Clase Auxilar 9: Test χ^2

Profesora Nancy Lacourly
Auxiliares: Andrés Iturriaga J. - Héctor Olivero Q.

12 de mayo de 2009

La distribución Multinomial:

Consideremos n repeticiones independientes de un experimento con k posibles resultados. Supongamos que la probabilidad de cada resultado está dada por $\{p_i\}_{i=1}^{k}$.
Si denotamos X_i como la cantidad de veces que se repite el resultado i en las n repeticiones, entonces tenemos:

$$\mathbb{P}(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1!n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

Siempre que $n_1 + \ldots + n_k = n$.

Ejemplo 1:

Si $k = 2$ entonces el modelo multinomial se reduce al modelo binomial. En efecto:
Sea: $p_1 = p$ entonces $p_2 = 1 - p$. Además si $n_1 + n_2 = n$, entonces $n_2 = n - n_1$. Con esto:

$$\mathbb{P}(X_1 = n_1, X_2 = n_2) = \frac{n!}{n_1!n_2!} p_1^{n_1} p_2^{n_2} \quad (1)$$

$$\mathbb{P}(X_1 = n_1, X_2 = n - n_1) = \frac{n!}{n_1!(n-n_1)!} p^{n_1}(1-p)^{n-n_1} \quad (2)$$

$$= \binom{n}{n_1} p^{n_1}(1-p)^{n-n_1} \quad (3)$$

Ejemplo 2:

Consideremos el caso de un dado de seis caras. ¿Cuál es la probabilidad de obtener dos unos, dos tres y dos seis, cuando se lanza un dado seis veces?.
Utilizando lo anterior tenemos:

$$\mathbb{P}(X_1 = 2, X_2 = 0, X_3 = 2, X_4 = 0, X_5 = 0, X_6 = 2) =$$

$$\frac{6!}{2!0!2!0!0!2!}(\frac{1}{6})^2(\frac{1}{6})^0(\frac{1}{6})^2(\frac{1}{6})^0(\frac{1}{6})^0(\frac{1}{6})^2$$

$$\mathbb{P}(X_1 = 2, X_2 = 0, X_3 = 2, X_4 = 0, X_5 = 0, X_6 = 2) = 0{,}0019$$

Ejemplo 3:

Sea X una v.a. que sigue una distribución $U(0,1)$ y sea:

$$Y = \begin{cases} 1 & X \in [0, \frac{1}{6}] \\ 2 & X \in (\frac{1}{6}, \frac{4}{6}] \\ 3 & X \in (\frac{4}{6}, 1] \end{cases} \tag{4}$$

Notemos que:

$$\mathbb{P}(Y=1) = \frac{1}{6}, \quad \mathbb{P}(Y=2) = \frac{1}{2}, \quad \mathbb{P}(Y=3) = \frac{1}{3}$$

Supongamos que se repite n veces el experimeno Y, entonces:

$$\mathbb{P}(Y_1 = n_1, Y_2 = n_2, Y_3 = n_3) = \frac{n!}{n_1!n_2!n_3!}\left(\frac{1}{6}\right)^{n_1}\left(\frac{1}{2}\right)^{n_2}\left(\frac{1}{3}\right)^{n_3}$$

Si $n_1 + n_2 + n_3 = n$.

Proposición

Sea:

$$Q = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$$

Entonces asintóticamente Q sigue una distribución χ^2_{k-1}

Observación: Si para calcular Q es necesario estimar l parámetros, entonces Q sigue una distribución χ^2_{k-l-1}

La idea:

Sea una v.a. X sigue cierta distribución de probabilidad. Si separamos en k subconjuntos los valores que puede tomar X (como en el ejemplo 3) y definimos una nueva v.a. Y que toma valores de 1 a k según el subconjunto en que está el resultado de la realización de X, entonces tenemos que al repetir n veces Y el número de apariciones conjuntas de los posible resultados[1] sigue una distribución multinomial.

Esto lo usamos para realizar el siguiente test de hipótesis:

H_0 : X sigue una distribución F_θ

H_1 : X no sigue una distribución F_θ

Podemos proceder de la siguiente manera: Suponemos que H_0 es cierta y construimos la variable aleatoria Y, con ella construimos el estadístico Q con el cual podemos calcular p-valores o regiones de rechazo.

[1] Ver ejemplo 3

La idea:

Sea una v.a. X sigue cierta distribución de probabilidad. Si separamos en k subconjuntos los valores que puede tomar X (como en el ejemplo 3) y definimos una nueva v.a. Y que toma valores de 1 a k según el subconjunto en que está el resultado de la realización de X, entonces tenemos que al repetir n veces Y el número de apariciones conjuntas de los posible resultados[1] sigue una distribución multinomial.

Esto lo usamos para realizar el siguiente test de hipótesis:

H_0 : X sigue una distribución F_θ

H_1 : X no sigue una distribución F_θ

Podemos proceder de la siguiente manera: Suponemos que H_0 es cierta y construimos la variable aleatoria Y, con ella construimos el estadístico Q con el cual podemos calcular p-valores o regiones de rechazo.

[1] Ver ejemplo 3

La idea:

Sea una v.a. X sigue cierta distribución de probabilidad. Si separamos en k subconjuntos los valores que puede tomar X (como en el ejemplo 3) y definimos una nueva v.a. Y que toma valores de 1 a k según el subconjunto en que está el resultado de la realización de X, entonces tenemos que al repetir n veces Y el número de apariciones conjuntas de los posible resultados[1] sigue una distribución multinomial.

Esto lo usamos para realizar el siguiente test de hipótesis:

H_0 : X sigue una distribución F_θ

H_1 : X no sigue una distribución F_θ

Podemos proceder de la siguiente manera: Suponemos que H_0 es cierta y construimos la variable aleatoria Y, con ella construimos el estadístico Q con el cual podemos calcular p-valores o regiones de rechazo.

[1] Ver ejemplo 3

Independencia

El test para detectar independencia entre dos variables X e Y es un caso particular de lo anterior. En este caso, lo que testeamos es que la distruibución conjunta de las dos variables sea el producto de las distribuciones marginales. El estadístico que usamos es:

$$Q = \sum_{i,j} \frac{(M_{i,j} - np_i p_j)^2}{np_i p_j}$$

Con $p_i = \mathbb{P}(X = i)$, $p_j = \mathbb{P}(Y = j)$ y $M_{i,j} = frec(X = i, Y = j)$, que asintóticamente sigue una distribución χ^2_{pq-1}. Si no conocemos p_i y p_j usamos estimaciones de los mismos. Y en este caso tenemos que Q sigue una distribución $\chi^2_{(p-1)(q-1)}$.

Problema

Los datos siguientes muestran las frecuencias de conteo para 400 observaciones acerca del número de colonias de bacterias por campo en un microscopio, utilizando muestras de una capa delgada de leche:

# bacterias (i)	0	1	2	3	4	5
# Campos con i bac.	56	104	80	62	42	27
# bacterias (i)	6	7	8	9	>=10	
# Campos con i bac.	9	9	5	3	3	

- Realice un test de hipótesis para H_0 : Los datos provienen de una Poisson con $\alpha = 0{,}05$.
- Calcule el p-valor del test.

Problema

Se considera la variable X el número de personas que se presentan a la boletería de una estación de metro durante un intervalo de diez minutos. Se mide esta variable sobre cien periodos de diez minutos y se obtuvieron los siguientes datos:

# personas que llegan (i)	3	4	5	6	7	8
# intervalos con i per.	7	9	14	15	17	13

# personas que llegan (i)	9	10	11	12	13	14
# intervalos con i per.	8	6	5	3	2	1

- Realice un test de hipótesis para H_0 : Los datos provienen de una Poisson con $\alpha = 0{,}05$. Recuerde que el EMV de λ es la media muestral.

Problema

Se quiere probar si hay una relación entre la aprobación de MA34B - 02 y el sexo de los alumnos. Para lo que se cuenta con los siguientes datos:

	No Aprueba	Aprueba	Total
Hombre	45	34	79
Mujer	9	7	16
Total	54	41	95

Estudie la independencia entre sexo y aprobación.

Problema

Se quiere testear la hipótesis de que la duración de una ampolleta sigue una distribución exponencial. Para esto se ha tomado una muestra de 300 ampolletas y se ha medido su duración, los resultados obtenidos aparecen en la hoja "Problema 4" del archivo Excel adjunto.

Construya un test que permita determinar la veracidad de la hipótesis nula. resultados

Problema

Se considera una muestra de 4000 delfines de tres especies diferentes $F1$, $F2$ y $F3$ y cuatro zonas diferentes del Pacifico $C1$, $C2$, $C3$ y $C4$. Se considera en la tabla que aparece en en la hoja "Problema 5" del archivo Excel las frecuencias de delfines por especie/zona.
¿Cómo se llama la tabla?
Busque mediante un test de hipótesis si existe o no una preferencia de ciertas especies por cierta zona.

Problema

Los datos que aparecen en la hoja "Problema 6" del archivo Excel adjunto muestran la frecuencia de conteo del número de vehículos que pasan por un peaje durante 1000 periodos de 15 minutos. Verifique el supuesto que los datos provienen de una distribución de Poisson con un error de tipo I igual al 0,05.