Nombre:	**CI:**	**Nro 2do informe**

SEGUNDA PRUEBA PARCIAL - SOLUCIONES

30 de Noviembre de 2019

Ejercicio 1 Queremos estudiar si sobrevivir o no al accidente del Titanic depende o no de la clase en la que se alojaba la persona en el barco. Para eso se cuenta con los siguientes datos

	Sobrevive	No sobrevive
Primera clase	100	111
Segunda clase	96	142
Tercera clase	114	98

Para realizar una prueba de hipótesis se tiene la siguiente salida de R

```
datos=matrix(c(100,96,114,111,142,98),ncol=2)
chisq.test(datos)

##
##  Pearson's Chi-squared test
##
## data:  datos
## X-squared = 8.16, df = 2, p-value = 0.01691
```

1. Plantear la prueba de hipótesis que se desea testear y hallar la región crítica a nivel $\alpha = 0{,}05$.[1]

   Para plantear la prueba de hipotesis a testear sea X la variable aleatoria que vale 1 si la persona sobrevive y 0 si no, sea Y la variable aleatoria que toma los valores 1, 2 y 3 según si la persona viajó en primera, segunda o tercera clase respectivamente. Queremos testear

$$\begin{cases} H_0: & X \text{ es independiente de } Y \\ H_1: & X \text{ no es independiente de } Y \end{cases}$$

   La región critica es $RC = \{661 \times T_{661} > 6\}$ siendo $661 \times T_{661}$ el estadistico visto en clase.

2. Realizar la prueba de hipótesis a nivel $\alpha = 0{,}05$. Justificar
   Se rechaza H_0, esto se puede ver o bien porque $8{,}16 > 6$ o porque el $0{,}0169 < 0{,}05$

3. Realizar la prueba de hipótesis a nivel $\alpha = 0{,}01$. Justificar
   No se rechaza ya que el p-valor no es menor que 0,01.

Ejercicio 2
Consideremos el siguiente conjunto de datos

```
datosej2=c(0.7,4.2,-0.2,3.1,0.15,2.0,3.9,0.6,-0.42,3.5,-4.3,-0.39,-2.1,-0.08,0.3)
```

Considere la siguiente salida de R

[1] recordar que qchisq(0.95,2)=6

```r
cor.test(datosej2,sort(datosej2),method="spearman")

##
##  Spearman's rank correlation rho
##
## data:  datosej2 and sort(datosej2)
## S = 840, p-value = 0.06021
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##  rho
## -0.5
```

1. ¿Qué se puede concluir de la misma si tomamos el nivel $\alpha = 0{,}05$?. No se rechaza la hipótesis nula (de que la muestra es iid) a nivel $\alpha = 0{,}05$ ya que el p-valor supera 0,05

2. Queremos además testear si los mismos se ajustan a una distribución normal o a una uniforme, para eso tenemos la siguiente salida de R

   ```r
   ks.test(datosej2,"pnorm",1,2)

   ##
   ##  One-sample Kolmogorov-Smirnov test
   ##
   ## data:  datosej2
   ## D = 0.22628, p-value = 0.3698
   ## alternative hypothesis: two-sided

   ks.test(datosej2,"punif",-1/2,1/2)

   ##
   ##  One-sample Kolmogorov-Smirnov test
   ##
   ## data:  datosej2
   ## D = 0.46667, p-value = 0.001629
   ## alternative hypothesis: two-sided
   ```

   ¿A que distribución se ajustan mejor los datos? justificar.
   Claramente se ajustan mejor a la normal con media 1 desvío 2 ya que por ejemplo una prueba de nivel 0.01 no rechaza dicha distribución (ya que $0{,}3698 > 0{,}01$) mientras que si se rechaza que sea uniforme entre -1/2 y 1/2 para ese nivel (ya que $0{,}00169 < 0{,}01$).

Ejercicio 3

Para estudiar si hay una relación de dependencia lineal entre dos variables x e y tenemos la siguiente salida de R

```r
      reg=lm(y~x)
summary(reg)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30403 -0.11803 -0.01099  0.10070  0.49680
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.001326   0.026384   -0.05     0.96
## x            0.963755   0.023751   40.58   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1829 on 48 degrees of freedom
## Multiple R-squared:  0.9717,Adjusted R-squared:  0.9711
## F-statistic:  1647 on 1 and 48 DF,  p-value: < 2.2e-16
```

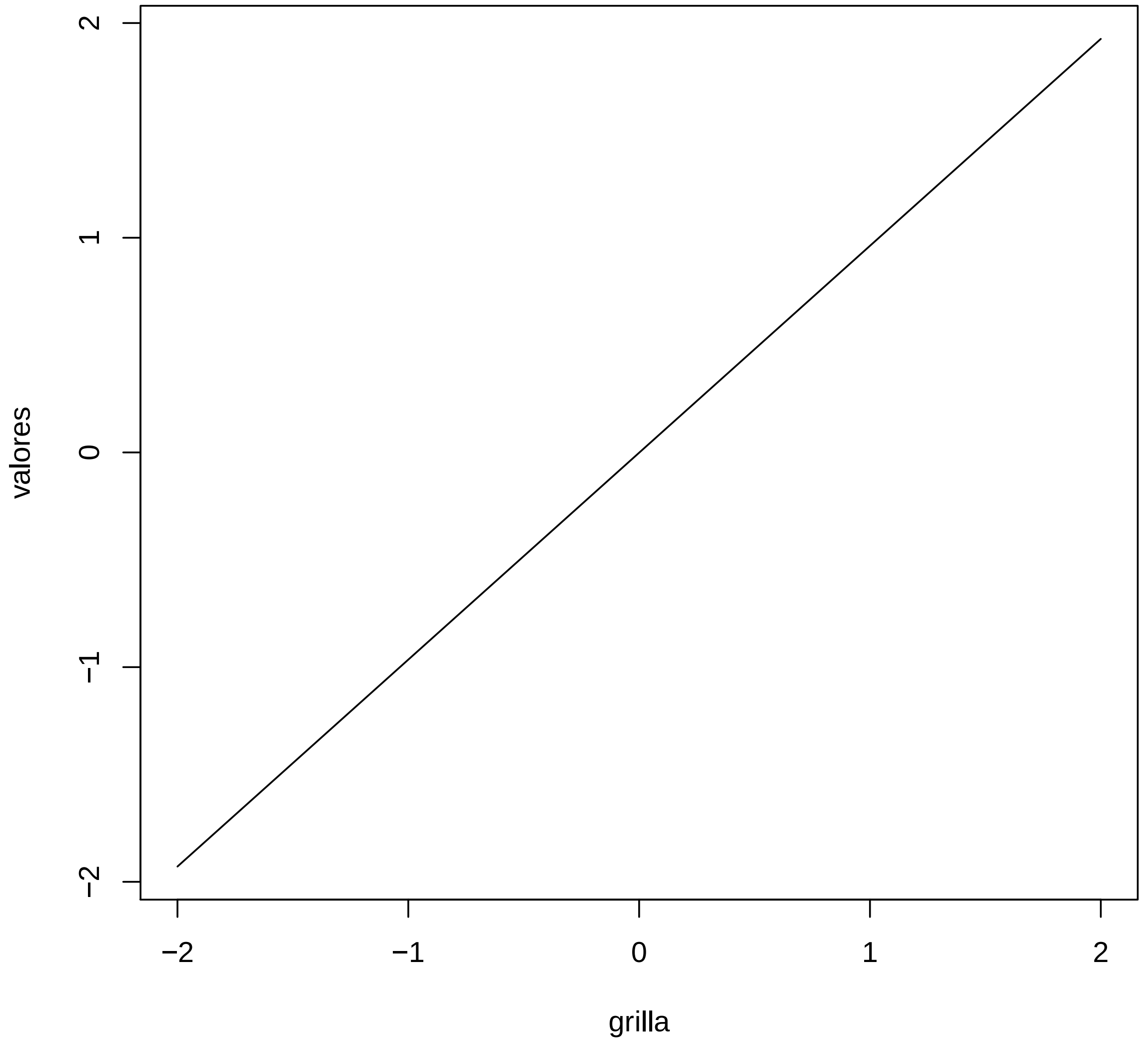
1. Plantear el modelo lineal teórico que se quiere estimar, y hallar los valores estimados de los parámetros.
   El modelo teórico es $y = ax + b + e$ siendo a, b coeficientes a determinar y e un error.

2. Dibujar la recta estimada.

```
grilla=seq(from=-2,to=2,by=0.1)
coef=reg$coefficients
coef

##  (Intercept)            x
## -0.001325764  0.963754584

valores=grilla*coef[2]+coef[1]
plot(grilla,valores,type='l')
```



3. Según la salida de R anterior, ¿es razonable suponer que hay una relación de dependencia lineal? Si ya que el p-valor de la prueba $H_0 : a = 0$ contra $H_0 : a \neq 0$ es menor que 2×10^{-16} y por lo tanto se rechaza H_0 a nivel por ejemplo 0,001.

4. Hallar el valor del coeficiente de determinación R^2. Interpretar dicho valor. El valor es 0.97, da próximo a 1 y por lo tanto la varianza total de las y es explicada casi en su totalidad por el modelo lineal.