UNIVERSIDAD CATOLICA BOLIVIANA “SAN PABLO”

MAESTRIA EN CIENCIA DE DATOS, TERCERA VERSION

Materia: ANALISIS ESTADÍSTICO I

Practica No.3

Maestrante: Ramón Wilder Serdán Cárdenas

Noviembre 2021

La Paz – Bolivia

Plantear y resolver 2 ejercicios de los temas avanzados en la clase del 19 y 20 de noviembre.

1. Distribución de probabilidad normal.
   a. El precio promedio de las acciones que pertenecen a S&P500 es de $30 y la desviación estándar es $8.20 (BusinessWeek, Special Annual Issue, primavera de 2003). Suponga que los precios de las acciones están distribuidos normalmente.
      i. ¿Cuál es la probabilidad de que el precio de las acciones de una empresa sea por lo menos de $40?

```
> x<-40
> probZ<-pnorm(x,mean,sd,lower.tail = FALSE)
> #pMayorzZ<-1-probZ
> #pMayorzZ
> palmenos<-probZ
> palmenos
[1] 0.1113249
```

      ii. ¿De que el precio de las acciones de una empresa no sea mayor a $20?

```
> x<-20
> probZ<-pnorm(x,mean,sd,lower.tail = TRUE)
> pAlomas<-probZ
> pAlomas
[1] 0.1113249
```

      iii. ¿De cuánto deben ser los precios de las acciones de una empresa para que esté entre las 10% mejores?

```
> probabilidad <- 0.9
> qnorm(probabilidad,mean,sd,lower.tail = TRUE)
[1] 40.50872
```

   b. El volumen de negociaciones en la Bolsa de Nueva York es más intenso en la primera media hora (en la mañana temprano) y la última media hora (al final de la tarde) de un día de trabajo. A continuación se presentan los volúmenes (en millones de acciones) de 13 días de enero y febrero (214 163 265 194 180 202 198 212 201 174 171 211 211 La distribución de probabilidad de los volúmenes de negociaciones es aproximadamente normal.
      i. Calcule la media y la desviación estándar a usar como estimaciones de la media y de la desviación estándar de la población.

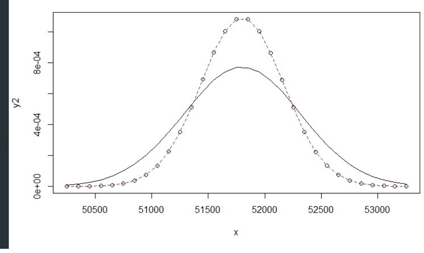
```
> volumenes<-c(214,202,174,163,198,171,265,212,211,194,201,211,180)
> #a)
> media<-mean(volumenes)
> media
[1] 199.6923
> ds<-sd(volumenes)
> ds
[1] 26.03966
```

      ii. ¿Cuál es la probabilidad de que, en un día elegido al azar, el volumen de negociaciones en la mañana temprano sea superior a 180 millones de acciones?

```
> x<-180
> probZ<-pnorm(x,media,ds,lower.tail = FALSE)
> probMayor<-probZ
> probMayor
[1] 0.7752482
```

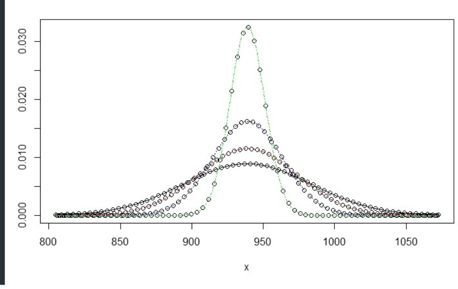
2. Distribuciones muestrales
   a. Al problema de los administradores de EAI. Suponga que se usa una muestra aleatoria simple de 60 administradores. ¿Dibuje la distribución muestral si se emplean muestras aleatorias simples de tamaño 60 y 120?

```
> media<-51800
> n1<-60
> n2<-120
> desv<-4000
> error.est1<-desv/sqrt(n1)
> error.est2<-desv/sqrt(n2)
> x1<-51300
> x2<-52300
> za<-(x1-media)/error.est1
> zb<-(x2-media)/error.est1
> 1-(pnorm(zb)-pnorm(za))
[1] 0.3329216
> # grafico
> x <- seq(media-3*error.est1,media+3*error.est1 , by = 100)
> y1 <- dnorm(x, mean = media, sd = error.est1)
> y2 <- dnorm(x, mean = media, sd = error.est2)
> plot(x,y2)
> lines(x,y1)
> lines(x,y2, col="red",lty=2)
```



b. El costo medio anual de un seguro para automóvil es de $939 (CNBC, 23 de febrero de 2006). Suponga que la desviación estándar es σ = $245. ¿Cuál es la probabilidad de que una muestra aleatoria simple de pólizas de seguros de automóvil la media muestral no difiera más de $25 de la media poblacional si el tamaño de la muestra es 30, 50, 100 y 400?

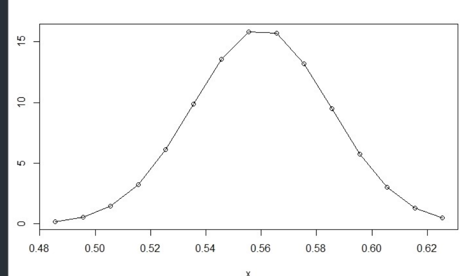
```
> media<-939
> n<-c(30,50,100,400)
> desv<-245
> error.est<-c(desv/sqrt(n))
> q<-25
> x1<-media -q
> x2<-media +q
> z1<-(x1-media)/error.est
> z2<-(x2-media)/error.est
> 1-(pnorm(z2)-pnorm(z1))
[1] 0.57622958 0.47057886 0.30753492 0.04126909
> # grafico
>
> x <- seq(media-3*error.est[1],media+3*error.est[1] , by = 1)
> y <- dnorm(x, mean = media, sd = error.est)
> y1 <- dnorm(x, mean = media, sd = error.est[1])
> y2 <- dnorm(x, mean = media, sd = error.est[2])
> y3 <- dnorm(x, mean = media, sd = error.est[3])
> y4 <- dnorm(x, mean = media, sd = error.est[4])
> plot(x,y)
> lines(x,y1, col="black",pch="o",lty=1)
> lines(x,y2, col="red",pch="*",lty=2)
> lines(x,y3, col="blue",pch="+",lty=3)
> lines(x,y4, col="green",pch="-",lty=4)
```



3. Distribución muestral de $\overline{p}$

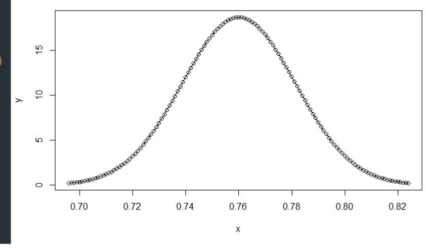
a. Roper ASW realizó una encuesta para obtener información acerca de la opinión de los estadounidenses respecto al dinero y la felicidad (Money, octubre de 2003). Cincuenta y seis por ciento de los entrevistados dijo revisar el estado de su bloc de cheques por lo menos una vez al mes. ¿Cuál es la probabilidad de que la diferencia entre las proporciones muestral y poblacional no sea mayor que 0.04?

```
> w<-0.04
> x1<-p - w
> x2<-p + w
> z1<-(x1-p)/error.est
> z2<-(x2-p)/error.est
> pnorm(z2)-pnorm(z1)
[1] 0.892961
> # grafico
> x <- seq(p-3*error.est[1],p+3*error.est[1] , by = .001)
> y <- dnorm(x, mean = p, sd = error.est)
> y1 <- dnorm(x, mean = p, sd = error.est[1])
> y2 <- dnorm(x, mean = p, sd = error.est[2])
>
> plot(x,y)
> lines(x,y1, col="black",pch="o",lty=1)
> lines(x,y2, col="red",pch="*",lty=2)
> # grafico
> x <- seq(p-3*error.est[1],p+3*error.est[1] , by = .01)
> y <- dnorm(x, mean = p, sd = error.est)
> y1 <- dnorm(x, mean = p, sd = error.est[1])
> y2 <- dnorm(x, mean = p, sd = error.est[2])
> plot(x,y)
```



b. The Grocery Manufacturers of America informa que 76% de los consumidores leen los ingredientes que se enumeran en la etiqueta de un producto. Suponga que la proporción poblacional es p = 0.76 y que de la población de consumidores se selecciona una muestra de 400 consumidores. ¿Cuál es la probabilidad de que la diferencia entre las proporciones muestral y poblacional no sea mayor que 0.03?

```
> w<-0.03
> x1<-p - w
> x2<-p + w
> z1<-(x1-p)/error.est
> z2<-(x2-p)/error.est
> pnorm(z2)-pnorm(z1)
[1] 0.8399427
> # grafico
>
> x <- seq(p-3*error.est[1],p+3*error.est[1] , by = .001)
> y <- dnorm(x, mean = p, sd = error.est)
> y1 <- dnorm(x, mean = p, sd = error.est[1])
> y2 <- dnorm(x, mean = p, sd = error.est[2])
>
> #y3 <- dnorm(x, mean = media, sd = error.est[3])
> #y4 <- dnorm(x, mean = media, sd = error.est[4])
> plot(x,y)
> lines(x,y1, col="black",pch="o",lty=1)
> lines(x,y2, col="red",pch="*",lty=2)
```



4. Intervalos de confianza
   a. En una muestra aleatoria simple con n = 54 la media muestral fue 22.5 y la desviación estándar muestral 4.4. Encuentre un intervalo de confianza de 90%, 95%, 99%.

```
media<-22.5
n<-54
desv<-4.4
NConfianza<-c(0.90,0.91, 0.92, 0.93,0.94,0.95,0.96,0.97,0.98,0.99)
Z<-qnorm((1-NConfianza)/2,lower.tail = FALSE)
error.est <- desv/sqrt(n)
margen.error <- Z * error.est
lim.inf <- media - margen.error#
lim.sup <- media + margen.error#
Medida<-c("Nivel de Confianza","Limite Inferior","Limite Superior")
tabla = data.frame(NConfianza,lim.inf,lim.sup)
tabla
 NConfianza  lim.inf  lim.sup
       0.90 21.51512 23.48488
       0.91 21.48486 23.51514
       0.92 21.45175 23.54825
       0.93 21.41509 23.58491
       0.94 21.37385 23.62615
       0.95 21.32644 23.67356
       0.96 21.27029 23.72971
       0.97 21.20063 23.79937
       0.98 21.10707 23.89293
       0.99 20.95769 24.04231
```

   b. La International Air Transport Association realiza encuestas entre los viajeros de negocios en las que se califica la calidad de los aeropuertos de salida internacional. La calificación máxima es 10. Se seleccionó una muestra aleatoria simple de 50 viajeros de negocios y a cada uno se le pidió su calificación para el aeropuerto internacional de Miami. Las calificaciones que dieron estos 50 viajeros se encuentran en “miami.xlsx”. Calcule el intervalo de confianza de 95% para la media poblacional de las calificaciones al aeropuerto de Miami.

```
> miami <- read_excel ("miami.xlsx")
> media<-mean(miami$calificaciones)
> n<-length(miami$calificaciones)
> desv<-sd(miami$calificaciones)
> NConfianza<-c(0.95)
> Z<-qnorm((1-NConfianza)/2,lower.tail = FALSE)
> error.est <- desv/sqrt(n)
> margen.error <- Z * error.est
> lim.inf <- media - margen.error
> lim.sup <- media + margen.error
> cat("limite inferior: ", lim.inf, ", limite superior: ",lim.sup)
limite inferior:  5.740497 , limite superior:  6.939503
```

5. Pruebas de hipótesis

La rentabilidad anual promedio de los fondos mutualistas U.S. Diversified Equity de 1999 a 2003 fue 4.1% (BusinessWeek, 26 de enero de 2004). Un investigador desea realizar una prueba de hipótesis para ver si los rendimientos de determinados fondos de crecimiento (mid-cap growth funds) difieren de manera significativa del promedio de los fondos U.S. Diversified Equity. En una muestra de 40 fondos de crecimiento el rendimiento medio fue 3.4%. Suponga que por estudios anteriores se sabe que la desviación estándar poblacional de estos fondos de crecimiento es σ = 2%. Use los resultados muestrales para calcular el estadístico de prueba y el valor-p para la prueba de hipótesis.

Por lo tanto, se rechaza a Hipótesis Nula de que la hipótesis es igual a la media poblacional.

```
mediamuestral <- 0.034
desvia <- 0.02
muestra <- 40
media <- 0.041
estadistico <- (mediamuestral - media) / (desvia / sqrt(muestra))
estadistico
-2.213594
pnorm(estadistico)
0.01342835
estadistico
-2.213594
2*pnorm(estadistico)
0.0268567
```