Capítulo 9

ANOVA INDEPENDIENTES

Utilizado para una variable independiente de tipo categórica

Centrado en la variabilidad de las muestras

Hipótesis:

H0: el estudio es igual para todos.
HA: el estudio es diferente para al menos uno.

CONDICIONES:

1. La escala de la variable dependiente es de intervalos iguales.
2. Las muestras son obtenidas de manera aleatoria e independiente desde población de origen.
3. Se puede suponer que la población de origen sigue una distribución normal.
4. Las muestras tienen varianzas aproximadamente iguales.

VERIFICACIÓN DE CONDICIONES:

1. Identificar si en cada valor, existe la misma diferencia con el siguiente.
2. Si una muestra no depende de una anterior, entonces son independientes
3. Realizar grafico q-q o test de Shapiro para cada muestra.
4. Prueba de homocedasticidad con Levene, la realiza ezANOVA

Prueba de LEVENE
H0: Las varianzas de las muestras son iguales.
HA: Al menos una de las muestras tiene varianza distinta.

Post Hoc:
Determinar donde se encuentran las diferencias usar Benfferonni, Holm o HSD Turkey.

Capítulo 10

ANOVA CORRELACIONADAS

Usado cuando las muestras están pareadas

Hipótesis:
Similar a la anterior

CONDICIONES:

1. La escala de la variable dependiente es de intervalos iguales.
2. Las mediciones son independientes al interior de cada grupo.
3. Se puede suponer que la población de origen sigue una distribución normal.
4. La matriz de varianzas-covarianzas es esférica las varianzas entre los diferentes niveles de las medidas repetidas deben ser iguales.

OBS: Toda magnitud física sigue una escala de intervalos

VERIFICACIÓN DE CONDICIONES:
Similar a la anterior, excepto por la 4.
4. Usar a prueba de esfericidad de Mauchly de ezANOVA()

Prueba de Mauchly
H0: Las varianzas de las muestras son iguales.
HA: Al menos una de las muestras tiene varianza distinta.

Capítulo 11

INFERENCIA NO PARAMÉTRICA CON MEDIANAS

Cuando no se cumplen los supuestos para utilizar las pruebas anteriores, es decir, la escala no es por intervalos y las variables no siguen una distribución normal.

PRUEBA PARA UNA O DOS MUESTRAS:
a. Prueba de suma de rangos de Wilcoxon
Usada con muestras independientes

CONDICIONES:
1. Las observaciones de ambas muestras son independientes.
2. La escala de medición empleada debe ser a lo menos ordinal, de modo que tenga sentido hablar de relaciones de orden ("igual que", "menor que", "mayor o igual que").

Hipótesis:
H0: No hay diferencias entre...
HA: Hay diferencias entre...

b. Prueba de rangos con signo de Wilcoxon:
Usado con muestras pareadas.

CONDICIONES:
1. Los pares de observaciones son independientes.
2. La escala de medición empleada para las observaciones es intrínsecamente continua.
3. La escala de medición empleada para ambas muestras debe ser a lo menos ordinal.

PRUEBAS PARA MÁS DE DOS MUESTRAS

a. Prueba de Kruskal-Wallis
CONDICIONES:

1. La variable independiente debe tener a lo mas de 2.

2. La escala de la variable dependiente debe ser, a lo menos, ordinal. 3. Las observaciones son independientes entre sí.

OBS: es una prueba de tipo ómnibus, por lo que se necesita hacer post hoc cuando se detectan diferencias.

b. Prueba de Friedman
CONDICIONES:
1. La variable independiente debe ser categórica y tener a lo menos tres niveles.

2. La escala de la variable dependiente debe ser, a lo menos, ordinal.
3. Los sujetos son una muestra aleatoria e independiente de la población.

Capítulo 12

REMUESTREO

BOOTSTRAPPING

Genera una mayor cantidad de muestras a partir de la original.

OBS: Para inferir sobre un valor nulo, se debe desplazar la distribución de Bootstrap

a. Para una muestra independiente

Hipótesis: Cuando se da un valor nulo.
H0: $\mu = 75$ [ms]
HA: $\mu > 75$ [ms]

b. Para dos muestras independientes
Hipótesis: Cuando se da un valor nulo.
H0: $\mu h - \mu m = 1,5$
HA: $\mu h - \mu m \neq 1,5$

MÉTODO

1. Fijar la cantidad B de repeticiones bootstrap.
2. En cada repetición, hacer un remuestreo con reposición de tamaño nA a partir de la muestra A y otro de tamaño nB a partir de la muestra B.
3. En cada repetición, calcular el estadístico de interés para generar la distribución bootstrap.
4. Construir el intervalo de confianza para el estadístico de interés.

OBS: Para muestras pareadas, el procedimiento es similar

PRUEBA DE PERMUTACIONES

Método de Montecarlo

1. Formular las hipótesis a contrastar (e identificar el estadístico de interés θ).
2. Crear una gran cantidad P de permutaciones (generalmente terminada en 9 para simplificar los cómputos) a partir de las muestras originales, usando muestreo sin reposición sobre la muestra combinada, y obtener el estadístico θ para cada una de las muestras.
3. Generar la distribución que el estadístico θ tendría si la hipótesis nula fuese cierta.
4. Determinar la probabilidad de encontrar un valor de θ al menos tan extremo como el observado en la distribución generada.
5. Comparar variable continua en dos muestras indep.

Capítulo 13

REGRESIÓN LINEAL SIMPLE

$y = \beta 0 + \beta 1x$

CORRELACIÓN

Fuerza de una relación lineal, esta es directa cuando $R > 0$ o inversa cuando $R < 0$.

Regresión lineal mínimos cuadrados:

CONDICIONES:

1. Los datos deben presentar una relación lineal.
2. La distribución de los residuos debe ser cercana a la normal.
3. La variabilidad de los puntos en torno a la línea de mínimos cuadrados debe ser aproximadamente constante.
4. Las observaciones deben ser independientes entre sí.

UN BUEN MODELO DE RLS

1. Tiene un gráfico en que los residuos se distribuyen aleatoriamente en torno a la línea de valor 0, sugiere que es razonable suponer que las variables presentan una relación lineal.
2. Cuando los residuos forman una banda horizontal en torno a la línea de valor 0, sugiere una variabilidad aproximadamente constante de los residuos.
3. La ausencia de residuos que se alejen del patrón que forman los demás sugiere la ausencia de valores atípicos.

INTERPRETACIÓN REGRESIÓN LINEAL:

La pendiente explica la diferencia esperada en el valor de la respuesta si el predictor x se incrementa en una unidad.

la intercepción corresponde a la respuesta que se obtendría en promedio si x fuese igual a 0.

OBS: Si la regresión usa un predictor categórico, este debe ser pasado a una variable numérica.
OBS: Luego de generar un modelo, este debe ser evaluado.

Capítulo 14

REGRESIÓN LINEAL MULTIPLE

CONDICIONES:

1. Las variables predictoras deben ser cuantitativas o dicotómicas (1 Y 0)
2. La variable de respuesta debe ser cuantitativa y continua, sin restricciones para su variabilidad.
3. Los predictores deben tener algún grado de variabilidad (su varianza no debe ser igual a cero).
4. No debe existir multicolinealidad.
5. Los residuos deben ser homocedásticos (con varianzas similares) para cada nivel de los predictores.
6. Los residuos deben seguir una distribución cercana a la normal centrada en cero.
7. Los valores de la variable de respuesta son independientes entre sí.
8. Cada predictor se relaciona linealmente con la variable de respuesta.

OBS: Cuando hay una variable categórica, esta debe ser transformada con dummy.

EVALUACIÓN DEL AJUSTE DE UNA RLM

AIC: un menor valor, indica un mejor modelo
ANOVA: usado para comparar si un modelo genera mejores modelos que otro

EVALUACIÓN RLM

Valores con influencia

Residuo estandarizado: 95 % de ellos se encuentre entre -1,96 y 1,96, y el 99 % entre -2,58 y 2,58. 2.

Valor predicho ajustado: corresponde al valor predicho si se excluyera dicho punto en el ajuste del modelo.

Residuo estudiantizado: indica cuánto influye la presencia de un punto en el conjunto de entrenamiento en su valor predicho

Diferencia en ajuste: más conocido como DFFit, es la diferencia entre el valor predicho para la observación evaluada cuando esta es considerada en el ajuste del modelo y cuando no lo es.

Diferencia en betas: más conocido como DFBeta, SON PREOCUPANTES observaciones en que este estimador es mayor a 1.

Distancia de Cook: cuando son mayor a 1 pueden ser considerados como potencialmente problemáticos.

Apalancamiento: Un apalancamiento igual a 0 señala que un punto no ejerce influencia alguna, mientras que un valor de 1 indica que la influencia ejercida por esa observación es total. Se consideran preocupantes aquellas observaciones para las cuales esta medida supere en dos o tres veces el apalancamiento promedio

Razón de covarianza: Aquellas observaciones para las cuales el valor de esta medida esté fuera del intervalo definido se consideran preocupantes.

VERIFICACIÓN CONDICIONES:

a. Independencia de los residuos:
Con Durbin-Watson:
H0: los residuos son independientes
Ha: los residuos están correlacionados

b. Distribución normal de los residuos:
Usando shapiro.test

c. Homocedasticidad de los residuos:
Usando ncvTest(model)
Ho: las varianzas son iguales
Ha: son distintas

d. Multicolinealidad
VIF : >5 problematicos
Tolerancia (1/Vif): < 0.2 problematicos

Capitulo 15

REGRESIÓN LOGÍSTICA

modelo lineal generalizado, que admite una variable de respuesta cuyos residuos sigan una distribución diferente a la normal.