

Análisis descriptivo e inferencia básica

ESTADISTICA DESCRIPTIVA

Con los datos del fichero ElPulso.rda calcular:

#a) Cálculo de estadísticos
summary(DatosPulso)

#1) Media de Pulso1
mean(DatosPulso\$Pulso1, na.rm=TRUE) # orden directa del R
numSummary(DatosPulso[, "Pulso1"], statistics=c("mean"), quantiles=c(0,.25,.5,.75,1)) # ventanas de R-cmdr

#2) Mediana de Pulso1
summary(DatosPulso) # para todas las variables
median(DatosPulso\$Pulso1, na.rm=TRUE) # orden directa del R

#3) Desviación típica de Pulso1
numSummary(DatosPulso[, "Pulso1"], statistics=c("sd"), quantiles=c(0,.25,.5,.75,1)) # ojo esta es la distribución muestral (n-1)

$\sqrt{\text{var}(\text{DatosPulso\$Pulso1}) * (\text{length}(\text{DatosPulso\$Pulso1}) - 1) / \text{length}(\text{DatosPulso\$Pulso1})}$ # la desviación transformada para que este dividida por n, comando en R

#también se pueden conseguir todas a la vez
numSummary(DatosPulso[, "Pulse1"], statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))

#4) Media de Peso.kg
mean(DatosPulso\$Peso.kg, na.rm=TRUE) # orden directa del R
numSummary(DatosPulso[, "Peso.kg"], statistics=c("mean"), quantiles=c(0,.25,.5,.75,1)) # ventanas de R-cmdr

#5) Mediana de Peso.kg
median(DatosPulso\$Peso) # orden directa del R

#6) Desviación típica de Peso.kg
numSummary(DatosPulso[, "Peso.kg"], statistics=c("sd"), quantiles=c(0,.25,.5,.75,1)) # ojo esta es la distribución muestral (n-1)
 $\sqrt{\text{var}(\text{DatosPulso\$Peso.kg}) * (\text{length}(\text{DatosPulso\$Peso.kg}) - 1) / \text{length}(\text{DatosPulso\$Peso.kg})}$ # la desviación transformada para que este dividida por n, comando en R

#-todos a la vez
numSummary(DatosPulso[, "Peso"], statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))

#7) Media de Altura.cm
mean(DatosPulso\$Altura.cm)
numSummary(DatosPulso[, "Altura.cm"], statistics=c("mean"), quantiles=c(0,.25,.5,.75,1)) # ventanas de R-cmdr

#8) Mediana de Altura

```
median(DatosPulso$Altura.cm)
```

#9) Desviación típica de Altura

```
numSummary(DatosPulso[, "Altura.cm"], statistics=c("sd"), quantiles=c(0,.25,.5,.75,1)) #ojo esta es la distribución muestral (n-1)
```

```
sqrt(var(DatosPulso$Altura.cm)*(length(DatosPulso$Altura.cm)-1)/length(DatosPulso$Altura.cm)) # la desviación transformada para que este dividida por n, comando en R
```

#Para todos

```
numSummary(DatosPulso[, "Altura.cm"], statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

#10) Media de la Altura.cm de las mujeres1

```
altura.cm.mujer <- subset(DatosPulso, subset=Sexo=='Mujer', select=c(Altura.cm)) # crea una nueva data solo con las mujeres y la variable Altura.cm
```

```
altura.cm.mujer
```

```
mean(altura.cm.mujer)
```

```
mujeres <- subset(DatosPulso, subset=Sexo=="Mujer") #Otra forma de hacerlo con todas las variables, pero luego hay que seleccionar "mujeres" como datos activos
```

```
mean(mujeres$Altura.cm)
```

#También que nos de el resumen numérico separado por sexo

```
numSummary(DatosPulso[, "Altura.cm"], groups=DatosPulso$Sexo, statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

#11) Mediana de la Altura de los fumadores

```
tapply(mujeres$Altura.cm, list(Fumar=mujeres$Fumar), median, na.rm=TRUE) # se obtiene la tabla de medianas para fumadores y no fumadores
```

#12) Desviación típica del Peso.kg de las mujeres.

```
tapply(DatosPulso$Peso.kg, list(Sexo=DatosPulso$Sexo), sd, na.rm=TRUE)
```

```
sd(mujeres$Peso.kg)
```

T-TEST E INTERVALOS DE CONFIANZA

#1) Calcular el intervalo de confianza para el peso.kg medio de # todos los individuos con (alfa = 0.05).

```
t.test(DatosPulso$Peso.kg, alternative='two.sided', mu=0.0, conf.level=.95) # 95 percent confidence interval:  
63.66709 68.13108
```

#2) Calcular el intervalo de confianza para el peso.kg medio de las #mujeres con (alfa = 0.05).

```
mujeres <- subset(DatosPulso, subset=Sexo=="Mujer") # Crear el filtro y luego el intervalo  
t.test(mujeres$Peso.kg, alternative='two.sided', mu=0.0, conf.level=.95) # 95 percent confidence interval:  
54.11977 58.29063
```

#3) Estudios recientes afirman que la altura media de las mujeres de #esta población es $\mu = 167$ cm. A la vista de estos datos, ¿podemos aceptar dicha hipótesis?

```
t.test(DatosPulso$Altura.cm, alternative='two.sided', mu=167, conf.level=.95) #
```

#4) Calcular el intervalo de confianza para el Pulso1 medio de las mujeres que no fuman.

```
mujeres.nofuma <- subset(DatosPulso, subset=Sexo=="Mujer"&Fumar=="No Fuma") # Primero se filtran los datos y se selecciona la base
```

```
t.test(mujeres.nofuma$Pulso1, alternative='two.sided', mu=0.0, conf.level=.95) # 95 percent confidence interval: 70.35997 78.82521
```

#5) Calcular el intervalo de confianza para la media del incremento del pulso (Pulso2-Pulso1) para los individuos que corrieron.

```
corrieron <- subset(DatosPulso, subset=Correr=="Si")
```

```
t.test(corrieron$difer.pulso, alternative='two.sided', mu=0.0, conf.level=.95) #95 percent confidence interval: 13.74454 24.08403
```

#Para datos pareados media de la diferencia.

```
t.test(DatosPulso$Pulso1, DatosPulso$Pulso2, alternative='two.sided', conf.level=.95, paired=TRUE)
```

#6) Existen diferencias en el pulso1 entre hombres y mujeres

```
t.test(Pulso1~Sexo, alternative='two.sided', conf.level=.95, var.equal=FALSE, data=DatosPulso)
```

#Con el argumento var.equal, estamos diciendo que las varianzas no #son iguales pero primero se puede probar si las varianzas son #iguales

```
tapply(DatosPulso$Pulso1, DatosPulso$Sexo, var, na.rm=TRUE)
```

```
var.test(Pulso1 ~ Sexo, alternative='two.sided', conf.level=.95, data=DatosPulso)
```

El p-valor es mayor que 0,05 por lo tanto no podemos rechazar #la hipótesis de que son iguales, entonces rehacemos el test, #colocando TRUE en el argumento

```
t.test(Pulso1~Sexo, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=DatosPulso)
```

TABLAS DE CONTINGENCIA Y MEDIDAS DE ASOCIACIÓN

Existe asociación entre la intención de voto y la educación

```
library(abind, pos=4)
```

```
.Table <- xtabs(~education+vote, data=Chile)
```

```
.Table
```

```
.Test <- chisq.test(.Table, correct=FALSE)
```

```
.Test
```

```
remove(.Test)
```

```
remove(.Table)
```

Se pueden pedir porcentajes por columnas

```
.Table <- xtabs(~education+vote, data=Chile)
```

```
.Table
```

```
colPercents(.Table) # Column Percentages
```

```
.Test <- chisq.test(.Table, correct=FALSE)
```

```
.Test
```

```
remove(.Test)
```

```
remove(.Table)
```

```

# Se pueden pedir porcentajes por filas y marcar todas las opciones
.Table <- xtabs(~education+vote, data=Chile)
.Table
rowPercents(.Table) # Row Percentages
.Test <- chisq.test(.Table, correct=FALSE)
.Test
.Test$expected # Expected Counts
round(.Test$residuals^2, 2) # Chi-square Components
remove(.Test)
fisher.test(.Table)
remove(.Table)

# También podemos evaluarlo para la región de Santiago
.Table <- xtabs(~education+vote, data=Chile, subset=region=="SA")
.Table
.Test <- chisq.test(.Table, correct=FALSE)
.Test
.Test$expected # Expected Counts

# tablas múltiples nos muestra la misma tabla para todas las regiones, si hacer el test

.Table <- xtabs(~education+vote+region, data=Chile)
.Table
remove(.Table)

#Existe asociación entre la variable ingreso, statusquo y edad??
cor(Chile[,c("age", "income", "statusquo")], use="complete.obs")
#luego se pueden Pruebas de hipótesis a pares para ver si es #significativamente diferente de cero.
cor.test(Chile$age, Chile$income, alternative="two.sided", method="pearson")
#Esto de todas formas se puede ver gráficamente, matríz de digrama de dispersión

```