

PRACTICANDO EL CONTRASTE DE HIPÓTESIS

Una de las técnicas más utilizadas en Estadística Inferencial es el contraste de hipótesis.

Esta metodología te va a permitir contestar preguntas que admitan un **verdadero o falso** como respuesta.

Se utiliza muchísimo en publicaciones científicas y de la salud.

En esta hoja de trabajo te ayudaré a entender el concepto de p-valor y la metodología del contraste.

¡A por ello!

EJEMPLO - ENTENDIENDO EL MÉTODO DE ANÁLISIS

Descarga la plantilla Excel del contraste de hipótesis.

Tenemos disponible la tabla de datos Iris.

Voy a ponerte un ejemplo de esta tabla de datos para que entiendas dónde se sitúa el contraste de hipótesis en la metodología que describí en el pretrainig.

PASO 1 - DEFINIENDO EL OBJETIVO – definiendo las hipótesis

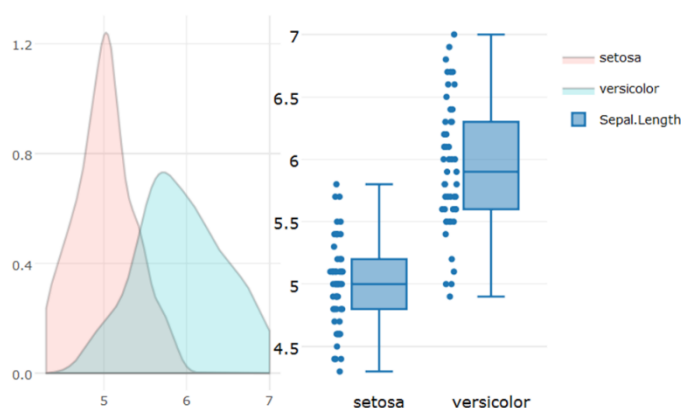
Se trata de observar diferencia entre grupos de especies. En este pequeño estudio se quiere dos especies:

Setosa y Versicolor.

Y queremos ver si hay diferencias en la longitud del pétalo.

PASO 2 - DESCRIBIENDO LA PREGUNTA

Antes de nada, es muy interesante visualizar la información que queremos analizar:



Con un histograma de densidad y un boxplot que ya conoces es suficiente para lograrlo.

PASO 3 - ANÁLISIS / CONTRASTE DE HIPÓTESIS

Para analizar este estudio podemos utilizar un contraste de hipótesis.

PRACTICANDO EL CONTRASTE DE HIPÓTESIS

Define la H1 o hipótesis de investigación:

H1: las poblaciones de setosa y versicolor tienen diferencias significativas en la longitud del sépalo

Define la H0 o hipótesis nula:

H0: las dos poblaciones son iguales

Umbral de contraste (alpha):

5% = 0.05

Test Estadístico (selección del test):

Los grupos son normales y tienen las mismas varianzas >> T-TEST para grupos independientes

Significación o p-valor:

p-valor < 2.2e-16

Respuesta:

El p-valor < 0.05 hay claras evidencias que las poblaciones de setosa y versicolor son diferentes en relación al sepal.length

PASO 4 - CONCLUSIÓN

Con la información que te dado en la exploración y el análisis, ¿qué más puedes afirmar?

*Pista: fíjate en la **centralidad** y la **dispersión** y observa las diferencias. Describe lo que ves y utiliza el poder la inferencia para dar conclusiones más potentes:*

El objetivo del análisis era determinar si hay diferencias en la longitud del pétalo de las especies versicolor y setosa. El análisis de datos, en concreto el contraste de hipótesis (en este caso el test estadístico T-student indica que la especie versicolor presenta una probabilidad significativa de tener un tamaño del sépalo más grande la setosa. La media de la longitud de pétalo de la setosa está en torno a los 5 mm y la de la versicolor está cerca de los 6mm.

En cuanto a la dispersión (SD e IQR), los datos indican una variabilidad un poco mayor en el caso del versicolor (0,352 (setosa) vs 0,516 (versicolor) y también un IQR un poco mayor(0,4 (setosa) vs 0,7 (versicolor). Todo esto indica una mayor uniformidad en la longitud de pétalo en el caso de la especie versicolor).

PRACTICANDO EL ROADMAP DEL CONTRASTE DE HIPÓTESIS Y EL P-VALOR

Te voy a mostrar los resultados de un análisis de un pequeño estudio. (de hecho ya lo has visto con anterioridad)

PASO 1 - DEFINIENDO EL OBJETIVO – definiendo las hipótesis

Se trata de saber si hay relación entre medidas. Peso y consumo del coche.

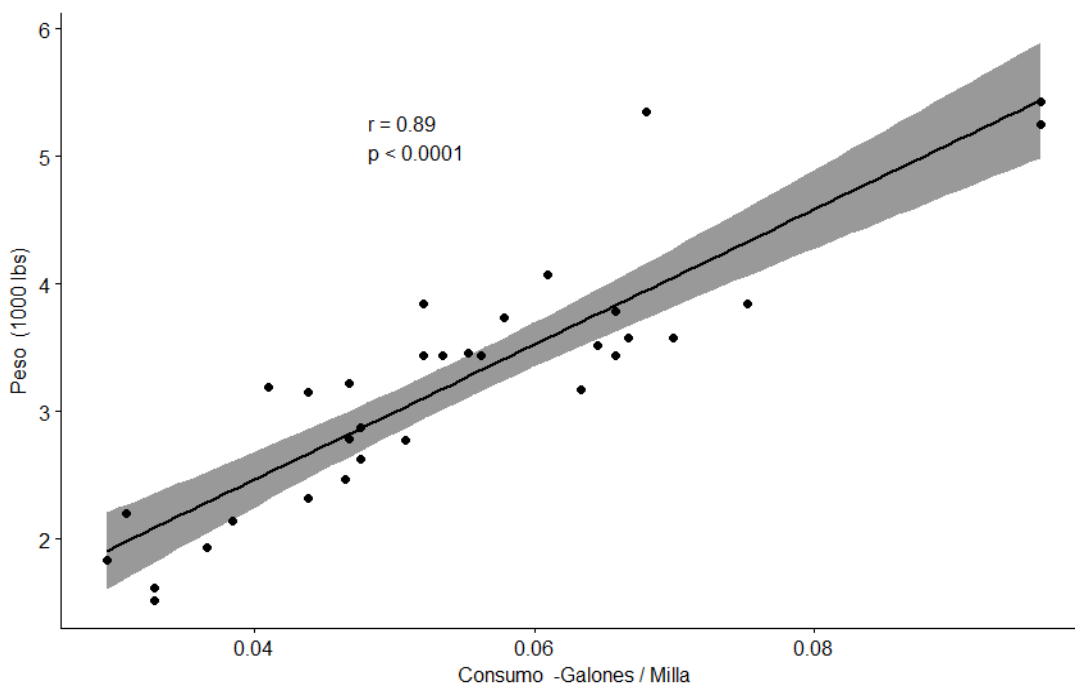
Tenemos dos variables: Consumo y peso

Son dos variables cuantitativas continuas.

El objetivo: saber si hay relación entre las dos medidas. ¿Te suena?

PASO 2 - DESCRIBIENDO LA PREGUNTA

Como ya has visto para describir dos medidas puedes utilizar un diagrama de dispersión e incluso dibujar la línea de tendencia.



Fíjate que tienes el resultado del p-valor.

¡Ahora ya puedes saber qué significa!

PASO 3 - ANÁLISIS / CONTRASTE DE HIPÓTESIS

PRACTICANDO EL CONTRASTE DE HIPÓTESIS

Para analizar este estudio podemos utilizar un **contraste de hipótesis** para saber si tenemos relación entre medidas.

Y después calcular el **coeficiente de correlación** (que ya sabes qué significa)

Con estos resultados y gráficos, rellena la tabla del contraste

Rellena la tabla del contraste de hipótesis:

Define la H1 o hipótesis de investigación: Hay relación entre las dos medidas (peso y consumo)
Define la H0 o hipótesis nula: No hay relación entre las dos medidas (peso y consumo)
Umbral de contraste (alpha): 5% = 0.05
Técnica Estadística: Coeficiente de correlación de Pearson
Significación o p-valor: 0,0001
Respuesta: Sí hay relación entre las dos medidas (peso y consumo))

PASO 4 - CONCLUSIÓN

Con la información que te dado en la exploración y el análisis y lo que sabes del p-valor y del coeficiente de correlación, ¿qué más puedes afirmar?

Pista: utiliza lo que sabes del coeficiente de correlación y la fuerza del p-valor para aportar conclusiones potentes de tus datos

El coeficiente de correlación es 0,89. Tiene signo positivo (lo cual indica una correlación directa) y un valor que es muy alta (está entre 0,8 y 1). El valor de p asociado a dicho análisis es muy significativa, e indica que podemos aceptar la hipótesis de que hay relación entre las dos medidas peso y consumo de coches (H1), con una probabilidad muy baja de estar aceptando la hipótesis erróneamente (0,1%). Aunque las dos variables están relacionadas, hay que puntualizar que no se puede concluir que haya causalidad entre ellas.

¡Bien!

Ya has visto dos ejemplos de contraste de hipótesis y has aplicado la metodología del contraste paso a paso.

Ahora ya tienes mucho más claro cómo puedes obtener conclusiones de los datos con el contraste de hipótesis.

Y ves mucho más claro los 4 pasos que vas a aprender a ejecutar durante el programa.

¡Let's go!