

11. INSTRUCCIONES PARA USAR RSTUDIO

Antes de empezar, instalar los paquetes:

```
install.packages(c("mosaic", "ggformula", "dplyr", "PropCIs", "ThresholdROC", "pROC", "survival"))
```

Después, cargar los paquetes:

```
library(mosaic)
library(ggformula)
library(dplyr)
library(PropCIs)
library(ThresholdROC)
library(pROC)
library(survival)
```

PREPARAR BASE DE DATOS

Pasar variable a factor

```
base$variable <- factor(base$variable, levels = 1:0, labels = c("valor1", "valor0"))
```

Cambiar el primer nivel de un factor

```
base$variable <- relevel(base$variable, ref = "valor")
```

Filtrar datos

```
nuevabase <- filter(base, variable == "valor")
```

Calcular variable

```
base$variablenueva <- ifelse(base$variable operacion, "si", "no")
```

ESTADÍSTICA DESCRIPTIVA

Tablas de frecuencias

```
tally(~variable, data=base)
tally(~variable, data=base, format = "percent")
```

Tablas de frecuencias por intervalos

```
base$variablenueva <- cut(base$variable, breaks = c(corte1, corte2, ...))
tally(~variablenueva, data=base)
```

Gráficos

```
gf_bar(~variable, data=base)
gf_histogram(~variable, data=base)
gf_density(~variable, data=base)
gf_boxplot(~variable, data=base)
```

Medidas de síntesis

```
favstats(~variable, data=base)
```

Análisis bivariate

```
tally(variable1 ~ variable2, data=base)
gf_boxplot(variable1 ~ variable2, data = base) + coord_flip()
favstats(variable1 ~ variable2, data = base)
```

PROBABILIDAD

Diagnóstico

```
Tablacontingencia <- tally(variablediagnostico ~ variablerealidad, data=base)
Tablacontingencia
diagnostic(Tablacontingencia)
roc(variablerealidad ~ variableCN, data=base, plot = TRUE, print.auc = TRUE,
legacy.axes=TRUE)
```

Distribuciones

```
dbinom (x=valor, size=n, prob =p)
pbinom (q=valormax, size=n, prob=p)
qbinom (p=probabilidadacumulada, size=n, prob =p)
dpois (x=valor, lambda=λ)
ppois (q=valormax, lambda=λ)
qpois (q=probabilidadacumulada, lambda=λ)
pnorm(q=valor, mean=mu, sd=sigma)
pnorm(q=valor, mean=mu, sd=sigma, lower.tail = FALSE)
qnorm(p=probabilidadacumulada, mean=mu, sd=sigma)
```

CONTRASTES

Conformidad de una proporción

```
prop.test(~variable, data=base, success = "valor", p=pi)
binom.test(~variable, data=base, success = "valor", p=pi)
```

Conformidad de una multinomial

```
tabla <- tally(~variable, data=base)
chisq.test(tabla, p = c(pi1, pi2, pi3))
```

Conformidad de una media

```
shapiro.test(base$variable)
t.test(~variable, data=base, mu=valor, conf.level=0.95)
```

Homogeneidad de proporciones en grupos independientes

```
tablacontingencia <- tally(variable1 ~ variable2, data=base)
tablacontingencia
res <- chisq.test(tablacontingencia)
res
res$expected

fisher.test(tablacontingencia)
```

Homogeneidad de proporciones en grupos emparejados (McNemar)

```
tablacontingencia <- tally(variableantes ~ variabledespues, data=base)
tablacontingencia
mcnemar.test(tablacontingencia)
```

Homogeneidad de medias en dos grupos independientes

```
favstats(variableCN ~ variableCL, data=base)
G1 <- filter(base, VariableCL == "G1")
G2 <- filter(base, VariableCL == "G2")
shapiro.test(G1$variableCN)
shapiro.test(G2$variableCN)
var.test(variableCN ~ variableCL, data=base)
t.test(variableCN ~ variableCL, data=base, var.equal=TRUE)

t.test(variableCN ~ variableCL, data=base, var.equal=FALSE)

wilcox.test(variableCN ~ variableCL, data=base)
```

Homogeneidad de dos medias emparejadas

```
favstats(variableA ~ variableD, data=base)
base$dif <- base$variableA - base$variableD
shapiro.test(base$dif)
t.test(~dif, data=base)

wilcox.test(base$dif)
```

Análisis de la varianza (ANOVA)

```
favstats(variableCN ~ variableCL, data=base)
bartlett.test(variableCN ~ variableCL, data=base)
p1.aov <- aov(variableCN ~ variableCL, data=base)
anova(p1.aov)
shapiro.test(p1.aov$residuals)
```

Si R_H en la ANOVA aplicamos pruebas t por parejas con la corrección de Bonferroni: α/n° comparaciones

Relación entre variables cuantitativas

```
plot1 <- gf_point(variableY ~ variableX, data=base)
plot1
cor.test(variableY ~ variableX, data=base)
cor.test(variableY ~ variableX, data=base, method = "spearman")

m1 <- lm(variableY ~ variableX, data=base)
summary(m1)
confint(m1)
prediccions <- predict(m1, interval = "prediction")
head(prediccions)
```

```
resid(m1)
shapiro.test(resid(m1))
predict(m1)
gf_point(resid(m1) ~ predict(m1))
gf_point(rstandard(m1) ~ predict(m1))
```

Regresión logística y análisis de supervivencia

```
levels(base$variableCL)
base$variableCL<-relevel(base$variableCL,ref="No")
m1 <- glm(variabley ~ variablex, data =base, family=binomial)
exp(confint(m1))

time=c(59,115,121,156,431,448,477)
event=c(1,1,0,1,1,0,0)
base=data.frame(time,event)
base$survival_object<-Surv(base$time,base$event)
head(base)
s1 <- survfit(survival_object ~ 1, data=base)
summary(s1)
plot(s1)
```

Tutorial R Studio

Carga la base de datos TutorialRStudioBEIR.csv en RStudio (Environment->Import Dataset-> From text (base)) poniendo como nombre "base" y marcando la opción de Heading "Yes". Después, carga los paquetes:

```
library(mosaic)
library(ggformula)
library(dplyr)
library(PropCIs)
library(ThresholdROC)
library(pROC)
library(survival)
```

1. La variable Trabajo1, no ha quedado bien codificada así que conviértela en factor asignándole los valores 0=Suspendido y 1=Aprobado. Después, filtra los datos para trabajar únicamente con los alumnos que han hecho evaluación continuada y describe las variables Notas_final y Sexo, con los estadísticos o gráficos que consideres adecuados.

```
base$Trabajo1<-factor(base$Trabajo1,levels=1:0,labels=c("Aprobado", "Suspendido"))
nuevabase<- filter(base,Evaluacion=="Ev. Continuada")
tally(~Sexo,data=nuevabase)
tally(~Sexo,data=nuevabase,format="percent")
nuevabase$Notaint<-cut(nuevabase$Notas_final,breaks=c(0, 5, 7, 9, 10))
tally(~Notaint,data=nuevabase)
gf_bar(~Sexo,data=nuevabase)
gf_histogram(~Notas_final,data=nuevabase)
gf_boxplot(~Notas_final,data=nuevabase)
favstats(~Notas_final,data=nuevabase)
```

2. Volvemos a la base original. Sabemos que el profesor utiliza el primer trabajo como un indicador de lo que pasará con el trabajo final. Si consideramos Trabajo1 como un test diagnóstico (+,-) y Trabajo2 como la realidad (E,S), calcula los índices epidemiológicos que hemos estudiado.

```
Tablacontingencia<- tally(Trabajo1~Trabajo2,data=base)
Tablacontingencia
diagnostic(Tablacontingencia)
```

3. Calcula la probabilidad de que...

a) ... exactamente 3 alumnos aprueben el trabajo final en una muestra de 8 alumnos elegida al azar.

Utiliza la probabilidad de aprobar de la muestra.

```
dbinom(x=3,size=8,prob=0.457)
```

b) ... un alumno haya asistido a menos de 15 clases si sabemos que las clases asistidas siguen una Poisson con media 12.

```
ppois(q=14,lambda=12)
```

c) ...un alumno suspenda el examen final, si sabemos que la nota en la población sigue una normal de media 6,28 y desviación típica 2,15. ¿Y qué nota limitará el 5% de las mejores clases?

```
pnorm(q=5,mean=6.28,sd=2.15)
```

```
qnorm(p=0.95,mean=6.28,sd=2.15)
```

4. Hay la creencia de que la proporción de mujeres es del 60% de la clase. ¿Podemos afirmar que es falso?

```
prop.test(~Sexo,data=base,success="Mujer",p=0.6)
```

5. ¿Podemos afirmar que las clases tienen el mismo número de alumnos en la población?

```
tabla<-tally(~Clase,data=base)
chisq.test(tabla,p=c(1/3, 1/3, 1/3))
```

6. ¿La nota media del examen parcial este año es distinta de 7,5? Un 7,5 fue la nota del año pasado.

```
shapiro.test(base$Notas_parcial)
t.test(~Notas_parcial,data=base,mu=7.5,conf.level=0.95)
```

7. ¿Hay relación entre el sexo y el tipo de evaluación?

```
tablacontingencia<-tally(Sexo~Evaluacion,data=base)
tablacontingencia
res<-chisq.test(tablacontingencia)
res
res$expected

fisher.test(tablacontingencia)
```

8. ¿Ha habido un cambio entre los aprobados del primer trabajo y del segundo? Trabaja con una significación del 10%.

```
tablacontingencia<-tally(Trabajo1 ~ Trabajo2, data = base)
tablacontingencia
mcnemar.test(tablacontingencia)
```

9. ¿Hay diferencias significativas en la nota del examen final entre los que van a academia y los que no van? ¿Llegaríamos a la misma conclusión si hacemos la prueba correspondiente no paramétrica?

```
favstats(Notas_final ~ Academia, data=base)
G1<-filter(base, Academia=="Si")
G2<-filter(base, Academia=="No")
shapiro.test(G1$Notas_final)
shapiro.test(G2$Notas_final)
var.test(Notas_final~Academia, data=base)
t.test(Notas_final~Academia, data=base, var.equal=TRUE)

t.test(Notas_final~Academia, data=base, var.equal=FALSE)

wilcox.test(Notas_final~Academia, data=base)
```

10. ¿Ha habido un cambio en las notas entre el examen parcial y el final? ¿Llegaríamos a la misma conclusión si hacemos la prueba correspondiente no paramétrica?

```
base$dif<-base$Notas_parcial-base$Notas_final
shapiro.test(base$dif)
t.test(~dif, data=base)

wilcox.test(base$dif)
```

11. ¿Hay diferencias en la asistencia en función de la clase?

```
favstats(Asistencia~Clase, data=base)
bartlett.test(Asistencia~Clase, data=base)
p1.aov<-aov(Asistencia~Clase, data=base)
anova(p1.aov)
shapiro.test(p1.aov$residuals)
```

12. Analiza la relación entre la asistencia a clase y la nota del examen final.

Asume normalidad de las variables

```
plot1<-gf_point(Notas_final~Asistencia, data=base)
plot1
cor.test(Notas_final~Asistencia, data=base)
m1<-lm(Notas_final~Asistencia, data=base)
summary(m1)
confint(m1)
prediccions<-predict(m1, interval="prediction")
head(prediccions)
resid(m1)
shapiro.test(resid(m1))
predict(m1)
gf_point(resid(m1)~predict(m1))
gf_point(rstandard(m1)~predict(m1))
```