

A continuación se presenta el tipo de combustión de dos cohetes químicos con formulaciones diferentes. Los ingenieros de diseño se interesan en la media como la varianza del tiempo de combustión.

	Tipo 1	Tipo 2	
65	82	64	56
81	67	71	69
57	59	83	74
66	75	59	82
82	70	65	79

Probar las hipótesis de que las y de que los tiempos de combustión promedio son iguales ($\alpha = 0{,}05$)

3.7. Ejercicios resueltos

Ejercicio resuelto 1.

Con la información proporcionada por el ejemplo 2.6.3.

- Cuál es el valor p para la prueba
  $P(z \leq 1{,}92) = 0{,}0274$ luego $P = 2(0{,}0274) = 0{,}0549$
- Encontrar un intervalo de confianza para la media al $95\,\%$

$$IC = \bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 812 \pm (1{,}96)\frac{25}{\sqrt{16}}$$

  con lo cual $\mu \in [799{,}75; 824{,}25]$ es decir que con una confianza del $95\,\%$ la viscosidad de un detergente promedia entre 799.75 y 824.25.

Ejercicio resuelto 2

Encontrar un intervalo de confianza para el ejemplo 2.6.4

$$IC(\mu_1 - \mu_2) = \bar{x_1} - \bar{x_2} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = [6{,}39; 8{,}60]$$

Ejercicio resuelto 3

Construir un intervalo de confianza de $99\,\%$ para la vida media de anaquel con los datos del ejercicio 2.6.5

$$IC = \bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = [110{,}914; 151{,}086]$$

Ejercicio resuelto 4

Probar las hipótesis de que las dos varianzas son iguales ($\alpha = 0{,}05$) del ejemplo 2.6.5. **Igualdad de varianzas**

(i.)

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

(ii.) $\alpha = 0{,}05$

(iii.) Estadístico de contraste. $F_0 = \frac{s_1^2}{s_2^2} = 0{,}97$

(iv.) Como el test es bilateral, $F_{\alpha/2,n_1-1,n_2-1} = F_{\alpha/2,9,9} = 0{,}024$ no se rechaza, la hipótesis nula, es decir que las varianzas poblacionales son iguales.

Ejercicio resuelto 5

En un artículo de Solid State Technology, "Diseño ortogonal para optimización de procesos y su aplicación en el grabado químico con plasma"se describe un experimento para determinar el efecto de la velocidad de flujo de C_2F_6 sobre la uniformidad del grabado de una oblea de sicilio usada en la fabricación de circuitos integrados. los datos de la velocidad de flujo son los siguientes:

Flujo	1	2	3	4	5	6
125	2.7	4.6	2.6	3	3.2	3.8
200	4.6	3.4	2.9	3.5	4.1	5.1

La velocidad de flujo afecta la uniformidad del grabado promedio?

(i.)

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

(ii.) $\alpha = 0{,}05$

(iii.) Estadístico de contraste. $t_0 = \frac{\bar{x_1} - \bar{x_2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -1{,}34$

(iv.) Como el test es bilateral, $t_{\alpha/2,n_1+n_2-2} = t_{\alpha/2,10} = -2{,}22$ no se rechaza, la hipótesis nula, es decir que La velocidad de flujo no afecta la uniformidad del grabado promedio.

La velocidad del flujo afecta la variablilidad de una oblea a otra en la uniformidad del grabado? (Usar $\alpha = 0{,}05$).

(i.)

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

(ii.) $\alpha = 0{,}05$

(iii.) Estadístico de contraste. $F_0 = \frac{s_1^2}{s_2^2} = 0{,}86$

(iv.) Como el test es bilateral, $F_{\alpha/2,n_1-1,n_2-1} = F_{\alpha/2,5,5} = 5{,}505$ no se rechaza, la hipótesis nula, es decir que la velocidad de flujo no afecta la variabilidad en la uniformidad de grabado.

Ejercicio resuelto 6

Doce inspectores midieron el diámetro de un cojinete de bolas, utilizando cada uno dos tipos diferentes de calibradores.

Inspector	Calibrador 1	Calibrador 2
1	0.265	0.264
2	0.265	0.265
3	0.266	0.264
4	0.267	0.266
5	0.267	0.267
6	0.265	0.268
7	0.267	0.264
8	0.267	0.265
9	0.265	0.265
10	0.268	0.267
11	0.268	0.268
12	0.265	0.269

Existe diferencia significativa entre las medias de la población de mediciones de las que se seleccionaron dos muestras?. Encontrar el p-valor y construir un intervalo de confianza para la diferencia de las mediciones.

(i.) $\begin{cases} H_0 : \mu_d = 0 \\ H_1 : \mu_d \neq 0 \end{cases}$

(ii.) $\alpha = 0{,}05$

(iii.) Estadístico de contraste $t_0 = \dfrac{\bar{d}}{s_d/\sqrt{n}} = 0{,}43$

(iv.) $t_{\alpha/2,11} = 2{,}20$ no hay evidencia estadística con lo cual no se rechaza la hipótesis nula; no existe diferencia significativa entre las medias de la población de mediciones

Ejercicio resuelto 7

Se están comparando dos populares analgésicos con base en la rapidez de absorción del cuerpo. Específicamente se afirma que la tableta 1 absorbe el doble de rapidez de la tableta 2. Suponer que σ_1^2, σ_2^2 se conocen. Desarollar un estadístico de prueba para

$$\begin{aligned} H_0 &: 2\mu_1 = \mu_2 \\ H_1 &: 2\mu_1 \neq \mu_2 \end{aligned}$$

En este caso, se conocen las varianzas poblacionales y se conoce que $E(\bar{x}) = \mu$ y $V(\bar{x}) = \frac{\sigma^2}{n}$, se tienen dos muestras donde la rapidez de la primera tableta absorbe el doble de la segunda, es decir que $2\bar{x_1} - \bar{x_2} \sim N(\mu, \sigma^2)$

$$E(2\bar{x_1} - \bar{x_2}) = E(2\bar{x_1}) - E(\bar{x_2}) = 2E(\bar{x_1}) - E(\bar{x_2}) = 2\mu_1 - \mu_2$$

$$\begin{aligned} V(2\bar{x_1} - \bar{x_2}) &= E((2\bar{x_1} - \bar{x_2})^2) - (E(2\bar{x_1} - \bar{x_2}))^2 \\ &= E(4\bar{x_1}^2 - 4\bar{x_1}\bar{x_2} + \bar{x_2}^2) - (4(E(\bar{x_1})^2 - 4E(\bar{x_1}\bar{x_2}) + (E(\bar{x_2})^2) \\ &= 4E(\bar{x_1}^2) - 4E(\bar{x_1}\bar{x_2}) + E(\bar{x_2}^2) - 4(E(\bar{x_1})^2 + 4E(\bar{x_1}\bar{x_2}) - (E(\bar{x_2})^2) \\ &= 4E(\bar{x_1}^2) - 4(E(\bar{x_1})^2 + E(\bar{x_2}^2) - E(\bar{x_2})^2 \\ &= \frac{4n_1}{n_1}(E(\bar{x_1}^2) - (E(\bar{x_1})^2) + \frac{n_2}{n_2}(E(\bar{x_2}^2) - E(\bar{x_2})^2) \\ &= 4\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \end{aligned}$$

Luego $Z_0 = \frac{2\bar{x_1} - \bar{x_2}}{\sqrt{4\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ rechazando la hipótesis nula si, $|z_0| > z_{\alpha/2}$

Ejercicio resuelto 8

Suponer que se está probando

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned}$$

donde σ_1^2, σ_2^2 se conocen. Los recursos para hacer el muestreo son limitados, por lo que $N = n_1 + n_2$. Cómo deberían asignarse las N observaciones entre las dos poblaciones para obtener la prueba con la potencia más alta?
Se debería maximizar $Z_0 = \frac{2\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ con $N = n_1 + n_2$ Considerar $L = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{N - n_1}$ y derivando con respecto a n_1, se tiene que $\dfrac{dL}{dn_1} = -\frac{\sigma_1^2}{n_1^2} + \frac{\sigma_2^2}{(N-n_1)^2} = 0$ con lo cual se tiene que $\frac{\sigma_2^2}{(N-n_1)^2} = \frac{\sigma_1^2}{n_1^2}$ y $\frac{\sigma_2^2}{(n_2)^2} = \frac{\sigma_1^2}{n_1^2}$ así $\frac{n_1}{n_2} = \frac{\sigma_1}{\sigma_2}$

Ejercicio resuelto 9

Desarrollar la ecuaciones

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$$

y

$$\frac{s_1^2}{s_2^2}F_{1-\alpha/2,n_1-1,n_2-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2}F_{\alpha/2,n_1-1,n_2-1}$$

para un intervalo de confianza de $100(1-\alpha)\,\%$ para la varianza(as) de una distribución normal. Ambas ecuaciones como lo indican las hipótesis provienen de una distribución normal $X \sim N(\mu, \sigma^2)$ de tal manera que $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$, de modo que

$$P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 100(1-\alpha)\,\%$$

Por otro lado, $\frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} \sim F_{n_1-1,n_2-1}$ de modo que

$$P\left(\frac{s_1^2}{s_2^2}F_{1-\alpha/2,n_1-1,n_2-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2}F_{\alpha/2,n_1-1,n_2-1}\right) = 100(1-\alpha)\,\%$$

Ejericio resuelto 10

Desarollar una ecuación para un intervalo de confianza de $100(1-\alpha)\,\%$ para la diferencia de medias de dos distribuciones normales donde $\sigma_1^2 \neq \sigma_2^2$

$t_0 = \frac{\bar{x_1}-\bar{x_2}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim t_{\alpha/2,\nu}$

$P\left(-t_{\alpha/2,\nu} \leq t_0 \leq t_{\alpha/2,\nu}\right) = 100(1-\alpha)\,\%$ de modo que

$-t_{\alpha/2,\nu} \leq t_0 \leq t_{\alpha/2,\nu}$

$-t_{\alpha/2,\nu} \leq \frac{\bar{x_1}-\bar{x_2}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \leq t_{\alpha/2,\nu}$

$-t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \leq \bar{x_1}-\bar{x_2}-(\mu_1-\mu_2) \leq t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

$\bar{x_1}-\bar{x_2}-t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \leq \mu_1-\mu_2 \leq \bar{x_1}-\bar{x_2}+t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

donde $\nu = \dfrac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_2-1}}$

Ejercicio resuelto 11

1. El tiempo medio que un estudiante de la asignatura de estadística emplea para prepararse para el examen final es de 50 horas cuatrimestal y desviación típica de 20 horas. En la siguiente tabla se presentan el tiempo empleado de 28 estudiantes para la preparación del examen.

68	49	45	76	65	50
54	92	24	36	60	66
57	74	52	75	36	40
62	56	94	57	64	
72	65	59	45	33	

a) Según estudios, indican que el tiempo medio de preparación media es inferior a 50 horas cuatrimestral. Está de acuerdo? ($\alpha = 0{,}01$)
Los datos del problema están dados $\mu = 50$, $\sigma = 20$, $n = 28$. se contrastará un test de hipótesis (prueba unilateral).

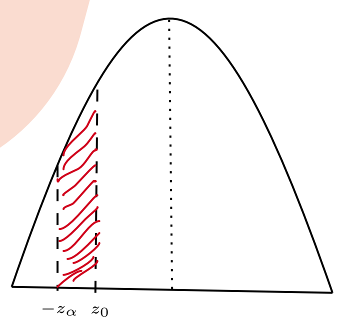
1. $\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu < 50 \end{cases}$

2. $\alpha = 0{,}01$

3. Estadístico de contraste. Cómo la desviación estándar es conocida, entonces $\bar{x} = 58{,}0714$:

$$Z_0 = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{58{,}0714 - 50}{20/\sqrt{28}} = 2{,}1354$$

4. Región de rechazo: $2\alpha = 0{,}02$ y $1 - 0{,}02 = 0{,}98$ (buscar en la tabla) luego $Z_{\alpha=0{,}02} = 2{,}33$



Cómo el test es unilateral y $Z_0 > -Z_\alpha$ No hay evidencia estadística para rechazar la hipótesis nula, con lo cual el tiempo medio de preparación no es inferior a 50 horas, es decir que no se está de acuerdo con los estudios realizados.

b) Encontrar un intervalo de confianza al $95\,\%$ para el tiempo medio de preparación para el examen.

(Confianza $95\,\%$ $\alpha = 0{,}05$)

$$\mu \in \bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
$$\mu \in 58{,}0714 \pm Z_{0{,}05/2}\frac{20}{\sqrt{28}}$$
$$\mu \in 58{,}0714 \pm 1{,}96\frac{20}{\sqrt{28}} \quad \text{(tabla sesión 7)}$$
$$\mu \in 58{,}0714 \pm 7{,}4081$$

con lo cual $\mu \in (58{,}0714 - 7{,}4081; 58{,}0714 + 7{,}4081) = (50{,}6633; 65{,}4795)$ es decir que con una confianza del $95\,\%$ el tiempo medio de preparación del examen se encuentra entre 50.6633 y 65.4795 horas.

c) Si se tiene un error estándar de 0.04. De acuerdo a la información anterior cuál es el tamaño de la población de los estudiantes que se preparan para el examen de estadística a nivel nacional?
$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2 \rightarrow n = \left(\frac{(1{,}96)(20)}{0{,}04}\right)^2 = 960400$

d) Si se desconoce la desviación típica (estándar), el tiempo medio que un estudiante emplea no sea 50 horas cuatrimestrales?($\alpha = 0{,}02$)
Los datos del problema están dados $\mu = 50$, σ(desconocida) $n = 28$. Se contrastará un test de hipótesis (prueba bilateral).

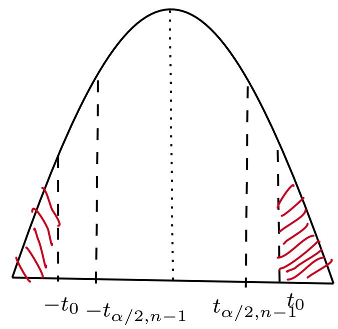
1. $\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$

2. $\alpha = 0{,}02$

3. Estadístico de contraste. Cómo la desviación estándar es desconocida, entonces $\bar{x} = 58{,}0714$ y $s = 16{,}6108$:

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{58{,}0714 - 50}{16{,}6108/\sqrt{28}} = 2{,}5712$$

4. Región de rechazo: $\frac{\alpha}{2} = 0{,}01$ y $1 - 0{,}01 = 0{,}99$ (buscar en la tabla t-student o comando en r $qt(1 - (0{,}02/2), 27)$) luego $t_{\alpha/2,n-1} = t_{0{,}02/2,28-1} = t_{0{,}01,27} = 2{,}4726$



Cómo el test es bilateral y $|t_0| > t_{\alpha/2,n-1}$ hay evidencia estadística para rechazar la hipótesis nula, con lo cual el tiempo medio de preparación es diferente a 50 horas.

e) Por medidas de precaución el día del examen se dividen los estudiantes en dos grupos: grupo A (correspondiente a las tres primeras columnas de la tabla, 15 estudiantes) grupo B: (tres últimas columnas de la tabla, 13 estudiantes). Hay diferencia en los tiempos medios de ambos grupos ? ($\alpha = 0{,}05$)
Con la información anterior se deben calcular los estadísticos descriptivos (media, varianza/desviación típica).

	A	B
n	15	13
$\bar{x}$	$61{,}5333$	$54{,}0769$
s^2	$310{,}4095$	$226{,}4103$
s	$17{,}6184$	$15{,}0469$

Es de destacar que las desviaciones típicas (estándar) poblacionales son desconocidas. Se contrastará un test de hipótesis (prueba bilateral).

1. $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$

2. $\alpha = 0{,}05$

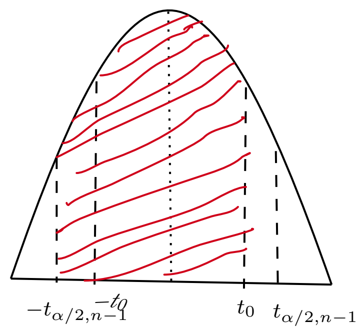
3. Estadístico de contraste:

$$t_0 = \frac{\bar{x_1} - \bar{x_2}}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{61{,}5333 - 54{,}0769}{(16{,}4815)\sqrt{\frac{1}{15} + \frac{1}{13}}} = 1{,}19401$$

donde

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(15 - 1)(310{,}4095) + (13 - 1)(226{,}4103)}{15 + 13 - 2}} = \sqrt{271{,}6406} = 1$$

4. Región de rechazo: $\frac{\alpha}{2} = 0{,}025$ (buscar en la tabla t-student o comando en r $qt(1-(0{,}05/2), 26)$) luego $t_{\alpha/2,n_1+n_2-2} = t_{0{,}05/2,15+13-2} = t_{0{,}025,26} = 2{,}0555$



Cómo el test es bilateral y $|t_0| < t_{\alpha/2,n_1+n_2-1}$ no hay evidencia estadística para rechazar la hipótesis nula, con lo cual el tiempo medio de preparación para los dos grupos es similar u homogéneo.