MA34B – Estadística
Tests De Hipótesis

Prof. Rodrigo Abt B.
rabt@dim.uchile.cl

Introducción (1)

- Muchas veces la idea de un investigador no sólo es estimar el(los) parámetro(s) de un población, sino que también responder preguntas de investigación o crear un procedimiento para decidir en base a información datos observados. Por ejemplo, un médico podría preguntarse si el consumo de café aumenta el riesgo de cáncer al estómago, un ingeniero podría preguntarse si existen diferencias entre la resistencia a la tracción de dos vigas, un sociólogo podría preguntarse si el ingreso de los hombres casados es mayor que el de sus cónyuges, o un matemático podría preguntarse si una moneda se encuentra cargada o no si después de 1000 lanzamientos salen 479 caras.
- En cada uno de estos casos el investigador *postula* o *conjetura* algo respecto de la población en estudio. En este sentido hace una afirmación, la cual pretenderá rechazar o aceptar, respecto del(los) parámetro(s) de una o más poblaciones, en cuyo caso planteará una ***hipótesis***.
- Usualmente el investigador no cuenta con toda la información de la población, por lo cual deberá hacer uso de una muestra para poder determinar la veracidad o falsedad de una hipótesis.

Introducción (2)

- En Estadística, las hipótesis son muy precisas. La aceptación de una hipótesis debe conducir al rechazo de las alternativas y viceversa. En Estadística se plantean dos tipos de hipótesis, una denominada *hipótesis nula* (H_0) y otra denominada *hipótesis alternativa* (H_1 o H_A).
- La hipótesis nula representa la afirmación de que *no existe efecto* o *no hay diferencia* en los parámetros involucrados.
- Por ejemplo en un juicio H_0 podría representar la afirmación "el acusado en inocente", o en el caso del consumo de café y el cáncer, H_0 sería "el café no aumenta el riesgo". Si μ_H y μ_M representan los ingresos promedios de hombres y mujeres casados respectivamente, la hipótesis nula H_0 podría ser "$\mu_H = \mu_M$"
- Por esto, si un investigador pretende buscar si existe un efecto o una diferencia, entonces debe tratar de rechazar H_0 a favor de H_1. Esta interpretación está muy relacionada con los errores de decisión que podemos cometer, como veremos más adelante.

Clasificación De Las Hipótesis

Hipótesi
s

No
Paramétrica
s

$$H_0: \quad X \sim F$$
$$H_1: \quad X \nsim F$$

Donde $\Theta = \{\theta_0, \theta_1\}$

Paramétri
cas

Simples

$$H_0: \quad \theta = \theta_0$$
$$H_1: \quad \theta = \theta_1$$

Donde $\Theta = \{\theta_0, \theta_1\}$

Compues
tas

$$H_0: \quad \theta \in \Theta_0$$
$$H_1: \quad \theta \in \Theta_1$$

Donde $\Theta_0 \cap \Theta_1 = \emptyset$ y

$\Theta_0 \cup \Theta_1 \subseteq \Theta$

Hipótesis Paramétricas

Reglas De Decisión

- Todo procedimiento de toma de decisiones requiere de un criterio de decisión.
- En este caso el criterio debe construirse en base a los valores muestrales. El criterio se puede expresar como una región o conjunto W de valores posibles tales que si la muestra pertenece dicho conjunto entonces rechazamos (o aceptamos) la hipótesis nula.
- En términos algebraicos podríamos buscar una región $W \subset R^n$ tal que si $x_1, x_2, \ldots, x_n \in W$, entonces rechazamos H_0.
- A dicha región W le denominaremos “Región Crítica” o “Región de Rechazo”

Errores De Decisión

- La pregunta es entonces cómo encontrar dicha región de rechazo.
- Debemos pensar entonces las características deseables que debiera tener una región de rechazo.
- Deberíamos escoger aquellas regiones que nos hagan tomar decisiones lo más acertadas (o lo menos equivocadas) posible.
- En este sentido, al decidir uno podría cometer dos tipos básicos de error:
  - *Rechazar la hipótesis nula cuando en realidad esta es verdadera*. En este caso diremos que se ha cometido un **Error de Tipo I**.
  - *Aceptar la hipótesis nula cuando en realidad esta era falsa*. En este caso diremos que se ha cometido un **Error de Tipo II**.

Medidas Para Los Tipos De Error

- Ciertamente la idea es equivocarse lo menos posible con ambos tipos de error al utilizar bajo una determinada región de rechazo.
- El hecho de trabajar con una muestra no nos garantiza estar libres del error. Podríamos contar la fracción de veces que nos equivocamos y tratar de alcanzar una cota máxima.
- Si definimos como:
  - α = P(Cometer Error de Tipo I) = P(Rechazar H_0/H_0 cierta) = $P(x_1,x_2,\ldots,x_n \in W/\theta \in \Omega_0)$
  - β = P(Cometer Error de Tipo II) = P(Aceptar H_0/H_0 falsa) = $P(x_1,x_2,\ldots,x_n \in W^c/\theta \notin \Omega_0)$
- Entonces buscaremos minimizar α y β.

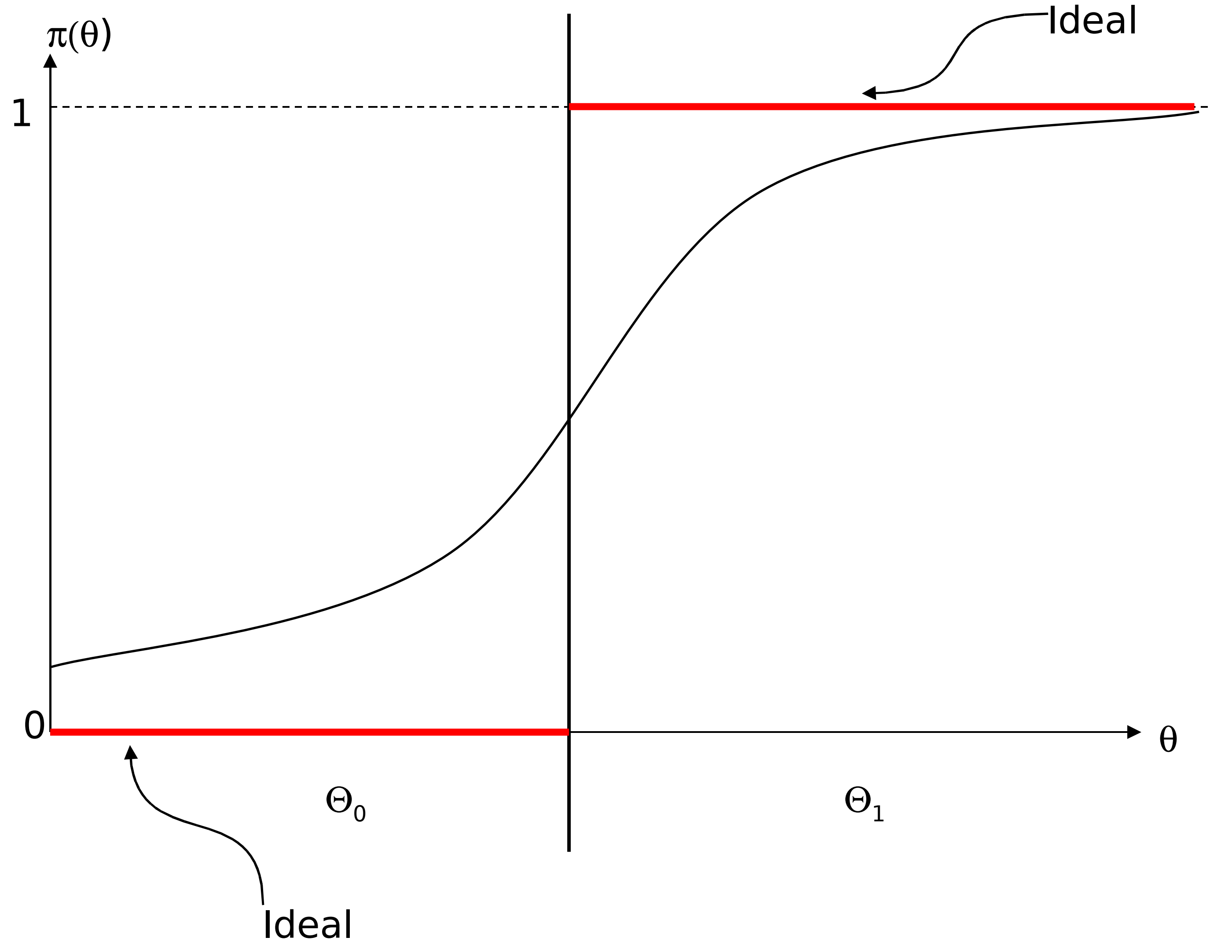
La Función De Potencia

- La Función de Potencia nos permite analizar el comportamiento de una región crítica en función de los parámetros involucrados en las hipótesis.
- Se define como:

$$\pi_W(\theta) = P(\text{Rechazar } H_0/\theta) = P(x_1, x_2, \ldots, x_n \in W/\theta)$$

- Una función de potencia ideal sería entonces aquella que para valores de θ en Θ_0 fuera 0, y para θ en Θ_1 fuera 1.
- Se denomina *tamaño del test* a sup $\{\pi(\theta)/\theta \in \Theta_0\}$.

Función De Potencia



Lema de Neyman-Pearson

- Si bien no es posible encontrar una función de potencia ideal, usualmente es posible encontrar una región de rechazo, de modo que cuando nos encontremos en el dominio de H_0, la probabilidad de cometer un error de tipo I esté acotada.
- El problema está en que si queremos reducir α, aumenta β, es decir, hay un trade-off entre los errores.
- El Lema de Neyman-Pearson nos provee de un procedimiento que permite encontrar aquella región de rechazo que minimiza los errores de decisión, acotando la probabilidad del Error de Tipo I.

Caso Hipótesis Paramétricas Simples

- Sea $X_1, X_2, \ldots, Xn$ una m.a.s. de una v.a. X con f.d.p. $f(x|\theta)$, y las hipótesis:
  - $H_0 : \theta = \theta_0$
  - $H_1 : \theta = \theta_1$
- Lema de Neyman-Pearson: Si δ^* es una regla de decisión tal que para un $k>0$ fijo, se rechaza H_0 si $f_1/f_0>k$, y se acepta H_0 si $f_1/f_0<k$, entonces para toda regla δ tal que $\alpha(\delta)<= \alpha(\delta^*)$ se tiene que $\beta(\delta)>=\beta(\delta^*)$.

Lema de Neyman-Pearson (cont.)

- Donde f_i corresponde a la función de verosimilitud de $x_1, x_2, \ldots, x_n$ evaluada en $\theta = \theta_i$ (i=0,1).
- Se denomina *nivel de significación* (α_0) a la cota máxima de la probabilidad del Error de Tipo I.
- Este procedimiento nos permite encontrar la FORMA de la región de rechazo.
- Para encontrar el valor crítico de la región encontrada, se combina el resultado anterior con la restricción de cota máxima para la probabilidad del Error de Tipo I, y se despeja.
- El lema de N-P es tal que minimiza $a\alpha + b\beta$, es decir, la suma ponderada de los errores, donde $a/b = k$.

Ejemplo: Caso de Media de una Normal (1)

- Sea $X_1,X_2,\ldots,X_n$ una m.a.s de una v.a. $X\sim N(\mu,\sigma^2)$, suponga las hipótesis:
  - $H_0 : \mu = \mu_0$
  - $H_1 : \mu = \mu_1$
- Considere que $\mu_1>\mu_0$, un nivel de significación α=5%, y que la varianza σ^2 es conocida.
- La forma de la región de rechazo se determina a través del Lema de Neyman-Pearson:

Ejemplo: Caso de Media de una Normal (2)

- En efecto se rechaza H_0 si $f_1/f_0 > k$, luego:

$$\begin{aligned}
\frac{f_1}{f_0} &> k \\
\frac{exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_1)^2\}}{exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2\}} &> k \\
\Rightarrow \sum_{i=1}^{n} x_i &> C
\end{aligned}$$

Luego la región de rechazo es de la forma $W = \{\sum_{i=1}^{n} x_i > C\}$ o $W = \{\bar{X} > C'\}$.

Ejemplo: Caso de Media de una Normal (3)

- Ahora para rechazar necesitamos el valor de C (o C'), el cual se puede obtener con la cota impuesta sobre el error de tipo I. Es decir, se tiene que:

$$\begin{aligned}
P(\text{Rechazar } H_0/H_0 \text{ cierta}) &= \alpha \\
P(\bar{X} > C'/\mu = \mu_0) &= 0.05 \\
P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{C'-\mu}{\sigma/\sqrt{n}}/\mu = \mu_0\right) &= 0.05 \\
P(N(0,1) > Z_{C'}) &= 0.05 \\
\Rightarrow Z_{C'} &= 1.645
\end{aligned}$$

- De donde podemos despejar C, y decidir si rechazar o no H_0 en función de la media muestral obtenida.

Discusión

- ¿Cómo habría cambiado el problema si en vez de tener estas hipótesis, hubiésemos considerado:
  - $H_0 : \mu = \mu_0$
  - $H_1 : \mu = \mu_A$

como las hipótesis en cuestión?

- Siendo μ_A otro posible valor para μ tal que $\mu_A > \mu_0$.

1. ***La región de rechazo es la misma***
2. ***El error de tipo I no depende de*** μ_A

Extensión del Lema de Neyman-Pearson (1)

- Podemos concluir que el problema se resuelve de la misma manera para cualquier valor μ_A tal que $\mu_A > \mu_0$, es decir, el procedimiento es válido para contrastar las hipótesis:
  - $H_0 : \mu = \mu_0$
  - $H_1 : \mu > \mu_0$
- Por otro lado podemos observar que la función de potencia es no decreciente, y por ende si $\pi_W(\theta_0) <= \alpha$ => $\pi_W(\theta) <= \alpha \quad \forall\ \theta <= \theta_0$.
- Por lo tanto el procedimiento es válido también para las hipótesis:
  - $H_0 : \mu <= \mu_0$
  - $H_1 : \mu > \mu_0$

Extensión del Lema de Neyman-Pearson (2)

- Concluimos que el Lema se puede aplicar entonces para hipótesis con alternativas unilaterales.

¿Qué sucede con las alternativas bilaterales?

Hipótesis con Alternativa Bilateral (1)

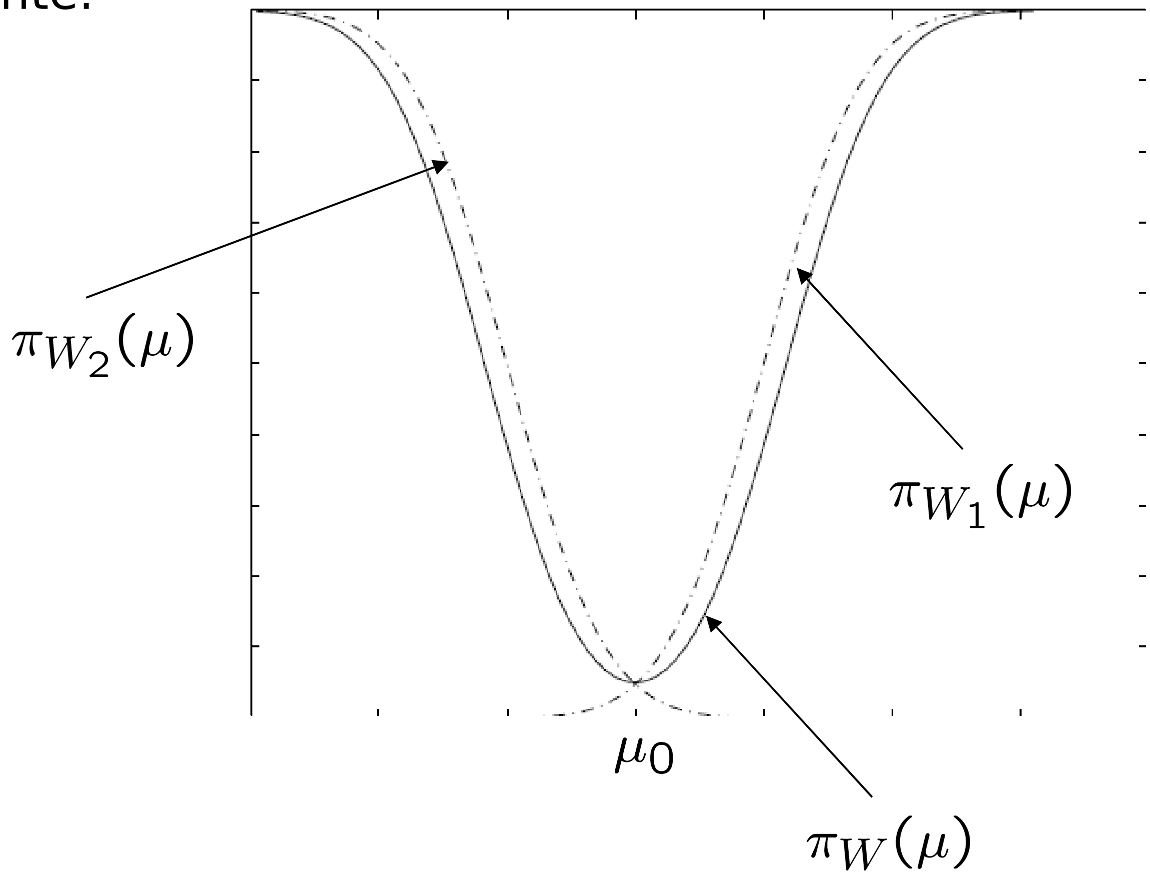
- Supongamos el mismo problema de la media de la Normal, pero esta vez consideremos las hipótesis:
  - $H_0 : \mu = \mu_0$
  - $H_1 : \mu \neq \mu_0$
- Sabemos que si $\mu>\mu_0$ entonces la región de rechazo que minimiza los errores de decisión es $W_1=\{\sum X>C_1\}$, pero si $\mu<\mu_0$, entonces la región crítica que minimiza los errores es $W_2=\{\sum X<C_2\}$.
- Si elegimos como región crítica para el problema a W_1 sabemos que la función de potencia π_{W_1} crece a medida que μ se aleja a la derecha de μ_0 (lo que es razonable), pero no se comporta bien cuando nos alejamos de μ_0 en la dirección contraria. De hecho la región que maximiza la función de potencia en este caso es W_2.

Hipótesis con Alternativa Bilateral (2)

- Cuando una región de crítica maximiza la función de potencia cuando toma los valores de la hipótesis alternativa se dice que dicha región es Uniformemente Más Potente (UMP).
- En este caso la región W_1 es UMP para $\mu>\mu_0$, y la región W_2 es UMP para $\mu<\mu_0$. Luego no es posible encontrar una región crítica más potente para la hipótesis $\mu \neq \mu_0$.
- Como no es posible encontrar una región UMP para alternativas bilaterales, se recurre a la siguiente estrategia: separar la hipótesis alternativa en dos y considerar dos tests de hipótesis separados, de modo que la región de rechazo final sea la unión de las regiones determinadas por separado.
- En el ejemplo $W = W_1 \cup W_2$, de donde $W=\{\sum X>C\_1$ ó $\sum X<C_2\}$.

Hipótesis con Alternativa Bilateral (3)

- Si graficásemos las funciones de potencia de las regiones por separado y la potencia de la región conjunta podríamos observar lo siguiente:



Observaciones Finales

- Cuando no es posible encontrar un test UMP se puede recurrir a otra técnica denomina Test de Razón de Verosimilitud.
- Otra alternativa para decidir una hipótesis es a través del denominado P-Valor, que es el mínimo nivel de significación al cual rechazo H_0 para la muestra observada, dicho de otro modo, corresponde a la probabilidad de rechazar H_0 con los valores observados. Corresponde a la de que el evento dado por la región crítica pueda haber sido producto del azar.

Hipótesis No Paramétricas

Introducción (1)

- Se tira un dado 102 veces, y se registra el número veces que se obtiene cada cara del mismo:

M_i	1	2	3	4	5	6	**Total**
f_i	12	11	22	20	16	21	**102**

- ¿Con la información presentada podemos concluir si el dado está cargado o no?
- Si llamamos X a la variable aleatoria que representa la aparición de una cara, entonces X={1,2,3,4,5,6}.
- Denotemos por p_i=P(X=i) (i=1,...,6).

Introducción (2)

- Pensemos ahora, qué debería pasar si el dado el dado NO estuviese cargado.
- Si el dado no estuviese cargado, deberíamos esperar que la probabilidad de aparición de cada cara fuese $p_i = 1/6$.
- Entonces, si en un dado no cargado el número de apariciones esperadas para cada cara, e_i, sería la misma: $102 \cdot 1/6 = 17$, i.e.:

M_i	1	2	3	4	5	6	**Total**
e_i	17	17	17	17	17	17	**102**

Introducción (3)

- ¿Cómo podríamos medir qué tan lejos están nuestras observaciones de las de un dado no cargado?
- Lo obvio sería considerar las diferencias entre las frecuencias reales y las esperadas.
- Si llamamos $d_i = e_i - f_i$, entonces tenemos:

M_i	1	2	3	4	5	6
d_i	5	6	-5	-3	1	-4

- Podríamos construir un indicador del error total que aporta cada diferencia podría ser $T = \sum d^2_i$.
- Si un dado no estuviese cargado, deberíamos esperar que T fuese pequeño.
- Ahora la pregunta es ¿qué tan pequeño podemos tolerar que sea T?

Condiciones de Borde

- Desde el punto de vista del a Estadística podríamos decidir rechazar la hipótesis de que el dado está cargado si T>C para algún valor C dado, de modo que la probabilidad de cometer un error de tipo I no fuese superior a un nivel α dado.
- Podemos plantear entonces la hipótesis H_0 : “El dado NO está cargado”, y una región de crítica preliminar de la forma T>C.
- Ahora bien, para poder resolver lo anterior necesitamos conocer la distribución bajo la cual se rige la hipótesis nula, es decir, conocer la distribución bajo la cual $p_i = 1/6$.
- Veremos que la distribución aludida es la ***Multinomial***

La Distribución Multinomial

- Es una generalización de la distribución Binomial, donde, en vez de existir solo dos posibles resultados para cada repetición de experimento hay “k” resulyados posibles.
- Si M_i representa el resultado para el experimento i, se tiene que:

$$P(M_1 = m_1, \ldots, M_k = m_k) = \frac{n! p_1^{m_1} p_2^{m_2} \ldots p_k^{m_k}}{m_1! m_2! \ldots m_k!}$$

- Donde $E(M_i)=np_i$

El Estadístico Chi-Cuadrado

- Desde el punto de vista de los tests de hipótesis, la hipótesis nula se puede escribir como:

$$H_0 : p_i = p_i^0$$

- Donde p_i^0 es la probabilidad a priori dad por la distribución Multinomial
- Para la construcción de la región de rechazo se puede recurrir al Test de Razón de Verosimilitud, y se puede probar que asintóticamente el estadístico:

$$Q = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} \sim \chi_{k-1}$$

Procedimiento De Contraste

- Luego la región de rechazo es de la forma W={Q>C}, donde C se obtiene acotando el error de tipo I.
- En el caso del dado, se tiene que $H_0 : p_i = 1/6$, y que Q = 6.589.
- Por otro lado $P(\chi_5^2 > C)=0,05 \Rightarrow C = 11,07$.
- De este modo NO se rechaza que el dado no está cargado.
- Una forma equivalente de decidir es con el P-Valor.

Contraste Para Otras Distribuciones (1)

- Si generalizamos el resultado, se puede utilizar este procedimiento para hacer tests sobre otras distribuciones.

$$
\begin{array}{l}
H_0 : F(X) = F_0(X) \\
H_1 : F(X) \neq F_0(X)
\end{array}
$$

- En el caso de distribuciones discretas, se procede a calcular $p_i = P(X=i)$ (suponiendo H_0 cierta) para todo el dominio de definición de X.
- Se calcula el estadístico Q y se decide.

Contraste Para Otras Distribuciones (2)

- En el caso de v.a.s. continuas se procede a discretizar la distribución tomando intervalos disjuntos equiespaciados (o equiprobables), y se calcula p_i como la probabilidad de caer en el intervalo “i”.
- Es importante considerar todos los posibles valores para la v.a. en el cálculo de los p_i.

Tablas De Contingencia

- El estadístico chi cuadrado no sólo es útil para hacer contrastes no paramétricos, sino que también para hacer tests de independencia entre v.a.s. discretas. (Se puede hacer en el caso continuo, siempre y cuando se discreticen las variables involucradas)
- Cuando dos v.a.s. discretas X e Y son independientes, se tiene que p_{ij} = P(X=i e Y=j) = P(X=i)P(Y=j)=$p_i p_j$ $\forall$ i,j $\in$ AxB.
- Se puede estimar p_i y p_j como

$$p_i = \sum_j f_{ij}/n \qquad p_j = \sum_i f_{ij}/n$$

- Se puede probar que entonces

$$Q = \sum_{i,j} \frac{(f_{ij} - np_ip_j)^2}{np_ip_j} \sim \chi^2_{(l-1)(m-1)}$$

- Donde l,m son el número de filas y columnas de las celdas con los valores del caso

Ejemplo (1)

- Supongamos que queremos ver si existe o no independencia entre el sexo de una persona y la carrera universitaria que esperan estudiar 500 alumnos de un colegio mixto, obteniéndose los siguientes resultados:

	Ingenieria	Medicina	Leyes	Total
Hombres	122	74	96	**292**
Mujeres	43	80	85	**208**
Total	**165**	**154**	**181**	**500**

Ejemplo (2)

- Si las variables X (sexo) e Y (carrera) fuesen independientes debería esperarse que:

p_{HI} = P(X=H e Y=I)=P(X=H)P(Y=I) =(292/500)*(165/500)

- Luego la cantidad esperada de Hombres que quieren Ingenieria debería ser np_{HI} = 96,36.
- Luego armamos el estadístico Chi-cuadrado de la manera usual, lo que nos da un valor de Q = 25,33 > C = 5,99 (χ^2 con 2 g.l. al 5% de significación).
- Por lo tanto se rechaza la hipótesis de independencia entre X e Y, es decir, la elección de carrera SI depende del sexo de la persona.
- El P-Valor asociado a 25,33 es 0,00, lo cual indica un fuerte rechazo de la hipótesis.