Pauta Auxiliar 14: Test de hipotesis, Tabla de contingencía.

MA3403 - Probabilidades y Estadistica
Profesor: Fernando Lema
Auxiliar: Martín Castillo & José Cereceda
6 de julio de 2014

P1. Sean $X \sim N\left(\mu_X, \sigma_X^2\right), Y \sim N\left(\mu_Y, \sigma_Y^2\right)$, y sean $X_1, ..., X_n$ y $Y_1, ..., Y_m$ m.a.s. de X, Y, se desea comparar las medias (a priori desconocidas) de ambas v.a.

a) Suponga $\sigma_X^2 \sigma_Y^2$ conocidos y plantee un test de hipotesis de comparación de medias.

b) Suponga $\sigma_X^2 \sigma_Y^2$ desconocidos y plantee lo mismo que antes, ¿Cómo resolvería el problema?

P2. Durante la última elección presidencial, los resultados fueron los esperados (MB 62,16 % y EM 37,84 %), sin embargo poco se sabe de las caracteristicas de los votantes. Para estudiar lo anterior se realizó una encuesta a 400 votantes.

a) Se postula que la edad promedio de los votantes es mayor a 45 años. Se obtiene una muestra que promedia 45,6 años. Suponiendo que la edad es normal con $\sigma = 10$ años , decida si la afirmación es cierta considerando un nivel de significancia $\alpha = 0{,}05$. Calcule p-valor.

Solución:

$$H_0 : \mu \leq 45$$
$$H_1 : \mu > 45$$

Luego, sabemos que el p-valor es de la forma:

$$p\text{-valor} = \mathbb{P}(Z > z_{obs}|H_0) = \mathbb{P}\left(Z > \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \mid H_0\right) = \mathbb{P}(Z > 1{,}2) = 1 - 0{,}88 = 0{,}12$$

Finalmente, como p-valor es $>$ que el nivel de significancia α, no se rechaza H_0.

b) Si σ es desconocido pero la muestra entrega $\hat{\sigma} = 10$ ¿Como cambia conceptualmente su respuesta en (a)? Determine un intervalo de confianza del 90 % para σ^2 del tipo $\sigma^2 < b$(cota superior).

Solución:
En primer lugar, en caso de no conocer la varianza, y estimarla, sólo cambia el estadístico usado. Pasa de normal a ser una t-student de $n - 1$ grados de libertad, y el t_{obs} es el mismo (ya que $\hat{\sigma} = 10$).

Para determinar un intervalo del tipo cota superior debemos imponer:

$$1 - \alpha = \mathbb{P}(a \leq \chi^2 \leq \infty) \tag{1}$$

Reemplazando $\alpha = 0{,}1$, se obtiene $a = 118{,}5$ (usando la fila de 100 grados de libertad). Ahora,

$$(1) = \mathbb{P}\left(0 \leq \sigma^2 \leq \frac{1}{a} \cdot \hat{\sigma}(n-1)\right)$$

Así, el intervalo es:

$$I.C. : (0; 33{,}67)$$

c) Los datos de la encuesta fueron tabulados según rango de edad y votación, obteniendose

	(18 - 30)	(30 - 60)	(60- ∞)	total
MB	55	130	65	250
EM	25	70	55	150
	80	200	120	400

¿Es posible aformar que la votación "depende" de la edad? Considere $\alpha = 0,1$
Asumiendo que la respuesta es positiva (aunque no lo sea) ¿que tipo de dependencia existe?

Solución:
Sabemos que el p-valor en este caso es de la forma: $\mathbb{P}(\chi^2_{(p-1)(q-1)} > Q)$, con $Q = \sum_{i,j} \frac{(M_{ij} - E_{ij})^2}{E_{ij}}$.
Además, en este caso $(p-1)(q-1) = 3$.
Calculando los valores, $E_{11} = \frac{250 \cdot 80}{400} = 50, E_{12} = 125, E_{13} = 75, E_{21} = 30, E_{22} = 75, E_{23} = 45$.
Finalmente, el p-valor corresponde a:

$$\mathbb{P}(\chi^2_3 > 5{,}34) > 0{,}1$$

Por lo tanto, no se rechaza H_0.

P3. Se investiga la duración de un compuesto de caucho para lo cual se construyen 16 llantas y se prueban hasta que fallan, obteniéndose una duración promedio de 61.014 km. Suponga normalidad.

a) ¿Es posible suponer que la duración promedio de la población supera los 60.000 km? Concluya con un nivel de significancia $\alpha = 0,05$. Calcule $p - valor$. Suponga $\sigma = 2{,}600$ km.
Solución:
Tenemos el siguiente test:

$$H_0 : \mu \leq 60{,}000$$
$$H_1 : \mu > 60{,}000$$

Además, $\overline{X} = 61{,}014$. Usamos el estadístico

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

luego el $p - valor$ es de la forma $\mathbb{P}\left(Z > Z_{obst} | H_0\right)$. Reemplazando los datos se obtiene que $Z_{obs} = 1,56$ por lo tanto se tiene que el $p - valor = 0,06$. Por lo tanto no se rechaza H_0.

b) Si σ es desconocido. ¿Cómo cambia conceptualmente la respuesta de *a*) si $\hat{\sigma} = 2{,}600$ km ? Construya un I.C. del 90 % (cota superior) para σ.
Solución:
Si σ es desconocido, enconces se debe estimar, lo que cambia el estadístico

$$t_{n-1} = \frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}}$$

y T es una $t - student$ de $n-1$ grados de libertad. Por lo tanto se tiene que el $p - valor$ en este caso corresponde a $\mathbb{P}\left(t_{n-1} > 1,56\right) = 0{,}07$. De igual manera que antes, no se rechaza H_0.
Sabemos que

$$(n-1)\frac{\sigma^2_{n-1}}{\sigma^2} \sim \chi^2_{n-1} \qquad n-1 = 15$$

Para construir el intervalo, buscamos a, b tales que $\mathbb{P}\left(a \leq \chi^2_{n-1} \leq b\right) = 0{,}9$. Como buscamos una cota superior para σ. Imponemos $b \to \infty$ y buscamos a tal que $\mathbb{P}\left(a \leq \chi^2_{15}\right) = 0{,}9$. se obtiene entonces $a = 5,23$. Así tenemos $\sigma \leq 4{,}403$.

c) Para profundizar la investigación se toma un grupo de 60 llantas y después que han fallado se los clasifica según: X duración Y tipo de conducción. Los datos obtenidos muestran a continuación:

Y/X	<60.000 km	≥ 60.000 km
Defensivo	8	12
Normal	11	14
Agresivo	8	7

¿Es posible suponer que la duración depende del tipo de conducción? Use $\alpha = 0{,}05$.

¿Cómo cambia el estadístico usado previamente si en vez de una muestra de tamaño 60 se toma una muestra de tamaño 600 ó 6000 y todas las proporciones en la tabla se mantienen?

Solución:

Para el test de independencia usamos el estadistico

$$Q = \sum_{i,j} \frac{(M_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(p-1(q-1))=2}$$

En este caso

$$E_{11} = n\mathbb{P}\left(< 60{,}000\right)\mathbb{P}\left(\text{Defensivo}\right) = 60\frac{27}{60}\frac{20}{60} = 9$$

y con el resto

$$E_{12} = 11 \quad E_{21} = 11,25 \quad E_{22} = 27,5 \quad E_{31} = 6{,}75 \quad E_{32} = 8,25$$

Luego $Q_{obs} = 7,24$ con lo que el $p - valos = \mathbb{P}\left(Q > 7,24\right) = 0,04$ por lo tanto se rechaza H_0 (independencia).

Finalmente si sólo se cambia el tamaño de la muestra manteniendo las proporciones $Q'_{obs} = nQ_{obs}$