

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

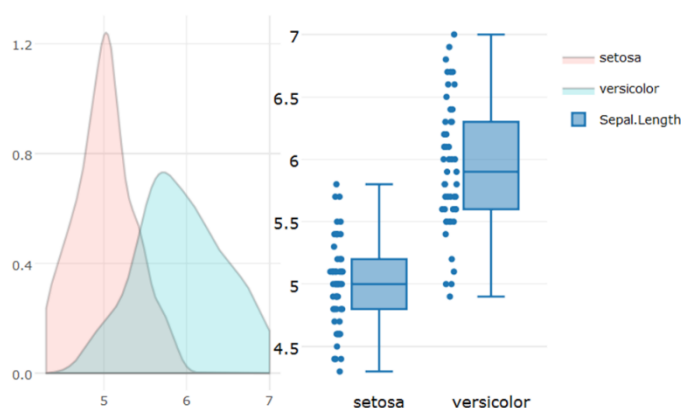
Se trata de observar diferencia entre grupos de especies. En este pequeño estudio se quiere dos especies:

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 pétalo

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:

En el análisis sobre la diferencia en la longitud del sépalo entre las especies de iris Setosa y Versicolor, se observan varios puntos clave a partir de los datos:

- **Centralidad y Dispersión:** Setosa generalmente presenta sépalos más cortos y con menos variabilidad que Versicolor. Esto es evidente en el boxplot, donde la mediana de Setosa es inferior, y en el histograma de densidad, donde su distribución es más estrecha y está más desplazada hacia medidas menores.
- **Distribución:** La distribución de la longitud del sépalo en Setosa es compacta y tiende hacia valores más bajos, mientras que Versicolor tiene una distribución más amplia y centrada en longitudes mayores. Esto implica una mayor dispersión y rango en Versicolor.

PRACTICANDO EL CONTRASTE DE HIPÓTESIS

- **Análisis inferencial:** Las diferencias marcadas en la longitud del sépalo pueden ser cruciales para la identificación precisa de estas especies y podrían reflejar adaptaciones distintas a sus respectivos ambientes o estrategias reproductivas.

En conclusión, el análisis confirma diferencias estadísticamente significativas en la longitud del sépalo entre Setosa y Versicolor, ofreciendo datos valiosos sobre su clasificación y adaptación ecológica.

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

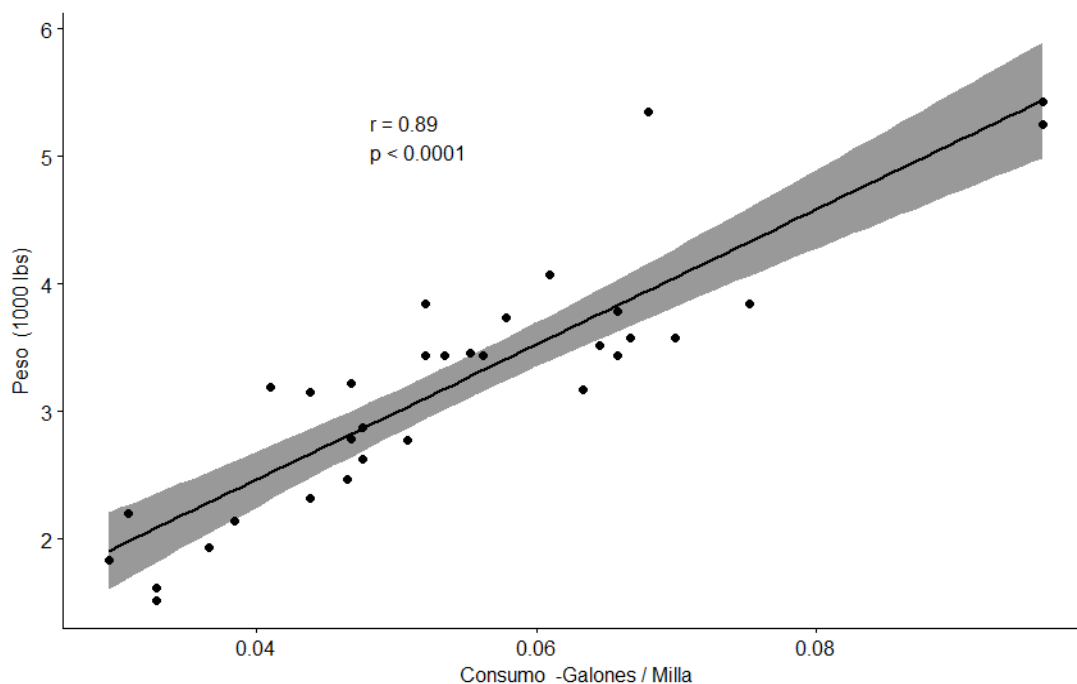
Son dos variables cuantitativas continuas.

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.

PRACTICANDO EL CONTRASTE DE HIPÓTESIS



Fíjate que tienes el resultado del p-valor.

¡Ahora ya puedes saber qué significa!

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: H1: Existe relación entre el consumo y el peso del coche
Define la H0 o hipótesis nula: H0: No existe relación entre el consumo y el peso del coche
Umbral de contraste (alpha): 5% = 0.05
Técnica Estadística: Coeficiente de correlación de Pearson
Significación o p-valor: p-valor < 0.0001
Respuesta: Se acepta la H1: Existe una fuerte correlación positiva entre el consumo y el peso del coche

PRACTICANDO EL CONTRASTE DE HIPÓTESIS

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

Analizando el gráfico de dispersión y los resultados estadísticos, podemos hacer varias afirmaciones:

- **Fuerte Correlación Positiva:** El coeficiente de correlación de Pearson ($r = 0.89$) indica una relación positiva muy fuerte entre el peso del coche y su consumo de combustible. Un valor tan alto sugiere que a medida que aumenta el peso del coche, también lo hace su consumo de combustible.
- **Significación Estadística:** Un p-valor menor a 0.0001 es extremadamente significativo lo que significa que es altamente improbable que la fuerte correlación observada sea debido al azar y por lo tanto, rechazamos la hipótesis nula (H_0) de no relación y aceptamos la hipótesis alternativa (H_1) que propone una relación significativa entre las dos variables.

En resumen, los datos no solo muestran una relación clara y significativa entre el peso y el consumo de combustible, sino que también ofrecen una base sólida para futuras investigaciones o decisiones en el diseño de vehículos, enfocándose en cómo el peso influye en la eficiencia del combustible.

¡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!