Telecommunications Engineering
Technical University of Cartagena
Entregable

Andrés Ruz Nieto

Problema 1

El motor de un cohete se fabrica al unir dos tipos de propulsor: uno de encendido y un impulsor. Se piensa que la resistencia al esfuerzo cortante de la junta y (en psi) es una función lineal de la edad x (en semanas) del propulsor cuando se arma el motor. En la tabla siguiente aparecen 20 observaciones:

x	2158.7	1678.15	2316	2061.3	2207.5	1708.3	1784.7	2575	2357.9	2277.7
y	15.5	23.75	8	17	5	19	24	2.5	7.5	11

x	2165.2	2399.55	1779.8	2336.75	1765.3	2053.5	2414.4	2200.5	2654.2	1753.7
y	13	3.75	25	9.75	22	18	6.0	12.5	2	21.5

El objetivo del estudio es analizar la posible relación entre x e y.

a) Representa la nube de puntos y discute su forma. ¿Te parece adecuado un modelo lineal para explicar y en función de x?

b) Suponiendo una relación lineal, ajusta una recta de regresión de y en función de x. Escribe la ecuación de la recta y discute la bondad del modelo.

c) Para una edad de $x = 5.5$ ¿Que valor de y darías como predicción?

d) Calcula el histograma de la variable residuos (representalo de forma aproximada y discute su forma) y la distribución de frecuencias asociada (toma los limites que por defecto calcula el R).¿Que porcentaje de valores son mmenores que cero? (ayuda: si el ajuste lo has volcado en el objeto recta, los residuos los puedes obtener mediante

```
residuos<-recta$residuals
```

Problema 2

Supón que las variables X e Y del problema 1 siguen una distribución normal.

a) Da un estimador puntual de las medias y varianzas poblacionales de X e Y.

b) ¿Se puede suponer que la media de la población μ_X es significativamente mayor que 2050?. Plantea el contraste y discute que concluyes en base al p-valor obtenido?

c) Suponiendo que $\sigma_Y = 7.7$, construye un intervalo de confianza para la media μ_Y de la población de la variable Y con una confianza del 94%.

Problema 3

Sea Y una variable aleatoria continua con función de distribución:

$$F(y) = \begin{cases} \frac{y^2+y}{2} & \text{if } 0 < y < 1 \\ 0 & \text{en otro caso} \end{cases}$$

Escribe un procedimiento para generar valores de Y con R usando el método de la transformada inversa. Genera 1000 valores, dibuja el histograma y construye la tabla de frecuencias.

Problema 4

Responde a las siguientes cuestiones . (Establece el modelo y la probabilidad que se busca. Escribe en este caso el comando R que usas)

1. El número de llamadas telefónicas que llegan a un local del 112 se sabe que siguen una distribución de Poisson de media dos llamadas por minuto. Calcula la probabilidad de

   a) Haya menos de tres llamadas en un periodo de un minuto.

   b) El tiempo entre dos llamadas sucesivas sea menor de 0.5 minutos.

   c) Haya al menos 2800 llamadas en 24 horas.

2. En una fábrica que suministra agua mineral, se ha establecido que el volumen de llenado está normalmente distribuido con media 150cl y desviación típica de 2cl.

   a) el criterio de calidad de la compañía implica no vender botellas que contengan menos de 147cl. Calcula la proporción de botellas que no se pueden vender.

   b) Calcula el valor de c tal que $P(|X - 150| < c) = 0.99$

3. La probabilidad de recibir erroneamente un bit transmitido por un canal de transmisión digital es 0.2. Los tests de transmisión son independientes.

   a) Calcula la probabilidad de que en los próximos seis bits transmitidos al menos haya 2 errores.

   b) la probabilidad de que el primer bit eeroneo trnamitido sea el cuarto.

4. Cuando se prueban tarjetas de circuito empleadas en la manufactura de reproductores de discos compactos, a la larga el porcentaje de partes defectuosas es del 5%. Sea X el número de tarjetas defectuosas en una muestra seleccionada al azar de tamaño 25.

   a) Calcula la probabilidad de que como mucho haya dos tarjetas defecuosas.

   b) Calcula la probabilidad de que haya al menos cinco tarjetas defecuosas.

   c) Calcula la probabilidad de que haya entre uno y cuatro tarjetas defecuosas.

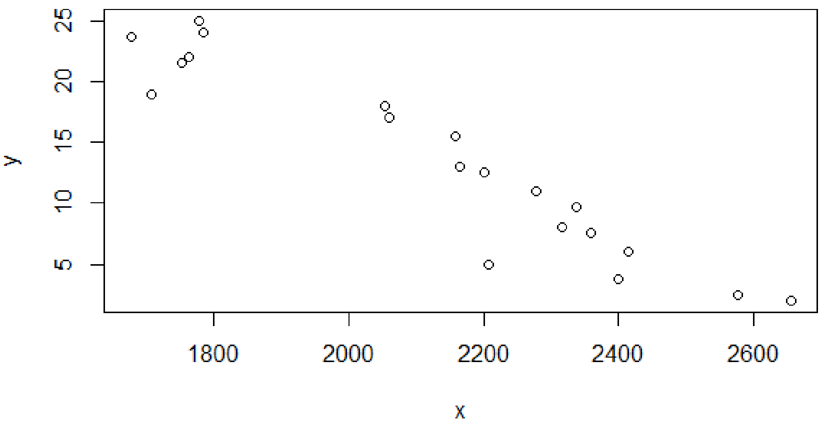
Entregable de prácticas: Estadística

Ejercicio 1

a) Con los comandos:

```r
x<-c(2158.7,1678.15,2316,2061.3,2207.5,1708.3,1784.7,2575,2357.9,2277.7,2165.2,
2399.55,1779.8,2336.75,1765.3,2053.5,2414.4,2200.5,2654.2,1753.7)
y<-c(15.5,23.75,8,17,5,19,24,2.5,7.5,11,13,3.75,25,9.75,22,18,6.0,12.5,2,21.5)
plot(x,y)
```

Representamos la siguiente la nube de puntos



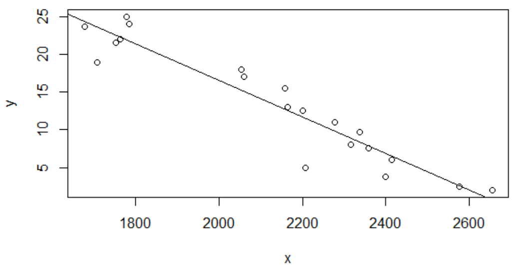
Como podemos ver la nube de puntos sigue un modelo lineal de asociación negativa o inversa

b) Usamos los comandos

```r
recta1<-lm(y~x)
summary(recta1)
abline(recta1)
```

Aplicamos el método de los mínimos cuadrados.

$$d = -\left(\sum_{i=1}^{N} X_i\right)^2 + N\left(\sum_{i=1}^{N} {X_i}^2\right) \quad a = \frac{-\sum_{i=1}^{N} X_i \sum_{i=1}^{N} Y_i + \sum_{i=1}^{N} X_i Y_i}{d} \quad b = \frac{-\sum_{i=1}^{N} X_i \sum_{i=1}^{N} X_i Y_i + \sum_{i=1}^{N} {X_i}^2 \sum_{i=1}^{N} Y_i}{d}$$

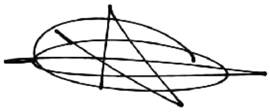



```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.5170 -0.6718  0.1750  1.5095  3.1141 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 65.034398   4.188147   15.53 7.21e-12 ***
x           -0.024243   0.001946  -12.46 2.75e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.537 on 18 degrees of freedom
Multiple R-squared:  0.8961,	Adjusted R-squared:  0.8903 
F-statistic: 155.2 on 1 and 18 DF,  p-value: 2.753e-10
```



R nos devuelve esta serie de datos por pantalla, de donde sacamos los siguientes valores:

a= -0.024243 y b= 65.034398 por lo que la recta será “**y= -0.024243x+65.034398**”

Multiple R-squared: 0.8961 > 0.8 por lo que el modelo es **fiable**

c) Simplemente tendríamos que sustituir el valor de x que se nos da (5.5), en la recta que hemos obtenido anteriormente

-0.024243*5.5+65.034398=**64.90106**

d) A través de los comandos

```
residuos<-recta1$residuals
hist(residuos,plot=T)
hist(residuos,plot=F)
```

Histograma asimétrico

Intervalos	Frecuencia
(-8,-6)	1
(-6,-4)	1
(-4,-2)	1
(-2,0)	7
(0,2)	6
(2,4)	4

Como podemos ver hay la misma cantidad de valores por encima que por debajo del 0, por lo cual por debajo del 0 tendremos un 50%

Ejercicio 2

a) Con los siguientes comandos, obtendremos la media y varianza (resultado a la derecha):

$$\sigma^2 = \frac{\sum_i^n (x_i - \bar{x})^2}{n} \qquad \bar{x} = \frac{\sum_i^n (x_i)}{n}$$

```
mean(x)  = 2132.407
var(x)   = 89443.21
mean(y)  = 13.3375
var(y)   = 58.66628
```

b) Usaremos los siguientes comandos:

```
t.test(x,alternative = "greater",mu=2050)
```

```
        One Sample t-test

data:  x
t = 1.2323, df = 19, p-value = 0.1164
alternative hypothesis: true mean is greater than 2050
95 percent confidence interval:
 2016.773      Inf
sample estimates:
mean of x
 2132.407
```

El p-valor es=0.1164 es superior al nivel de significación más alto (0.1) así que podemos afirmar de que la media de x es mayor que 2050.

c) Para construir el intervalo de confianza usaremos los siguientes comandos:

$$\mu \in [\bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma^2}{\sqrt{n}}]$$

```
alpha<-1-0.94
sigma<-7.7
cuantil<-qnorm(1-alpha/2) (mirar en la tabla de distribución normal)
lim_inf<-mean(y)-cuantil*sigma/sqrt(20)
lim_sup<-mean(y)+cuantil*sigma/sqrt(20)
```

Donde llegaremos a la siguiente conclusión:

[10.0992 < 13.375 < 16.5757] = 0.94

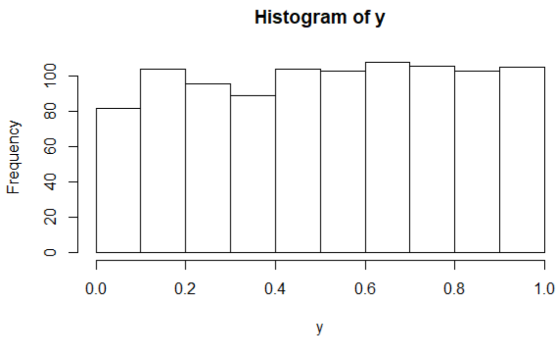
Ejercicio 3

a) En primer lugar, igualaremos nuestra función a “u” y posteriormente despejaremos “y”. Al hacer esta operación nos quedan las dos siguientes funciones.

y=2u
y=2u-1

A continuación, aplicamos el siguiente script para hallar el histograma.

```
y<-numeric(1000)
u<-runif(1000,0,1)
for (i in 1:1000){
   if(u[i]<0.5)
   y[i]<-2*u[i]
   else y[i]<-2*u[i]-1
}
hist(y)
```



b) Para construir la tabla de frecuencias por intervalos usaremos los siguientes códigos:

```
hist(y,plot=F)
```

También podemos usar:

```
nclass.Sturges(y)
seq(0,1,length=nclass.Sturges(y))
intervalosY=cut(y,breaks=seq(0,1,length=nclass.Sturges(y)),include.lowest=TRUE)
table(intervalosY)
```

Invervalos	Frecuencia
[0,0.1]	82
(0.1,0.2]	104
(0.2,0.3]	96
(0.3,0.4]	89
(0.4,0.5]	104
(0.5,0.6]	103
(0.6,0.7]	108
(0.7,0.8]	106
(0.8,0.9]	103
(0.9,1]	105

Ejercicio 4

Simplemente tenemos que sustituir los valores que nos dan, en la función correspondiente.

a) Comandos:

a. `ppois(3,2) = 0.8571235`

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \\ 0 \; en \; otro \; caso \end{cases}$$

b. `pexp(0.5) = 0.3934693`

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} \\ 0 \; en \; otro \; caso \end{cases}$$

c. `1-ppois(2800,2880) = 0.9312506`

b) Comandos:

a. `pnorm(147,150,2) = 0.0668072`

$$\varphi\left(\frac{x-\bar{x}}{\sigma}\right)$$

b. `pnorm((0.99+1)/2)*sqrt(2) el valor de c será 1.188126`

$$\varphi\left(\frac{c}{\sqrt{2}}\right) - \varphi\left(-\frac{c}{\sqrt{2}}\right)$$

$$2\varphi\left(\frac{c}{\sqrt{2}}\right) - 1 = 0.99$$

c) Comandos:

a. `1-pbinom(2,6,0.2) = 0.09888`

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} \\ 0 \; en \; otro \; caso \end{cases}$$

b. `pgeom(4,0.2) = 0.67232`

$$f(x) = \begin{cases} p(1-p)^x \\ 0 \; en \; otro \; caso \end{cases}$$

d) Comandos:

a. `pbinom(2,25,0.05) = 0.8728935`

b. `1-pbinom(5,25,0.05) = 0.001212961`

c. `pbinom(4,25,0.05)-pbinom(1,25,0.05) = 0.3504592`

Estadística. Grado en Ingeniería en Sistemas de Telecomunicación (GIST) Telemática (GIT).

Práctica 7 Entregable

1. Nombre..

Problema 1

Una señal telemétrica, T, se transmite . desde un sensor de temperatura en aun satelite de comunicaciones al receptor en el control de la misión, que recibe una señal R. Se observaron 10 valores de ambas señales:

R	-2.09	-1.57	-8.01	8.25	2.71	-2.31	2.18	2.75	5.53	1.97
T	0.10	-2.26	-7.60	6.41	2.43	-1.79	-0.14	2.31	3.40	-0.66

1. Representa la nube de puntos y discute su forma. ¿Te parece adecuado un modelo lineal para explicar T en función de R?

2. Suponiendo una relación lineal, ajusta una recta de T en función de R. Escribe la ecuación de la recta y discute la bondad del modelo.

3. Para una se señal recibida de $r = 6$ que valor de T darías como predicción.

4. Calcula el histograma de la variable residuos (representalo de forma aproximada y discute su forma) y la distribución de frecuencias asociada (toma los limites que por defecto calcula el R).¿Que porcentaje de valores son mayores que cero? (ayuda: si el ajuste lo has volcado en el objeto recta, los residuos los puedes obtener mediante

```
residuos<-recta$residuals
```

Problema 2

Suponemos que las variables R y T siguen distribuciones normales.

1. Da un estimador puntual de las medias y varianzas poblacionales de R e T.

2. ¿Se puede suponer que la media de la población μ_T para la señal transmitida T es significativamente menor que cero?. Plantea el contraste y discute que concluyes en base al p-valor obtenido?

3. Construye un intervalo de confianza para la media μ_R de la población de la variable R con una confianza del 99%.

Problema 3

Supón que la señal telemétrica, T del problema 1 sigue una distribución normal con $\mu = 0$ y $\sigma = 2$

1. ¿Cuál es la probabilidad de que el valor de la señal transmitida este entre -2 y 1.5?
2. ¿Qué valor de c es tal que el intervalo $(-c, +c)$ incluya el 98% de todos lo valores de T?
3. Si consideramos una muestra aleatoria $T_1, \ldots, T_{10}$ de la variable T. Calcular:

$$P(\sum_{i=1}^{10} T_i^2) \leq 100)$$

Problema 4

Sea la varible aleatoria continua con función de densidad:

$$f(y) = \begin{cases} \frac{1}{2}(2-y) & \text{if } 0 < y < 2 \\ 0 & \text{en otro caso} \end{cases}$$

Escribe un procedimiento para generar valores usando el R por el método de la transformada inversa. Genera 1000 de ellos y representa el histograma y la distribución de frecuencias.

Problema 5

Responde a las siguientes cuestiones:

a) Un triplista de un equipo de baloncesto hace canasta de tres puntos en el 80% de los intentos que realiza. ¿Cuál es la probabilidad de que en un partido logre su primer triple al quinto intento?

b) La probabilidad de que una cámara de bicicleta se pinche al hacer una ruta urbana es de 0.6. Si comprobamos 15 cámaras, calcula la probabilidad de que resistan como mucho 8 cámaras.

c) Se desea describir el comportamiento de la variable X esfuerzo vibratorio (Ib/pulg^2) en la paleta de una turbina de viento, a una velocidad particular en un túnel de viento. Supongamos que X sigue una distribución de Rayleigh, con parámetro $\sigma^2 = 100^2$ ¿Cuál es la probabilidad de que X sea menor de 90? ¿Y de que X esté entre 90 y 130?

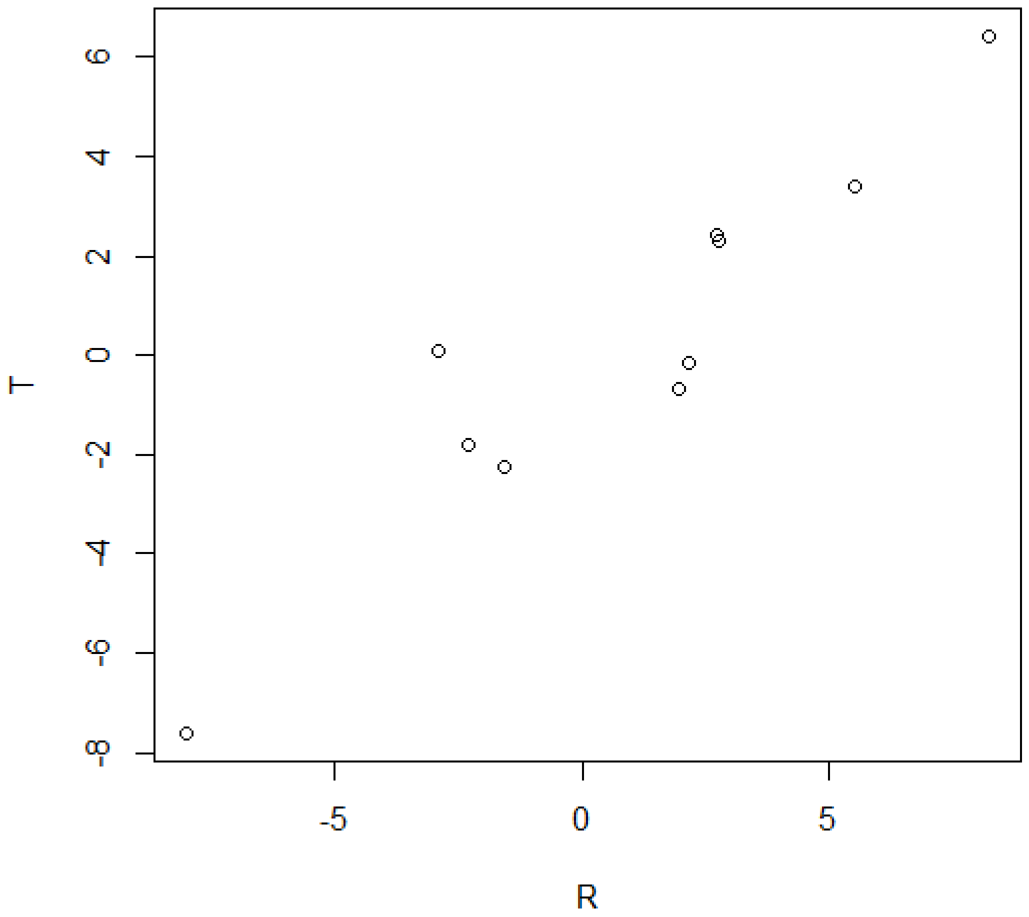
Cuestionario de prácticas

Ejercicio 1

1)

Realizando los comandos

R<-c(-2.9,-1.57,-8.01,8.25,2.71,-2.31,2.18,2.75,5.53,1.97)
T<-c(0.10,-2.26,-7.6,6.41,2.43,-1.79,-0.14,2.31,3.4,-0.66)
plot(R,T)
Obtenemos la siguiente gráfica



En la que, aparentemente, R y T siguen una relación lineal aproximada.

2)
Para ver cómo es la recta que relaciona R y T usamos los siguientes comandos:

recta<-lm(T~R)
summary(recta)

Nos saldría lo siguiente:

```
> recta<-lm(T~R)
> summary(recta)

Call:
lm(formula = T ~ R)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.72971 -0.92746  0.01087  0.61564  2.75828 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -0.4383     0.4496  -0.975    0.358    
R             0.7655     0.0999   7.663 5.95e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.396 on 8 degrees of freedom
Multiple R-squared:  0.8801,	Adjusted R-squared:  0.8651 
F-statistic: 58.72 on 1 and 8 DF,  p-value: 5.945e-05
```

Nos interesan los valores que aparecen en la columna Estimate Std.

Una recta tiene siempre la forma y = ax+b. En la columna de Estimate Std. Encontramos el valor de a y b de forma que

a = 0.7655

b = -0.4383

Por lo tanto, podemos establecer la siguiente relación lineal:

T = 0.7655·R – 0.4383

Según el programa vemos que R^2es igual a 0.8801>0.8, por lo tanto es una estimación fiable.

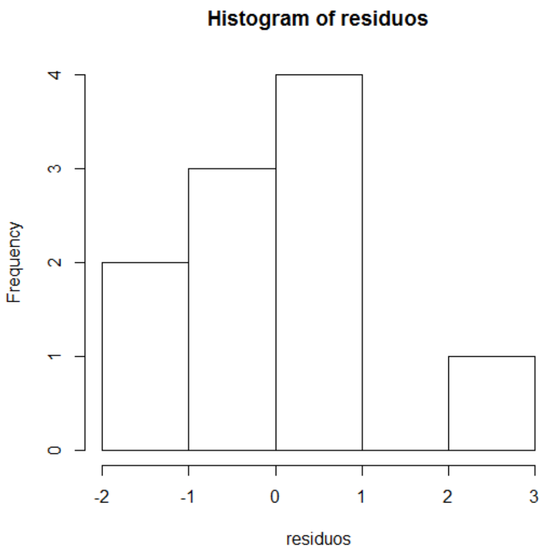
3)

Siguiendo nuestra estimación, si sustituimos R por 6 obtendríamos el siguiente valor de T:

T=0.7655·6 – 0.4383 =4.1547

Si miramos la línea, este resultado tiene sentido.

4)



La tabla la podemos extraer del histograma, pero también podemos calcularla usando el comando ***hist(residuos,plot=F)***

X_i	n_i	N_i	f_i	F_i
2.284610	1	1	0.1	0.1
-0.4793348	1	2	0.1	0.2
-0.8184414	1	3	0.1	0.3
0.4125949	1	4	0.1	0.4
0.8106305	1	5	0.1	0.5
0.5767927	1	6	0.1	0.6
-1.3295755	1	7	0.1	0.7
0.6689480	1	8	0.1	0.8
-0.4429905	1	9	0.1	0.9
-1.6832420	1	10	0.1	1

Intervalos	Marcas	n_i	N_i	f_i	F_i
[-2,-1]	-1.5	2	2	0.2	0.2
[-1,0]	-0.5	3	5	0.3	0.5
[0,1]	0.5	4	9	0.4	0.9
[1,2]	1.5	0	9	0	0.9
[2,3]	2.5	1	10	0.1	1

Como vemos, hay 5 valores que son mayores que cero, lo cual supone la mitad de los valores. Por lo tanto el porcentaje de valores mayores que 0 es el 50%.

Ejercicio 2

1)

Calcularemos las medias de R y T usando los comandos mean() y var().

Usando estos comandos comprobamos que…

mean(R) = 0.942; var(R) = 21.06635

mean(T) = 0.209; var(T) = 14.60785

Aunque observamos que la media de T es mayor que 0, realizaremos el contraste para reafirmarlo.

Para ello usaremos el comando ***t.test(T,alternative="less",mu=0)***

Tras escribir esto me sale lo siguiente:

```
> t.test(T,alternative="less",mu=0)

        One Sample t-test

data:  T
t = 0.1729, df = 9, p-value = 0.5667
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
     -Inf 2.424555
sample estimates:
mean of x
    0.209
```

El p-valor es 0.5667

No tenemos ningún nivel de significación así que comparamos el p-valor con el nivel de significación usual más alto (0.1)

Como el p-valor es superior a 0.1, podemos rechazar la hipótesis alternativa $\mu_T < 0$.

Esto es totalmente lógico ya que, como hemos visto $\mu_T = 0.209$, que es mayor que 0.

3)

Para calcularlo usaremos el siguiente comando

t.test(R,alternative="two.sided",mu=0.942,conf.level=0.99)(donde mu es la media de R)

```
> t.test(R,alternative="two.sided",mu=0.942,conf.level=0.99)

        One Sample t-test

data:  R
t = 0, df = 9, p-value = 1
alternative hypothesis: true mean is not equal to 0.942
99 percent confidence interval:
 -3.774893  5.658893
sample estimates:
mean of x
    0.942
```

De acuerdo con lo que vemos en la tabla, el intervalo de confianza es [-3.774893,5.658893]

Ejercicio 3

1)

Para hacer la probabilidad entre -2 y 1.5 calcularemos F(1.5) y F(-2) y hacemos la resta con el comando ***pnorm(1.5,0,4)-pnorm(-2,0,4)***

Esto da 0.3376322

2)

Sabemos que $F(c) - F(-c) = 0.98$

Por lo tanto…

$$\phi\left(\frac{c}{2}\right) - \phi\left(-\frac{c}{2}\right) = 0.98 \rightarrow 2 \cdot \phi\left(\frac{c}{2}\right) = 1.98 \rightarrow \phi\left(\frac{c}{2}\right) = 0.99 \rightarrow \frac{c}{2} = 2.326 \rightarrow c = 4.652$$

Para calcular esto último tenemos que escribir

Y después multiplicar el resultado por 2.

qnorm(0.99)

3)

$$\sum_{i=1}^{10} T_i^2 = Y \; sigue \; una \; distribución \; \chi^2_{10}$$

Para calcular $P(Y \leq 100)$ primero tipificamos

$$P(Y \leq 100) = P\left(\frac{Y}{4} \leq 25\right)$$

Y después escribimos el comando siguiente:

pchisq(25,10), donde 50 es el valor con el que comparamos $\frac{Y}{2}$ y 10 son los grados de libertad.

El resultado que nos da es 0.9946545, con lo cual $\sum_{i=1}^{10} T_i^2$Y≤100 prácticamente

Ejercicio 4

Para aplicar este método primero tendremos que calcular F(y). Para ello integramos.

$$F(y) = \begin{cases} \int_0^2 \frac{1}{2}(2-y) = y \cdot \left(1 - \frac{y}{4}\right) \; si \; 0 < y < 2 \\ 0 \; si \; y \leq 0 \\ 1 \; si \; y \geq 2 \end{cases}$$

Ahora lo igualamos a “u” y despejamos y.

$$u = y \cdot \left(1 - \frac{y}{4}\right) \rightarrow u = y - \frac{y^2}{4} \rightarrow y = 2\left(1 \pm \sqrt{1 - \frac{4}{4}u}\right) = 2 \pm 2\sqrt{1-u}$$

Tomamos sólo el valor negativo porque y no puede ser mayor que 2:

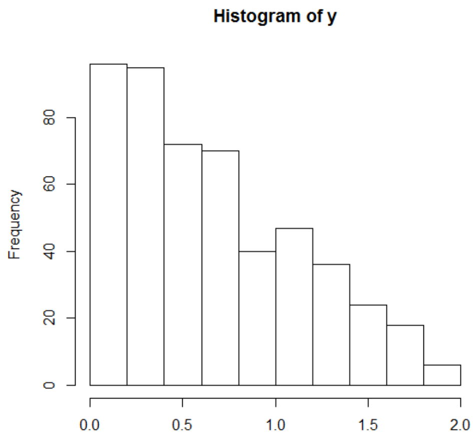
$$y = 2\left(1 - \sqrt{1-u}\right)$$

A continuación escribimos los siguientes comandos:

u<-runif(1000,0,2)

y<-2*(1-sqrt(1-u)) → Al hacer esto, algunos valores nos dan NaN porque toma valores de u mayores de 1.

El histograma obtenido es el siguiente:



Y la tabla de frecuencias la siguiente (obtenida con los comandos table() y cumsum())

La haremos por intervalos:

Intervalos	Marcas	n_i	N_i	f_i	F_i
[0.0,0.2]	0.1	97			
[0.2,0.4]	0.3	95			
[0.4,0.6]	0.5				
[0.6,0.8]	0.7				
[0.8,1]	0.9				
[1,1.2]	1.1				
[1.2,1.4]	1.3				
[1.4,1.6]	1.5				
[1.6,1.8]	1.7				
[1.8,2]	1.9				

(no supe terminar la tabla)

Ejercicio 5

a)

Digamos que el éxito es que meta una canasta de 3 puntos. Por lo tanto, la probabilidad de éxito es de 0.8 (p=0.8)
Como hablamos del número de veces que tiene que tirar para conseguir su primer triple, usamos el modelo geométrico.

X="número de fracasos hasta el primer éxito". Por lo tanto, queremos calcular P(X=4)
Para ello, usaremos el comando ***dgeom(4,0.8)***

Haciéndolo, concluimos que P(X=4) = 0.00128

b)

Digamos que el éxito es que la cámara resista.
Estamos hablando de un número de experimentos n en el que cada uno tiene la misma probabilidad de éxito, por lo tanto usaremos el modelo binomial

P(resistir) = 1-P(pinchar)=0.4 → p=0.4

X="número de cámaras que resisten". Queremos calcular $P(X \leq 8)$
Para ello, usaremos el comando ***pbinom(8,size=15,prob=0.4)***
El resultado que nos da es 0.9049526.

c)

Para ver la probabilidad de que sea menor o igual a 90, es decir, F(90) se escribe lo siguiente lo siguiente:
pweibull(90,shape=2,scale=sqrt(2)*100)
El resultado es el siguiente: 0.3330232

Para calcular $P(90 < X < 130)$ calcularemos F(130)-F(90). Para ello haremos el siguiente comando:
pweibull(130,shape=2,scale=sqrt(2)*100) - pweibull(90,shape=2,scale=sqrt(2)*100)
El resultado que da es: 0.2374195