UNIVERSIDAD DEL NORTE

http://www.uninorte .edu .co

REPASO DE ESTADÍSTICA II

Antalcides Olivo

Julio del 2003

DISTRIBUCIONES MUESTRALES

Aproximación a una distribución normal estándar

Teorema 1.1 (Demoivre-Laplace.) ***Supóngase que Y tiene una distribución binomial con n pruebas y con probabilidad de tener éxito en cualquier prueba denotada por p. Por lo que podemos considerar a Y, como el número de éxitos en n pruebas, como una suma de una muestra formada por ceros y unos. Entonces***

$$\bar{X} = \frac{Y}{n}$$

tendrá aproximadamente una distribución normal con

$$E(X_i) = p$$
$$V(X_i) = \frac{p(1-p)}{n}$$

Ahora indicaremos como obtener una buena aproximación.

- **Si p se acerca a 0.5 y $n > 10$ la aproximación es muy buena.**

- **En general si $np > 5$ para0.5$\geq$ p o cuando $nq > 5$ si** $p > 0{,}5$

- **Si p se acerca al 0 o al 1 la aproximación no es buena**

Ejemplo 1 ***Durante una epidemia de gripa, se enfermaron el 30 % de la población. En un caserio de 2000 per-***

sonas, ¿Cuál es la probabilidad de que al menos 40 padezcan de la enfermedad?, calcular la probabilidad de que 60 personas padezcan con gripa.

Solución 1.1 ***Tenemos que la variable aleatoria que representa las personas con gripa es binomial***

$$
\begin{aligned}
X &\rightsquigarrow B\,(n = 200, p = 0{,}3) \\
\mu &= np = 60 \\
\sigma^2 &= npq = 42 \\
X &\approx X_N \rightsquigarrow N(60, 42)
\end{aligned}
$$

$$P\,(40 \geq X) = P\left(\frac{40-60}{\sqrt{42}} \geq \frac{X-60}{\sqrt{42}}\right) = 0{,}999$$

$$P\,(X = 60) = \binom{200}{60} p^{60} q^{140} \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(60-\mu)^2}{\sigma^2}} = 0{,}063$$

Otra forma de realizarlo es

$$P\,(X = 60) \approx P\,(60{,}5 \geq X_N \geq 59{,}5) = 0{,}0638$$

Distribución ji- cuadrado

Teorema 1.2 ***Sean*** $Z_1, Z_2, Z_3, \cdots, Z_k$ ***variables aleatorias distribuidas normalmente e independientes con media cero y varianza uno, entonces la variable aleatoria***

$$X^2 = \sum_{i=1}^{k} Z_i^2$$

tiene una dsitribución e probabilidad llamada JI-CUADRADO con k grados de libertad.
Donde $\mu = k$ ***y*** $\sigma^2 = 2k,$ ***si***

$$k \longrightarrow \infty \Longrightarrow \chi^2 \sim N\left(\mu, \sigma^2\right)$$

Y su función de densidad es :

$$f_{\chi_{\alpha,k}} \begin{cases} 0 & si \;\; x \in (-\infty, 0] \\ \frac{1}{2^{\frac{k}{2}}} \Gamma\left(\frac{k}{2}\right) x^{\frac{k}{2}-1} e^{-\frac{x}{2}} & si \;\; x \in (0, \infty) \end{cases}$$

Si $k \to \infty \Rightarrow \mathcal{X}^2 \sim N(\mu, \sigma^2)$

Ejemplo 2 ***Un ejemplo de una estadística ji-cuadrada es: supóngase que*** $X_1, X_2, X_3, \cdots, X_n$ ***es una muestra aleatoria de una población normal. con media*** μ ***y varianza*** σ^2 ***la función de varianza muestral :***

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

se distribuye como $\mathcal{X}^2_{n-1}$

Ejemplo 3 ***Encuentre el valor*** $\chi^2_{0,10,13}$ ***el valor es 19.812***

Ejemplo 4 ***Determine la probabilidad P***$\left(38{,}582 \geq \chi^2 \geq 6{,}844\right)$

Distribución t o Student

Teorema 1.3 ***Sea*** $Z \sim N(0{,}1)$ ***y V una variable ji-cuadrada con k grados de libertad ,si Z y V son independientes,***

entonces la variable aleatoria .

$$T = \frac{Z}{\sqrt{\frac{V}{k}}}$$

Se dice que tiene una distribución t con k grados de libertad y se abrevia t_k

La media y la varianza de la distribución t son $\mu = 0$ $\sigma^2 = \frac{k}{k-2}$ **para** $k > 2$**.**
Por ejemplo si queremos encontrar los valores $t_{\alpha,k}$ **para los cuales**
$\alpha \leq 0{,}05$ **tenemos :**

$$P\{t \geq t_{0,05,10}\} = P\{t \geq 1{,}813\} = 0{,}05$$

del mismo modo para la cola inferior $t_{0,95,10}$ **y** $-t_{0,05,10}$ **es -1.813.**

Ejemplo 5 ***Supóngase que*** $X_1, X_2, X_3, \cdots, X_n,$ ***es una muestra aleatoria de una distribución normal con media*** μ ***y***

Back

Close

varianza σ^2, y sean $\overline{X}$ y S^2 la media y la varianza de la muestra . considerese la estadística .

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \rightsquigarrow t_{n-1}$$

tiene una distribución t con n-1 grados de libertad.

Ejemplo 6 *Suponga que de una población con una media de 14 se toma una muestra de tamaño 11, si la media muestral es 18 y la desviación muetral es de 14.3 determine la probabilidad de que la media muestral sea mayor que la poblacional.*

Ejemplo 7 *Suponga que de una población n0rmal con media $\mu = 20$ se toma una muestra de tamaño 16. Si la desviación estándar S = 4, encuentre P$(\bar{X} < 21,753)$*

Solución 7.1 *Encontramos un valor t para 21,753. Como*

$$t = \frac{21,753-20}{4/4} = 1,753$$

Así

$$P\left(\bar{X} < 1,753\right) = 1 - P\left(\bar{X} \geq 1\right)$$

buscando en la tabla obtenemos

$$P\left(\bar{X} < 1,753\right) = 1 - 0{,}05$$

		α		
gl	**0.01**	**0.05**	**0.025**	**0.010**
12	**1.356**	**1.782**	**2.179**	**2.681**
13	**1.350**	**1.771**	**2.160**	**2.650**
14	**1.345**	**1.761**	**2.145**	**2.624**
15	**1.341**	**1.753**	**2.131**	**2.602**
16	**1.337**	**1.746**	**2.120**	**2.583**

Por tanto P($\bar{\mathbf{X}}$ > 1.753) =.05

Distribución F

Teorema 1.4 ***Sean W y Y variables aleatorias JI-cuadrado independientes con u y v grados de libertad respectiva-***

mente . EL cociente

$$F = \frac{W}{u} \Big/ \frac{V}{v}$$

se dice que tiene una distribución f co u grados de libertad en el numerador y v grados de libertad en el denominador y se denota:

$$F_{u,v}$$

La media y la varianza son:La variable aleatoria F es no negativa y la distribución es asimétrica a la derecha, como podemos observar F se asemeja a la ji-cuadrado, pero F está centrada alrededor de 1 y los paramétros u y v lo cual le proporcionan mayor flexibilidad en cuanto a la forma de su gráfica .

Ejemplo 8 ***Si tenemos dos poblaciones normales con varianzas σ_1^2 y σ_2^2 y si tomamos dos muestras aleatorias*** *independientes de tamaño* n_1 *y* n_2 *de las poblaciones 1 y 2 respectivamentes y que* S_1^2 *y* S_2^2 *son las varianzas muestrales entonces:*

$$\frac{S_1^2}{\sigma_1^2} \Bigg/ \frac{S_2^2}{\sigma_2^2}$$

Tiene una distribución F con $n_1 - 1$ **y** $n_2 - 1$ **grados de libertad en el numerador y en el denominador respectivamente.**

Los puntos porcentuales de la cola inferior $\mathbf{F}_{1-\alpha,u,v} = \frac{1}{F_{\alpha,v,u}}$

Ejemplo 9 ***En una prueba sobre la efectividad de dos píldoras para dormir, A y B se utilizaron dos grupos independientes de personas con insomnio, a un grupo de tamaño 40 se le administraró la píldora A y al otro***

grupo B de tamaño 60 se la admisnistró la píldora B, registrandose el número de horas de sueño de quienes usan cada tipo de píldora Si se supone que las distribuciones da cada una de las muestras es normal y que $\sigma_1^2 = \sigma_2^2$ ***calcule el valor del estadístico F y determine*** $P\left(\frac{S_1^2}{S_2^2} > 1{,}8\right)$

Ejemplo 10 ***Si*** S_1^2 ***y*** S_1^2 ***representan las varianzas de variables aleatorias independientes de tamaño*** $n_1 = 8$ ***y*** $n_2 = 12$***, tomadas de poblaciones normales con iguales varianzas, encuentre la***

$P\left(\frac{S_1^2}{S_2^2} < 4{,}89\right)$

Ejemplo 11 ***Determine*** $f_{0{,}95}$ ***con v =19 y u = 24***

Intervalos de confianza

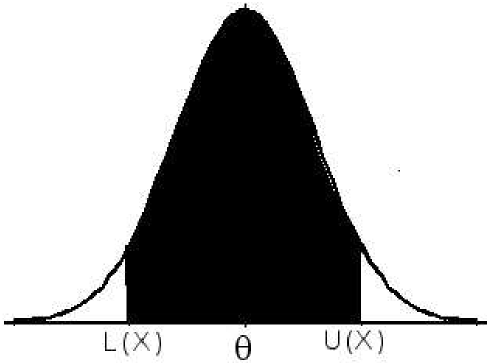
En las secciones anteriores se trató los métodos para estimar un parámetro puntual, usando una estadística adecuada $\widehat{\theta}(x)$ **cuyo valor** $\widehat{\theta}$ **se toma como el valor estimado**

del valor desconocido θ, pero en muchas situaciones éste método no nos proporciona suficiente información acerca del parámetro de interés, ya que sólo el número puede no tener mucho significado, entonces hay que hacer lo siguiente:

Definición 1 ***Teniendo una muestra aleatoria*** $X_1, X_2, X_3, \ldots, X_n$ ***que tiene una distribución*** $f(x, \theta)$***, la cual es conocida para cada parámetro*** θ ***dado un dato*** $x = (x_1, x_2, x_3, \ldots, x_n)$***, se procede a buscar dos estadísticas*** $L(x)$ ***y*** $U(x)$ ***que cumplan***
$L(X) \leq U(x)$ ***para cada posible dato*** x ***tal que :***
$P(L(X) < \theta < U(x)) \geq 1 - \alpha_0$ ***pra cierto*** $\alpha_0 \epsilon (0, 1)$ ***fijo lo más cercano posible a*** $1 - \alpha_0$ ***.***
A $I(x) = (L(x), U(x))$ ***se le llama intervalo de confianza para*** θ ***a un nivel de confianza*** $1 - \alpha_0$***.***

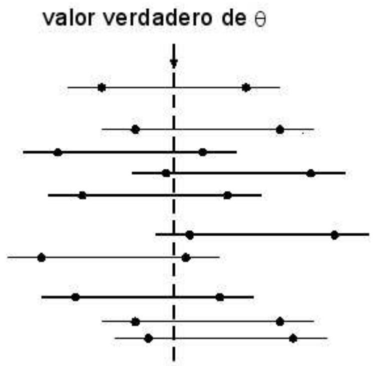
En la práctica se usan niveles $\alpha_0 = 5\%, 1\%, 0,1\%$ ya que la idea es encontrar un intervalo de confianza tal que la

probabilidad de que el parámetro de interés se encuentre dentro sea alto.
Por esta razón nos interesa un intervalo con límites muy próximos y con un nivel de confianza elevado.



Una interpretación correcta de la confianza´´$1-\alpha_0$ ´´tiene que ver con el concepto de probabilidad estocástica, es decir que si un experimento A se realiza una y otra vez de frecuencia relativa se acerca más al valor de la probabilidad cuantas más veces se realice.

O sea lo que obtenemos es la probabilidad de que el $100\left(1-\alpha_0\right)\,\%$ de los intervalos posibles con amplitud $U\left(X\right)-L\left(X\right)$ contienen al parámetro θ.



En la gráfica estamos suponiendo que solo podemos construir 10 intervalos de longitud $U\left(X\right)-L\left(X\right)$. (En realidad se pueden construir infinitos) Observamos en la gráfica que sólo un intervalo no contiene el parámetro θ, es decir, que el nivel de confianza es del $90\,\%$.

Intervalo de confianza como estimación

En muchas aplicaciones se usa una estimación $\widehat{\Psi} = q(\widehat{\theta})$, de $\Psi = q(\theta)$ para tomar $L(x) = \widehat{\Psi} - D$, $U(x) = \widehat{\Psi} + D$ con $D = D(x)$ una estadística escogida tal que cumpla la definición de intervalo de confianza, es decir:

$$I(x) = (\widehat{\Psi} - D, \widehat{\Psi} + D)$$

Entre más pequeño sea $D(x)$ o la longitud del intervalo más precisa será la estimación
Nota: D puede depender o no de x.

Ejemplo 12 ***Se tiene una variable de interés con esperanza μ desconocida y varianza σ^2 conocida, entonces se toma como parámetro $\theta = \mu$.***
Hay que suponer que las variables muestrales $X_i \sim N(\mu, \sigma^2)$ forman una muestra X de tamaño n para la variable de

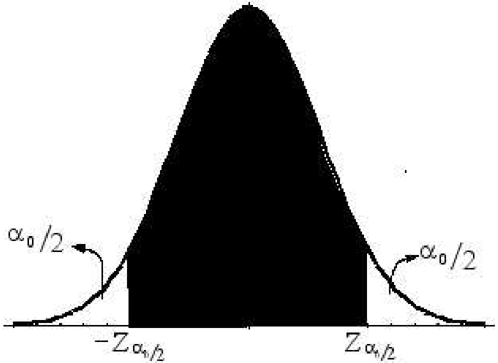
interés, por tanto
$\widehat{\mu} = \overline{X}$ ***es una estimación puntual razonable para*** μ ***por lo cual se buscará un intervalo de confianza donde se debe derterminar la estadística*** D ***tal que:***

$$P(\overline{X} - D < \mu < \overline{X} + D) = 1 - \alpha_0 \quad \textbf{dado} \quad \alpha_0$$

y para ello se hace lo siguiente :

$$\overline{X} - D < \mu < \overline{X} + D$$
$$-D < \overline{X} - \mu < D$$
$$-\frac{Dn^{1/2}}{\sigma} < \frac{n^{1/2}(\overline{X} - \mu)}{\sigma} < \frac{Dn^{1/2}}{\sigma}$$

Ahora si hacemos $Z = \frac{n^{1/2}(\overline{X}-\mu)}{\sigma} \sim N(0,1)$ ***y*** $z = \frac{n^{1/2}D}{\sigma}$ ***entonces***
$P(\overline{X} - D < \mu < \overline{X} + D) = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1 = 1 - \alpha_0$
fijando α_0 ***se obtiene el valor de*** z ***y luego*** $D = \frac{\sigma z}{n^{1/2}}$



NOTA : Si aumentamos el tamaño de la muestra se puede obtener un intervalo de confianza de una longitud tan pequeña como se quiera.

Ejemplo 13 ***Consideremos ahora que σ no se conoce, entonces en este caso $\theta = (\mu, \sigma)$ con $\Psi = q(\theta) = \mu$, que es lo que se va a estimar .***
En este caso $Z = \frac{n^{1/2}(\overline{X}-\mu)}{S} \sim T(n-1)$ o sea que σ^2 es reemplazada por su estimación insesgada .
Como T_{n-1} es simétrica entonces

$P(Z \geq z) = \alpha_0/2$ ***obteniendo así la estadística*** $D = \dfrac{Sz}{n^{1/2}}$, ***que ahora si depende de la muestra***

NOTA: Para una muestra fija de tamaño n ***y una desviación estándar*** σ, ***cuanto más alto es la nivel de confianza*** $100(1-\alpha_0)\,\%$, ***es mayor la longitud del intervalo de confianza resultante.***
Como la longitud del intervalo de confianza mide la presición de la estimación, observamos que la presición se relaciona inversamente con el nivel de confianza resultante.

Elección del tamaño de la muestra

La presición del del intervalo de confianza en el ejercicio 9 es $\frac{\sigma z}{n^{1/2}}$, **es decir el error** e **que se determina al usar** $\overline{X}$ **para estimar** μ **debe ser** $e = |\overline{X} - \mu| < D$, **con un nivel de**

confianza $100(1-\alpha_0)\,\%$ **lo que implica, que si el tamaño de la muestra se puede controlar, podemos escoger en** n **tal que el error especifico** e **sea mayor que el error al estimar** μ **para un nivel de confianza** $100(1-\alpha_0)$**.**
De lo que se deduce

$$\frac{z\sigma}{\sqrt{n}} \leq e \Rightarrow (\frac{z\sigma}{e})^2 \leq n$$

Intervalos de confianza sobre la diferencia de medias

Ejemplo 14 ***Considerese dos variables aleatorias*** X_{1j} ***y*** X_{2k} ***con medias*** μ_1 ***y*** μ_2***, respectivamente, conocidas y con varianzas*** σ_1^2 ***y*** σ_2^2***, respectivamente, conocidas. Encontrar un intervalo de confianza del*** $100(1-\alpha_0)\,\%$ ***para la diferencia de las medias*** μ_1 ***y*** μ_2***.***
Sea $X_{11}, X_{12}, X_{13}, \ldots, X_{1n_1}$, ***una muestra aleatoria de tamaño*** n_1 ***y***
$X_{21}, X_{22}, X_{23}, \ldots, X_{2n_2}$, ***una muestra aleatoria de tamaño*** n_2

Si $\overline{X}_{1*}$ ***y*** $\overline{X}_{2*}$ ***son las medias de las muestras, la estadística:***

$$Z = \frac{\overline{X}_{1*} - \overline{X}_{2*} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Si X_{1j} ***y*** X_{2k} ***son normales o aproximadamente normales de acuerdo con el teorema central del límite, entonces:***

$$P\left(\overline{X}_{1*} - \overline{X}_{2*} - D < \mu_1 - \mu_2 < \overline{X}_{1*} - \overline{X}_{2*} + D\right) = 1 - \alpha_0$$

De lo que se deduce que:

$$P\left[\frac{-D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < Z < \frac{D}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right] = 2\phi(z) - 1$$

Y como $\phi(z) = 1 - \alpha_0/2$ ***si fijamos*** α_0 ***tenemos que*** $z = \frac{D}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$***, entonces*** $D = z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ ***y si*** $n_1 = n_2$ ***tenemos que;***

el tamaño de la muestra se acota :

$$(\frac{z}{e})^2(\sigma_1^2+\sigma_2^2) \leq n$$

Ejemplo 15 ***Sean dos variables aleatorias independientes X_{1j} y X_{2k} con medias μ_1 y μ_2, respectivamente, conocidas, y con varianzas σ_1^2 y σ_2^2, respectivamente ,desconocidas. Encontrar un intervalo de confianza del $100(1-\alpha_0)\,\%$ para la diferencia de las medias μ_1 y μ_2.***

Sea $X_{11}, X_{12}, X_{13}, \ldots, X_{1n_1}$, una muestra aleatoria de tamaño n_1 y

$X_{21}, X_{22}, X_{23}, \ldots, X_{2n_2}$, una muestra aleatoria de tamaño n_2

Si $\overline{X}_{1*}$ y $\overline{X}_{2*}$ son las medias de las muestras y S_1^2 y S_2^2 las varianzas muestrales, entonces S_1^2 y S_2^2 son estimadores de σ_1^2 y σ_2^2 respectivamente, por lo que hay que considerar dos casos :

- ***si $\sigma_1^2 = \sigma_2^2 = \sigma$, entonces podemos encontrar un esti-***

mador ponderado de σ ***es decir :***

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

De lo que podemos deducir :

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

que es la suma de dos ji-cuadradas con n_1-1 ***y*** n_2-1 ***grados de libertad .***

Ahora podemos determinar el intervalo de confianza

$$P\{\overline{X}_{1*}-\overline{X}_{2*}-D<\mu_1-\mu_2<\overline{X}_{1*}-\overline{X}_{2*}+D\}=1-\alpha_0$$

después de ciertas operaciones llegamos a que podemos transformar la expresión en una que depende de:

$$\frac{\overline{X}_{1*}-\overline{X}_{2*}-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

si fijamos α_0 ***obtenemos la estadística*** $D = t_{\alpha_0/2} S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

NOTA : en los ejemplos 10 y 12 tenemos que si n_1 ***y*** n_2 ***son mayores que 30, la distribución*** t ***se aproxima a*** $N(0,1)$ ***por lo que el intervalo obtenido en el ejemplo 9 es equivalente al de estos dos casos .***

- ***Si las varianzas poblacionales son diferentes obtenemos una estadística :***

$$t = \frac{\overline{X}_{1*} - \overline{X}_{2*} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_v$$

donde:

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

de lo que se deduce que :

$$D = t_{\alpha/2,v}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Intervalo de confianza sobre la varianza de una distribución normal

Ejemplo 16 ***Se tiene una variable de interés que se distribuye normalmente con esperanza*** μ ***desconocida y varianza*** σ^2 ***desconocida, entonces se toma como parámetro*** $\theta = \sigma^2$ ***.***
Hay que suponer que las variables muestrales $X_i \sim N(\mu, \sigma^2)$ ***forman una muestra*** X ***de tamaño*** n ***para la variable de interés, por tanto si*** S^2 ***es la varianza muestral. determine el intervalo de confianza*** $100(1-\alpha_0)\,\%$ ***respecto a*** σ^2
Tenemos que determinar $P(h(z) \leq \sigma^2 \leq g(z)) = 1-\alpha_0$ ***donde*** $g(z)$ ***y*** $h(z)$ ***son funciones positivas .***

Como $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ ***si fijamos*** α_0 ***encontramos que :***

$$h(z) = \frac{(n-1)S^2}{\chi^2_{\alpha_0/2,n-1}} \quad y \quad g(z) = \frac{(n-1)S^2}{\chi^2_{1-\alpha_0/2,n-1}}$$

Intervalo de confianza sobre la razón de varianzas de dos distribuciones normales

Ejemplo 17 ***Sean dos variables aleatorias independientes X_{1j} y X_{2k} con medias μ_1 y μ_2, respectivamente,desconocidas, y con varianzas σ_1^2 y σ_2^2, respectivamente ,desconocidas. Encontrar un intervalo de confianza del $100(1-\alpha_0)\,\%$ respecto al cociente σ_1^2/σ_2^2***

Sea $X_{11}, X_{12}, X_{13}, \ldots, X_{1n_1}$, una muestra aleatoria de tamaño n_1 y

$X_{21}, X_{22}, X_{23}, \ldots, X_{2n_2}$, una muestra aleatoria de tamaño n_2 notamos que la distribución de muestreo

$$F = \frac{\frac{S_2^2}{\sigma_2^2}}{\frac{S_1^2}{\sigma_1^2}} \sim F_{n_2-1,n_1-1}$$

por lo que se puede demostrar que el intervalo de confianza es :

$$\frac{S_1^2}{S_2^2} F_{1-\alpha_0/2,n_2-1,n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{S_1^2}{S_2^2} F_{\alpha_0/2,n_2-1,n_1-1}$$

Intervalo de confianza sobre una proporción

Ejemplo 18 ***Teniendo una muestra aleatoria*** $X_1, X_2, X_3, \ldots, X_n$ ***que cada una de las variables aleatorias tienen una distribución binomial,y que*** $\widehat{p} = \overline{X}$ ***es el estimador puntual de*** p ***. Determine el intervalo de confianza del*** $100(1-\alpha_0)\,\%$ ***para*** p
Si p ***no está demasiada cerca de 0 ó 1, y si*** n ***es relativamente grande tenemos que***

$$Z = \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Por lo que podemos demostrar que el intervalo de confianza para p ***es :***

$$\widehat{p} - Z_{\alpha_0/2}\sqrt{\frac{p(1-p)}{n}} \leq p \leq \widehat{p} + Z_{\alpha_0/2}\sqrt{\frac{p(1-p)}{n}}$$

El tamaño de la muestra para que el error $e = |p - \widehat{p}|$ ***al estimar*** p ***sea menor que*** $Z_{\alpha_0/2}\sqrt{\frac{p(1-p)}{n}}$ ***es:***

$$n = (\frac{Z_{\alpha_0/2}}{e})^2 p(1-p)$$

Intervalo de confianza sobre la diferencia en dos proporciones

Ejemplo 19 ***Sean dos variables aleatorias independientes*** X_{1j} ***y*** X_{2k} ***tal que cada una de las variables aleatorias está distribuida binomialmente y*** $\widehat{p}_1 = \overline{X_{1*}}$ ***y*** $\widehat{p}_2 = \overline{X_{2*}}$ ***son estimadores de*** p_1 ***y*** p_2 ***respectivamente. Determinar el intervalo de confianza a un nivel de*** $100(1-\alpha)\,\%$
Tenemos que la variable:

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$$

Entonces el intervalo de confianza para $p_1 - p_2$ ***es:***

$$\widehat{p}_1 - \widehat{p}_2 - D \leq (p_1 - p_2) \leq \widehat{p}_1 - \widehat{p}_2 + D$$

Donde :

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

Prueba de hipótesis

Error de tipo I

Se escoge un valor crítico c tal que, para cierto $\alpha_0 \in (0,1)$ fijo se cumpla

$$P(T \geq c|H_0) = \sup_{\theta} P(\geq c|\theta) \leq \alpha_0$$

y que esta probabilidad tienda a α_0 .
Cometer un error de tipo I. Significa que podemos rechazar H_0 a pesar de que H_0 fuera correcta, entonces escogiendo c como se indica se controla la probabilidad del error de tipo I por medio de α_0 . Por lo que a veces se dice que se rechaza la hipótesis nula a un nivel de α_0 .

Back

Close

Criterio del p-valor

El valor p es la probabilidad de obtener una estadística de prueba igual o más exacta que el resultado obtenido a partir de los datos de la muestra, dado que la hipótesis, H_0 es realmente verdadera. Por lo que este críterio se conoce como el nivel de significación observado.

El procedimiento en este caso es calcular un $\alpha = P(T \geq t|H_0)$ para un $t = T(x)$ que resulta de los datos x .
Al valor $\alpha = \alpha(T(x))$ se le llama el p-valor correspondiente a la hipótesis H_0 y a la observación x.
En concordancia con el error de tipo I, a menor p-valor, mayor fuerza toma la decisión de rechazar H_0

Desviación del p-valor

Si α se encuentra en,

- $5\,\% \geq \alpha > 1\,\%$ **Se dice que el p-valor tiene una desviación casi significativa de** H_0
- $1\,\% \geq \alpha > 0,1\,\%$ **Se dice que el p-valor tiene una desviación significativa de** H_0
- $0,1\,\% \geq \alpha$ **Se dice que el p-valor tiene una desviación muy significativa de** H_0
- $\alpha > 5\,\%$ **Se acepta la hipótesis** H_0 **en el sentido que no se pudo encontrar una desviación adecuada para rechazar** H_0

Debido a que rechazar H_0 **siempre es una conclusión fuerte en cambio no rechazar generalmente es una conclusión débil, a menos que sepamos que** β **es pequeña, debemos en la mayor parte de los casos construir hipótesis tal que**

el enunciado en el cual se desea una conclusión fuerte, este en la hipótesis alternativa, es decir la hipótesis nula debe ser siempre de la forma $H_0 : \theta = \theta_0$, mientrás que la alternativa debe ser de una de las siguientes formas,

- $H_1 : \theta \neq \theta_0$
- $H_1 : \theta \geq \theta_0$
- $H_1 : \theta \leq \theta_0$

Pruebas de bondad de ajuste

Prueba chi cuadrado de bondad de ajuste

Suponemos ahora un problema que puede ser caracterizado por una variable aleatoria discreta cuyos valores representan K posibles categorias y ocurren con probabilidades

$p_k : k = 1, 2, 3, \cdots K$ **en el cual nos interesa la hipótesis** $\mathbf{H}_0 : p_k = p_{k_0}$; **Siendo** p_{k_0} **valores fijos, Contra la alternativa lógica** $\mathbf{H}_{1.}$

Como estadístico de prueba se escoge la llamada χ^2

$$X^2 = \sum_{k=1}^{K} \frac{(N_k - np_{k_0})^2}{np_{k_0}}$$

Se rechaza la hipótesis H_0 **con base en los valores concre-**

tos $n_1, \cdots, n_k$ si y sólo si para el valos x^2 de X^2 vale

$$x^2 \geq c,$$

donde c es algún valos crítico.

Ahora se usa el echo de que, bajo H_0, la estadística $X^2 \rightsquigarrow \chi^2_{K-1}$ en algunos libros se utiliza como notación para el estadístico χ^2

$$X^2 = \sum_{i=1}^{K} \frac{(O_k - E_k)^2}{E_k}$$

donde O_k es la frecuencia observada y E_k es la frecuencia esperada.

En el caso de que la distribución tenga un parámetro la hipótesis es

$$H_0 = p_k = p_k(\theta), k = 1, 2, \cdots, K$$

siendo $\theta = (\theta_1, \cdots, \theta_s)$, $s < K - 1$, un parámeto; Así

$$X^2 = \sum_{k=1}^{K} \frac{\left(N_k - np_{k_0}\left(\hat{\theta}\right)\right)^2}{np_{k_0}\left(\hat{\theta}\right)} \rightsquigarrow \chi^2_{K-1-s}$$

bajo H_0, así los grados de libertad se reducen exactamente un número igual al número de párametros, y rechazariamos la hipótesis si $X^2 > \chi^2_{K-1-s}$.

Ejemplo 20 ***Un científico de computadores ha desarrollado un algoritmo para generar enteros seudoaleatorios sobre un intervalo 0-9. codifica el algoritmo y genera 100 dígitos seudoaleatorios . Los datos se muestran en la tabla ¿Existe evidencia de que el generador de núme-***

ros aleatorios está trabajando correctamente ?

n	O_k	E_k
0	94	100
1	93	100
2	112	100
3	101	100
4	104	.
5	95	.
6	100	.
7	99	.
8	108	.
9	94	100
total	1000	1000

(tabla. 1)

Si el generador está trabajando correctamente , entonces los valores 0-9 deben seguir una distribución uniforme discreta . lo que indica que cada uno de los dígitos debe

ocurrir exactamente 100 veces , esto es $E_k = 10$, **para** $k = 0, 1, \cdots 9$, **entonces los grados de libertad son 10-1= 9 y el estadísitici de prueba tiene un valor**

$$X^2 = \frac{(O_k - 100)^2}{100} = 3,72$$

si tomamos un nivel de confianza de $\alpha = 0{,}05$ **tenemos que** $\chi^2_{0{,}05{,}9} = 16,92$ **por lo que no podemos rechazar la hipótesis, por lo que el generador de números parece estar trabajando satisfactoriamente**

Ejemplo 21 ***Se considera en forma hipótetica que el número de llamadas a una central teléfonica en un mismo interalo de tiempo sigue una distribución de Poisson. Una muestra aleatoria de 60 intervalos obteniendo los***

siguientes resultados

n^o de llamadas	O_k	E_k
0	32	
1	15	
2	9	
3	4	
total	60	

(tabla 2)

La media de la supuesta distribución de Pisson se desconoce y puede estimarse a partir de los datos de la muestra , así $\lambda = 0{,}7$, ***calculamos ahora las frecuencias esperadas como*** $E_k = np_k$ ***donde*** p_k ***es la probabilidad hipótetica asiciada a cada k-ésimo intervalo y n es el número total de observaciones . Las hipótesis apropia-***

das son

$$H_0 : p(x) = \frac{e^{-0{,}75}(0{,}75)^x}{x!} \quad x = 1, 2, 3, \cdots$$

$H_1 : p(x)$ ***no es de Poisson con parámetro*** λ

calculemos las frecuencias esperadas

n^o de llamadas	p_k	E_k
0	***0.42***	
1	***0.54***	
2	***0.133***	
3	***0.033***	

(tabla 3)

Los grados de libertad son $k-1-s$ ***=1 y el estadístico de pruba tiene un valor de*** X^2 ***=3.24 , por lo que*** $\chi^2_{0{,}05{,}1} = 3{,}84$, ***así nos queda que no podemos rechazar la hipótesis***

Pruebas de tablas de contingencia

Prueba de independencia **En un estudio estadístico es importante averiguar si dos variables de clasificación , ya sean cualitativas o cuantitativas son independientes.**

En esta prueba los n elementos de muestra de una población pueden clasificarse de acuerdo con dos criterios diferemtes. Por ello es interesante saber si los dos métodos de calsificación son estadísticamente independientes. Supongamos por ejemplo que el primer método tiene r niveles y que el segundo método c niveles . Sea O_{ij} la frecuencia observada para el nivel i y para el nivel j de los dos métodos de clasificación , por lo que los datos aparecerán como una tabla de r renglones y c columnas

Para La prueba de independencia se usan las siguientes hipótesis

H_0 : Las dos variables de clasificación son independientes.

$\mathbf{H}_1$: **Las dos variables de clasificación son dependientes.**
Usandose el estadístico de prueba

$$X^2 = \sum_{i,j=1}^{r,c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i,j=1}^{r,c} \frac{O_{ij}^2}{E_{ij}} - n$$

Donde $p_{ij} = u_i v_j$ **y**

$$\hat{u}_i = \frac{1}{n} \sum_{j=1}^{c} O_{ij}$$

$$\hat{v}_j = \frac{1}{n} \sum_{i=1}^{r} O_{ij}$$

$$E_{ij} = n\hat{u}\hat{v} = \frac{1}{n} \sum_{i=1}^{r} O_{ij} \sum_{j=1}^{c} O_{ij}$$

Así tenemos que $X^2 \rightsquigarrow \chi^2_{(r-1)(c-1)}$ **y se rechazará la hipótesis si** $X^2 > \chi^2_{(r-1)(c-1)}$

Ejemplo 22 ***En un estudio para determinar si las opiniones sobre el cierre de escuelas están relacionadas con la profesión, se obtuvo la tabla de contingencia siguiente***

$profesión$

Opinión	***Profesores***	***Abogados***	***Médicos***	$\sum$
A favor	***3***	***67***	***15***	***85***
Se opone	***105***	***50***	***75***	***230***
No opinó	***4***	***3***	***8***	***15***
$\sum$	***112***	***120***		

(tabla 6)

Haga una prueba con $\alpha = 0{,}05$. ***Para determianr si la opinión está relacionada con la profesión***

Solución 22.1 H_0 : ***La opinión y las profesiones son independientes.***
H_1 : ***La opinión y las profesiones son dependientes.***

Encontramos organizados los resultados en la tabla 7

celda ij	O_{ij}	E_{ij}	$\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$
1,1		28.85	23.16
1,2		30.91	42.14
1,3		25.24	4.15
2,1		78.06	9.29
2,2		83.64	13.53
2,3		68.30	0.66
3,1		5.09	0.23
3,2		5.45	1.10
3,3		4.46	2.81
$\sum$			97.084

(tabla 7)

gl=4 $\boldsymbol{X}^2 = 97{,}08$ $\chi^2_{0{,}05{,}4} = 4{,}48$ ***por tant, rechazamps*** $\boldsymbol{H}_0$ ***por que tenemos evidencia estadística de la opinión del cierre de las escuelas estárelacionada con la profesión.***

Tipo de error posible : Tipo I , $\alpha = 0{,}05$.
Valor p : $p < 0{,}005$

Anova

Anova con un factor

El método de ANOVA es un críterio que requiere del cálculo de dos estimaciones independientes para σ^2**, la varianza poblacional común.**

estas dos estimaciones las denominaremos S_b^2, S_w^2 **,donde** $\mathbf{S}_b^2$ **es la estimación de la varianza entre las muestras y** $\mathbf{S}_w^2$ **es la estimación de la varianza interior de las muestras; con lo que resulta el estadístico**

$$F = \frac{S_b^2}{S_w^2}.$$

En este caso tenemos k **muestras como se ilustra en la**

tabla

$$\begin{array}{ccccc} \bar{X}_i & \bar{X}_1 & \bar{X}_1 & \cdots & \bar{X}_k \\ S_i & S_1 & S_2 & \cdots & S_k \\ n_i & n_1 & n_2 & \cdots & n_k \end{array}$$

Para simplificar los cálculos suponemos que

$$n = n_1 = n_2 = \cdots = n_k,$$

entonces hallamos S_w^2 **la estimación ponderada para** σ^2

$$S_w^2 = \sum_{i=1}^{k} \frac{S_i^2}{k},$$

como esta estimación se basa en k **muestras cada una de tamaño** n **y cada una tiene** $(n-1)$ **grados de libertad asociados ellas entonces los grados de libertad** gl_w **asociados a** $\mathbf{S}_w^2$ **es**

$$gl_w = \sum_{i=1}^{k} (n-1) = k\,(n-1)$$

Back
Close

generalizando

$$S_w^2 = \frac{\sum_{i=1}^{k} (n_i - 1)\, S_i^2}{\sum_{i=1}^{k} n_i - k}$$

$$gl_w = \sum_{i=1}^{k} n_i - k$$

$$\bar{X} = \sum_{i=1}^{k} \frac{\bar{X}_i}{k}$$

$$S_b^2 = n \sum_{i=1}^{k} \frac{\left(\bar{X}_i - \bar{X}\right)}{k-1}$$

donde los grados de libertad asociados a S_b^2 son

$$gl_b = k - 1$$

Ejemplo 23 ***Suponga que estamos interesados en determinar cual de los tres automóviles medianos con el mismo cilindraje es más eficiente en términos de millas***

recorridas, si los tres tienen las mismas características técnicas

Citation	*Le Baron*	*Reliant*
23	*30*	*27*
24	*31*	*27*
24	*32*	*26*
24	*32*	*26*
25	*30*	*24*

(tabla 3.1)

Solución 23.1 ***Si denotamos por*** μ_1, μ_2, μ_3 ***las medias poblacionales del millaje por camino de gasolina para las marcas citadas, podemos escribir las hipótesis estadísticas como***

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : ***Al menos dos medias poblacionales son distintas***

Coloquemos en una tabla los datos estadisticos

$Citation$	$Lebaron$	$Reliant$
$n = 5$	$n = 5$	$n = 5$
$\bar{X} = 24$	$\bar{X} = 31$	$\bar{X} = 26$
$S^2 = 0{,}5$	$S^2 = 1$	$S^2 = 1{,}5$

entonces

$$S_w^2 = 1, \;\; \bar{X} = 27, \;\; S_b^2 = 65$$

$$F = \frac{65}{1} = 65$$

$$F_{0{,}05{,}2{,}12} = 3,89$$

Por lo que se rechaza la hipótesis nula

Modelos lineales

Definición 2 ***Sean Y y X variables aleatorias y supongamos que la relación existente entre ellas es***

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{1}$$

donde $\epsilon \sim N\left(0, \sigma^2\right)$ ***es un error aleatorio, es decir***

$$E\left(Y|X\right) = \beta_0 + \beta_1 X \tag{2}$$

Notese que el modelo es lineal con relación a los llamados coeficientes de regresión β_0, β_1, ***es decir el modelo***

$$Y = \beta_0 + \beta_1 X^n + \epsilon, \tag{3}$$

Donde$n \in IN$ **es también un modelo de regresión lineal.**

En la definición 5.1 se establecieron 4 supuestos los cuales son

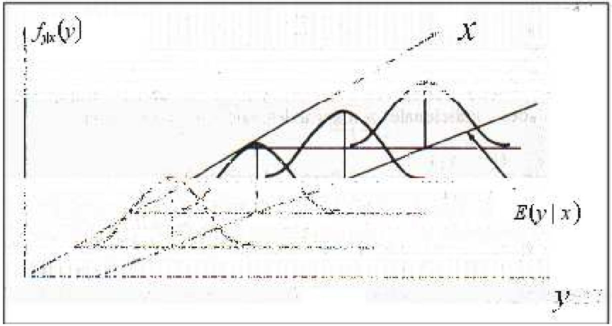
1. **Para cada valor de x_i, las variables aleatorias ϵ_i se distribuyen noermalmente**
2. **Para cada valos x_i, la media o valor esperado de ϵ_i es cero**
3. **Para cada valos x_i, la varianza de ϵ_i es contante**
4. **Los ϵ_i son independientes**

Como consecuencia de los cuatro supuestos se pueden hacer las siguientes observaciones

1. **Los valores de x son fijos**
2. **Los valores de los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$ son constanes, pero desconocidas, sim embargo pueden estimarse basandonos en los cuatro supuestos y el método de los mínimos cuadrados**
3. **Como el valor de y para un x fijo está determinado por la relación $y = \beta_0 + \beta_1 x + \epsilon$. Por tanto los valores de**

y dependeran de los valores de ϵ. Por tanto y es una variable aleatoria

Para un valor fijo de x, la distribución muestral de y es normal, porque sus valores dependen de ϵ , y los valores de ϵ se distribuyen normalmente. Como muestra la figura.



La distribución muestral de y para un valor fijo de x tiene una media denotada por $\mu_{y|x}$, donde la ecuación $E(y|x) = \beta_0 + \beta_1 x$ se llama ecuación de regresión poblacional.

Ahora determinaremos los LS-Estimadores $\hat{\beta}_0$ y $\hat{\beta}_1$ de β_0 y β_1 respectivamente

Un modelo lineal de acuerdo con la definición 5.1 se llama modelo lineal simple, para este modelo supòngamos que tenemos n pares de observaciones, por ejemplo

$$(y_1, x_1), (y_2, x_2), (y_3, x_3), \cdots, (y_n, x_n).$$

Entonces empleamos estos datos para estimar los parámetros desconocidos β_0 y β_1 por medio del método de los mínimos cuadrados explicado en el capítulo 2.2.3. Por lo que

$$\begin{aligned} W(\beta_0, \beta_1) &= \sum_{i=1}^{n} e_i^2 \\ &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned}$$

Back
Close

Entonces los LS-Estimadores de β_0 y β_1 deben satisfacer

$$\left.\frac{\partial W}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\left.\frac{\partial W}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) x_i = 0$$

resolviendo este sistema de ecuaciones se obtienen $\hat{\beta}_0$ y $\hat{\beta}_1$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}$$

Entonces el modelo de regresión lineal simple ajustado es

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Back
Close

Para simplificar la notación notaremos

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n} y_i x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}$$

Así $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$.

Estimación puntual **analizaremos ahora las propiedades de los estimadores** $\hat{\beta}_0$ **y** $\hat{\beta}_1$

- $E\left(\hat{\beta}_1\right) = E\left(\frac{S_{xy}}{S_{xx}}\right) = \beta_1$
- $V\left(\hat{\beta}_1\right) = V\left(\frac{S_{xy}}{S_{xx}}\right) = \frac{\sigma^2}{S_{xx}}$
- $E\left(\hat{\beta}_0\right) = \beta_0$
- $V\left(\hat{\beta}_0\right) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$

- $cov = \left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\sigma^2 \bar{x}}{S_{xx}}$

Queda como ejercicio la prueba de estas propiedades .

De acuerdo con esto observamos que $\hat{\beta}_0$ y $\hat{\beta}_1$ son estimadores insesgados .

Ahora si notamos

$$SS_E = \sum_{i=1}^{n} e_i^2 = S_{yy} - \hat{\beta}_1 S_{xy}$$

entonces $\hat{\sigma}^2 = \frac{SS_E}{n-2} \equiv MS_E$ es un estimador insesgado de σ^2 y $S_{yy} \equiv MS_R$

Prueba de hipótesis

Nos queda presentar los estadísticos que utilizaremos para una prueba de hipótesis

1. **Si tenemos las hipótesis**

$$H_0 : \beta_1 = \beta_{1,0}$$
$$H_1 : \beta_1 \neq \beta_{1,0}$$

entonces se usa el estadístico

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{MS_E}{S_{xx}}}} \rightsquigarrow t_{\alpha/2,n-2}$$

2. **Si**

$$H_0 : \beta_0 = \beta_{0,0}$$
$$H_1 : \beta_0 \neq \beta_{0,0}$$

Se usa el estadístico

$$t = \frac{\hat{\beta}_0 - \beta_{0,0}}{\sqrt{\frac{MS_E}{S_{xx}}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \rightsquigarrow t_{\alpha/2,n-2}$$

3. **Si**

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Se usa el estadístico

$$F = \frac{MS_R}{MS_E} \rightsquigarrow F_{\alpha,1,n-2} \rightsquigarrow t^2_{\alpha/2,n-2}$$

Intervalos de confianza

Es posible encontrar los límites para un intervalo de confianza del $(1-\alpha)\,100\,\%$ **para cada caso**

1. **Para** β_1 **es**

$$\hat{\beta} \pm t_{\alpha/2}\sqrt{\frac{MS_E}{S_{xx}}}$$

donde

$$t_{\alpha/2} \rightsquigarrow t_{\alpha/2,n-2}$$

2. Una vez que se ha encontrado la ecuación de regresión muestral y que se ha determinado el modelo $E\left(y|x\right) = \beta_0 + \beta_1 x$, podemos usar la ecuación de regresión $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$ para realizar predicciones; al hacerlo queremos estimar el valor promedio para y dado x, es decir $E\left(y|x\right)$. Así como predecir el valor aleatorio y para un valor x dado.

   Los límites del intervalo de confianza para $E\left(y|x_0\right)$ estan dados por la fórmula

$$\hat{y} \pm t_{\alpha/2}\sqrt{SM_E\left(\frac{1}{n} + \frac{\left(x_0 - \bar{x}\right)^2}{S_{xx}}\right)}$$

   donde

$$t_{\alpha/2} \rightsquigarrow t_{\alpha/2,n-2}$$

3. Los límites del intervalo de confianza para un valor de un solo valor aleatorio y dado un valor particular $x = x_0$

están dados por la expresión

$$\hat{y} \pm t_{\alpha/2}\sqrt{SM_E\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

donde

$$t_{\alpha/2} \rightsquigarrow t_{\alpha/2,n-2}$$

Ejercicios 7.1 ***En un estudio para determinar si las tasas de desempleo son distintas para las ciudades del norte, centro y occidente del país, se eligieron muestras aleatorias en las principales ciudades de cada región para analizarlas. Los datos de la tabla 5 representan los porcentajes de las tasas de desempleo. Si suponemos que los***

supuestos para el anova se cumplen y usamos $\alpha = 0{,}05$

Norte	*Centro*	*Occidente*
5.2	*11.4*	*7.2*
11.5	*9.1*	*15.3*
6.3	*6.6*	*10.3*
6.6	*10.5*	*9.5*
7.7	*3.6*	
3.8		
7.6		

(4)

Ejercicios 7.2 *La agencia de protección del medio ambiente proporcionó la información de la tabla5, que compara el tamaño del motor en pulgadas cúbicas de desplazamiento (cid), y las millas por galón (mpg) estimadas* *para 8 vehiculos compactos modelos 1984*

Coches	*Tamaño del motor*	
Cavalier	*121*	*30*
Datson	*120*	*31*
Omni	*97*	*34*
Escort	*98*	*27*
Mazda 626	*122*	*29*
Horzión	*97*	*34*
Corola	*85*	*38*
Alliance	*122*	*32*

(5)

1. a) *Realice el diagrama de dispersión*

   b) *Determine las estimacuones de* $\beta_0, \beta_1, \sigma^2,$ *y* , $\hat{y}(98)$

   c) *Realice una prueba de hipótesis en cada caso*

      1) $\beta_1 = 0$

      2) $\beta_1 = -0{,}15$

      3) $\beta_0 = 47$

4) $\hat{y}$ si $x_0 = 98$

Referencias

[1] **Williams Mendenhall.** ***Estadística matemática con aplicaciones***
Grupo editorial iberoamericano . 1990

[2] **Murray R . Spiegel .** ***Estadística*** **Mac Graw-Hill .1991**

[3] **Paul Meyer .** ***Probabilidad y aplicaciones estadísticas***
Addison Wesley, 1992

[4] **Walpole Ronald .** ***Probabilidad y estadística***
Mac Graw-Hill . 1995

[5] **Hines William.** ***Probabilidad y estadística***
CECSA. 1993