Segundo informe

Grupo 74

Ana Inés García Pintos, 5.186.144-1

Federico Ramos Maltez, 5.247.724-7

Nicolás Fernández Sauleda, 5.242.510-3

22/11/2021

Índice

Ejercicio 1: *Intervalos de confianza*

Primeramente aclarar que se utiliza una semilla de valor *7474* para facilitar la reproductibilidad. Y también agregar que nos tocó una distribución $F \sim U(2, 10)$

1.a) *Media y varianza*

```
# Valores que se utilizan para realizar los cálculos.
n1 <- 150
min <- 2
max <- 10
# Cálculo de la media y de la varianza teórica respectivamente
mu1 <- (min+max)/2
var1 <- (((max-min)^2)/12/(n1))
```

El valor de la μ_0 teórica es de *6* y de la σ_0^2 teórica es igual a *0.036*.

1.b) *Intervalo de confianza para F*

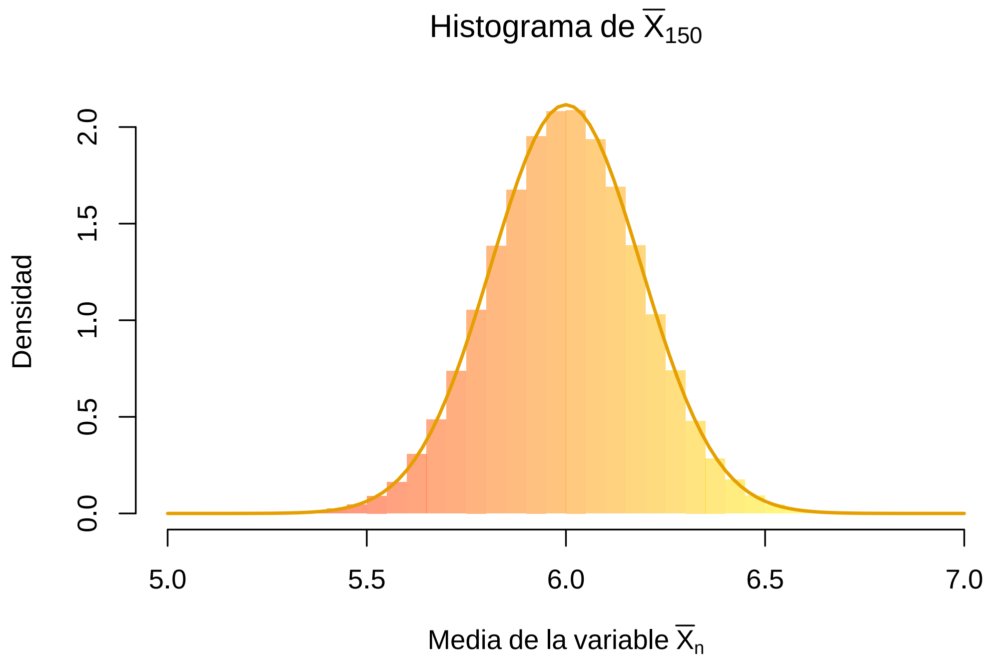
Para realizar los intervalos de confianza se construye una función llamada `unif.inter`, que se utilizará en la parte *(d)*.

```
# Construcción de la función
unif.inter <- function(datos, alpha = 0.05){
  n = length(datos)
  z = qnorm(1-alpha/2)
  xn = mean(datos)
  k = sd(datos)/sqrt(n)
  intervalo = c(xn-k*z,xn+k*z)
  return (intervalo)
}
# Aplicando la función
intervalosB <- unif.inter(runif(n1, min, max))
```

Se puede ver que el intervalo de confianza teórico a nivel 95 % para la media teórica de la distribución F es $I(F) =$ [5.69, 6.36]

1.c) *Simulaciones de* $\overline{X}_{150}$

```
# Construye un vector de largo 10^4 y realiza un bucle para crear simulaciones de X150.
v <- c()
for (j in 1:100000) {
  v <- c(v, mean(runif(n1, min, max)))
}
# Realiza histograma y superpone una normal de parámetros mu y sigma
hist(v, freq = FALSE, xlab = expression(Media ~ de ~ la ~ variable ~ bar(X)[n]), main=
↪  expression(Histograma ~ de ~ bar(X)[150]), ylab = "Densidad", breaks=40, xlim=c(5,7),
↪  col=heat.colors(40, 0.5), border=F)
curve(dnorm(x,mean=mu1, sd = sqrt(var1)), add = TRUE, lwd=2, col="#E69F00")
```



1.d) *Proporción de simulaciones que no contienen* μ_0

```
media_en_intervalo <- c()
# Bucle donde genera variables aleatorias y testea si su media cae en el intervalo de confianza
↪  al 95%, si cae almacena el valor 0, y si no cae, almacena el valor 1 dentro del vector
↪  media_en_intervalo.
for (i in 1:n1) {
  muestra <- runif(n1, min, max)
  intervalo <- unif.inter(muestra)
  if  ((intervalo[1] <= mu1)&(intervalo[2]>=mu1)) {
    media_en_intervalo <- c(media_en_intervalo, 0)
  }
  else {
    media_en_intervalo <- c(media_en_intervalo, 1)
  }
}
# Suma el vector y divide sobre el total de casos para conocer la proporción.
proporcion <- sum(media_en_intervalo)/150
```

Resultado de la proporción 0.06. Este valor debería tender al α utilizado que es de 0.05.

Ejercicio 2: *p-valor*

2.a) *Generación de la muestra y cálculo de p-valor.*

```
datos2 <- runif(n1, min, max)
x150 <- mean(datos2)
```

```r
# Se realiza una función para reutilizarla en la parte b
calcula_pvalor <- function(datos, min, max) {
  mean <- mean(datos)
  sd <- sd(datos)
  n <- length(datos)
  mu0 <- (min+max)/2
  p_valor <- 2*(1-(pnorm(sqrt(n)/sd * abs(mean-mu0))))
  return(p_valor)
}
calcula_pvalor(datos2, min, max)
```

Se obtiene un *p-valor*$= 0.832$ por lo que no hay evidencia suficiente para rechazar H_0 con un $\alpha = 0.05$.

2.b y c) *p-valor para mil muestras y test de Kolmogorov-Smirnov*

```r
promedio_muestras <- c()
pvalores_muestras <- c()
for (i in 1:1000) {
  muestrai <- runif(n1, min, max)
  promedio_muestras <- c(promedio_muestras, mean(muestrai))
  pvalores_muestras <- c(pvalores_muestras, calcula_pvalor(muestrai, min, max))
}
test_ks1 <- ks.test(pvalores_muestras, punif)
test_ks1
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  pvalores_muestras
## D = 0.024565, p-value = 0.5823
## alternative hypothesis: two-sided
```

El test de *Kolmogorov-Smirnov* nos dió un *p-valor*$= 0.582$ a un $\alpha = 0.01$, se puede observar para distancia cercanas a 0 p-valores mayores, por lo tanto existe evidencia significativa para no rechazar H_o. Esto significa que los los mil *p-valores* se aproximan a una $U(0, 1)$.

Ejercicio 3 *Comparación de dos muestras*

3.a) *Carga y análisis de la tabla*

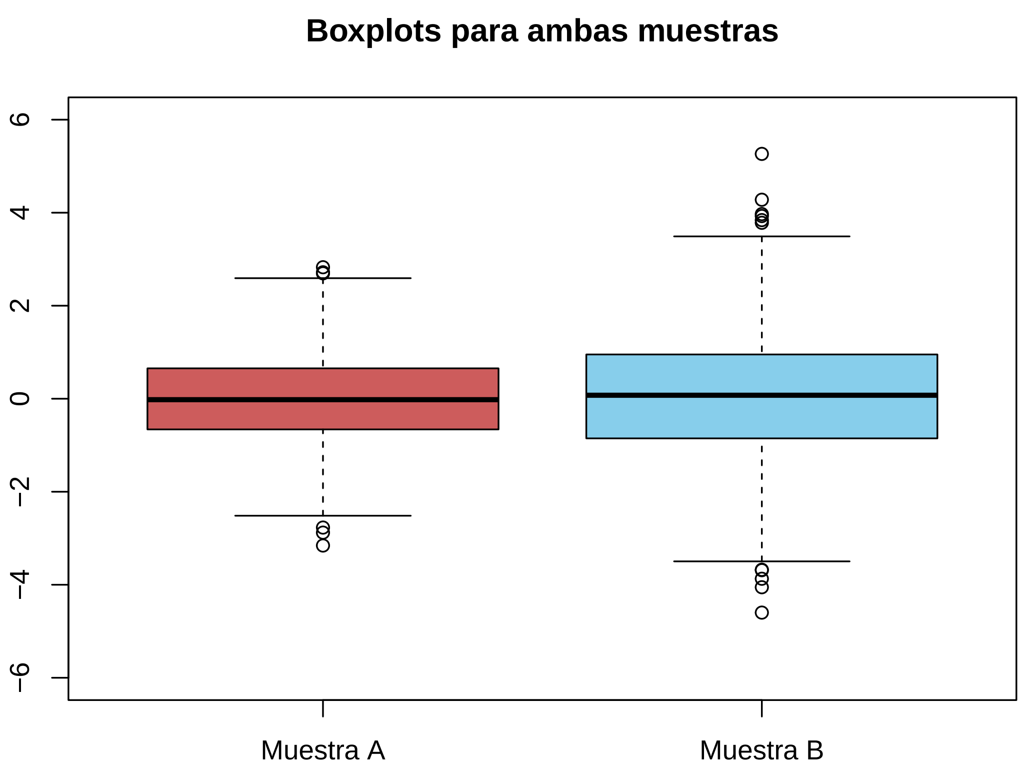
```r
# Importa las muestras
datos3 <- read.table("datosej3.txt", sep=" ", header=TRUE)
muestraA <- datos3$Muestra_A
muestraB <- datos3$Muestra_B
# Promedio y desvío estándar para ambas muestras
promedio3 <- c(mean(muestraA), mean(muestraB))
sd3 <- c(sd(muestraA), sd(muestraB))
summary(datos3)
```

```
##    Muestra_A          Muestra_B
##  Min.   :-3.15672   Min.   :-4.59902
##  1st Qu.:-0.65920   1st Qu.:-0.85106
##  Median :-0.02074   Median : 0.07489
##  Mean   :-0.02522   Mean   : 0.06478
##  3rd Qu.: 0.65275   3rd Qu.: 0.95103
##  Max.   : 2.82717   Max.   : 5.26403
```

Muestras	$\overline{X}$	σ
A	-0.025	0.065
B	0.974	1.377

3.b) *Boxplot para ambas muestras*

```
boxplot(datos3,ylim=c(-6,6), names=c("Muestra A", "Muestra B"), main="Boxplots para ambas
↪  muestras", col=c("indianred","skyblue"))
```



3.c) *Test de aleatoriedad*

A continuación se testea la aleatoriedad de cada una de las muestras utilizando el test de spearman.

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

```
##
##  Spearman's rank correlation rho
##
## data:  muestraA and sort(muestraA)
## S = 165797362, p-value = 0.8692
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## 0.005214833
```

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

```
##
##  Spearman's rank correlation rho
##
## data:  muestraB and sort(muestraB)
## S = 170068706, p-value = 0.519
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.02041326
```

Para ambas muestras obtuvimos *p-valores*>0.05, por lo tanto no hay evidencia suficiente para rechazar H_0, es decir, de que sean *i.i.d*

3.d) *Test de independencia*

Para testear la independencia de las muestras se utiliza el test de spearman, pero utilizando ambas muestras.

```
cortestAB <- cor.test(muestraA, muestraB, method = "spearman")
```

No hay evidencia suficiente para rechazar H_0 bajo $\alpha = 0.05$, ya que tenemos un *p-valor*$=0.874$. Por lo tanto, ambas muestras provienen de distribuciones diferentes.

3.e) *Kolmogorov-Smirnov*

```
test_ks <- ks.test(muestraA, muestraB)
test_ks
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  muestraA and muestraB
## D = 0.1, p-value = 9.08e-05
## alternative hypothesis: two-sided
```

Ya que *p-valor*$=$ 9.1e-05 hay evidencia suficiente para rechazar H_o, es decir que no provienen de la misma distribución.

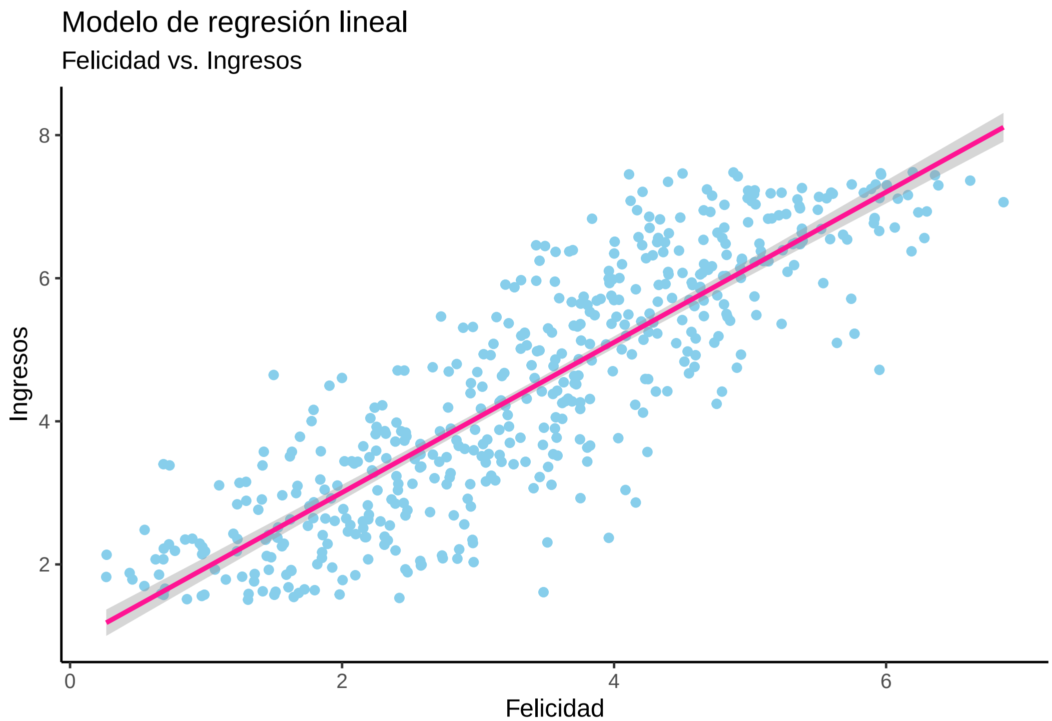
Ejercicio 4 *Regresión lineal*

```
# Importa y limpia los datos
datos4 <- subset(read.csv("income.data.csv"), select=-c(X))
hapiness <- datos4$happiness
income <- datos4$income
# Ajusta un modelo de regresión lineal
modelo <- summary(lm(hapiness ~ income))
```

```
##
## Call:
## lm(formula = hapiness ~ income)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.02479 -0.48526  0.04078  0.45898  2.37805
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.20427    0.08884   2.299   0.0219 *
## income       0.71383    0.01854  38.505   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7181 on 496 degrees of freedom
## Multiple R-squared:  0.7493, Adjusted R-squared:  0.7488
## F-statistic:  1483 on 1 and 496 DF,  p-value: < 2.2e-16
```

Se observa que el valor de $R^2 = 0.749$ por lo que el ajuste es bueno ya que explica el 75 % de los casos. Nuestra h_0 es que $\widehat{\beta_0}$ y $\widehat{\beta_1}$ sean 0, en nuestro caso el $\widehat{\beta_0} = 0.204$ y el $\widehat{\beta_1} = 0.714$. Por otro lado los *p-valores* nos dieron 0.02 y 2e-16 respectivamente. Por lo tanto podemos decir que rechazamos h_0 a un nivel de confianza igual a 5 % y concluímos que nuestros coeficientes son significativos.

```
ggplot(datos4, aes(x = hapiness, y = income)) +
  geom_point(color="skyblue") +
  labs(x="Felicidad", y="Ingresos", title="Modelo de regresión lineal", subtitle="Felicidad vs.
  ↪  Ingresos")+
  stat_smooth(method = "lm", col = "deeppink")+
  theme_classic()
```



La pendiente es positiva y de valor 0.204 , existe una relación de proporcionalidad entre los ingresos y la felicidad, un aumento de ingresos provoca aumento de felicidad.