HW3 - ANOVA

Shiloh Bradley

6/10/2020

```
S <- read.csv("restaurant4anova.csv",header=TRUE)
head(S)
```

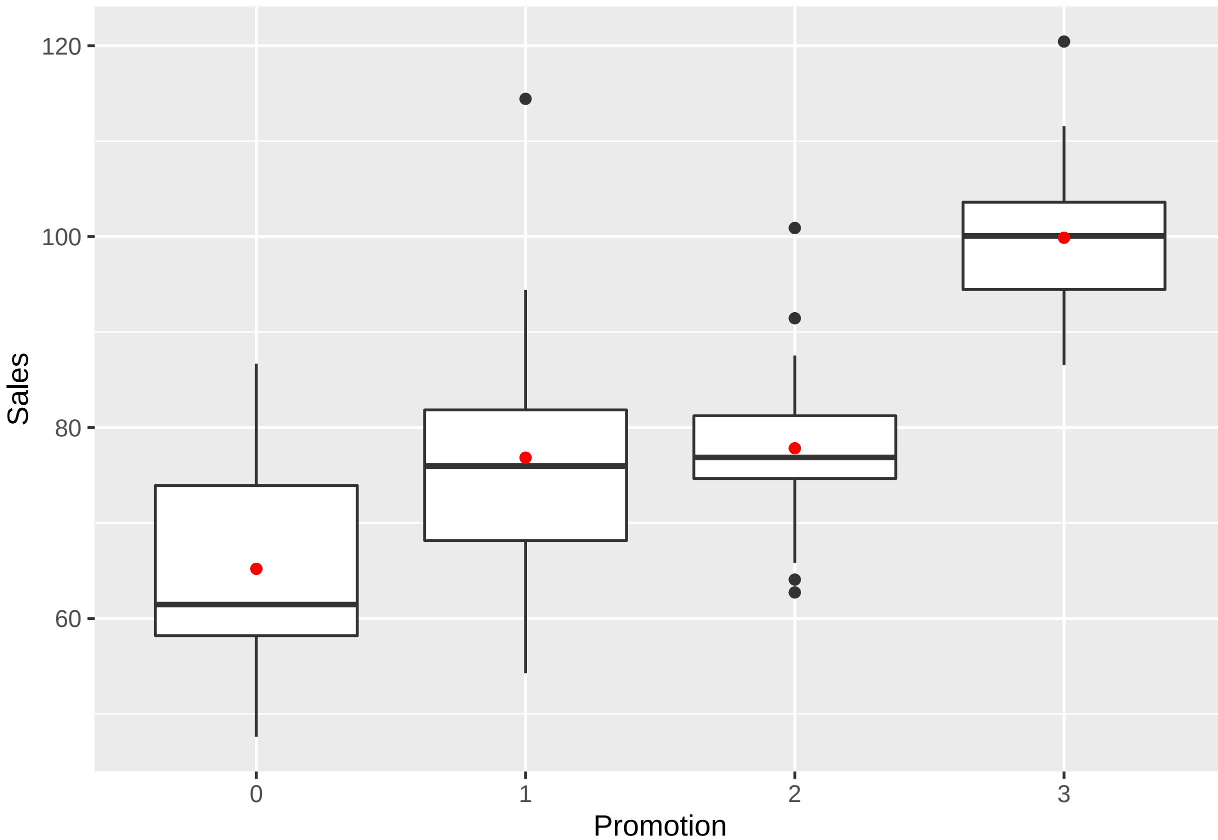
```
##   Promo   Sales
## 1     0 47.6032
## 2     0 48.9218
## 3     0 86.6995
## 4     0 60.9114
## 5     0 74.0193
## 6     0 58.4517
```

```
tail(S)
```

```
##    Promo   Sales
## 80     3  88.961
## 81     3  94.453
## 82     3 100.070
## 83     3  91.999
## 84     3 102.255
## 85     3 102.313
```

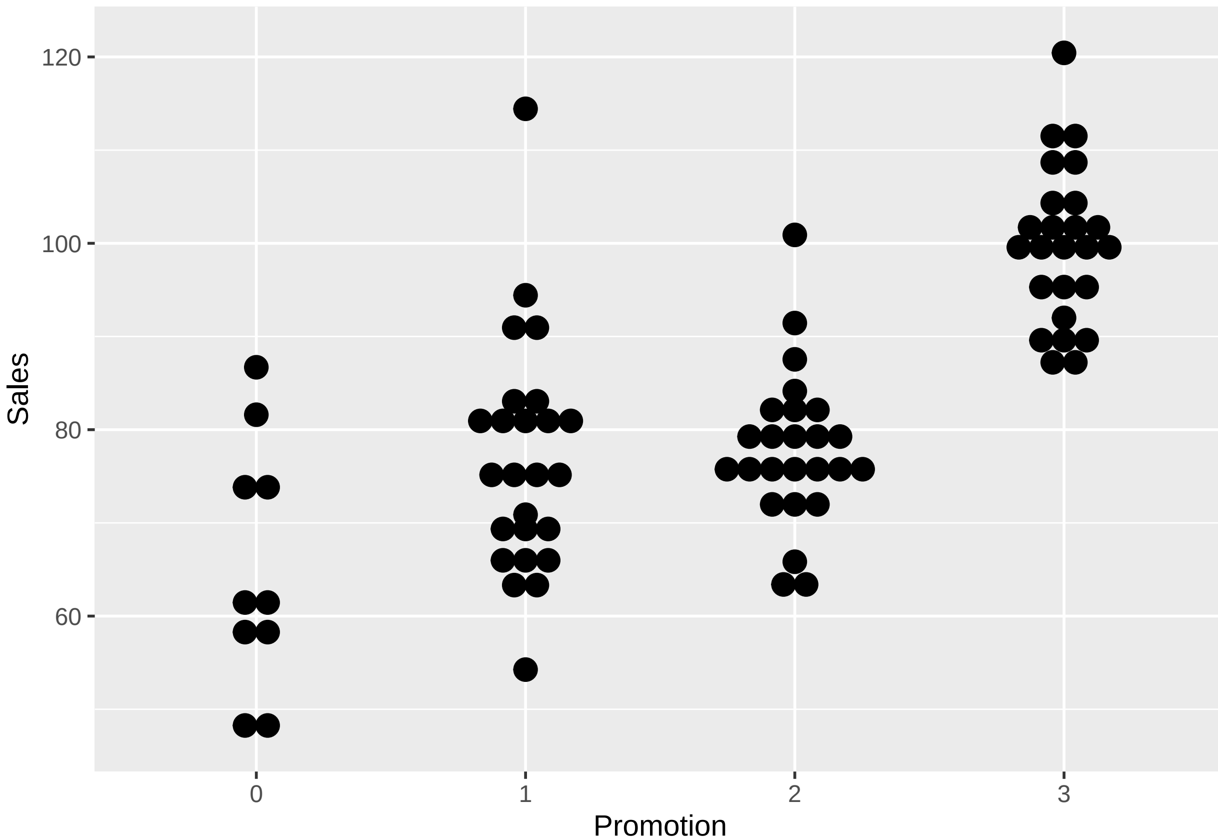
These box plots show us that promo 3 is the best. No promotion is the worst. There is an overlap between these three though. We wil lse

```
BX2 <- ggplot(S,aes(x=factor(Promo),y=Sales)) + geom_boxplot()+
       stat_summary(fun = "mean", geom="point",color="red")
BX2 + xlab("Promotion")
```



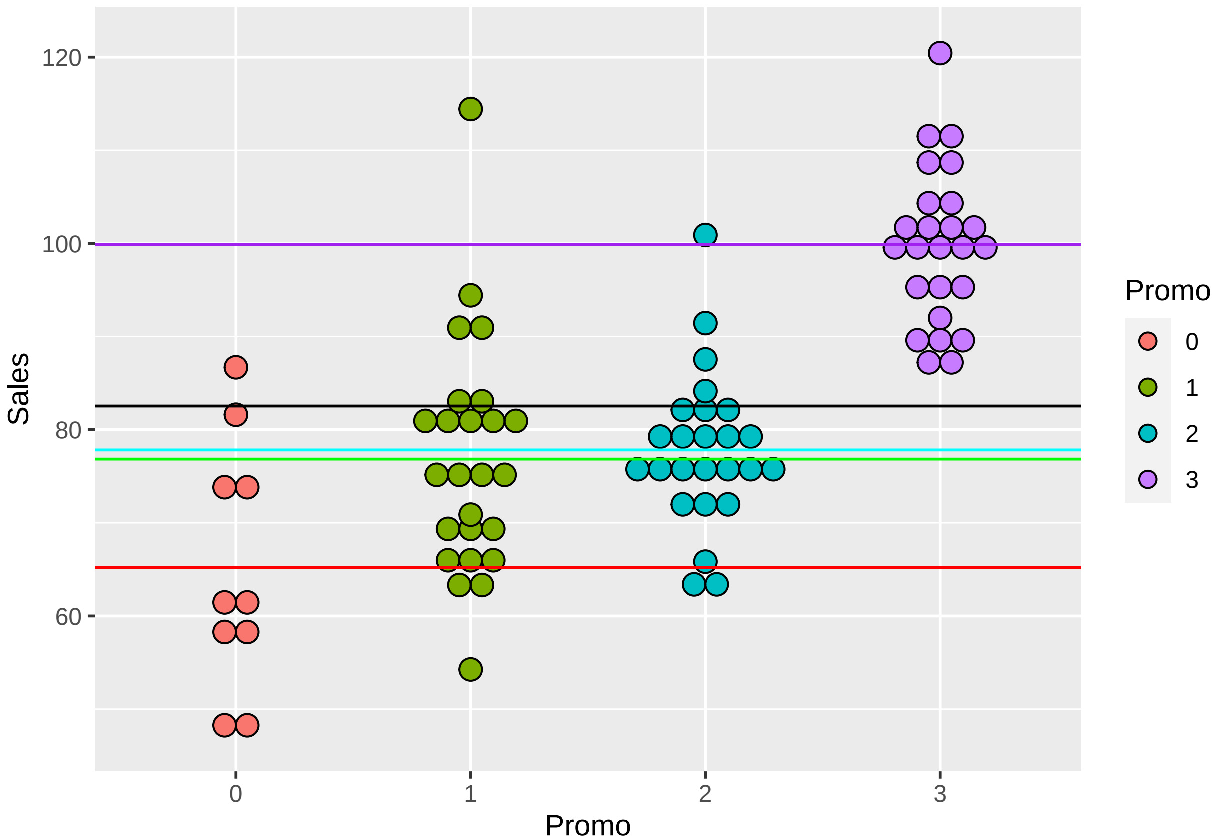

```
p.dot <- ggplot(S,aes(x=factor(Promo),y=Sales)) +
  geom_dotplot(binaxis='y', stackdir='center')
p.dot + xlab("Promotion")
```

```
## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#xbari. 65.19592    76.84204    77.82536    99.882
#xbar.. 82.53757882
S$Promo <- factor(S$Promo)
p.dot <- ggplot(S,aes(x=Promo,y=Sales,fill=Promo)) +
  geom_dotplot(binaxis='y', stackdir='center') +
  geom_hline(yintercept = c(65.19592,76.84204,77.82536,99.882), color=c("red","green","cyan  ","purple")
  geom_hline(yintercept = 82.54)
p.dot
```

```
## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```




```
# summary stats
xbar <- tapply(S$Sales, S$Promo, mean)
SD <- tapply(S$Sales, S$Promo, sd)
round(xbar,2)
```

```
##     0     1     2     3
## 65.20 76.84 77.83 99.88
```

```
round(SD,2)
```

```
##     0     1     2     3
## 13.25 12.44  8.26  8.32
```

Sales is the response variable. Promo is a categorical predictor of Sales.

```
a1 <- aov(data=S,Sales~factor(Promo))
summary(a1) ## Shows us the ANOVA table
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## factor(Promo)  3  11894    3965   37.39 2.87e-15 ***
## Residuals     81   8588     106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
names(a1)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "contrasts"     "xlevels"       "call"          "terms"
## [13] "model"
```

Since P-Value is less than .05, we reject H0 and conclude that not all means are equal. Q-Q plot shows that: residuals plot along the normal line, suggesting residuals are normal. In addition, P-value of Shapiro-Wilks' normality test is > 0.05, so normality is verified. This makes the P-value from F-test accurate.

```
#qq plot with normal line (normality test)
df <- as.data.frame(a1$residuals)
colnames(df)[1] <- "residuals"

shapiro.test(df$residuals) # W = 0.97087, p-value = 0.05123
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$residuals
## W = 0.97087, p-value = 0.05123
```

```
ggplot(df)+stat_qq(aes(sample=residuals)) +
  geom_qq_line(aes(sample=residuals))+
  geom_text(aes(x=0.5, y=-20, label="Shapiro-test p-value = 0.05123"))
```