



CONTRASTE DE HIPÓTESIS

Comparación de dos grupos

UTILIDAD

Es común necesitar comparar dos grupos de estudios, un ejemplo común es el estudio de fármacos, comparando el efecto en un grupo que recibió el tratamiento y uno que recibió un placebo.

Anteriormente contrastamos hipótesis que comparaban un parámetro estimado de la población con un valor elegido, ahora se expande el concepto para comparar un parámetro estimado para dos poblaciones e inferir si estas son iguales o diferentes

INDEPENDENCIA Y MUESTRAS PAREADAS

Dos grupos se definen como independientes cuando los valores de una muestra elegida de una población no condicionan ni son condicionados por los valores de la otra.

Las muestras pareadas no son independientes y generalmente surgen de obtener los resultados del mismo individuo (o gemelos) en distintas condiciones.

CONSIDERACIONES GENERALES RESPECTO A INDEPENDENCIA

Cuando trabajamos con muestras pareadas siempre comparamos medias, cuando son independientes podemos comparar medias o proporciones.

El n en muestras pareadas siempre va a ser igual entre los 2 grupos.

El n en muestras independientes puede ser igual o diferente entre los 2 grupos.

Si el n de las dos muestras es diferente siempre son independientes, si son iguales los n pueden ser independientes o pareadas.

VARIABLE DE RESPUESTA Y FACTOR DE ESTUDIO

En todos los casos de estudio para tener dos muestras que comparar necesitamos de dos variables, aquella a la que mediremos la media o proporción para comparar y la que utilizaremos para dividir en dos grupos.

La primera es llamada variable de respuesta y será cuantitativa si estamos comparando medias o cualitativa si comparamos proporciones (media de glicemia, media de peso, proporción de sanos/enfermos)

La segunda es el factor de estudio y siempre será cualitativa dicotómica, ya que estamos categorizando en dos grupos para comparar (sano/enfermo, hombre/mujer, tratamiento/placebo)

COMPARAR PROPORCIONES

ASUNCIONES

Partimos de dos asunciones:

- Las dos muestras son aleatorias simples e independientes.
- El número de aciertos y fallos son por lo menos 5 para cada uno por cada muestra o np y nq es mayor o igual a 15 en las dos muestras.

Es común querer comparar dos proporciones, si las proporciones estimadas son diferentes puede ser por una diferencia en la población o por azar. El test de hipótesis nos permite saber si la diferencia es significativa.

OTRAS GENERALIDADES PROPORCIONES

La diferencia entre dos proporciones muestrales sigue una distribución normal.

Generalmente la hipótesis nula afirma que las dos proporciones son iguales.

Se utiliza para el estadístico de prueba una media entre las diferencias de las proporciones que se calcula:

$$\bar{p} = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$$

El estadístico de prueba es el siguiente:

$$Z_c = \frac{(p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

COMPARAR MEDIAS MUESTRAS PAREADAS

Dos muestras con datos pareados se pueden reducir analíticamente a una sola muestra conteniendo la diferencia de cada par como variable de estudio, por lo que el problema de comparar medias de muestras pareadas es igual a un test de un grupo en el que planteamos como hipótesis nula que la diferencia de medias es igual a 0.

Por esta razón este procedimiento sigue los mismos supuestos que el test de hipótesis para una media.

COMPARACIÓN MUESTRAS PAREADAS

GENERALIDADES

Para el cálculo del estadístico debemos calcular las diferencias entre todos los datos pareados, su media y su desviación y calculamos usando el siguiente estadístico:

| Antes | Después | Diferencia |
|-------|---------|------------|
| 4 | 5 | 1 |
| 5 | 5 | 0 |
| 3 | 4 | 1 |
| 2 | 4 | 2 |
| 4 | 4 | 0 |

$$t_c = \frac{\bar{d}}{\hat{s}_d / \sqrt{n}} \quad gl = n - 1$$

COMPARAR MEDIAS MUESTRAS INDEPENDIENTES

Para comparar dos muestras independientes también utilizamos la diferencia entre los datos de ambas muestras dividido el error estándar, mismo procedimiento teórico que para los casos anteriores.

La complejidad en estos casos surgen de necesitar estimar el error estándar poblacional a partir de los muestrales.

DIFERENCIA DE CASOS VARIANZA POBLACIONAL CONOCIDA

La primera posibilidad, y mas sencilla, es conocer la varianza poblacional y el estadístico se calcula de la siguiente manera:

$$Z_c = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Si las varianzas coinciden podemos simplificar el cálculo en:

$$Z_c = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{1/n_1 + 1/n_2}}$$

VARIANZA POBLACIONAL DESCONOCIDA Y DIFERENTE ENTRE MUESTRAS

En este caso es igual que el anterior, solo que sigue una distribución t, utilizando las varianzas muestrales.

$$t' = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}}}$$

El cálculo de los grados de libertad del estadístico se realiza de la siguiente manera:

$$gl = \frac{\left(\frac{\hat{s}_1^2}{n_1} + \frac{\hat{s}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\hat{s}_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{\hat{s}_2^2}{n_2}\right)^2}{n_2 - 1}}$$

VARIANZA POBLACIONAL DESCONOCIDA E IGUAL ENTRE MUESTRAS

El estadístico en esta situación es:

$$t_c = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
$$gl = n_1 + n_2 - 2$$

El valor S_p se calcula:

$$s_p = \sqrt{\frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2}}$$

COMPARAR LA VARIANZA DE DOS MUESTRAS

TEST F DE IGUALDAD DE VARIANZAS

En los casos en los que la varianza poblacional no es conocida no alcanza con ver si el valor puntual de la varianza de cada muestra son iguales para compararlos, ya que las diferencias pueden ser atribuidas al muestreo aleatorio es necesario un nuevo contraste de hipótesis para saber si las varianzas son significativamente iguales o diferentes.

$$F_c = \frac{\hat{S}_{mayor}^2}{\hat{S}_{menor}^2}$$
$$gl = (n_a - 1), (n_b - 1)$$

EJERCICIO

Dos tipos de medicación para la urticaria se están estudiando para determinar si existe una diferencia en la proporción de reacciones en pacientes adultos.

20 de 200 adultos de una muestra aleatoria que fueron tratados con la medicación A todavía presentaban urticaria luego de 30 minutos de recibir la medicación.

12 de 200 adultos que recibieron la medicación B presentaban urticaria 30 minutos luego de la aplicación de la medicación.

Comparar a un nivel de significación del 99%.

EJERCICIO

Se investiga la efectividad del hipnotismo en la reducción del dolor. Se elige una muestra aleatoria de individuos que reportan resultados subjetivos de dolor con hipnosis y antes.

¿La hipnosis reduce las sensación de dolor? Probar a un 5% de error.

| Con hipnosis | Sin hipnosis | Diferencia |
|--------------|--------------|------------|
| 6,8 | 6,6 | 0,2 |
| 2,4 | 6,5 | -4,1 |
| 7,4 | 9 | -1,6 |
| 8,5 | 10,3 | -1,8 |
| 8,1 | 11,3 | -3,2 |
| 6,1 | 8,1 | -2 |
| 3,4 | 6,3 | -2,9 |
| 2 | 11,6 | -9,6 |

EJERCICIO

Se cree que el promedio de tiempo que juegan deportes por días niños y niñas entre 7 y 11 años es el mismo. De un experimento realizado se extrae la siguiente información:

| | Tamaño muestral | Promedio de horas | Varianza muestral |
|-------|-----------------|-------------------|-------------------|
| Niñas | 9 | 2 | 0,75 |
| Niños | 16 | 3,2 | 1,2 |

1. Probar a un 95% de confianza si existe diferencia (asumiendo varianza poblacional de niñas de 0,8 y niños 1,05)
2. Si no conociéramos las varianzas poblacionales, podríamos considerar estas muestras como de igual varianza? (alfa 0,05)
3. Probar a un 95% de confianza si existe diferencia (asumiendo varianza poblacional desconocida)
4. Probar a un 90% de confianza si existe diferencia (varianza pob desconocida y test F con alfa=0,01)