Técnicas de Inferencia Estadística II

Tema 2. Contrastes de hipótesis paramétricos

Parte II. Contrastes para dos muestras

M. Concepción Ausín
Universidad Carlos III de Madrid

Grado en Estadística y Empresa
Curso 2014/15

Contenidos

Distinguiremos dos casos:

- Dos muestras independientes: Suponemos dos muestras aleatorias simples $(X_1, X_2, \ldots, X_n)$ e $(Y_1, Y_2, \ldots, Y_m)$ de dos poblaciones normales independientes:

$$X \sim N(\mu_1, \sigma_1^2) \quad \text{e} \quad Y \sim N(\mu_2, \sigma_2^2).$$

- Una muestra bivariante: Suponemos una muestra bivariante, $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ de una población normal bivariante,

$$(X, Y) \sim N\left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right), \left(\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right) \right).$$

En el primer caso las variables X e Y son independientes. En el segundo, X e Y son dependientes (a no ser que $\sigma_{12} = 0$).

Supondremos siempre que las medias (μ_1 y μ_2), las varianzas (σ_1 y σ_2), y en su caso, la covarianza (σ_{12}), son desconocidas.

Ejemplo 2.6.

Distinguir si se trata de dos muestras independientes de dos variables X e Y independientes o de una muestra bivariante de una variable (X, Y) de modo que X e Y puedan ser dependientes.

1. $(X_1, X_2, \dots, X_{10})$ *e* $(Y_1, Y_2, \dots, Y_{15})$ *representan los salarios de* 10 *mujeres y* 15 *hombres, respectivamente.*
2. $(X_1, X_2, \dots, X_8)$ *e* $(Y_1, Y_2, \dots, Y_8)$ *son las calificaciones de* 8 *estudiantes en matemáticas y estadística, respectivamente.*
3. $(X_1, X_2, \dots, X_{16})$ *e* $(Y_1, Y_2, \dots, Y_{16})$ *son las edades de 16 fumadores y 16 no fumadores.*
4. $(X_1, X_2, \dots, X_{20})$ *e* $(Y_1, Y_2, \dots, Y_{20})$ *representan el número de parados en 20 ciudades de dos paises distintos.*
5. $(X_1, X_2, \dots, X_{32})$ *e* $(Y_1, Y_2, \dots, Y_{32})$ *representan el peso de 32 pacientes antes y después de un tratamiento de adelgazamiento.*

Contrastes para dos muestras independientes de dos poblaciones normales

Suponemos ahora la primera situación en la que disponemos de dos muestras aleatorias simples $(X_1, X_2, \ldots, X_n)$ e $(Y_1, Y_2, \ldots, Y_m)$ de dos poblaciones normales, $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, independientes.

Queremos resolver contrastes del tipo:

$$\begin{array}{lll} H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 & H_1 : \mu_1 > \mu_2 & H_1 : \mu_1 < \mu_2 \end{array}$$

$$\begin{array}{lll} H_0 : \sigma_1^2 = \sigma_2^2 & H_0 : \sigma_1^2 = \sigma_2^2 & H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 & H_1 : \sigma_1^2 > \sigma_2^2 & H_1 : \sigma_1^2 < \sigma_2^2 \end{array}$$

Contrastes para la igualdad de varianzas

Se tienen dos muestras $(X_1, X_2, \ldots, X_n)$ e $(Y_1, Y_2, \ldots, Y_m)$ de dos poblaciones normales e independientes, $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, con medias, μ_1 y μ_2, desconocidas.

Queremos resolver contrastes del tipo:

$$\begin{array}{lll} H_0 : \sigma_1^2 = \sigma_2^2 & H_0 : \sigma_1^2 = \sigma_2^2 & H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 & H_1 : \sigma_1^2 > \sigma_2^2 & H_1 : \sigma_1^2 < \sigma_2^2 \end{array}$$

El estadístico de contraste es:

$$\frac{S_1^2}{S_2^2} \sim_{H_0} F_{n-1,m-1}$$

Las regiones de rechazo y el cálculo del p-valor se realiza de manera análoga al contraste para la varianza, teniendo en cuenta que la distribución de $F_{n-1,m-1}$ es asimétrica.

Ejemplo 2.7.

Se conjetura que las acciones de una compañía sufrirían más variación en una industria con competencia en precios que en una en la que existiera un duopolio y colusión tácita.

En un estudio sobre la industria de generadores mediante turbinas de vapor, se halló que en 4 años de competencia en precios la variación de las acciones de la General Electric fue de 114.09. En los siguientes 7 años, en los cuales hubo un duopolio y colusión tácita, esta varianza fue de 16.08.

Asumir que los datos pueden considerarse muestras aleatorias independientes de dos poblaciones normales y contrastar al 5 % la conjetura anterior.

Contrastes para la igualdad de medias

Se tienen dos muestras $(X_1, X_2, \ldots, X_n)$ e $(Y_1, Y_2, \ldots, Y_m)$ de dos poblaciones normales e independientes, $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$, con varianzas, σ_1^2 y σ_2^2 , desconocidas.

Queremos resolver contrastes del tipo:

$$\begin{array}{lll} H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 & H_1 : \mu_1 > \mu_2 & H_1 : \mu_1 < \mu_2 \end{array}$$

Si las varianzas son iguales: $\sigma_1^2 = \sigma_2^2$

El estadístico de contraste es:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim_{H_0} t_{n+m-2}$$

Si las varianzas son distintas: $\sigma_1^2 \neq \sigma_2^2$

El estadístico de contraste es:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n}+\frac{S_2^2}{m}}} \sim_{H_0} t_f$$

donde: $$f = \frac{\left(\frac{S_1^2}{n}+\frac{S_2^2}{m}\right)^2}{\frac{1}{n-1}\left(\frac{S_1^2}{n}\right)^2+\frac{1}{m-1}\left(\frac{S_2^2}{m}\right)^2}$$

Ejemplo 2.8.

De una muestra aleatoria de 12 licenciados en Económicas en una Universidad pública, los sueldos de su primer empleo fueron los siguientes (expresados en miles de dólares):

26.2,	*29.3,*	*31.3,*	*28.7,*	*27.4 ,*	*25.1,*
26.0,	*27.2,*	*27.5,*	*29.8,*	*32.6,*	*34.6*

De otra muestra aleatoria independiente de 10 licenciados en Económicas en una Universidad privada los primeros sueldos fueron los siguientes:

25.3,	*28.2,*	*29.2,*	*27.1,*	*26.8,*
26.5,	*30.7,*	*31.3,*	*26.3,*	*24.2*

Asumiendo normalidad en los datos, discutir si existen diferencias entre los sueldos de los licenciados de Universidades públicas y privadas.

Contrastes para una muestra bivariante de una población normal bivariante

Consideramos que se tiene una muestra $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ de una población normal bidimensional:

$$(X, Y) \sim N\left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right), \left(\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array}\right)\right).$$

Queremos resolver contrastes del tipo:

$$\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array} \qquad \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{array} \qquad \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{array}$$

Contrastes para una muestra bivariante de una población normal bivariante

Transformamos el problema en otro en el que se tiene una sola muestra $(D_1 = X_1 - Y_1, D_2 = X_2 - Y_2, \ldots, D_n = X_n - Y_n)$ de la variable:

$$D = X - Y \sim N(\mu_D, \sigma_D^2)$$

donde $\mu_D = E[X - Y] = \mu_1 - \mu_2$.

El problema se convierte en resolver los contrastes:

$$\begin{array}{lll} H_0 : \mu_D = 0 & H_0 : \mu_D = 0 & H_0 : \mu_D = 0 \\ H_1 : \mu_D \neq 0 & H_1 : \mu_D > 0 & H_1 : \mu_D < 0 \end{array}$$

El estadístico de contraste es:

$$\frac{\bar{D}}{S_D/\sqrt{n}} \sim_{H_0} t_{n-1}$$

donde S_D^2 es la cuasivarianza muestral de $(D_1, \ldots, D_n)$.

Ejemplo 2.9.

Antes de lanzar una promoción muy agresiva de un cierto producto dirigida a los hipermercados de grandes superficies, la directora de marketing de la empresa quiere saber si es o no rentable. Para ello se seleccionan al azar 5 hipermercados de Madrid para llevar a cabo la promoción y se recogen datos de las ventas en miles de euros antes y después de la promoción. Se supone que las ventas se distribuyen normalmente.

Antes	*102*	*120*	*135*	*114*	*175*
Después	*110*	*125*	*141*	*113*	*182*

- *Contrastar la hipótesis de que dicha promoción sea rentable, teniendo en cuenta que se trata de datos apareados.*
- *Contrastar la misma hipótesis, pero asumiendo que son muestras independientes.*
- *Comparar y explicar las diferencias en los dos apartados anteriores.*

Contrastes para la igualdad de medias de dos muestras independientes

Consideramos que se tienen dos muestras $(X_1, X_2, \ldots, X_n)$ e $(Y_1, Y_2, \ldots, Y_m)$ de dos poblaciones no necesariamente normales de medias μ_1 y μ_2 y varianzas, σ_1^2 y σ_2^2, resp., tales que n y m sean grandes ($n, m >$30).

Aunque las poblaciones no sean normales, se pueden resolver contrastes para la diferencia de medias:

$$\begin{array}{lll} H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 & H_1 : \mu_1 > \mu_2 & H_1 : \mu_1 < \mu_2 \end{array}$$

utilizando el Teorema Central del Límite, que garantiza que:

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \to_{H_0} N(0, 1)$$

Ejemplo 2.11.

El método MATWES fue diseñado para medir las actitudes hacia las mujeres ejecutivas. Una puntuación alta indica actitudes negativas hacia las mujeres ejecutivas. Se conjetura que la actitud hacia las mujeres ejecutivas cambia en función del sexo.

Para contrastar esta hipótesis se tomaron muestras aleatorias independientes de 151 hombres y de 108 mujeres estudiantes de M.B.A. En el grupo de los hombres se obtuvo una puntuación media de 85.8 con una desviación típica de 19.3. En el de mujeres se obtuvo una puntuación media de 71.5 con una desviación típica de 12.2.

- *Plantea el contraste oportuno y resuélvelo para* $\alpha = 0.01$.
- *¿Cómo se construiría un intervalo de confianza al 99 % para la diferencia de puntuaciones medias? ¿contendría al 0?*

Contrastes para la igualdad de medias de dos muestras emparejadas

Consideramos que se tiene una muestra $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ de una población bivariante, no necesariamente normal, de media (μ_1, μ_2) y matriz de varianzas desconocida, tal que n es grande ($n >$30).

Aunque la población no sea normal bivariante, se pueden resolver contrastes del tipo:

$$\begin{array}{lll} H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 & H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 & H_1 : \mu_1 > \mu_2 & H_1 : \mu_1 < \mu_2 \end{array}$$

Tomando la muestra de diferencias, $\{D_1 = X_1 - Y_1, \ldots, D_n = X_n - Y_n\}$ y utilizando el Teorema Central del Límite, que garantiza que:

$$\frac{\bar{D}}{S_D/\sqrt{n}} \rightarrow_{H_0} N(0, 1)$$