Equipo 9

Proyecto Estadística
Fase 1

Introducción

El objetivo general del informe es argumentar la realización de los ejercicios donde se ponen en práctica las habilidades dadas en conferencia.

1. Análisis de la población Normal

A partir de generar una población normal se seleccionarán 8 muestras(4 con reemplazo y 4 sin reemplazo) con diferentes tamaños mayores que 20, 30,30 y 40 respectivamente, con el

objetivo de comparar y graficar los estadísticos descriptivos obtenidos en las muestras y en la población original.

1.1 Estadísticos Descriptivos de la Población

Primeramente se generó una población normal de tamaño 500 $\mu = 0$ y $\sigma = 1$, recogiéndose los mismos en la siguiente cuadro faltando solamente la moda por tratarse de datos continuos.

```
[1] "Population results"
> descriptiveStatistics(population)
        min         max        mean      median    variance   standDesv     coefVar
-3.68193500  2.82644485  0.05699128  0.12123372  1.04168548  1.02062994 18.27797876
> fivenum(population)
[1] -3.6819350 -0.6884687  0.1212337  0.7282026  2.8264448
> quantile(population)
        0%        25%        50%        75%       100%
-3.6819350 -0.6846786  0.1212337  0.7276103  2.8264448
```

1.2 Estadísticos Descriptivos de la Población

Se tomaron 8 muestras 4 con reemplazo y 4 sin reemplazo

Estadísticos descriptivos de las muestras con reemplazo:

Muestra 1:

```
[1] "Sample for size=25 and REPLACE=TRUE"
> samplesWithReplacement1 <- sample(population, 25, TRUE)
> descriptiveStatistics(samplesWithReplacement1)
       min        max       mean     median   variance  standDesv    coefVar
-2.7518635  2.4163475 -0.1941647 -0.4011475  1.8955740  1.3767985 -9.7627119
> summary(samplesWithReplacement1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.7519 -1.1516 -0.4011 -0.1942  0.5147  2.4163
> quantile(samplesWithReplacement1)
        0%        25%        50%        75%       100%
-2.7518635 -1.1516042 -0.4011475  0.5147433  2.4163475
```

Muestra 2:

```
[1] "Sample for size=60 and REPLACE=TRUE"
> samplesWithReplacement2 <- sample(population, 60, TRUE)
> descriptiveStatistics(samplesWithReplacement2)
         min          max         mean       median     variance    standDesv      coefVar
 -2.03512161   2.82644485  -0.07253089  -0.13882127   0.87052752   0.93302065 -12.00216190
> quantile(samplesWithReplacement2)
        0%        25%        50%        75%       100%
-2.0351216 -0.6830231 -0.1388213  0.5551215  2.8264448
```

Muestra 3:

```
[1] "Sample for size=129 and REPLACE=TRUE"
> samplesWithReplacement3 <- sample(population, 129, TRUE)
> descriptiveStatistics(samplesWithReplacement3)
        min         max        mean      median    variance   standDesv     coefVar
-3.68193500  2.42478625  0.03661206  0.13809277  1.01781149  1.00886644 27.79989387
> quantile(samplesWithReplacement3)
        0%        25%        50%        75%       100%
-3.6819350 -0.6578996  0.1380928  0.6172234  2.4247863
```

Muestra 4:

```
[1] "Sample for size=350 and REPLACE=TRUE"
> samplesWithReplacement4 <- sample(population, 350, TRUE)
> descriptiveStatistics(samplesWithReplacement4)
       min        max       mean     median   variance  standDesv    coefVar
-2.8656151  2.4568510  0.0279136  0.1380928  1.1029773  1.0502273 39.5139798
> quantile(samplesWithReplacement4)
        0%        25%        50%        75%       100%
-2.8656151 -0.7712907  0.1380928  0.7287950  2.4568510
```

Estadísticos descriptivos de las muestras sin reemplazo:

Muestra 5:

```
[1] "Sample for size=25 and REPLACE=FALSE"
> samplesWithoutReplacement1 <- sample(population, 25)
> descriptiveStatistics(samplesWithoutReplacement1)
         min          max         mean       median     variance    standDesv      coefVar
 -1.44941496   2.03963411  -0.06854977  -0.23549160   0.88088146   0.93855286 -12.85024718
> quantile(samplesWithoutReplacement1)
        0%        25%        50%        75%       100%
-1.4494150 -0.7169969 -0.2354916  0.5901093  2.0396341
```

Muestra 6:

```
[1] "Sample for size=60 and REPLACE=FALSE"
> samplesWithoutReplacement2 <- sample(population, 60)
> descriptiveStatistics(samplesWithoutReplacement2)
       min        max       mean     median   variance  standDesv    coefVar
-1.8102102  2.3284059  0.1558302  0.2150612  0.8848640  0.9406721  5.6783858
> quantile(samplesWithoutReplacement2)
        0%        25%        50%        75%       100%
-1.8102102 -0.4726132  0.2150612  0.6355280  2.3284059
```

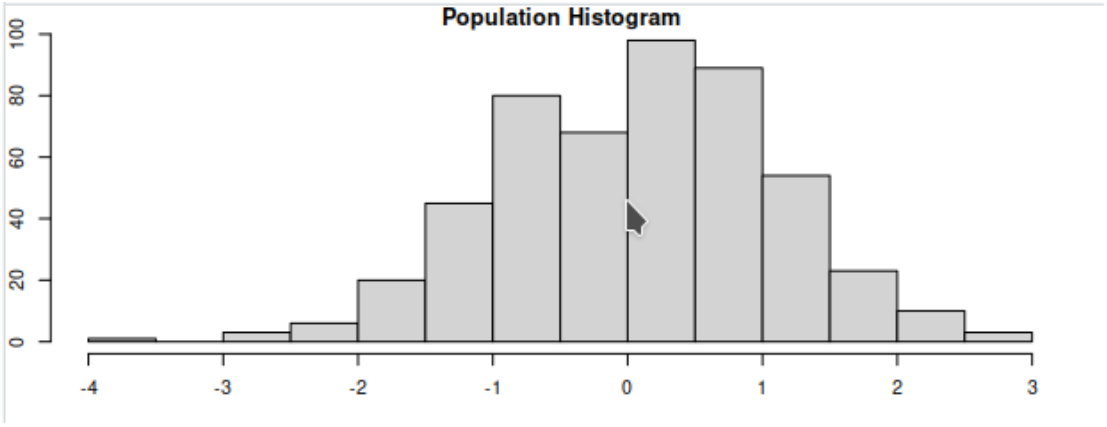
Muestra 7:

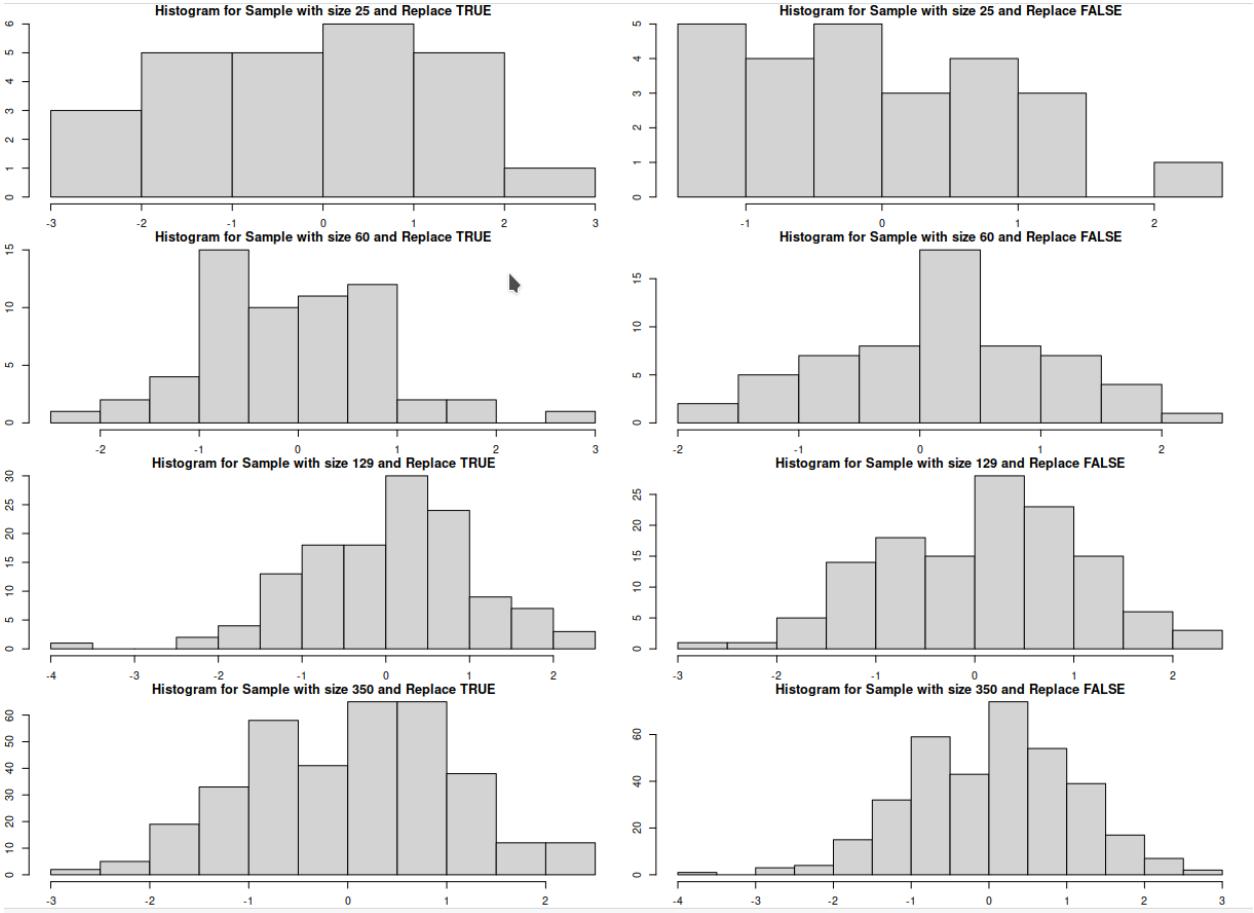
```
[1] "Sample for size=129 and REPLACE=FALSE"
> samplesWithoutReplacement3 <- sample(population, 129)
> descriptiveStatistics(samplesWithoutReplacement3)
        min         max        mean      median    variance   standDesv     coefVar
-2.75186349  2.41634750  0.06634387  0.17497125  1.02260870  1.01124117 15.41376234
> quantile(samplesWithoutReplacement3)
        0%        25%        50%        75%       100%
-2.7518635 -0.7589969  0.1749713  0.7213425  2.4163475
>
```

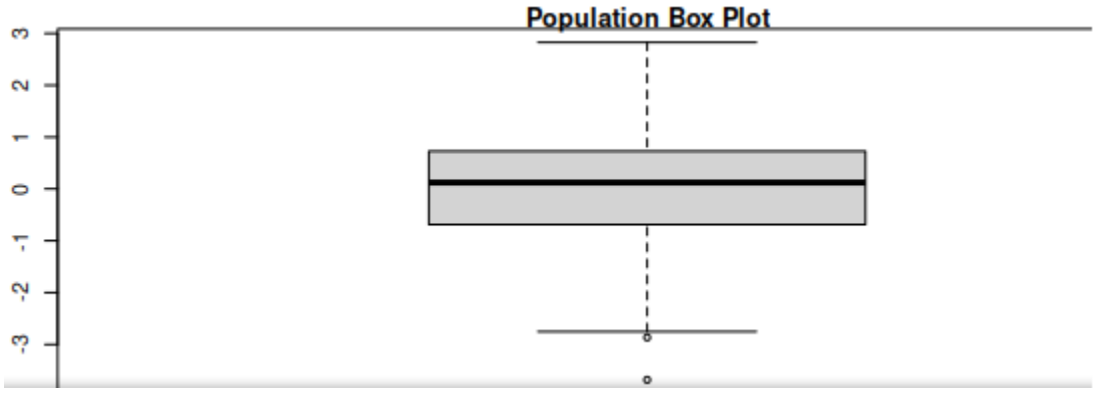
Muestra 8:

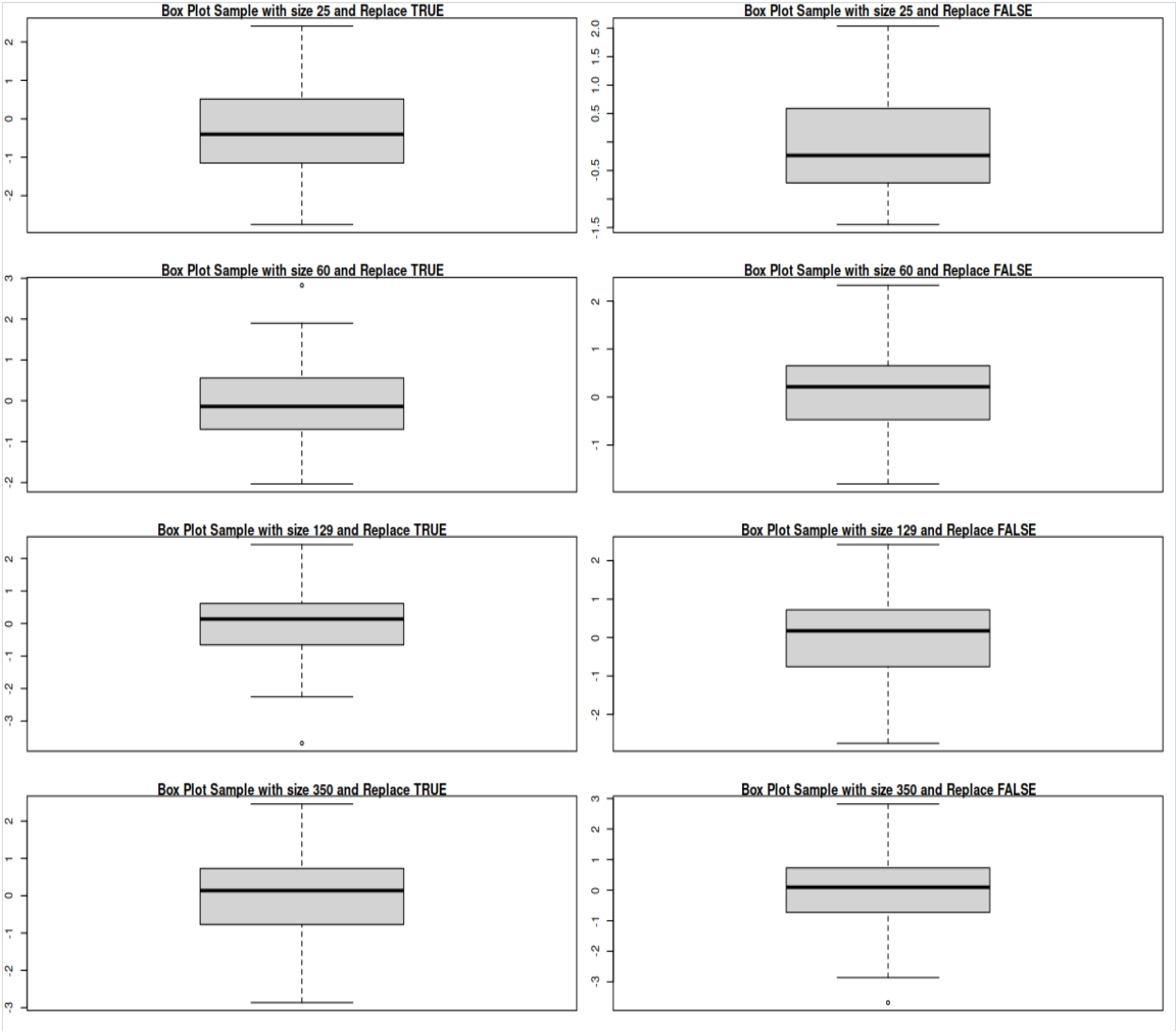
```
[1] "Sample for size=350 and REPLACE=FALSE"
> samplesWithoutReplacement4 <- sample(population, 350)
> descriptiveStatistics(samplesWithoutReplacement4)
        min         max        mean      median    variance   standDesv     coefVar
-3.68193500  2.81866909  0.03460122  0.09361822  1.09666688  1.04721864 31.69445516
> quantile(samplesWithoutReplacement4)
         0%         25%         50%         75%        100%
-3.68193500 -0.72696263  0.09361822  0.72879496  2.81866909
```

1.3 Gráficos de los resultados(Histogramas de frecuencia y diagrama de cajas)




Histogram for Sample with size 25 and Replace TRUE
Histogram for Sample with size 25 and Replace FALSE
Histogram for Sample with size 60 and Replace TRUE
Histogram for Sample with size 60 and Replace FALSE
Histogram for Sample with size 129 and Replace TRUE
Histogram for Sample with size 129 and Replace FALSE
Histogram for Sample with size 350 and Replace TRUE
Histogram for Sample with size 350 and Replace FALSE


Population Box Plot



1.4 Intervalos de Confianza para la media y la variancia de cada muestra

Los datos a continuación se calcularon con un nivel de confianza del 95%

```
[1] "Confidence Intervals for Sample with size 25 and Replace TRUE"
> c("Mean:", meanConfidenceInterval(samplesWithReplacement1))
[1] "Mean:"              "-0.762479181787802" "0.374149777713114"
> c("Variance: ", varianceConfidenceInterval(samplesWithReplacement1))
[1] "Variance: "       "1.155718122318"   "3.66851270973016"
>
> paste("Confidence Intervals for Sample with size", length(samplesWithoutReplacement1), "and Replace FALSE")
[1] "Confidence Intervals for Sample with size 25 and Replace FALSE"
> c("Mean:", meanConfidenceInterval(samplesWithoutReplacement1))
[1] "Mean:"              "-0.455965346286591" "0.318865810058873"
> c("Variance: ", varianceConfidenceInterval(samplesWithoutReplacement1))
[1] "Variance: "        "0.537067213322228" "1.704773733322"
>
> paste("Confidence Intervals for Sample with size", length(samplesWithReplacement2), "and Replace TRUE")
[1] "Confidence Intervals for Sample with size 60 and Replace TRUE"
> c("Mean:", meanConfidenceInterval(samplesWithReplacement2))
[1] "Mean:"              "-0.308613351989738" "0.163551565341561"
> c("Variance: ", varianceConfidenceInterval(samplesWithReplacement2))
[1] "Variance: "        "0.625459649550117" "1.29497519231618"
>
> paste("Confidence Intervals for Sample with size", length(samplesWithoutReplacement2), "and Replace FALSE")
[1] "Confidence Intervals for Sample with size 60 and Replace FALSE"
> c("Mean:", meanConfidenceInterval(samplesWithoutReplacement2))
[1] "Mean:"               "-0.0821883074953893" "0.393848705662254"
> c("Variance: ", varianceConfidenceInterval(samplesWithoutReplacement2))
[1] "Variance: "        "0.635760161227407" "1.31630175926554"
>
> paste("Confidence Intervals for Sample with size", length(samplesWithReplacement3), "and Replace TRUE")
[1] "Confidence Intervals for Sample with size 129 and Replace TRUE"
> c("Mean:", meanConfidenceInterval(samplesWithReplacement3))
[1] "Mean:"              "-0.137483181448126" "0.210707309086594"
> c("Variance: ", varianceConfidenceInterval(samplesWithReplacement3))
[1] "Variance: "        "0.808143993491349" "1.32162395549002"
>
> paste("Confidence Intervals for Sample with size", length(samplesWithoutReplacement3), "and Replace FALSE")
[1] "Confidence Intervals for Sample with size 129 and Replace FALSE"
> c("Mean:", meanConfidenceInterval(samplesWithoutReplacement3))
[1] "Mean:"              "-0.10816116762446" "0.240848914936494"
> c("Variance: ", varianceConfidenceInterval(samplesWithoutReplacement3))
[1] "Variance: "       "0.81195298844757" "1.32785311640816"
>
> paste("Confidence Intervals for Sample with size", length(samplesWithReplacement4), "and Replace TRUE")
[1] "Confidence Intervals for Sample with size 350 and Replace TRUE"
> c("Mean:", meanConfidenceInterval(samplesWithReplacement4))
[1] "Mean:"               "-0.0821129190552388" "0.137940115021839"
> c("Variance: ", varianceConfidenceInterval(samplesWithReplacement4))
[1] "Variance: "       "0.95601613949891" "1.28682827591965"
>
> paste("Confidence Intervals for Sample with size", length(samplesWithoutReplacement4), "and Replace FALSE")
[1] "Confidence Intervals for Sample with size 350 and Replace FALSE"
> c("Mean:", meanConfidenceInterval(samplesWithoutReplacement4))
[1] "Mean:"               "-0.0751100960265683" "0.14431254003796"
> c("Variance: ", varianceConfidenceInterval(samplesWithoutReplacement4))
[1] "Variance: "        "0.950546481469069" "1.27946594141335"
```

1.5 Intervalos de Confianza para la media y la variancia de cada muestra

Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí

2. De acuerdo a sus set de datos:

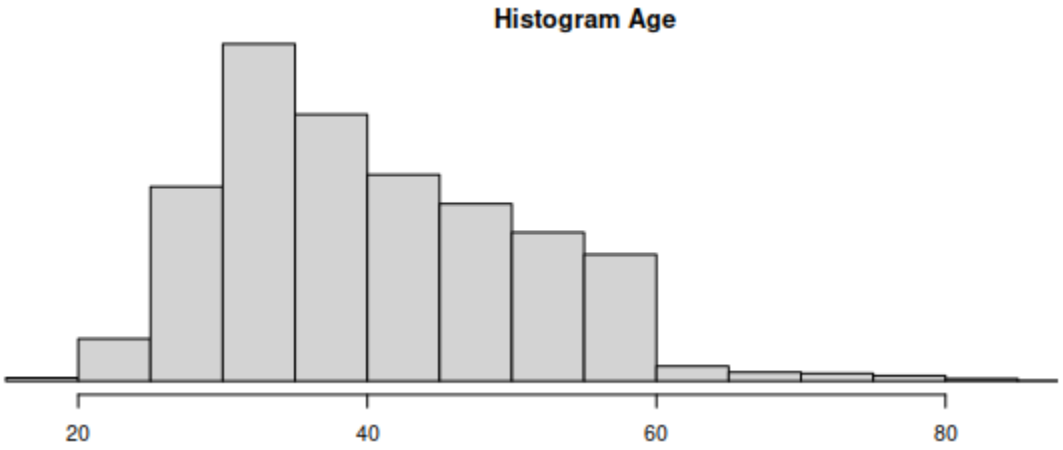
Lo primero para aclarar, es que los planes e ideas para captar o suscribir clientes no funcionan de la misma forma en todas las empresas o bancos.

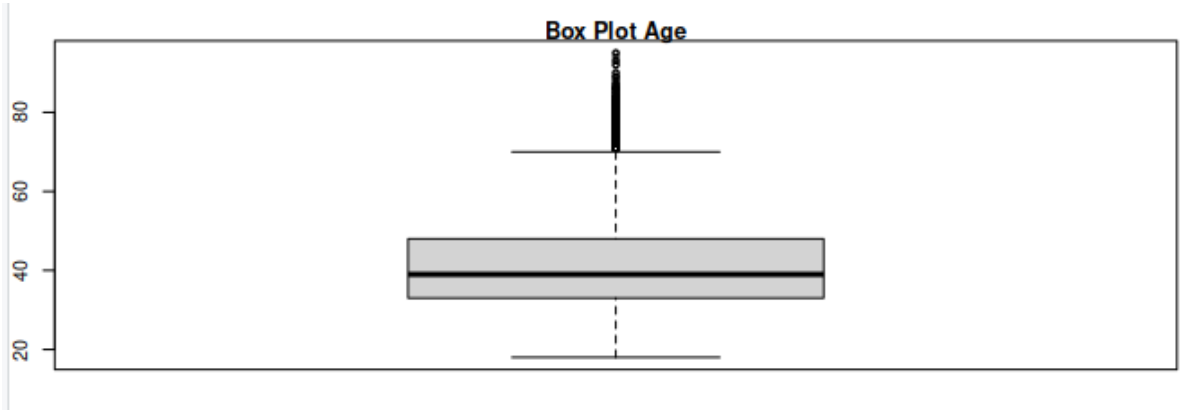
Entonces si la pregunta es **cómo suscribir clientes, la respuesta va a depender del tamaño y tipo del banco, presupuesto, capital humano y tecnológico**, entre otros factores.:

2.1 Estadísticos descriptivos de los 3 más importantes(explicar importancia):

Teniendo en cuenta lo anterior decidimos optar por por los campos “job”, “age” y “education” cuyos estadísticos se pueden obtener al correr el código.

2.2 Graficar Resultados





2.2 Interpretar Resultados en términos del problema

3. Existen diferencias en la duración promedio de las campanas que siguen las personas casadas contra las que siguen las personas solteras?

(Error guardándose)Para solucionar el problema deberíamos comparar la media de la duración de las campañas que siguen las personas casadas contra las que siguen las solteras.

Pero si analizamos el problema podemos percatarnos que estamos en presencia de dos muestras de una población en la que se quiere comparar los valores de parámetros de estas muestras, en nuestro caso la duración de las campañas, y para esto en la práctica se realizan las pruebas de hipótesis con comparación como vimos en la conferencia 3.

Ahora teniendo en cuenta que no conocemos la varianza, antes necesitamos realizar una prueba de igualdad contra la diferencia en las varianzas y dependiendo del resultado de la misma realizamos la prueba de medias con la varianza desconocida correspondiente.

Entonces:

Sea y las varianzas correspondientes a la duración de las campañas que siguen las personas solteras y casadas.

Planteamos, la prueba de hipótesis para la comparación de las varianzas

Calculamos los estimadores puntuales para las varianzas:

Luego fijamos un el cual será nuestro nivel de significación.

Calculamos el estadígrafo

Elegimos la región crítica

o ,

Siendo los tamaños de la muestra de personas solteras y casadas correspondiente.

Y luego comprobamos si es posible rechazar la hipótesis nula para α=0.05

Ahora, para facilitar los cálculos, en R contamos con una función var.test() que recibe como parámetros las muestras y nos devuelve un p-value el cual es el menor nivel de significación para el cual la hipótesis nula es rechazada y si además este p-value es menor que α=0.05 entonces se cumple la región crítica.

Al calcularlo obtenemos un p-value de 4.262e-08 = 0.000000043 < 0.05 = α , por lo que se cumple que las varianzas son iguales para un nivel de significación de 0.05.

Luego pasamos a realizar el de las medias con varianzas desconocidas iguales.

Sean y los parámetros de las poblaciones normales de personas casadas y solteras correspondientes

Planteamos la prueba de hipótesis para este caso

Calculamos los estimadores puntuales para las varianzas

El estadígrafo quedaría de la siguiente manera:

Elegimos como región crítica

En R podemos encontrar una función llamada t.test() el cual recibe las muestras a analizar, el valor calculado en la varianza y una alternativa (parámetro opcional), en nuestro caso le pasamos “less”.

Comparamos el p-value obtenido 2.296602e-07 = 0.0000002297 < 0.05 = α.

El p-value es menor, por tanto, podemos decir que se rechaza la hipótesis nula. Esto quiere decir en términos de nuestro problema que el promedio de duración de las campañas que siguen las personas casadas es menor que el que siguen las personas solteras, con un nivel de significación del 5%.

Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí Inserta tu texto aquí.

Integrantes

- Juan José López Martínez
- Juan Carlos Esquivel Lamis
- Yandy Sanchez Orosa

Colaboración

Juan José realizó el ejercicio 1 excepto el 1.4 Incluyendo la parte del Informe del 1.1, 1.2, 1.5 y 2.3

Yandy realizó el ejercicio 1.4 y el ejercicio 2(Incluyendo la parte del Informe 1.3, 1.4, 2.1, 2.2)

Ejercicio 3 Juan Carlos Incluyendo la parte del Informe