

# Diseño de Experimentos

## Tarea 1

Andrey Arguedas Espinoza - 2020426569

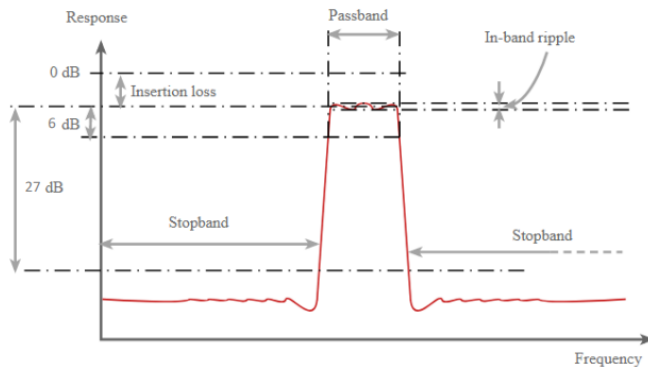
**Abstract**—Resolución de la tarea 1 del curso de Diseño de Experimentos - Tecnológico de Costa Rica.

**Index Terms**—Experimentos, Datos, Gráficos, Histogramas, Cajas, R.

Para este fin se tomó un mismo lote de materias primas y se dividió en 6 partes, la primera parte es el "lote de control" y se procesó en un equipo ya calificado, es decir, ya se utiliza en producción, este lote servirá como patrón para realizar la comparación. Las otras cinco partes corrieron en los equipos nuevos, un lote por equipo. Todos los 6 lotes fueron probados eléctricamente al 100%, obteniendo datos eléctricos para realizar una comparación estadística.

### I. INTRODUCCIÓN

La empresa RF Electronics Ltd, produce filtros de radio frecuencia tipo pasa-banda, abajo se muestra la respuesta típica de uno de estos filtros:



RF Electronics Ltd adquirió cinco nuevos equipos para su planta de manufactura en Costa Rica, dos de los equipos nuevos son marca VTek y los otros tres son marca Ismeca, antes de utilizarlas para producción se debe asegurar que los dispositivos ensamblados en cualquier máquina nueva se comporten de manera similar a los dispositivos que ya se producen.

Estos equipos nuevos se necesitan con mucha urgencia pues los clientes están requiriendo más productos y la empresa quiere aprovechar esta "bonanza", se hizo una inversión de varios millones de dólares para adquirirlos y hay mucha presión por parte de la alta gerencia para ponerlos a producir cuanto antes.

A los filtros producidos por esa empresa se le miden eléctricamente más de 30 parámetros (frecuencia central, pérdida por inserción, ancho de banda, rizo en la banda de paso, entre varios otros). El parámetro crítico del filtro que se escogió para realizar la comparación es uno que tiene que ver con la atenuación en la banda de rechazo (stopband) superior, pues es el parámetro que más causa problemas a los clientes.

#	Nombre	Descripción
1	Control	Orden de control, corrió en un equipo ya calificado
2	Exp 1	Orden experimental 1, corrió en el equipo nuevo #1
3	Exp 2	Orden experimental 2, corrió en el equipo nuevo #2
4	Exp 3	Orden experimental 3, corrió en el equipo nuevo #3
5	Exp 4	Orden experimental 4, corrió en el equipo nuevo #4
6	Exp 5	Orden experimental 5, corrió en el equipo nuevo #5

Usted debe analizar los datos obtenidos y dar su dictamen sobre si los equipos nuevos producen dispositivos estadísticamente similares a los producidos por los equipos ya calificados y en uso en la fábrica (representados en este experimento por el lote de control). Al final de la tarea usted debe responder asertivamente, respaldado en un análisis de datos la siguiente pregunta.

### II. PROCEDIMIENTO

#### 1) Inicialización:

##### a) Inclusión de paquetes y librerías a utilizar:

```

4  if(!require(psych)){install.packages("psych")}
5  if(!require(FSA)){install.packages("FSA")}
6  if(!require(Rmisc)){install.packages("Rmisc")}
7  if(!require(ggplot2)){install.packages("ggplot2")}
8  if(!require(rcompanion)){install.packages("rcompanion")}
9  if(!require(car)){install.packages("car")}
10 if(!require(multcompview)){install.packages("multcompview")}
11 if(!require(multcomp)){install.packages("multcomp")}
12 if(!require(lsmmeans)){install.packages("lsmmeans")}
13
14 library(rcompanion)
15 library(ggplot2)
16 library(car)
17 library(repr)

```

## b) Carga de los datos desde el csv:

```

19 # Directorio donde se encuentra el archivo
20 setwd(this.path::here())
21
22 # Se leen los datos y se guardan en la variable Data
23 my_data <- read.csv("Datos tarea 1.csv")
24 Data = my_data
25
26 # Se eliminan los datos originales de memoria
27 rm(my_data)
28
29 headTail(Data)

```

## c) Datos originales cargados:

```

1      Date      Time Employee      part_test_id      Lot      real_barcode_batch Sublot      start_datetime      manifest_rxture Analyzer_id      test_step      sequence      falcode S21..1      Stpbnd 2400.2482..S21..1
2 1/10/2017 18:18:47 30926 QP2502 -cna- Control AL50549 215819 8 2017-05-18 18:17:55 C322 4 100408 PT 342 0 26.11
3 1/10/2017 18:18:47 30926 QP2502 -cna- Control AL50549 215819 8 2017-05-18 18:17:55 C322 4 100408 PT 343 0 27.06
4 1/10/2017 18:18:42 30926 QP2502 -cna- Control AL50549 215819 7 2017-05-18 18:17:44 C322 4 100408 PT 509 0 27.79

```

## 2) Modificación del nombre de columnas (facilidad para trabajar):

## a) Modificación de nombre a columna a estudiar:

```

29 #Modificamos el nombre de la columna para que sea más sencillo trabajar
30 names(Data)[names(Data) == "Stpbnd.2400.2482..S21..1"] <- "Stpbnd"
31
32 headTail(Data)

```

```

1      Date      Time Employee      part_test_id      Lot      real_barcode_batch Sublot      start_datetime      manifest_rxture Analyzer_id      test_step      sequence      falcode S21..1      Stpbnd
2 1/10/2017 18:18:47 30926 QP2502 -cna- Control AL50549 215819 8 2017-05-18 18:17:55 C322 4 100408 PT 342 0 26.11
3 1/10/2017 18:18:47 30926 QP2502 -cna- Control AL50549 215819 8 2017-05-18 18:17:55 C322 4 100408 PT 343 0 27.06
4 1/10/2017 18:18:42 30926 QP2502 -cna- Control AL50549 215819 7 2017-05-18 18:17:44 C322 4 100408 PT 509 0 27.79

```

## 3) Obtenga para cada lote: tamaño de muestra, mínimo, máximo, rango, media, promedio y desviación estándar

```

34 #Informacion basica
35 Summarize(Stpbnd ~ Lot, data=Data, digits=4)
36
26/46 (Top Level)

```

```

Console Terminal Background Jobs
R 4.2.2 - C:/Users/Andrey/Desktop/ExperimentDesign/Tareas/Tarea 1/
> #Informacion basica
> Summarize(Stpbnd ~ Lot, data=Data, digits=4)
  Lot      n mean      sd      min      Q1 median      Q3      max
1 Control 21039 28.0186 2.0786 12.932 27.6325 27.808 27.988 49.570
2 Exp 1 42479 26.4612 2.0947  4.258 25.9710 26.236 26.524 49.307
3 Exp 2 46350 26.1855 2.0192 12.644 25.8810 26.057 26.223 54.794
4 Exp 3 44311 25.5702 1.8834  4.315 25.2220 25.398 25.579 49.690
5 Exp 4 45805 26.5679 1.9331 13.495 26.2740 26.461 26.619 55.897
6 Exp 5 44700 25.5246 1.9478 13.269 25.1590 25.334 25.510 49.982
>

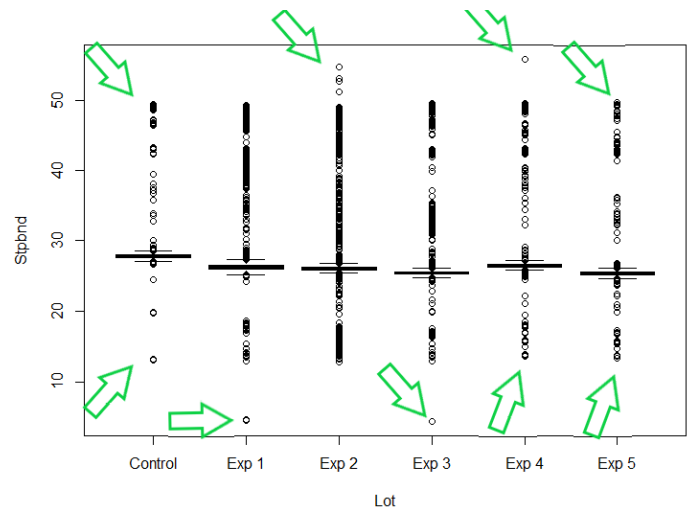
```

## 4) Con la información del punto anterior responda: ¿hay outliers?, en caso de haberlos, elimínelos y vuelva a calcular los indicadores del punto anterior. Viendo los datos podemos ver que para todos los lotes se encuentran los valores minimos y maximos a mucha más distancia del promedio comparados a lo que tenemos como desviacion estandar, solo con eso podemos saber que sí hay outliers. Para verlos aun más facilmente podemos graficarlos:

```

37 # Grafico de cajas y bigotes de los datos originales
38 boxplot(Stpbnd ~ Lot, data = Data)

```



Lot

Seguido vamos a eliminar los outliers y volver a graficar sin estos:

```

49 #Encontrar los outliers
50 outliers_control <- boxplot(control$Stpbnd, plot=FALSE)$out
51 outliers_lot1 <- boxplot(lot1$Stpbnd, plot=FALSE)$out
52 outliers_lot2 <- boxplot(lot2$Stpbnd, plot=FALSE)$out
53 outliers_lot3 <- boxplot(lot3$Stpbnd, plot=FALSE)$out
54 outliers_lot4 <- boxplot(lot4$Stpbnd, plot=FALSE)$out
55 outliers_lot5 <- boxplot(lot5$Stpbnd, plot=FALSE)$out
56
57 #Eliminando los outliers
58 control_no_outliers <- control[which(control$Stpbnd %in% outliers_control),]
59 lot1_no_outliers <- lot1[which(lot1$Stpbnd %in% outliers_lot1),]
60 lot2_no_outliers <- lot2[which(lot2$Stpbnd %in% outliers_lot2),]
61 lot3_no_outliers <- lot3[which(lot3$Stpbnd %in% outliers_lot3),]
62 lot4_no_outliers <- lot4[which(lot4$Stpbnd %in% outliers_lot4),]
63 lot5_no_outliers <- lot5[which(lot5$Stpbnd %in% outliers_lot5),]
64
65 #Unir todos los datos pero sin outliers
66 data_no_outliers <- rbind(control_no_outliers, lot1_no_outliers, lot2_no_outliers, lot3_no_outliers,
67 lot4_no_outliers, lot5_no_outliers)
68
69
70 #Visualizacion de los datos sin outliers
71 options(repr.plot.width=20, repr.plot.height=20)
72 par(mfrow=c(3,2))
73 boxplot(Stpbnd ~ Lot, data = control_no_outliers, main = "Control")
74 boxplot(Stpbnd ~ Lot, data = lot1_no_outliers, main = "Exp 1")
75 boxplot(Stpbnd ~ Lot, data = lot2_no_outliers, main = "Exp 2")
76 boxplot(Stpbnd ~ Lot, data = lot3_no_outliers, main = "Exp 3")
77 boxplot(Stpbnd ~ Lot, data = lot4_no_outliers, main = "Exp 4")
78 boxplot(Stpbnd ~ Lot, data = lot5_no_outliers, main = "Exp 5")
79

```

El resumen de los datos sin los outliers es:

```

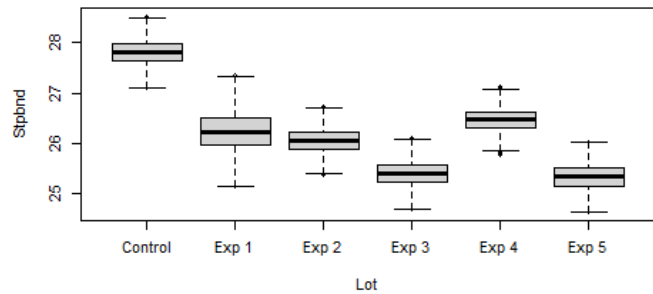
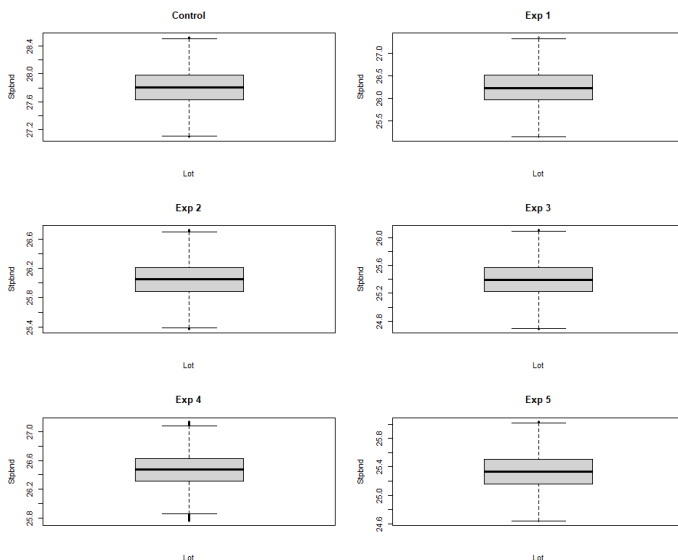
> #Info sobre los datos sin outliers
> Summarize(Stpbnd ~ Lot, data=data_no_outliers, digits=4)

```

	Lot	n	mean	sd	min	Q1	median	Q3	max
1	Control	20602	27.8069	0.2601	27.100	27.633	27.806	27.980	28.521
2	Exp 1	41748	26.2366	0.3593	25.148	25.968	26.229	26.511	27.353
3	Exp 2	44892	26.0450	0.2404	25.370	25.883	26.055	26.213	26.734
4	Exp 3	43482	25.3978	0.2500	24.687	25.221	25.394	25.570	26.114
5	Exp 4	42155	26.4674	0.2257	25.757	26.319	26.479	26.626	27.136
6	Exp 5	43949	25.3329	0.2490	24.633	25.158	25.331	25.503	26.036

Como podemos observar ahora los mínimos y máximos están muy cerca del promedio, además la desviación estandar es mucho menor.

Podemos observar esto más claramente en nuevos gráficos para los datos sin outliers:



5) Analice los datos obtenidos, indique si desde ya observa diferencias significativas contra el lote de control.

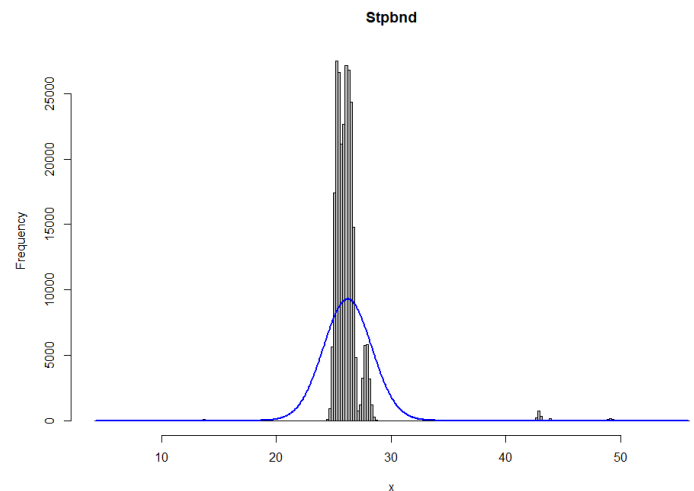
Efectivamente podemos ver las diferencias claras y como el lote de control está mayormente distanciado de los demás, solamente algunos datos (los más altos) del lote Exp1 tienen los mismos valores que el de control. Los demás lotes son visualizados como estadísticamente distintos al lote de control.

6) Grafique un histograma de la especificación Stpbnd(2400-2482)(S21)(1) con la totalidad de los datos del archivo (si ha eliminado outliers, vuelva a incluirlos).

```

86
87 #Histograma de Stpbnd pero con los datos originales
88 plotNormalHistogram(Data$Stpbnd, main = "Stpbnd", breaks = 300)
89

```



7) Comente acerca de la forma del histograma que obtuvo. ¿Puede desde ya concluir algo?

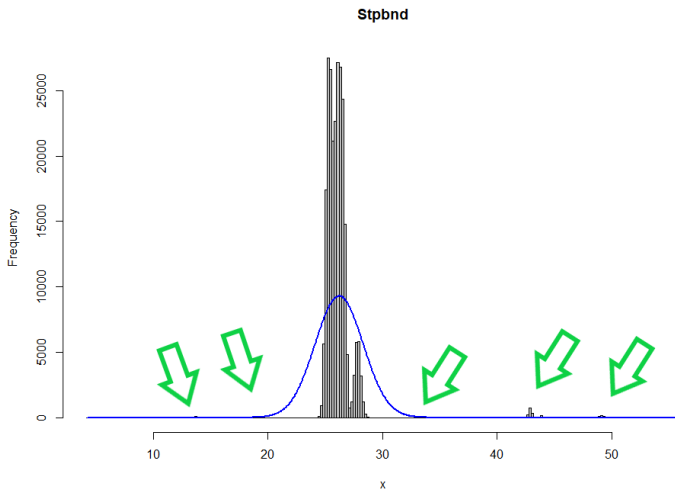
El histograma parece seguir la distribución normal sin embargo se observan zonas fuera de este con bastantes outliers, por lo que sería difícil dar una conclusión con solo este gráfico.

8) **Responda: ¿Se observan outliers?, ¿Cuántos?, ¿Cuál criterio utilizó para definir qué es un outlier?.**

Sí se observan (ver flechas verdes en el histograma), en total son 7856 outliers y se utilizó el criterio de obtenerlos usando el boxplot y los que estuvieran fuera de la zona eran considerados outliers.

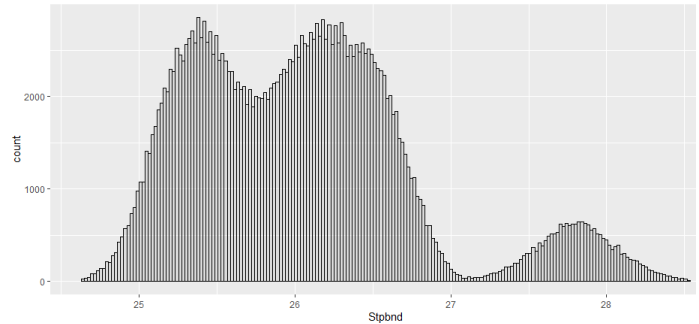
```
40 #Lotes
41 control <- Data[Data$Lot == "Control",]
42 lot1 <- Data[Data$Lot == "Exp 1",]
43 lot2 <- Data[Data$Lot == "Exp 2",]
44 lot3 <- Data[Data$Lot == "Exp 3",]
45 lot4 <- Data[Data$Lot == "Exp 4",]
46 lot5 <- Data[Data$Lot == "Exp 5",]
47 lot6 <- Data[Data$Lot == "Exp 6",]
48
49 #Encontrar los outliers
50 outliers_control <- boxplot(control$Stpbnd, plot=FALSE)$out
51 outliers_lot1 <- boxplot(lot1$Stpbnd, plot=FALSE)$out
52 outliers_lot2 <- boxplot(lot2$Stpbnd, plot=FALSE)$out
53 outliers_lot3 <- boxplot(lot3$Stpbnd, plot=FALSE)$out
54 outliers_lot4 <- boxplot(lot4$Stpbnd, plot=FALSE)$out
55 outliers_lot5 <- boxplot(lot5$Stpbnd, plot=FALSE)$out
```

values	
outliers_control	num [1:437] 49.2 48.8 26.9 49.1 49.3 ...
outliers_lot1	num [1:731] 14 28.9 39.7 40.3 40.3 ...
outliers_lot2	num [1:1458] 25.3 25.4 42.9 43.4 25.3 ...
outliers_lot3	num [1:829] 42.9 43 13.8 48.8 42.7 ...
outliers_lot4	num [1:3650] 25.3 25.7 25.2 25.1 25.6 ...
outliers_lot5	num [1:751] 24.6 47.4 13.6 24.6 42.9 ...



9) **Elimine los outliers (en caso de haber) y luego vuelva a graficar el histograma, ajuste la escala del eje X para sólo observar la parte que contiene datos. ¿Se ve diferente?, ¿Provee ahora menos o más información? ¿Por qué?**

```
89 #Histograma de Stpbnd pero sin outliers
90 ggplot(data_no_outliers, aes(x=Stpbnd, fill=Stpbnd)) +
91   geom_histogram(bins=200, alpha=0.1, position="dodge", colour = "black") +
92   NULL
```



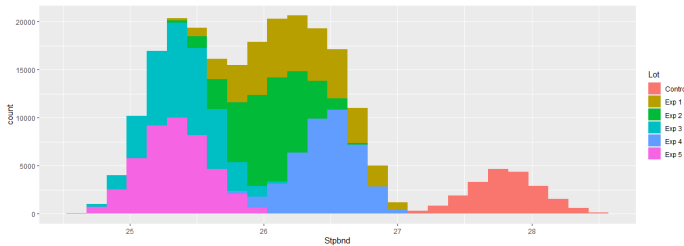
Podemos ver la información de manera más compacta, sin embargo no se ve tan clara forma normal, lo que si es más fácil de entender es que la mayoría de datos se encuentran en el rango de 25 y 27

- 10) Asigne un color por cada lote diferente y obtenga nuevamente el histograma, comente.

```

99 #Colores diferentes para cada lot en el histograma sin outliers
100 options(repr.plot.width=15, repr.plot.height=10)
101 ggplot(data_no_outliers, aes(Stpbnd, fill = Lot)) +
102   geom_histogram(binwidth = 0.15)

```



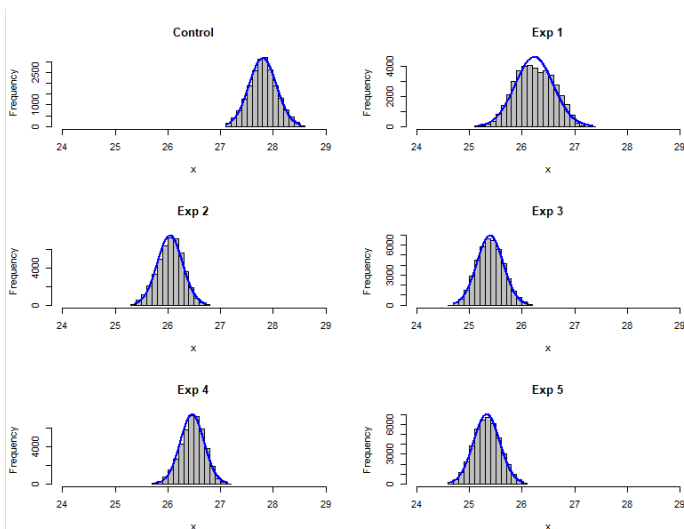
Podemos observar como el lote de control está un poco más distanciado que los demás, además en el eje x vemos como el Exp 3 y el Exp 5 se sitúan en prácticamente la misma región y tienen una forma muy similar lo que cambia es la cantidad de elementos, mientras que el Exp 1, Exp 2 y Exp 4 se colocan un poco más a la derecha del eje x e igualmente comparten una forma similar.

- 11) Como alternativa, puede desplegar los histogramas en “cascada”, un histograma por cada lote por separado, pero puestos en una misma página uno debajo del otro, todos los histogramas deben utilizar el mismo eje X, es decir, todos debe tener exactamente el mismo mínimo y el mismo máximo, comente.

```

106 #Histogramas en cascada
107 options(repr.plot.width=20, repr.plot.height=20)
108 par(mfrow=c(3,2))
109 plotNormalHistogram(control_no_outliers$Stpbnd, xlim = c(24,29))
110 plotNormalHistogram(lot1_no_outliers$Stpbnd, xlim = c(24,29), main = "Exp 1")
111 plotNormalHistogram(lot2_no_outliers$Stpbnd, xlim = c(24,29), main = "Exp 2")
112 plotNormalHistogram(lot3_no_outliers$Stpbnd, xlim = c(24,29), main = "Exp 3")
113 plotNormalHistogram(lot4_no_outliers$Stpbnd, xlim = c(24,29), main = "Exp 4")
114 plotNormalHistogram(lot5_no_outliers$Stpbnd, xlim = c(24,29), main = "Exp 5")

```

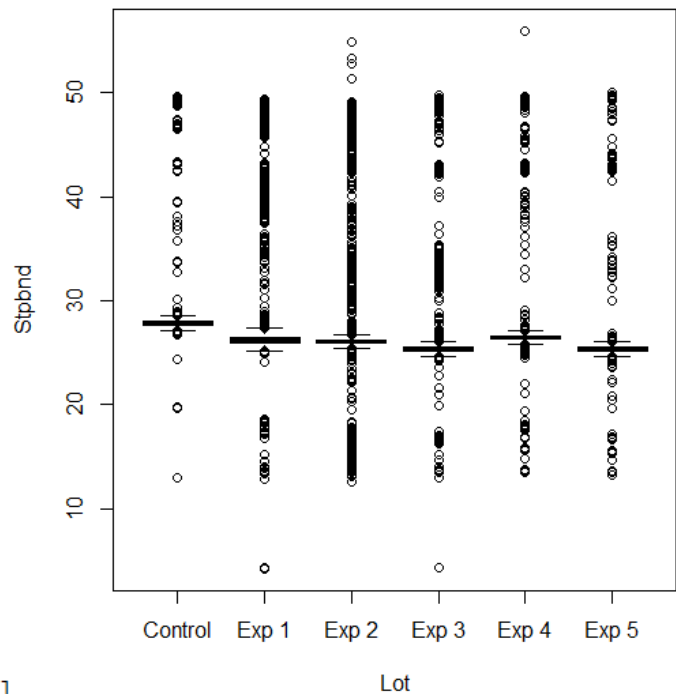


- 12) Obtenga un gráfico de cajas y bigotes, las categorías del eje X serán los nombres de cada lote de evaluación y el eje Y será la especificación Stpbnd(2400-2482)(S21)(1). Utilice todos los datos (incluya los outliers en caso de haberlos eliminado antes).

```

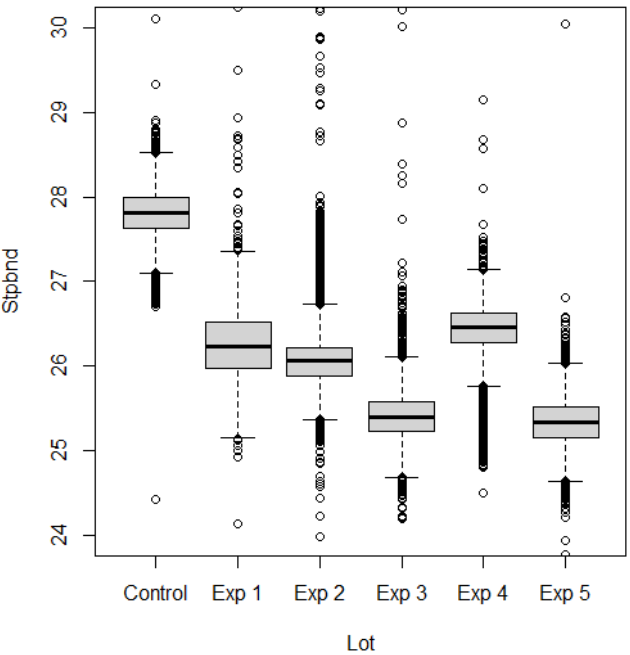
117 par(mfrow=c(1,2))
118 #Grafico de cajas y bigotes incluyendo outliers
119 boxplot(Stpbnd ~ Lot, data = Data)

```

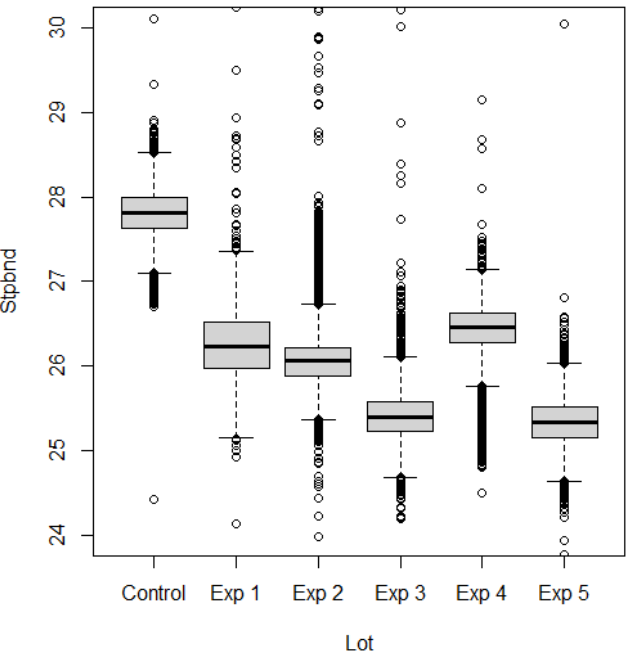
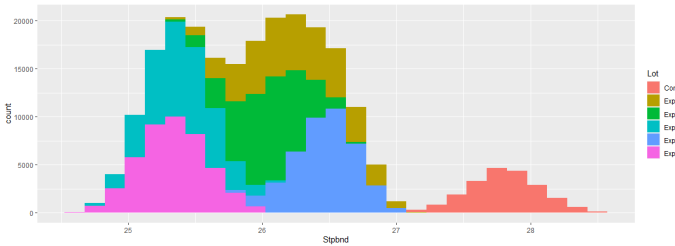


13) Ajuste la escala del eje Y para que tenga un mínimo igual a 24 y un máximo igual a 30.

```
121 #Ajustando el grafico para un min de 24 y max de 30
122 boxplot <- boxplot(Stpbnd ~ Lot, data = Data, ylim=c(24, 30))
```



14) Compare los gráficos del punto 11 y del punto 14, note como el gráfico del punto 14 tiene más información.



- 15) **Realice su conclusión, exprese claramente cuál es su criterio acerca de la aceptación o no de los nuevos equipos.**

En conclusión cualquiera de los equipos puede ser aceptado ya que con todos se obtienen resultados mejores al lote de control estadísticamente hablando.

- 16) **Responda: Cuál método para obtener su conclusión se le hizo más fácil o intuitivo: 1) con los datos numéricos, 2) histogramas, 3) diagrama de cajas y bigotes.**

Personalmente el método de diagrama de cajas y bigotes junto con el del histograma coloreado fueron los que se me hicieron más fáciles de entender.

### III. CONCLUSION

- 1) **¿Cuál o cuáles de los equipos nuevos aprueba usted para ser utilizado para la fabricación de dispositivos?**

Todos los equipos nuevos pueden aprobarse ya que son estadísticamente diferentes al equipo con el que se corrió el lote de control y generan valores más bajos de stopband.

- 2) **Educated guess**

Señor jefe, los lotes experimentales número 3 y 5 corrieron en las máquinas marca VTek, mientras que los lotes experimentales número 1, 2 y 4, corrieron en los equipos marca Ismeca. Este educated guess se hace bajo la observación de los gráficos de los lotes y sus formas tanto como similitudes.