Probabilidades y Estadística

Clase Auxiliar

Profesor: Fernando Lema
Auxiliares: Víctor Carmi - Abelino Jiménez G.

Universidad de Chile

Ingredientes

- Hipótesis Nula (generalmente se quiere rechazar)
- Hipótesis Alternativa (alternativa complementaria)
- Estadístico (acorde a las hipótesis)
- Región de Rechazo

Caso Varianza Conocida

$$H_0: \quad \theta \in \Omega_0$$
$$H_1: \quad \theta \in \Omega_1$$

Definición FUNCIÓN DE POTENCIA

$\pi(\theta) = P(decidir\ rechazar\ H_0 \mid \theta)$

Observación. θ no es variable aleatoria!!!

una probabilidad de error tipo I. Valor de $\pi(\theta)$ para $\theta \in \Omega_0$.

una probabilidad de error tipo II. Valor de $1 - \pi(\theta)$ para $\theta \in \Omega_1$.

Caso Varianza Conocida

$H_0 : \quad \theta \in \Omega_0$
$H_1 : \quad \theta \in \Omega_1$

Definición FUNCIÓN DE POTENCIA

$$\pi(\theta) = P(decidir\ rechazar\ H_0 \mid \theta)$$

Observación. θ no es variable aleatoria!!!

una probabilidad de error tipo I. Valor de $\pi(\theta)$ para $\theta \in \Omega_0$.

una probabilidad de error tipo II. Valor de $1 - \pi(\theta)$ para $\theta \in \Omega_1$.

Caso Varianza Conocida

$H_0 : \quad \theta \in \Omega_0$
$H_1 : \quad \theta \in \Omega_1$

Definición FUNCIÓN DE POTENCIA

$$\pi(\theta) = P(decidir\ rechazar\ H_0 \mid \theta)$$

Observación. θ no es variable aleatoria!!!

una probabilidad de error tipo I. Valor de $\pi(\theta)$ para $\theta \in \Omega_0$.

una probabilidad de error tipo II. Valor de $1 - \pi(\theta)$ para $\theta \in \Omega_1$.

Caso Varianza Conocida

$$H_0 : \quad \mu = \mu_0$$
$$H_1 : \quad \mu > \mu_0$$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} > C$$

¿Cómo determinamos C ?

$$P(\overline{X} > C \mid \mu = \mu_0) = \alpha$$

Caso Varianza Conocida

$H_0: \quad \mu = \mu_0$
$H_1: \quad \mu > \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} > C$$

¿Cómo determinamos C ?

$$P(\overline{X} > C \mid \mu = \mu_0) = \alpha$$

Caso Varianza Conocida

$H_0: \quad \mu = \mu_0$
$H_1: \quad \mu > \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} > C$$

¿Cómo determinamos C ?

$$P(\overline{X} > C \mid \mu = \mu_0) = \alpha$$

Caso Varianza Conocida

$H_0 : \quad \mu = \mu_0$
$H_1 : \quad \mu > \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} > C$$

¿Cómo determinamos C ?

$$P(\overline{X} > C \mid \mu = \mu_0) = \alpha$$

Ejemplo

$X \sim N(\mu, 36^2)$ y una muestra aleatoria de tamaño $n = 9$

$H_0: \quad \mu = 180$

$H_0: \quad \mu > 180$

nivel de significación: $\alpha = 0,05$

Luego, debemos encontrar C tal que

$$P\left(\overline{X} \geq C \mid \mu = 180\right) = 0,05$$

Como conocemos la varianza, tenemos que

$$\overline{X} \sim N(\mu, 144) \quad \Longleftrightarrow \quad \frac{\overline{X} - \mu}{12} \sim N(0,1)$$

Bajo H_0, se tiene

$$\frac{\overline{X} - 180}{12} \sim N(0,1)$$

Ejemplo

$X \sim N(\mu, 36^2)$ y una muestra aleatoria de tamaño $n = 9$

$H_0: \quad \mu = 180$
$H_0: \quad \mu > 180$
nivel de significación: $\alpha = 0,05$

Luego, debemos encontrar C tal que

$$P\left(\overline{X} \geq C \mid \mu = 180\right) = 0,05$$

Como conocemos la varianza, tenemos que

$$\overline{X} \sim N(\mu, 144) \iff \frac{\overline{X} - \mu}{12} \sim N(0,1)$$

Bajo H_0, se tiene

$$\frac{\overline{X} - 180}{12} \sim N(0,1)$$

Ejemplo

$X \sim N(\mu, 36^2)$ y una muestra aleatoria de tamaño $n = 9$

$H_0 : \quad \mu = 180$
$H_0 : \quad \mu > 180$
nivel de significación: $\alpha = 0,05$

Luego, debemos encontrar C tal que

$$P\left(\overline{X} \geq C \mid \mu = 180\right) = 0,05$$

Como conocemos la varianza, tenemos que

$$\overline{X} \sim N(\mu, 144) \quad \Longleftrightarrow \quad \frac{\overline{X} - \mu}{12} \sim N(0,1)$$

Bajo H_0, se tiene

$$\frac{\overline{X} - 180}{12} \sim N(0,1)$$

Ejemplo

$X \sim N(\mu, 36^2)$ y una muestra aleatoria de tamaño $n = 9$

$H_0: \quad \mu = 180$
$H_0: \quad \mu > 180$
nivel de significación: $\alpha = 0,05$

Luego, debemos encontrar C tal que

$$P\left(\overline{X} \geq C \mid \mu = 180\right) = 0,05$$

Como conocemos la varianza, tenemos que

$$\overline{X} \sim N(\mu, 144) \quad \Longleftrightarrow \quad \frac{\overline{X} - \mu}{12} \sim N(0,1)$$

Bajo H_0, se tiene

$$\frac{\overline{X} - 180}{12} \sim N(0,1)$$

Luego, queremos

$$P\left(\frac{\overline{X}-180}{12} \geq \frac{C-180}{12}\right) = 0,05$$

Por tablas, se tiene

$$\frac{C-180}{12} = 1,65 \quad \Rightarrow \quad C = 200$$

Luego, queremos

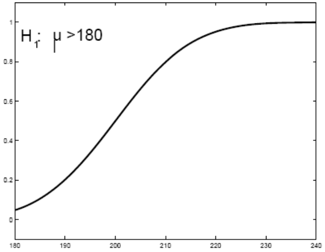
$$P\left(\frac{\overline{X}-180}{12} \geq \frac{C-180}{12}\right) = 0,05$$

Por tablas, se tiene

$$\frac{C-180}{12} = 1,65 \quad \Rightarrow \quad C = 200$$

Veamos el error de tipo II

μ	180	185	190	200	210	220	230
$\pi(\mu)$	0.05	0.11	0.20	0.50	0.80	0.95	0.994
$1-\pi(\mu)$	0.95	0.89	0.80	0.50	0.20	0.05	0.006



Caso Varianza Conocida

$H_0 : \quad \mu = \mu_0$
$H_1 : \quad \mu < \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C$$

¿Cómo determinamos C ?

$$P(\overline{X} < C \mid \mu = \mu_0) = \alpha$$

Caso Varianza Conocida

$H_0: \quad \mu = \mu_0$
$H_1: \quad \mu < \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C$$

¿Cómo determinamos C ?

$$P(\overline{X} < C \mid \mu = \mu_0) = \alpha$$

Caso Varianza Conocida

$H_0: \quad \mu = \mu_0$
$H_1: \quad \mu < \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C$$

¿Cómo determinamos C ?

$$P(\overline{X} < C \mid \mu = \mu_0) = \alpha$$

Caso Varianza Conocida

$H_0 : \quad \mu = \mu_0$
$H_1 : \quad \mu < \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C$$

¿Cómo determinamos C ?

$$P(\overline{X} < C \mid \mu = \mu_0) = \alpha$$

Caso Varianza Conocida

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \end{aligned}$$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C_1 \quad \vee \quad \overline{X} > C_2$$

¿Cómo determinamos C_1 y C_2 ?

$$P(\overline{X} < C_1 \mid \mu = \mu_0) = \frac{\alpha}{2}$$

$$P(\overline{X} > C_2 \mid \mu = \mu_0) = \frac{\alpha}{2}$$

Caso Varianza Conocida

$H_0: \quad \mu = \mu_0$
$H_1: \quad \mu \neq \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C_1 \quad \vee \quad \overline{X} > C_2$$

¿Cómo determinamos C_1 y C_2 ?

$$P(\overline{X} < C_1 \,|\, \mu = \mu_0) = \frac{\alpha}{2}$$

$$P(\overline{X} > C_2 \,|\, \mu = \mu_0) = \frac{\alpha}{2}$$

Caso Varianza Conocida

$H_0: \quad \mu = \mu_0$
$H_1: \quad \mu \neq \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C_1 \quad \vee \quad \overline{X} > C_2$$

¿Cómo determinamos C_1 y C_2 ?

$$P(\overline{X} < C_1 \mid \mu = \mu_0) = \frac{\alpha}{2}$$

$$P(\overline{X} > C_2 \mid \mu = \mu_0) = \frac{\alpha}{2}$$

Caso Varianza Conocida

$H_0 : \quad \mu = \mu_0$
$H_1 : \quad \mu \neq \mu_0$

α: Nivel de Significación

Forma de la Región de Rechazo.

$$\overline{X} < C_1 \quad \vee \quad \overline{X} > C_2$$

¿Cómo determinamos C_1 y C_2 ?

$$P(\overline{X} < C_1 \mid \mu = \mu_0) = \frac{\alpha}{2}$$

$$P(\overline{X} > C_2 \mid \mu = \mu_0) = \frac{\alpha}{2}$$

Caso Varianza Desconocida

Al igual que para intervalos de confianza, en el caso de no conocer σ, tenemos S_n

$$\frac{(\overline{X}-\mu)\cdot\sqrt{n-1}}{S_n} \sim t_{n-1}$$

Se procede de la misma forma que en el caso anterior.

Caso Varianza Desconocida

Al igual que para intervalos de confianza, en el caso de no conocer σ, tenemos S_n

$$\frac{(\overline{X}-\mu)\cdot\sqrt{n-1}}{S_n} \sim t_{n-1}$$

Se procede de la misma forma que en el caso anterior.

Caso Varianza Desconocida

Al igual que para intervalos de confianza, en el caso de no conocer σ, tenemos S_n

$$\frac{(\overline{X}-\mu)\cdot\sqrt{n-1}}{S_n} \sim t_{n-1}$$

Se procede de la misma forma que en el caso anterior.

Problema 1

Se sabe que en una muestra aleatoria de 10 vigas de acero, la resistencia promedio a la composición es:

$$\overline{X_n} = 57498 \ \frac{libras}{pulg^2}$$

$$S_n = 539 \ \frac{libras}{pulg^2}$$

$\alpha = 0,01$
Realice el test, suponiendo que $X \sim N(\mu, \sigma^2)$

$$H_0 : \ \mu = 57000$$

$$H_1 : \ \mu < 57000$$

Solución.

Región Crítica de la forma

$$\overline{X} \leq C$$

Queremos

$$P\left(\overline{X} \leq C \mid \mu = 57000\right) = 0,01$$

Sabemos que

$$z = \frac{(\overline{X} - 57000) \cdot \sqrt{9}}{539} \sim t_9$$

Luego

$$P(z < a) = 0,01 \Rightarrow \quad a = -2{,}821 = \frac{(C - 57000) \cdot 3}{539}$$

Solución.

Región Crítica de la forma

$$\overline{X} \leq C$$

Queremos

$$P\left(\overline{X} \leq C \mid \mu = 57000\right) = 0,01$$

Sabemos que

$$z = \frac{(\overline{X} - 57000) \cdot \sqrt{9}}{539} \sim t_9$$

Luego

$$P(z < a) = 0,01 \Rightarrow \quad a = -2{,}821 = \frac{(C - 57000) \cdot 3}{539}$$

Solución.

Región Crítica de la forma

$$\overline{X} \leq C$$

Queremos

$$P\left(\overline{X} \leq C \mid \mu = 57000\right) = 0,01$$

Sabemos que

$$z = \frac{(\overline{X} - 57000) \cdot \sqrt{9}}{539} \sim t_9$$

Luego

$$P(z < a) = 0,01 \Rightarrow \quad a = -2,821 = \frac{(C - 57000) \cdot 3}{539}$$

Solución.

Región Crítica de la forma

$$\overline{X} \leq C$$

Queremos

$$P\left(\overline{X} \leq C \mid \mu = 57000\right) = 0,01$$

Sabemos que

$$z = \frac{(\overline{X} - 57000) \cdot \sqrt{9}}{539} \sim t_9$$

Luego

$$P(z < a) = 0,01 \Rightarrow \quad a = -2{,}821 = \frac{(C - 57000) \cdot 3}{539}$$

Así, la región crítica es:

$$\overline{X} \leq 56493{,}16$$

Se dice que $\overline{X_{obs}} = 57498$.

Luego, **NO SE RECHAZA!!!!**

Así, la región crítica es:

$$\overline{X} \leq 56493{,}16$$

Se dice que $\overline{X_{obs}} = 57498$.

Luego, **NO SE RECHAZA!!!!**

Test de Comparación de Medias

Se quiere comparar la media de dos poblaciones.

Ejemplos

- Comparación de promedio de notas entre hombres y mujeres.
- Comparación de sueldos entre hombres y mujeres.
- Comparación de resultados entre dos pruebas independientes.

Las Hipótesis para el Test son:

$$H_0 : \mu_1 - \mu_2 = d_0$$
$$H_1 : \mu_1 - \mu_2 \neq d_0$$

Normalmente se toma $d_0 = 0$

Test de Comparación de Medias

Se quiere comparar la media de dos poblaciones.

Ejemplos

- Comparación de promedio de notas entre hombres y mujeres.
- Comparación de sueldos entre hombres y mujeres.
- Comparación de resultados entre dos pruebas independientes.

Las Hipótesis para el Test son:

$$H_0 : \ \mu_1 - \mu_2 = d_0$$
$$H_1 : \ \mu_1 - \mu_2 \neq d_0$$

Normalmente se toma $d_0 = 0$

Test de Comparación de Medias

Se tiene

$$X \sim \begin{cases} N(\mu_1, \sigma_1^2) & \textit{si la observacion pertenece a la poblacion I} \\ N(\mu_2, \sigma_2^2) & \textit{si la observacion pertenece a la poblacion II} \end{cases}$$

Si tomamos una muestra aleatoria en la Población I de tamaño n_1, entonces, se tiene

$$\overline{X_1} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$$

Del mismo modo, si tomamos una muestra aleatoria en la Población II de tamaño n_2, entonces, se tiene

$$\overline{X_2} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

Suponiendo que las muestras de ambas poblaciones se toman de forma independiente, se tiene

$$\overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2 \ , \ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Test de Comparación de Medias

Se tiene

$$X \sim \begin{cases} N(\mu_1, \sigma_1^2) & \textit{si la observacion pertenece a la poblacion I} \\ N(\mu_2, \sigma_2^2) & \textit{si la observacion pertenece a la poblacion II} \end{cases}$$

Si tomamos una muestra aleatoria en la Población I de tamaño n_1, entonces, se tiene

$$\overline{X_1} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$$

Del mismo modo, si tomamos una muestra aleatoria en la Población II de tamaño n_2, entonces, se tiene

$$\overline{X_2} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

Suponiendo que las muestras de ambas poblaciones se toman de forma independiente, se tiene

$$\overline{X_1} - \overline{X_2} \sim N\left(\mu_1 - \mu_2 \ , \ \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

Tenemos 2 Casos.

Si las varianzas son conocidas$\Rightarrow$trabajamos como siempre

Si no conocemos las varianzas $\Rightarrow$trabajamos con la t-Student. Para simplificar, se suponen varianzas iguales.

Siendo $H_o: \ \mu_1 = \mu_2$ y $H_1: \ \mu_1 > \mu_2$, se tiene que la región crítica es

$$\overline{X_1} - \overline{X_2} \geq t_\alpha \sqrt{\left(\frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}$$

donde $P(t_{n_1+n_2-2} \geq t_\alpha) = \alpha$
Claramente esto se puede extender a otros tipo de hipótesis alternativas.

Tenemos 2 Casos.

Si las varianzas son conocidas$\Rightarrow$trabajamos como siempre

Si no conocemos las varianzas $\Rightarrow$trabajamos con la t-Student. Para simplificar, se suponen varianzas iguales.

Siendo $H_o: \ \mu_1 = \mu_2$ y $H_1: \ \mu_1 > \mu_2$, se tiene que la región crítica es

$$\overline{X_1} - \overline{X_2} \geq t_\alpha \sqrt{\left(\frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}$$

donde $P(t_{n_1+n_2-2} \geq t_\alpha) = \alpha$
Claramente esto se puede extender a otros tipo de hipótesis alternativas.

Tenemos 2 Casos.

Si las varianzas son conocidas⇒trabajamos como siempre

Si no conocemos las varianzas ⇒trabajamos con la t-Student. Para simplificar, se suponen varianzas iguales.

Siendo $H_o: \ \mu_1 = \mu_2$ y $H_1: \ \mu_1 > \mu_2$, se tiene que la región crítica es

$$\overline{X_1} - \overline{X_2} \geq t_\alpha \sqrt{\left(\frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}$$

donde $P(t_{n_1+n_2-2} \geq t_\alpha) = \alpha$

Claramente esto se puede extender a otros tipo de hipótesis alternativas.

Test de Comparación de Medias

Tenemos 2 Casos.

Si las varianzas son conocidas$\Rightarrow$trabajamos como siempre

Si no conocemos las varianzas $\Rightarrow$trabajamos con la t-Student. Para simplificar, se suponen varianzas iguales.

Siendo $H_o: \ \mu_1 = \mu_2$ y $H_1: \ \mu_1 > \mu_2$, se tiene que la región crítica es

$$\overline{X_1} - \overline{X_2} \geq t_\alpha \sqrt{\left(\frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2 - 2}\right)\left(\frac{n_1 + n_2}{n_1 \cdot n_2}\right)}$$

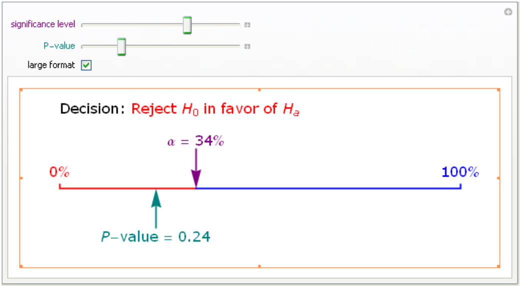
donde $P(t_{n_1+n_2-2} \geq t_\alpha) = \alpha$
Claramente esto se puede extender a otros tipo de hipótesis alternativas.

Problema 2

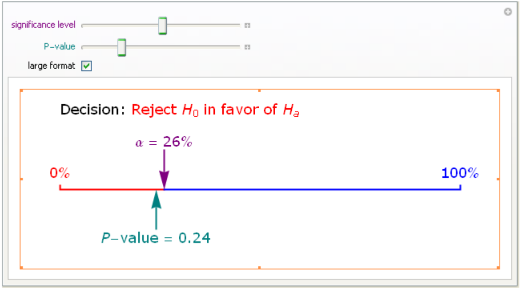
Se quiere comparar la velocidad de respuesta de dos discos. Se toman las hipótesis
$H_0: \ \mu_1 = \mu_2$ y $H_1: \ \mu_1 > \mu_2$, donde μ_k, es la media de la velocidad para el disco k.
Para el disco 1 se tomó una muestra de tamaño $n_1 = 13$, cuya media es $\overline{X_1} = 68,2$ y $S_1 = 18,6$.
Para el disco 2 se tomó una muestra de tamaño $n_2 = 15$, cuya media es $\overline{X_2} = 53,2$ y $S_1 = 15,8$.
Con $\alpha = 0,05$, hacer test.

Definición p-valor

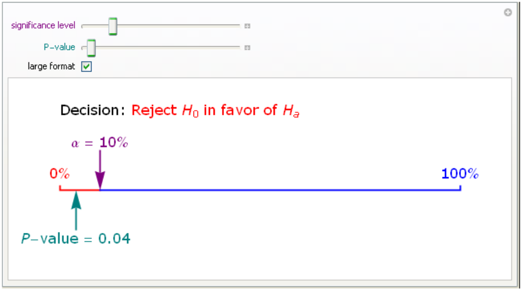
el p-valor está definido como la probabilidad de obtener un resultado al menos tan extremo como el que realmente se ha obtenido, suponiendo que la hipótesis nula es cierta.

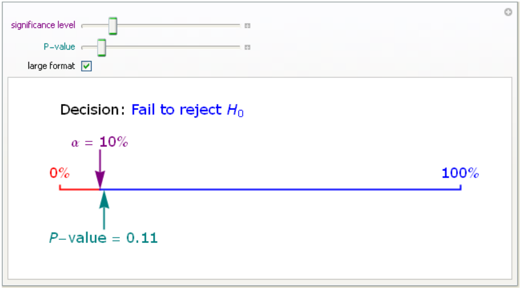


Concepto de p-valor.



Concepto de p-valor




significance level
P-value
large format
Decision: Fail to reject H_0
α = 10%
0%
100%
P-value = 0.11

Supongamos que queremos medir la evolución de una determinada variable.

Ejemplos.

- Se quiere medir la eficiencia de un medicamento.
- Se quiere medir la eficiencia de un determinado método de enseñanza.

El problema de este tipo de Test, es que las variables **NO son independientes**.

Esto se soluciona tomando la diferencia como variable.

Test para Pares de Observaciones

Supongamos que queremos medir la evolución de una determinada variable.

Ejemplos.

- Se quiere medir la eficiencia de un medicamento.
- Se quiere medir la eficiencia de un determinado método de enseñanza.

El problema de este tipo de Test, es que las variables **NO son independientes**.

Esto se soluciona tomando la diferencia como variable.

Se tiene que

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\hat{\sigma}_{X-Y}/\sqrt{n-1}} \sim t_{n-1}$$

donde

$$\hat{\sigma}^2_{X-Y} = \frac{\sum (d_i - \overline{d})^2}{n}$$

Test para Pares de Observaciones

Problema 3

Consideremos el experimento de “sembrado” de nubes para comparar la precipitación mensual en las dos áreas agrícolas. ¿Se puede decir que la precipitación mensual del área agrícola “sembrada” excede a la del área agrícola “sin sembrar”?

Area Agrícola	1	2	3	4	5
Sembrada	1.75	2.12	1.53	1.10	1.7
Sin Sembrar	1.62	1.83	1.40	0.75	1.71
D	0.13	0.29	0.13	0.35	-0.01

Test de Independencia

Tenemos dos variables Nominales, y queremos saber si son independientes o no.

Tabla de Contingencia

	Y_1	Y_2	...	Y_q	
X_1	n_{11}	n_{12}	$\cdots$	n_{1q}	$\sum_j n_{1j}$
X_2	n_{21}	n_{22}	$\cdots$	n_{2q}	$\sum_j n_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
X_p	n_{p1}	n_{p2}	$\cdots$	n_{pq}	$\sum_j n_{pj}$
	$\sum_i n_{i1}$	$\sum_i n_{i2}$	$\cdots$	$\sum_i n_{iq}$	n

Test de Independencia

Se tiene que el estadístico

$$Q = \sum_{ij} \frac{(n_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}} \sim \chi^2_{(p-1)(q-1)}$$

Acá H_0 es independencia.

Problema 4

Se quiere probar si hay diferencia de ingreso entre hombres y mujeres médicos. Para lo cual se hizo una encuesta a 200 médicos alazar y de manera independiente, obteniéndose la siguiente información:

	Ingreso Bajo	Ingreso Alto	Total
Mujeres	20	100	120
Hombres	70	10	80
Total	90	110	200

Estudie la independencia entre las variables Género o Ingreso.

Solución

A partir de la misma tabla, calculamos las probabilidades marginales

	Ingreso Bajo	Ingreso Alto	Marginal
Mujeres	20	100	$\frac{120}{200}$
Hombres	70	10	$\frac{80}{200}$
Marginal	$\frac{90}{200}$	$\frac{110}{200}$	1

Tenemos que calcular

$$\sum_{ij} \frac{(n_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}}$$

Solución

Para eso construimos una tabla donde en cada espacio contenga uno de los sumando

En cada posición ponemos el término

$$\frac{(n_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}}$$

Esto es

	Ingreso Bajo	Ingreso Alto
Mujeres	$\frac{\left(20-20\cdot\frac{120}{200}\cdot\frac{90}{200}\right)^2}{20\cdot\frac{120}{200}\cdot\frac{90}{200}}$	$\frac{\left(100-100\cdot\frac{120}{200}\cdot\frac{110}{200}\right)^2}{100\cdot\frac{120}{200}\cdot\frac{110}{200}}$
Hombres	$\frac{\left(70-70\cdot\frac{80}{200}\cdot\frac{90}{200}\right)^2}{70\cdot\frac{80}{200}\cdot\frac{90}{200}}$	$\frac{\left(10-10\cdot\frac{80}{200}\cdot\frac{110}{200}\right)^2}{10\cdot\frac{80}{200}\cdot\frac{110}{200}}$

Y sumando se tiene un valor cercano a 97,3.

$$\sum_{ij} \frac{(n_{ij} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}} \sim \chi^2_{(p-1)(q-1)}$$

Y Calculando el p-valor, se tiene que
$P(\chi^2_1 > 97{,}3) < 0,05$
Con esto, RECHAZAMOS H_0, y las variables Género e Ingreso no son independientes.

Cuando n es grande, no hay problema en aplicar el Teorema Central del Límite, y reducirse al caso normal.

El problema es cuando n es pequeño. En ese caso, hay que calcular "a mano" la distribución del estadístico a usar.

Problema 5

Sea $X \sim exp(\lambda)$ es decir

$$f_X(x) = \frac{1}{\lambda} e^{\frac{-x}{\lambda}}$$

Queremos encontrar una región crítica para el test.
$H_0: \lambda > \lambda_0$
$H_1: \lambda < \lambda_0$
cuando n es pequeño.

Tenemos que el estimador de máxima verosimilitud para λ es $\overline{X}$. De modo que la región crítica que tenemos que construir es de la forma

$$\overline{X} < C$$

¿Cómo determinamos C? Imponiendo:

$$P(\overline{X} < C \ \mid \lambda = \lambda_0) = \alpha$$

El problema que tenemos es que en este caso n es pequeño, luego hacer el cálculo "a mano"

Caso muestra pequeña

Recordemos que

$$X \sim exp(\lambda) \Longleftrightarrow X \sim Gamma(1, \lambda)$$

Por otra parte, si $X_i \sim exp(\lambda)$, donde la secuencia de v.a. son independientes, entonces

$$\sum_i^n X_i \sim Gamma(n\ ,\ \lambda)$$

Además, tenemos que si

$$X \sim Gamma(n, \lambda) \Longleftrightarrow \frac{2X}{\lambda} \sim Gamma(n, \frac{1}{2})$$

ie

$$\frac{2X}{\lambda} \sim \chi^2_{2n}$$

(PROBARLO!!!)

Recordemos que

$$X \sim exp(\lambda) \Longleftrightarrow X \sim Gamma(1, \lambda)$$

Por otra parte, si $X_i \sim exp(\lambda)$, donde la secuencia de v.a. son independientes, entonces

$$\sum_{i}^{n} X_i \sim Gamma(n\ ,\ \lambda)$$

Además, tenemos que si

$$X \sim Gamma(n, \lambda) \Longleftrightarrow \frac{2X}{\lambda} \sim Gamma(n, \frac{1}{2})$$

ie

$$\frac{2X}{\lambda} \sim \chi^2_{2n}$$

(PROBARLO!!!)

Caso muestra pequeña

Tenemos entonces que, bajo H_0

$$\frac{2 \cdot (X_1 + X_2 + ... + X_n)}{\lambda_0} \sim \chi^2_{2n}$$

Entonces, para obtener lo que se quería calcular, se tiene que

$$P\left(\frac{2 \cdot (X_1 + X_2 + ... + X_n)}{\lambda_0} < \frac{2 \cdot n \cdot C}{\lambda_0}\right) = \alpha$$

conociendo n y dándonos un nivel de significancia, por medio de tablas, podemos calcular $\frac{2 \cdot n \cdot C}{\lambda_0}$, es decir, podemos calcular C.