

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

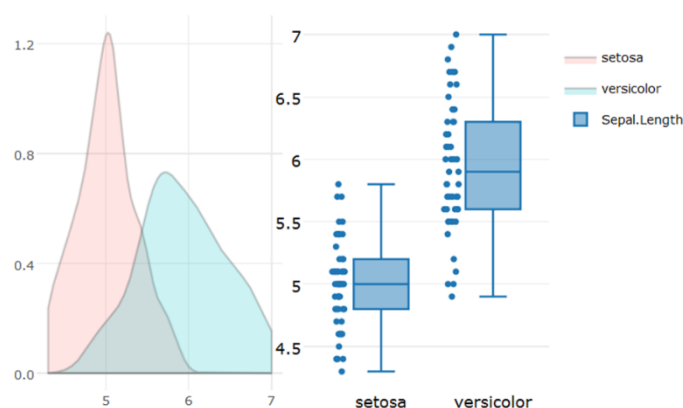
Se trata de observar diferencia entre grupos de especies. En este pequeño estudio se quiere dos especies:

Setosa y Versicolor.

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

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:

Con un porcentaje de probabilidad de error inferior a 0,05 ($p = 2,2 \cdot 10^{-16}$), aceptamos nuestra hipótesis de investigación (H1) por lo que las poblaciones de setosa y versicolor tienen diferencias significativas en la longitud del sépalo.

Claramente observamos que la media y desviación estándar en la longitud del sépalo para la variante Setosa (5,006 +/- 0,352) es inferior a la variante Versicolor (5,936 +/- 0,516).

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:

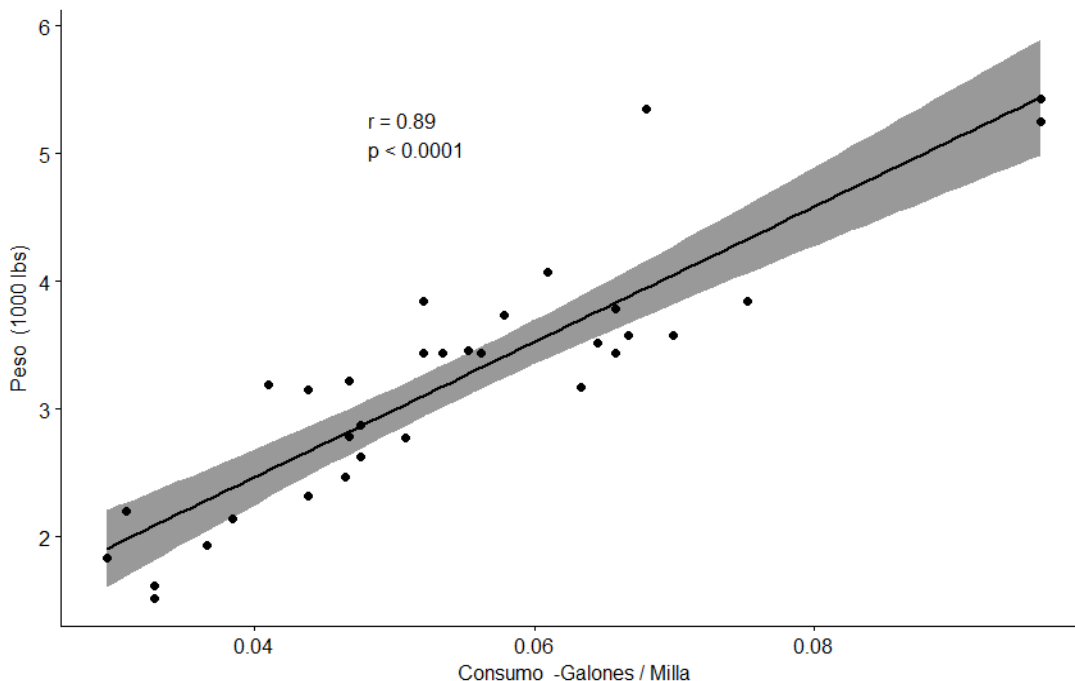
Consumo y peso

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!



PRACTICANDO EL CONTRASTE DE HIPÓTESIS

PASO 3 - ANÁLISIS / 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:

Existe relación entre las variables Peso y Consumo del coche.

Define la H0 o hipótesis nula:

No existe relación entre las variables peso y consumo del coche.

Umbral de contraste (alpha):

5% = 0.05

Técnica Estadística:

Coeficiente de correlación de Pearson

Significación o p-valor:

$p < 0,0001$

Respuesta:

El valor del coeficiente de $R = 0,89$, esto nos indica que la correlación entre el peso y el consumo del coche es muy alta y además positiva; por tanto, hay una relación directa entre consumo del vehículo y su peso.

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

Aceptamos la hipótesis de investigación H1, dado el valor tan bajo que hemos obtenido para p, y por otro lado, el valor tan alto obtenido para el coeficiente de correlación. El consumo del coche depende directamente de su peso, es decir, cuanto más pese el coche más consumirá.

PRACTICANDO EL CONTRASTE DE HIPÓTESIS

¡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!