

Como aprobar la Mierda de R

1. Abre la mierda del programa, paquete, cargar paquete, Rcmdr

Abierto el programa metemos los datos, miramos si los decimales son en punto o coma, y así lo determinamos. Le damos a datos, importar datos, desde archivo de texto etc.

Hacer

Realizar un histograma de ambos conjuntos de datos. Comentar las características más relevantes de ambos gráficos.

1. Se hace un histograma, y luego el otro, damos en gráficas, histogramas, seleccionamos una u otra y le damos aceptar.

Tenemos que comentar, unimodal, bimodal, asimetría a la izquierda o a la derecha, presenta o no atípicos

Hacer

Realizar, en gráficos separados, un diagrama de caja y bigotes para cada variable e identificar cada una de las líneas que lo constituyen, así como los valores numéricos correspondientes. ¿Existen datos atípicos? ¿Cuáles serían los valores admisibles entre los que se encontrarían los datos no atípicos para cada una de las dos variables?

Tomamos estadísticos, resúmenes, conjunto de datos activo y luego, mismo procedimiento hasta resúmenes numéricos.

Después hacemos el diagrama de caja y bigotes, le damos a gráficas, diagrama de caja, seleccionamos una variable u otra y con los resúmenes numéricos de antes, ponemos la media, RIQ, Xmax y mín, Q₁ y Q₃. Si en el diagrama de caja y bigotes salen atípicos, pondrá un n° al lado y los tenemos que visualizar en visualizar conjunto de datos  y ver qué valor tiene el n°.

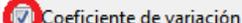
Para ver los valores admisibles en los que se encontrarían los datos NO ATÍPICOS, hacemos los límites

$$L_{\text{inferior}} = Q_1 - 1,5RIQ; L_{\text{superior}} = Q_3 + 1,5RIQ$$

A partir de los resultados obtenidos en los apartados anteriores, ¿qué medidas de centralización y dispersión consideras más adecuadas para resumir cada conjunto de datos?

Si es disperso, mediana y rango intercuartílico, si es centralizado media y desviación típica

¿Podemos asumir que la dispersión es similar en ambas variables? Comentar los resultados que consideres más relevantes así como los estadísticos que ha utilizado para dar respuesta a la cuestión anterior.

Para calcular el coeficiente de variación C_v, le damos a estadísticos, resúmenes, resúmenes numéricos, y le damos  a estadísticos y seleccionamos  y miramos cual tiene menor dispersión relativa. Lo hacemos para una y otra variable.

Si nos piden qué modelo de distribución de probabilidad continua te parece más adecuado para describir el comportamiento de la variable aleatoria?

Si la gráfica tiene forma de campana de gauss, **distribución normal**, si tiene una asimetría muy pronunciada hacia uno de los lados con valores muy pequeños al principio y muy grandes al final, **exponencial**.

Calcular el parámetro de distribución para modelos exponenciales

$$\lambda = \frac{1}{\text{mean}} = \text{lo que dé}$$

Si es para un modelo de distribución normal es directamente la media

$$\mu = \text{mean} = \text{lo que dé}$$

Calcular intervalos en una exponencial dada λ

Le damos a distribuciones, distribuciones continuas, distribución exponencial, probabilidades exponenciales, metemos los valores de las variables, y el parámetro exponencial λ , el que nos den, tenemos que saber para donde va el signo, si $<$ cola a la izquierda debemos seleccionar, si $>$ cola a la derecha. TENEMOS QUE RESTAR LOS VALORES!! restar

En el caso de que solo sea un único valor, se pone esa probabilidad.

Proporcionar un intervalo de confianza e indicar la distribución de probabilidad utilizada para construir dicho intervalo

Damos en estadísticos, medias, test t para una muestra, se selecciona el nivel de confianza por el valor que den (0.9, 0.98, el que sea), el que aparece señalado es el intervalo.

```
One Sample t-test
data: X
t = 7.2795, df = 33, p-value = 2.367e-08
alternative hypothesis: true mean is not equal to 0
98 percent confidence interval:
 24.30994 48.89594
sample estimates:
mean of x
 36.60294
```

Contrastes de hipótesis con confianza dada y valor promedio que no puede superar, comentar el p-valor y decir si parece adecuado el modelo

Damos en estadísticos, medias, test t para una muestra, se selecciona el nivel de confianza que nos den, y debemos poner en la hipótesis alternativa, dependiendo del signo $<$ ó $>$ si la media poblacional $<\mu_0$ ó media poblacional $>\mu_0$

Tenemos que observar si el p-valor es inferior a 1- el nivel de confianza, si es menor rechazamos H_0 con una gran confianza (en nosotros mismos) de lo que se deduce (si se ha hecho un contraste para inferior o superior) que es significativamente menor o mayor a lo que la hipótesis contrastó.

Representar los datos observados por ambas características mediante el diagrama de dispersión correspondiente, comentar si es adecuado o no y justificar

Le damos a gráficas, gráficas XY, seleccionados las variables explicativas (la x) y las explicadas (la y), debemos observar si hay correlación entre ellas, si la dispersión tiene alguna tendencia predecible.

Realizar un ajuste por mínimos cuadrados con el fin de predecir el comportamiento de la variable a partir de sus datos. Indicar la ecuación del modelo propuesto y dar una medida de bondad de ajuste justificando el mismo

Damos en estadísticos, ajuste de modelos, regresión lineal, seleccionamos la variable explicada y explicativa, y tomamos los siguientes datos.

```
lm(formula = Y ~ X, data = Datos)
Residuals:
    Min       1Q   Median       3Q      Max
-0.076516 -0.020012 -0.004806  0.027094  0.073747

Coefficients:
(Intercept)  1.000e+02  1.089e-02  9184.91  < 2e-16 ***
X            -2.903e-03  2.335e-04  -12.44  8.54e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03933 on 32 degrees of freedom
Multiple R-squared:  0.8285, Adjusted R-squared:  0.8232
F-statistic: 154.6 on 1 and 32 DF, p-value: 8.538e-14
```

$$y = 100 - 0,002903 \cdot x$$
$$R^2 = 0,8285 \sim 0,83 \text{ ajuste bueno}$$

Obtener predicciones para un modelo dado un determinado valor, decir si es fiable y razonarlo

Se sustituye el valor dado en la ecuación, como $\hat{y} = 100 - 0,002903 \cdot \boxed{60} = 99,82$

Dado un determinado R^2 si es bueno, es fiable, y si este valor pertenece al rango de las observaciones realizadas.