PRACTICAS DE ESTADISTICA: R COMANDER

1.Practica- Datos 🡪 Nuevo conjunto de datos🡪 darle nombre al conjunto🡪introducir variables( indicando si son variables o números) Conjuntos de datos activo🡪 variables las q sean. Modificar variables del conjunto de datos activo🡪 calcular nueva variable Para guardar Fichero 🡪 guardar instrucciones como…(pen)🡪 archivo de salida

2. PRACTICA 2 TABLA DE VARIABLES: R🡪Estadisticos 🡪 resúmenes 🡪 distribución de frecuencias 🡪 se señalan las variables (ctrl+) GRAFICO DE PORCENTAJES: Graficas 🡪grafica de sectores 🡪 se señala la variable 🡪 los porcentajes aparecen en gf DIAFRAMA DE BARRAS: Graficas🡪 grafica de barras 🡪 variable a elegir CONVERSION DE VARIABLES : Datos🡪 modificar variables de conjunto activo 🡪 convertir variable numérica en factor PARA UNA NUEVA VARIABLE: Datos🡪 modificar variables del conjunto de datos activo🡪 calcular nueva variable🡪le damso nombre a la variable . Ejemplo nueva variable (dif­\_ev)🡪expresion a calcular ( espidaf-espevidam) Ejercicio: Agrupar en 5 intervalos de igual amplitud, comenzando en 1500 y acabándolo en 4000 Amplitud de cada intervalo = (4000/1500)/5 = 500 🡪 para AGRUPAR: Graficas 🡪 histograma 🡪 elegir variable y para los intervalos en la ventana de las instrucciones indicamos breaks=c (1500,2000,2500,3000,3500,4000)🡪 se ejecuta a continuación. Para obtener una tabla de frecuencias que refleje la agrupación que hemos elegido, recodificamos la variable: Datos 🡪 modificar variables del conjunto de datos activo🡪 recodificar variables ( o segmentar variable numérica, pero esta opción no indica el valor inicial de segmentación) MEDIDAS DESCRIPTIVAS: Estadísticos 🡪 resúmenes 🡪 conjunto de datos activo, con esto tendremos el resumen del conjunto de datos activo con las siguientes variables: variables cuantitativas: ( mínimo, primer cuartil, mediana (=percentil 50), media, tercer cuartil, máximo y nº de datos perdidos) y variables cualitativas ( nombre de las modalidades de la variable con su frecuencia absoluta y además el nº de datos perdidos). En R destacar que a la desviación típica se le corresponde con la cuasidesviacion típica y a la varianza con la cuasivarianza. Ejercicio: Calcular e interpretar sobre la variable “calorías” las siguientes medidas descriptivas: percentil 30, media, d, típica, asimetría y apuntamiento. El percentol 30 = cuartil 30 = 0,3 = C0,3 / mean = media/ skewness = asimetría / Kurtosis = apuntamiento / Me= mediana / Para calcular la moda ( valor que mas veces se repite), convertir “calorías” a factor (factorcalorias) 🡪 Modificar variables del conjunto de datos activo 🡪 convertir variable numérica en factor para hallar el coeficiente de variación (CV) = cuasidesviacion típica / media : CV<10% ( media representativa) : 10%< CV< 30% ( media moderamente representativa) CV > 30% ( media poco representativa)

3. PRACTICA 3

Para ver si dos variables numéricas están relacionadas calculamos el coeficiente de correlación lineal de Pearson y dibujamos la nube de puntos en base al valor resultante ( -1 < r < 1 ) 🡪 estadísticos 🡪 resúmenes 🡪 matriz de correlaciones, y para la grafica XY ( nube de puntos 🡪 Graficas 🡪 grafica XY. Por otra parte, para calcular los resúmenes numéricos, pasamos a Estadísticos 🡪 Resúmenes 🡪 resúmenes numéricos 🡪 resumir por grupos ( por ejemplo, si pide calcular la media, la desviación típica, asimetría). Lo mismo pasa con los histogramas por grupos. Para calcular calcular las medidas descriptivas de variables numéricas según una o mas variables cualitativas o factores, elegimos Estadísticos🡪 Resúmenes 🡪 tabla de estadísticas ( si vemos que hay alguna variable que presenta muchos valores, podemos visualizar dicha información mas resumido en graficas 🡪 grafica de medias 🡪 sin barras de errores ( en factores seleccionamos la variable que convenga) Para el estudio conjunto de dos o mas variables cualitativas se utiliza una tabla de contingencia: Estadísticos 🡪 tablas de contingencia 🡪 tablas de doble entrada 🡪 seleccionamos la opción sin porcentajes y demarcamos la opción Test de Chi-cuadrado. Si queremos ver como se distribuye alguna variable en base a las otras dos variables ( empleados en base a nivel educativo dentro de cada categoría laboral), seleccionamos la opción de porcentajes por columnas. Análogamente, se puede estudiar como se distribuye la categoría laboral según el nivel educativo seleccionado Porcentaje por filas

4. PRACTICA 4 (RESOLUCION DE PROBLEMAS DE PROGRAMACON LINEAL) La función de lp del paquete IpSolve permite resolver problemas de programación lineal. Una vez instalado el paquete, lo cargamos(library(IpSolve)). Las instrucciones a seguir pueden ser: “min o max” si el problema es de minimizar o maximizar, vector de coeficiente de coste, un vector que recoja las direcciones de las restricciones “<=”, “=” o “>=”, entre otras. Ejemplo en una garanja se crían conejos y pollos, pero se desea que no se superen los 180 horas mensuales destinadas a esta actividad …………… Sean x=r de conejos e y=nº de pollos, queremos maximizar la función: f(x,y)= 3x+1,8y sujeta a las siguientes restricciones: 20x+10y=1000 y 3x+2y= 180

5. PRACTICA 5 ( DISTRIBUCIONES DE VARIABLES ALEATORIAS)

Debemos identificar (si no nos lo da) de que distribución se trata: Binomial, exponencial, poisson o geométricas. En los menus Distribuciones🡪 Distribuciones continuas/ distribuciones discretas, donde hallamos un conjunto de opciones que permiten trabajar con estos modelos. Para cada modelo discreto y continuo aparecen una serie de opciones: Cuantiles probabilidades acumuladas (P(X<x) y (P> x)), probabilidades (calcula la probabilidad de los valores que toma la variable) , grafica de la distribución ( de probabilidad o de distribución del modelo considerado). También tendremos que seleccionar cola derecha o cola izquierda. Ejemplo: la demanda de un articulo sigue la distribución normal, X-N (100,20), cde cuantas unidades debemos disponer al inicio de semana para satisfacer la demanda con una probabilidad de 0,99 🡪 Solución: nos pide el valor de “t”, tal que P(x=t)= 0,99🡪 distribuciones 🡪 distribuciones continuas 🡪 distribuciones normales 🡪 cuantiles normales 🡪 probabilidad (0,99) / media=100……….

6. PRACTICA 6 Aquí solamente se comprueban gráficamente los resultados teóricos del tema 4.

7. PRACTICA 7: INTERVALOS DE CONFIANZA Y CONTRASTES DE HIPOTESIS EN POBLACIONES NORMALES

* Se rechaza la hipótesis nula Ho si el p-value obtenido es menor o igual que a (significación)
* Para el estudio de la normalidad de los datos usamos el Test de Shapiro-wilk (estadísticos🡪resumenes🡪test de normalidad) ( siempre que las variables bajo estudio siguen distribuciones nromales

CH e I C para la media y para la varianza de una población normal se obtiene a través de estadísticos🡪 medias 🡪 test para una ,muestra y a través de Estadísticos 🡪 Varianzas 🡪 test de varianza para una muestra. Ejemplo: una empresa debe producir diariamente 800tn de un producto. El gerente sospecha que se esta produciendo menos de esa cantidad y por eso se observa la producción de los últimos 5 dias con estos datos: 802, 795, 753, 810 y 783. ¿hay evidencia para afirmar al 5% de significación, que la sospecha del gerente es correcta? Construir intervalos de confianza al 95% para la media y para la varianza poblacional. Tenemos……Para plasmar esta información , seleccionamos Estsdisticos 🡪 medias 🡪 test para una muestra 🡪 elegimos la variable, marcamos como hipótesis alternativa la opción media poblacional y escribimos la hipótesis nula: mu=800. Obtenemos p.valor=0.1528 al ser mayor que 0,05 nos lleva a q no se rechaza Ho y por tanto no podemos afirmar que se rechaze

El intervalo obtenido para la media es unilateral, como queremos un intervalo bilateral, tenemos que repetir el procedimiento marcando como hipótesis alternativa la opción media poblacional!=muo, así obtenemos el nuevo intervalo. De esta forma, para calcular un intervalo de confianza para la varianza poblacional al 95%, seleccionamos Estadísticos 🡪 Varianzas 🡪 Test de varianza para una muestra. En el caso de que tengamos muestras relacionadas o apareadas ( 2 columnas) seguimos Esrtadisticos 🡪 medias 🡪 test para datos relacionados: una vez ahí podremos seleccionar las variables que tengamos en 1 y 2 variable

8. PRACTICA 8: INTERVALOS DE CONFIANZA Y CONTRATES DE HIPOTESIS EN POBLACIONES NORMALES (II)

Aquí tenemos el segundo caso, el de dos poblaciones normales e independientes. Concretamente, trataremos sobre la diferencia de las medias de dichas poblaciones. Si las varianzas poblacionales son desconocidas, primero tendremos que contastar la igualdad de las mismas para ello: Estadísticos🡪varianzas🡪test F para 2 varianzas. A continuación elegimos la variable explicada y elegiremos la hipótesis alternativa bilateral. Una vez obtenido el resultado de si se rechaza o no, vemos el contraste sobre la diferencia de las medias en Estadísticos🡪medias🡪test I para muestras independientes. Por otra parte, para el análisis de la varianza de un factor🡪 estadísticos🡪medias🡪 ANOVA de un factor. Para comprobaciones usamos el test de tukey