Algunos Contrastes Paramétricos

Introducción

A diferencia del tema anterior que era fundamentalmente teórico, y ha servido para introducir los conceptos generales de la contrastación de hipótesis estadísticas que se aplicarán a todo tipo de tests, el presente tema se ocupa de presentar, de forma práctica, como realizar algunos de los contrastes paramétricos mas habituales.

A excepción de los contrastes relativos a proporciones, los demás contrastes que se presentan parten de que la población que se muestrea, sigue una distribución Normal y contando con este supuesto, deducen la distribución del estadístico de contraste. No obstante, como la media muestral sigue una distribución aproximadamente Normal, aunque la población de procedencia no lo sea, muchos de estos contrastes pueden utilizarse de forma más amplia con poblaciones no Normales, al menos como una aproximación. Esta aproximación será tanto mejor cuanto mayor sea la muestra y cuanto menor sea la discrepancia de la distribución poblacional con la Normal.

Desde un punto de vista practico, recomendamos seguir las siguientes etapas para la realización de un contraste a partir de unos datos o en una situación dada:

1ª Establezca la Hipótesis Nula que corresponda al problema en cuestión. Recuerde que debe ser una hipótesis de no diferencias o de efecto nulo de los tratamientos.

2ª Determine la Hipótesis Alternativa que va a utilizar, en general esta hipótesis reflejará el efecto de cual se quiere probar su existencia. Debe tenerse en cuenta que las hipótesis unidireccionales conducen a tests más potentes y en este sentido son preferibles.

3ª En función de la Hipótesis Nula y de las condiciones del muestreo, seleccione el estadístico de contraste más adecuado.

4ª A partir de los datos muestrales, calcule el valor del estadístico de contraste.

5ª Tomando en consideración el nivel de confianza que vaya a utilizar, y la Hipótesis Alternativa fijada en el punto 2º, determine el valor crítico.

6ª Compare el valor del estadístico con el valor crítico y vea cual es la decisión que debe adoptar.

7ª Interprete la decisión, adoptada en el punto anterior, en el contexto del problema planteado.

El seguimiento de estos pasos en el orden establecido, fijándonos en la resolución concreta del punto en que nos encontramos, sin prestar atención a los posteriores, sistematiza el procedimiento y facilita una correcta ejecución. No debe olvidarse nunca el último punto, sin cuya realización carecen de sentido los anteriores.

Contrastes de Significación

Se llama contrastes de significación a aquellos tests en los cuales la Hipótesis Nula es una hipótesis simple, es decir, consta de un solo punto del espacio paramétrico. Son pruebas en las que se pretende ver si un parámetro de la población difiere significativamente de un valor que representa la ausencia de efecto, o que venía considerándose tradicionalmente. La mayoría de estas pruebas conllevan la utilización de una sola muestra, por lo que a veces son denominadas, contraste para una media, contraste para una proporción, etc.

<u>Contraste de significación para la media</u>

Sea una variable aleatoria X que sigue una distribución Normal de media μ y desviación típica σ y queremos contrastar la hipótesis de que la media es igual a un determinado número μ_0. Por consiguiente, nuestra Hipótesis Nula será:

$$H_0: \mu = \mu_0$$

frente a una Hipótesis Alternativa que puede adoptar una de estas tres formas:

$$H_1: \mu \neq \mu_0$$

$$H_1: \mu > \mu_0$$

$$H_1: \mu < \mu_0$$

El estadístico de contraste que se utiliza para contrastar la Hipótesis Nula es:

$$t = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

que como vimos en el tema anterior, sigue una t de Student con n-1 grados de libertad, cuando se verifica la Hipótesis Nula. Cuando se verifica la Hipótesis Alternativa, este estadístico seguirá una distribución que estará desplazada a la derecha o a la izquierda de la t de Student según sea la verdadera media mayor o menor que el valor especificado μ_0.

<u>Contraste de significación para una proporción</u>

Sea una población en la que una determinada proporción de individuos P, presentan una cierta característica y queremos contrastar la hipótesis de que esa proporción es igual a una cierta cantidad P_0. Tendremos por consiguiente una Hipótesis Nula de la forma:

$$H_0: P = P_0$$

y una Hipótesis Alternativa que puede adoptar una de las siguientes expresiones:

$$H_1: P \neq P_0 \qquad H_1: P > P_0 \qquad H_1: P < P_0$$

Para contrastar esta hipótesis extraeremos una muestra de la población de tamaño n, y verificaremos si sus datos son compatibles con la hipótesis establecida. Sea p la proporción de individuos que presentan la característica en cuestión, en la muestra. Sabemos que esta proporción muestral sigue aproximadamente una distribución Normal de media P y de desviación típica $\sqrt{\frac{P \cdot Q}{n}}$ por consiguiente si tipificamos esta variable tendremos que:

$$\frac{p - P}{\sqrt{\frac{P \cdot Q}{n}}} \rightarrow N(0,1)$$

Por tanto si se verifica la Hipótesis Nula de que P = P_0 tendremos que el estadístico de contraste:

$$\frac{p - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$

seguirá una Normal cero uno. Por el contrario cuando se verifique la Hipótesis Alternativa es te estadístico seguirá una distribución Normal de media P-P_0 que estará situada a la derecha o a la izquierda de la Normal(0, 1) según P sea mayor o menor que P_0.

<u>Contraste de significación para el coeficiente de correlación</u>

Se trata de contrastar si dos variables Normales son incorreladas o si por el contrario existe una correlación significativa entre ambas. Por ello la Hipótesis Nula de este contraste es:

$$H_0: \rho = 0$$

frente a una de las Hipótesis Alternativas siguientes:

$$H_0: \rho \neq 0 \qquad H_0: \rho > 0 \qquad H_0: \rho < 0$$

Como en todos los contrastes se extrae una muestra de tamaño n y se comprueba si los resultados son compatibles con la Hipótesis Nula. En este caso el estadístico de contraste es:

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

donde r es el coeficiente de correlación de Pearson entre ambas variables, calculado a partir de los datos de la muestra. Si se verifica la Hipótesis Nula de que la correlación, en la población, entre ambas variables es cero. Entonces, este estadístico sigue una distribución t de Student con n-2 grados de libertad.

Contraste de significación para la varianza

Sea una variable aleatoria que sigue una distribución Normal(μ, σ) y se quiere contrastar si la varianza de esa variable es igual a una cierta cantidad especificada. Tendremos que entonces la Hipótesis Nula sometida a contraste es:

$$H_0 : \sigma^2 = \sigma_0^2$$

frente a una de las Hipótesis Alternativas de que la varianza sea distinta, mayor o menor que la cantidad especificada. Para establecer el estadístico de contraste, partimos del resultado del teorema de Fisher que nos dice que el estadístico:

$$\frac{nS^2}{\sigma^2} \rightarrow \chi^2_{n-1}$$

sigue una distribución Ji-cuadrado con n-1 grados de libertad, por consiguiente cuando se verifique la Hipótesis Nula tendremos que el estadístico de contraste:

$$\frac{nS^2}{\sigma_0^2} \rightarrow \chi^2_{n-1}$$

seguirá esa misma distribución. Mientras que si se verifica la Hipótesis Alternativa, al estar dividiendo por un número menor o mayor que la verdadera varianza, la distribución del estadístico estará desplazada a la derecha o a la izquierda de la distribución Ji-cuadrado.

Contraste de igualdad de medias

La situación a la que nos enfrentamos en este tipo de contrastes es la siguiente. Tenemos una variable X_1 que sigue una distribución Normal de media μ_1 y varianza σ_1^2. Tenemos otra variable X_2 que sigue una distribución Normal de media μ_2 y varianza σ_2^2.

$$X_1 \rightarrow N(\mu_1, \sigma_1^2) \qquad X_2 \rightarrow N(\mu_2, \sigma_2^2)$$

y deseamos contrastar si las medias de estas dos poblaciones son iguales o diferentes. La Hipótesis Nula será por consiguiente:

$$H_0 : \mu_1 = \mu_2$$

frente a una Hipótesis Alternativa que podrá adoptar una de las siguientes expresiones:

$$H_1 : \mu_1 \neq \mu_2 \qquad H_1 : \mu_1 > \mu_2 \qquad H_1 : \mu_1 < \mu_2$$

para ver si la Hipótesis Nula debe ser aceptada o rechazada, extraeremos sendas muestras de X_1 y X_2 y comprobaremos si los datos que nos proporcionan son compatibles con esa hipótesis. La muestra de X_1 será :

$$X_1 \quad \rightarrow \quad \left(x_{11}, x_{12}, \ldots x_{1n_1}\right)$$

$\bar{x}_1 \quad S_1^2 \quad \widetilde{S}_1^2$ serán respectivamente, la media, la varianza y la cuasivarianza de esta muestra. Análogamente la muestra de X_2 será:

$$X_2 \quad \rightarrow \quad \left(x_{21}, x_{22}, \ldots x_{2n_2}\right)$$

siendo $\bar{x}_2 \quad S_2^2 \quad \widetilde{S}_2^2$ las correspondientes media, varianza y cuasivarianza. A partir de esta situación, estableceremos los distintos estadísticos de contraste que se utilizarán, en función de las distintas condiciones que se den en las muestras.

Muestras Independientes

Esta situación se produce cuando ambas variables no presentan correlación alguna, es la habitual cuando dos tratamientos son aplicados a sujetos distintos, entre los cuales no existe ningún vínculo y que no han sido emparejados con respecto a ninguna variable. Responden a este esquema los planteamientos que enfrentan los resultados de un grupo experimental que se somete a tratamiento, con los de un grupo control. También las comparaciones en los valores de dos grupo de sujetos distintos, como son el estudio de diferencias significativas en una variable entre hombres y mujeres, etc. Dentro de esta situación podemos a su vez considerar dos casos, cuando las varianzas poblacionales, aunque desconocidas, pueden considerarse iguales y el más general de que sean distintas.

Varianzas iguales

Esta condición es más frecuente de lo que pudiera parecer en un primer momento, y se produce cuando los tratamientos tienen un efecto aditivo que incrementa o disminuye la media pero no afecta a la dispersión de los mismos. Por otra parte, esta situación se halla muy bien resuelta desde un punto de vista teórico, disponiéndose de un estadístico de contraste del cual se conoce su distribución exacta, independientemente de cual sea el tamaño muestral, por lo que siempre que sea posible será preferible utilizarlo.

Para determinar si las varianzas poblacionales pueden considerarse iguales, deberá atenderse tanto a consideraciones teóricas sobre la naturaleza de los efectos de los tratamientos, como a la evidencia empírica que proporciona el grado de discrepancia entre los valores de las varianzas muestrales, pudiéndose proceder a la contrastación previa de dicha hipótesis.

Como vimos en el tema de las distribuciones en el muestreo, el estadístico:

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \rightarrow \quad t_{n_1+n_2-2}$$

sigue una distribución t de Student con $n_1 + n_2 - 2$ grados de libertad, por consiguiente cuando se verifique la Hipótesis Nula de que ambas media son iguales, la diferencia $\mu_1 - \mu_2$ será cero y el estadístico de contraste:

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

seguirá la mencionada distribución t de Student. Por el contrario si se verifica la Hipótesis Alternativa la diferencia de las medias poblacionales será distinta de cero y la distribución del estadístico de contraste estará a la derecha o a la izquierda de la t de Student.

Varianzas desiguales

Cuando los tratamientos no tengan efectos aditivos o bien inicialmente las dos poblaciones posean dispersiones sensiblemente diferentes, tendremos que resignarnos a no emplear el estadístico anteriormente descrito. En este caso debemos a su vez distinguir dos situaciones que la muestra sean grandes o que sean pequeñas.

Muestras grandes

Si las muestras son lo suficientemente grandes como para proporcionar estimaciones bastante precisas de las varianzas poblacionales podremos afirmar que el estadístico:

$$\frac{x_1 - x_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}}$$

seguirá aproximadamente una distribución Normal cero, uno y en consecuencia el estadístico de contraste que vamos a utilizar:

$$\frac{x_1 - x_2}{\sqrt{\frac{\widetilde{S}_1^2}{n_1} + \frac{\widetilde{S}_2^2}{n_2}}}$$

seguirá, cuando se verifique la Hipótesis Nula, una distribución N(0.1), al menos de forma aproximada, y una distribución que estará desplazada a la derecha o a la izquierda de la misma cuando se verifique la Hipótesis Alternativa.

Muestras pequeñas

Cuando ocurra que las varianzas poblacionales son significativamente distintas y las muestras son pequeñas, hemos de decir que no tenemos una solución enteramente satisfactoria y por consiguiente sería preferible evitar esta situación, bien aumentando el tamaño de las muestras o bien replanteando el diseño de la experiencia. No obstante como solución paliativa puede recurrirse al siguiente resultado que señala que en estas condiciones, el mismo estadístico anterior:

$$\frac{x_1 - x_2}{\sqrt{\dfrac{\tilde{S}_1^2}{n_1} + \dfrac{\tilde{S}_2^2}{n_2}}}$$

sigue, bajo la Hipótesis Nula, aproximadamente, una distribución t de Student con k grados de libertad, donde k viene dado por la expresión:

$$k = \frac{\left(\dfrac{\tilde{S}_1^2}{n_1} + \dfrac{\tilde{S}_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{\tilde{S}_1^2}{n_1}\right)^2}{n_1 + 1} + \dfrac{\left(\dfrac{\tilde{S}_2^2}{n_2}\right)^2}{n_2 + 1}} - 2$$

Muestras Correlacionadas

Dentro de las situaciones en que las variables X_1 y X_2 están correlacionadas y por consiguiente las muestras de ambas poblaciones no son independientes, vamos a considerar el caso de que los datos de ambas muestras están apareados. Este tipo de datos se presenta en los diseños de medidas repetidas, es decir cuando los individuos son medidos en dos condiciones experimentales, también cuando existe un vínculo entre los sujetos de ambas muestras que hace que sus valores estén correlacionados, o bien cuando los sujetos, previamente a la administración de los tratamiento, han sido emparejados de acuerdo con alguna variable relacionada con la característica que se estudia.

El procedimiento para abordar la contrastación de la igualdad de medias en este caso, consiste en considerar la distribución de la diferencia de las dos variables aleatorias, es decir de la variable $X_d = X_1 - X_2$. En virtud de las propiedades de la distribución Normal tendremos que esta variable seguirá una distribución también Normal:

$$X_d \quad \rightarrow \quad N(\mu_d, \sigma_d)$$

verificándose que:

$$\mu_d = \mu_1 - \mu_2 \qquad \sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

Entonces nuestra Hipótesis Nula puede enunciarse de la forma siguiente:

$$H_0 : \mu_1 = \mu_2 \rightarrow H_0 : \mu_d = 0$$

transformándose de igual manera las Hipótesis Alternativas:

$$H_1 : \mu_1 \neq \mu_2 \rightarrow H_1 : \mu_d \neq 0$$

$$H_1 : \mu_1 > \mu_2 \rightarrow H_1 : \mu_d > 0$$

$$H_1 : \mu_1 < \mu_2 \rightarrow H_1 : \mu_d < 0$$

De esta forma, nuestro contraste de igualdad de medias se ha transformado en un contraste de significación para las diferencias. Por consiguiente el estadístico de contraste será el ya conocido:

$$\frac{\bar{x}_d}{\frac{S_d}{\sqrt{n-1}}} \rightarrow t_{n-1}$$

que bajo la Hipótesis Nula, seguirá una distribución t de Student con n-1 grados de libertad .