Apuntes practicas estadistica

FRECUENCIAS
Estadísticos → Resúmenes → Distribución de frecuencias

mediana (separa el 50% datos)
desviación típica (como de dispersos están los datos con respecto de la media)
coeficiente de variación(homogéneo +grande o heterogéneo +chico, datos en %)
Estadísticos → Resúmenes → Resúmenes numéricos…

DISTRIBUCIÓN DE FRECUENCIA absolutas y relativas
Estadísticos → Resúmenes → Distribución de frecuencias…

DIAGRAMA de barras y sectores (solo variables cualitativas de tipo factor)

HISTOGRAMA (representa los intervalos de frecuencias)
Gráficas → Histograma…

DIAGRAMA DE CAJA(detección de valores atípicos)

DISTRIBUCIÓN CONDICIONADA
Estadísticos → Tablas de contingencia → Tabla de doble entrada…

FILTRADO DE DATOS
Datos → Conjunto de datos activo → Filtrar el conjunto de datos activo…)

diagrama de DISPERSIÓN(relación de dos variables)
Gráficas → Diagrama de dispersión…

CORRELACIÓN de pearson o lineal(cuantifica la dependencia de dos variables)
Estadísticos → Resúmenes → Matriz de correlaciones…

ANALISIS DE REGRESION

COVARIANZA(grado de variación con respecto a la media, relación lineal entre 2 variables)
cov(datos[,c("Altura","Edad","Peso")], use="complete")

RELACIÓN LINEAL(si en la tabla tiene signo negativo inversa, es. positivo directa)
cor(datosc[,c("Altura","Edad","Peso")], use="complete")

REGRESIÓN LINEAL(relación entre variables, dependiente e independiente)***
Estadísticos → Ajuste de modelos → Regresión lineal…

RECTA DE REGRESIÓN

RegModel.2 <- lm(Peso~Altura, data=datosc, subset=SexoF=="Hombre")
summary(RegModel.2)

PREDICCIONES(valores de la recta de regresión para un valor dado)
predict(modelo,data.frame(Var Ind=valor))
\\lm(formula = Peso ~ Altura, data = Dataset)
predict(Lineal,data.frame(Altura=170))

BONDADES de ajuste(fiabilidada de las predicciones dadas)
es el r cuadrado de los analisis de regresion

MODELO REGRESIÓN(relaciones entre variables)
Estadísticos → Ajuste de modelos→ Modelo lineal…

Predicciones(de un modelo lineal)
predict(modelo,data.frame(Var Ind=valor))

Bondad del ajuste(es el r cuadrado) -> cuanto más se ajuste al 1 y al -1
mejor se ajusta al modelo
+++
+TIPOS DE MODELOS

MODELO MÚLTIPLE -> depende varias variables --> y = b0 + b1x1 + b2x2
\\ estadística --> ajuste de modelo --> modelo lineal (peso~altura+edad)

MODELO EXPONENCIAL -> y = exp(a + bx) --> log(y) = a + bx
*log es logaritmo neperiano
\\ Exponencial <- lm(log(Peso) ~ Altura, data=Dataset)
summary(Exponencial)

prediccion --> predict(Exponencial,data.frame(Altura=170))
exp(predict(Exponencial,data.frame(Altura=170)))

MODELO LOGARÍTMICO -> Y = a + b*log(x)
\\Logaritmo <- lm(Peso ~ log(Altura), data=Dataset)
summary(Logaritmo)

predict(Logaritmo,data.frame(Altura=170))

MODELO MULTIPLICATIVO -> y = a*x^b --> log(y) = log(a) + b*log(x)
\\Moltiplicativo <- lm(log(Peso) ~ log(Altura), data=Dataset)
summary(Multiplicativo)

exp(predict(Multiplicativo,data.frame(Altura=170)))

MODELO INVERSO -> y = a + b/x
\\Inverso <- lm(Peso ~ 1/Altura, data=Dataset)
summary(Inverso)

INTERVALOS Y CONTRASTES DE HIPÓTESIS

INTERVALOS Y CONTRASTES SOBRE LA MEDIA

Estadísticos → Medias → Test t para una muestra

INTERVALOS Y CONTRASTES SOBRE LA PROPORCIÓN

test t para una muestra: hacer preguntas sobre la media para que sea mayor o menor, todo intervalo tiene que cumplir esta característica

// altura (95% confianza)

// 95 percent confidence interval

//171.1291 175.3209

-para un subconjunto, primero se tiene que seleccionar y guardar como un conjunto nuevo

-para dos variables: altura media hombres y mujeres (x -> h, y -> m)

-test t para muestras indp.: para diferencia de medias

1º saber si las varianzas son iguales (si el intervalo contiene el 1):

- estadísticos, varianza, test f para varianzas

\\sexo F -> altura nivel confianza 95

ratio of variances 3.37588

cociente entre cuasivarianzas

95 percent confidence interval:

1.171317 8.357349 -> 1 no está dentro, así que no consideramos iguales

2º altura media 4.218553-10.539689 -> la altura media de los hombres es mayor que el de las mujeres

INTERVALO CONFIANZA SOBRE LA PROPORCIÓN

si la variable es una binomial -> es el número de veces que ocurre el suceso entre n

proporción de hombres?

estadísticos -> proporciones para una muestra

DESVIACIÓN ESTÁNDAR

tengo una variable con una distribución normal ->

1º veo cuantos datos tengo (n-1) n = 40

2º sd -> standard deviation

ei varianza=39*(sd(datosc$Altura)^2)/qchisq(0.975,39)-> cuantil de la chi cuadrado

-> probabilidad que deja a la izquierda, extremo izquierdo

es varianza=39*(sd(datos$Altura)^2)/qchisq(0.025,39)

-> deja a la derecha

DISTRIBUCIONES

	Densidad o masa	Distribución	Cuantil	Muestras aleatorias
Binomial	dbinom(x,size,prob)	pbinom(q,size,prob, lower.tail=TRUE)	qbinom(p,size,prob, lower.tail=TRUE)	rbinom(n,size,prob)
Poisson	dpois(x,lambda)	ppois(q, lambda,lower.tail=TRUE)	qpois(p, lambda, lower.tail=TRUE)	rpois(n, lambda)
Normal	dnorm(x,mean=0,sd=1)	pnorm(q, mean=0,sd=1, lower.tail=TRUE)	qnorm(p, mean=0,sd=1, lower.tail=TRUE)	rnorm(n, mean=0,sd=1)
Chi-cuadrado	dchisq(x,df)	pchisq(q,df, lower.tail=TRUE)	qchisq(p,df, lower.tail=TRUE)	rchisq(n,df)
t	dt(x,df)	pt(q,df, lower.tail=TRUE)	qt(p,df, lower.tail=TRUE)	rt(n,df)
F	df(x,df1,df2)	pf(q,df1,df2,lower.tail=TRUE)	qf(p,df1,df2, lower.tail=TRUE)	rf(n,df1,df2)

PROBABILIDADES
calcular:
a) p[X = 3] = [1] 0.2502823
// dist - discretas- binomial - prob bin.

para un solo valor
> dbinom(x,10,0.25) x-> es el valor

b)P [X <= 5] = [1] 0.9802723
> pbinom(c(5), size=10, prob=0.25, lower.tail=TRUE)

// dist - discretas- binomial - prob bin. acum- 5,10,0.25 cola izq

c)P [X > 4] = [1] 0.07812691
> pbinom(c(4), size=10, prob=0.25, lower.tail=FALSE

d)P [2 <= X <= 5] = p[x <= 5] - p[x <= 1] = [1] 0.7362471 ->
pbinom(c(5), size=10, prob=0.25, lower.tail=TRUE) -pbinom(c(1), size=10, prob=0.25, lower.tail=TRUE)

CUANTIL BINOMIAL
e) Determinar x, para que P[X <= x] = 0.75 = 3 ->
qbinom(c(0.75), size=10, prob=0.25, lower.tail=TRUE)
// cuantiles binomiales

MEDIANA DISTRIBUCIÓN
f) Determinar la mediana de la distribución = 2 -->
qbinom(c(0.5), size=10, prob=0.25, lower.tail=TRUE)

FUNCION MASA PROBABILIDAD (probabilidades de que ocurra cada caso)
g) Dibujar la función masa de probabilidad y la función de distribución de dicha distribución.
// binomial - graf distribución

h) Generar una muestra aleatoria de tamaño 500 de los datos, obtener la mediana de esta muestra aleatoria y comparar con la mediana obtenida de forma teórica.--> datos 500=rbinom(500,10,0.25)
datos 500 --> para verlos

| CONTRASTE DE LA HIPÓTESIS |
 ----

1. QUIENES SON LAS HIPÓTESIS
    - NULA (DONDE ESTA LA IGUALDAD // =, =>, =<) ej mu = 5
    -ALTERNATIVA (desigual, >, <) ej mu > 5

2. ESTADÍSTICO
    T=(MEDIA -MU/S/raíz n) t-student

3. CONSTRUIR LA REGIÓN CRÍTICA -> alfa medios
    se calcula una probabilidad, la prob. de que ese valor experimental pertenezca a esa región crítica

4.DECISIÓN
    Si p value < alfa -> se rechaza H0
    Si p value > alfa -> se acepta H0

~~~~~~

Altura = x distri. normal

- hip = media de la altura > 175
  hip. alternativa = media de la altura > 175
    -estadísticos -> medias -> test t para una muestra
        -med prob < mu0
        -cuadro 1 175
        -cuadro 2 .90

~~~~~~

ej 1. data: Altura
    t = 7.0444, df = 535, p-value = 1
    alternative hypothesis: true mean is less than 175
    90 percent confidence interval:
     -Inf 177.7459
    sample estimates:
    mean of x -> 177.3228

p-value = 1 --> 1 > 0.10, nos quedamos con la hipótesis nula, podemos decir con un error del 10% que la media es menor que 175

----

ej 2. porcentaje de hombres que estudian informática es superior al 80%, error = 0.01
X = nº hombres del total de la muestra (536) ---> BINOMIAL p desconocida

H0 = prop. <= 0.8
H1 = prop > 0.8

test prop. para una muestra
    prop. > p0
    niv. confianza = .99
    hip nula p = .8

p value < 0.01
p = 4,6 x 10^-9 < 0.01 rechazamos H0

----

ej 3. Consideramos dos variables
X -> peso hombres
Y -> peso mujeres DISTR. NORMAL

error 0.02
se quiere contrastar si no existen diferencias entre el peso de las mujeres y los hombres
H0 = media X = media Y
HI = media X distinto media Y

estadistica media test t muestras indep.
    est var test f
1. podemos resolver que las varianzas son iguales? (contraste o intervalo)
    sexo - peso
    bilateral
    niv. conf = .98

p-value = 0.7293 > 0.02 --> se acepta la hipótesis nula, las podemos considerar iguales

est medias test t muestr indp.
sexo- peso
niv. conf 0.98
si muestras independientes
bilateral

p-value = 2.665e-14 < 0.02 --> las medias son distintas