STA303/1004 - Week 3 R Markdown

January 21-25, 2019

Case Study 1: The Spock Conspiracy Trial Data

Get the data (from R library):

```
#load Sleuth3 R data library; see case0502
library(Sleuth3)
#Juries data
jury = case0502
#attach(jury)
head(jury)
```

1st 6 rows

```
##   Percent   Judge
## 1     6.4 Spock's
## 2     8.7 Spock's
## 3    13.3 Spock's
## 4    13.6 Spock's
## 5    15.0 Spock's
## 6    15.2 Spock's
```

```
Percent=jury$Percent
Judge=jury$Judge
```

Case Study 1: How many venires for each Judge?

factor

```
table(Judge)
```

```
## Judge
##        A      B      C      D      E      F Spock's
##        5      6      9      2      6      9      9
```

R function

```
with(jury, tapply(Percent,Judge,mean))
```

X factor

```
##        A        B        C        D        E        F  Spock's
## 34.12000 33.61667 29.10000 27.00000 26.96667 26.80000 14.62222
```

$\bar{y}_A$ $\bar{y}_B$

$\hat{\beta}_1 = \bar{y}_B - \bar{y}_A$

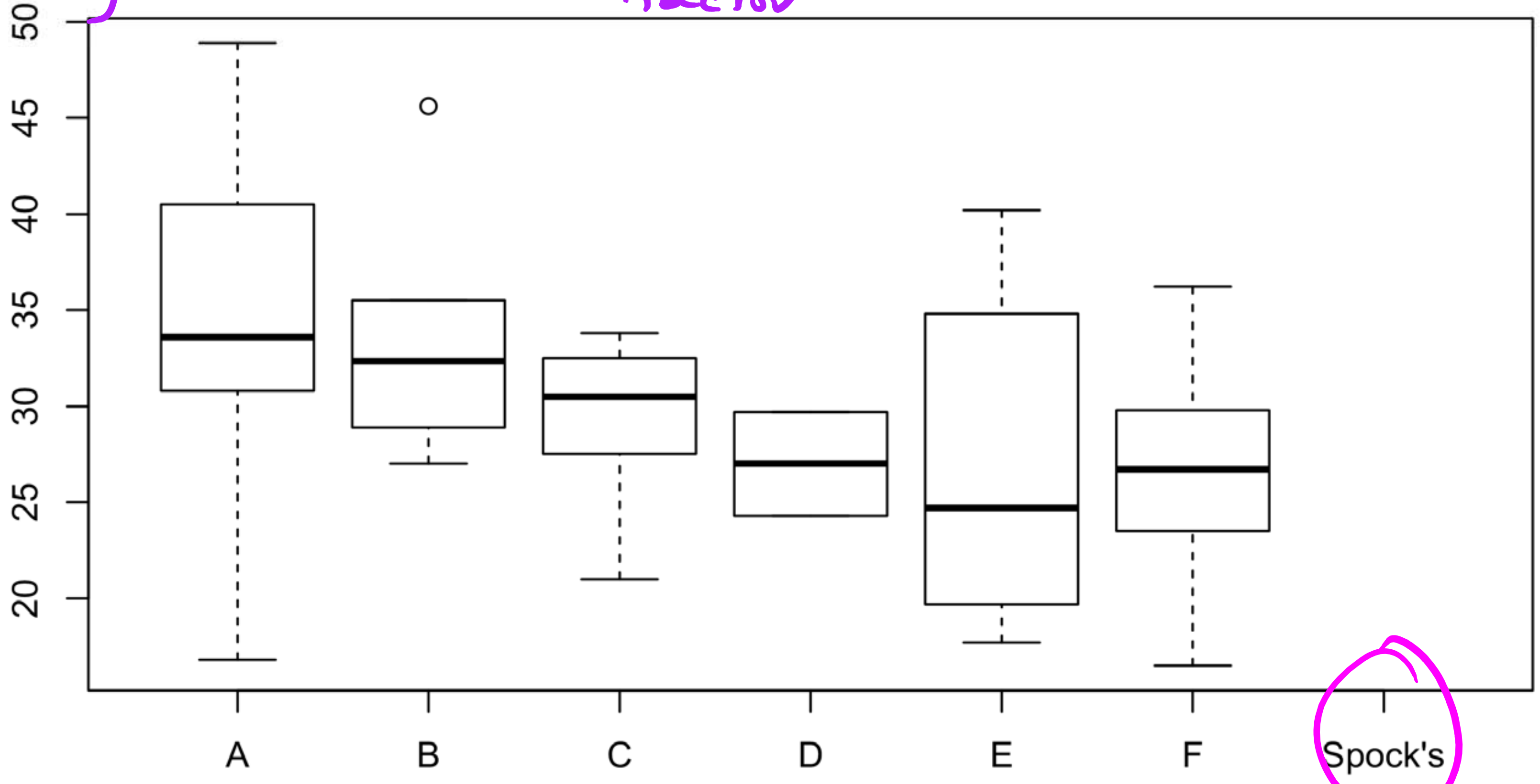
$E(y_i) = \beta_0 + \beta_i$

$\hat{y}_i = \bar{y}_i = b_0 + b_i$

Case Study 1: Boxplot of Judges

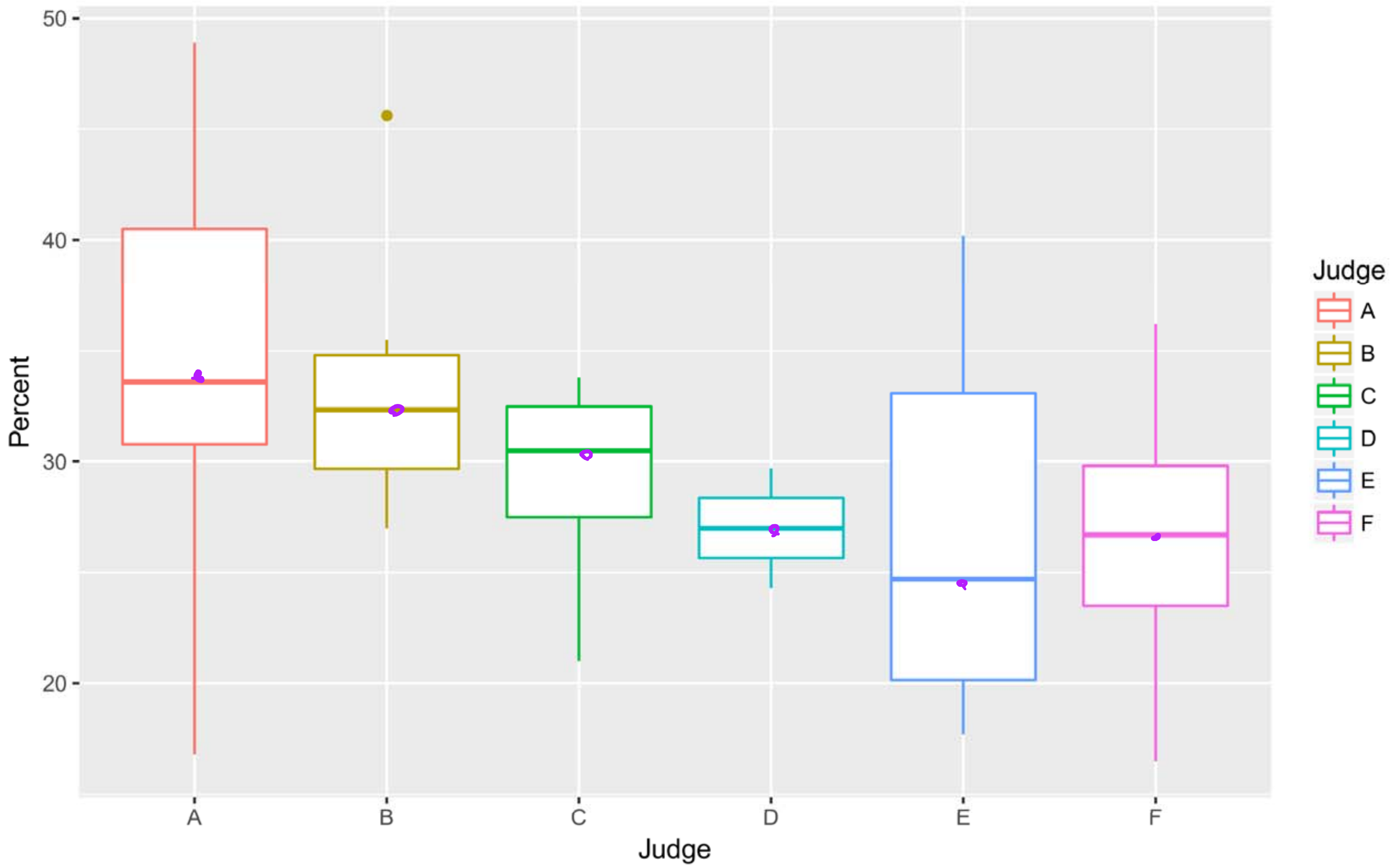
```r
# Get data subset of other judges
Others <- subset(jury, Judge != "Spock's")
boxplot(Percent~Judge, data=Others)
```

y ~ factor



Case Study 1: Boxplot of Judges

```
#install.packages("ggplot2")
library(ggplot2)
ggplot(Others, aes(x=Judge,y=Percent, color=Judge))+geom_boxplot()
```



Case Study 1: Q2-Compare the 6 other judges

① $H_0: \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F$

H_a: at least 2 means differ from each other

$^6C_2 = \binom{6}{2} = \frac{6(5)}{2} = 15$

y

factor

```
summary(aov(Percent~Judge,data=Others))
```

6−1

```
##            Df Sum Sq Mean Sq F value Pr(>F)
## Judge       5  326.5   65.29   1.218  0.324
## Residuals  31 1661.3   53.59
```

37 − 6

② Test Statistic = 1.218

$\sim F_{5,31}$

④ Conclusion

Evidence to support equality of means among the 6 other judges.

③ p-value = 0.324

large p-value

Compare variances of 6 other judges: Rule of thumb

```
sss<-with(Others, tapply(Percent,Judge,sd))
sss
```

```
##         A         B         C         D         E         F    Spock'
## 11.941817  6.582224  4.592929  3.818377  9.010142  5.968878         N
```

```
dim(sss)
```

```
## [1] 7
```

```
max(sss, na.rm=T)
```

```
## [1] 11.94182
```

```
min(sss, na.rm=T)
```

```
## [1] 3.818377
```

```
isTRUE((max(sss, na.rm=T)/min(sss, na.rm=T))>2)
```

```
## [1] TRUE
```

$$\frac{11.94}{3.81} > 2$$

Compare variances of 6 other judges: Bartlett's

```
bartlett.test(Percent~Judge, data=Others)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Percent by Judge
## Bartlett's K-squared = 6.3125, df = 5, p-value = 0.277
```

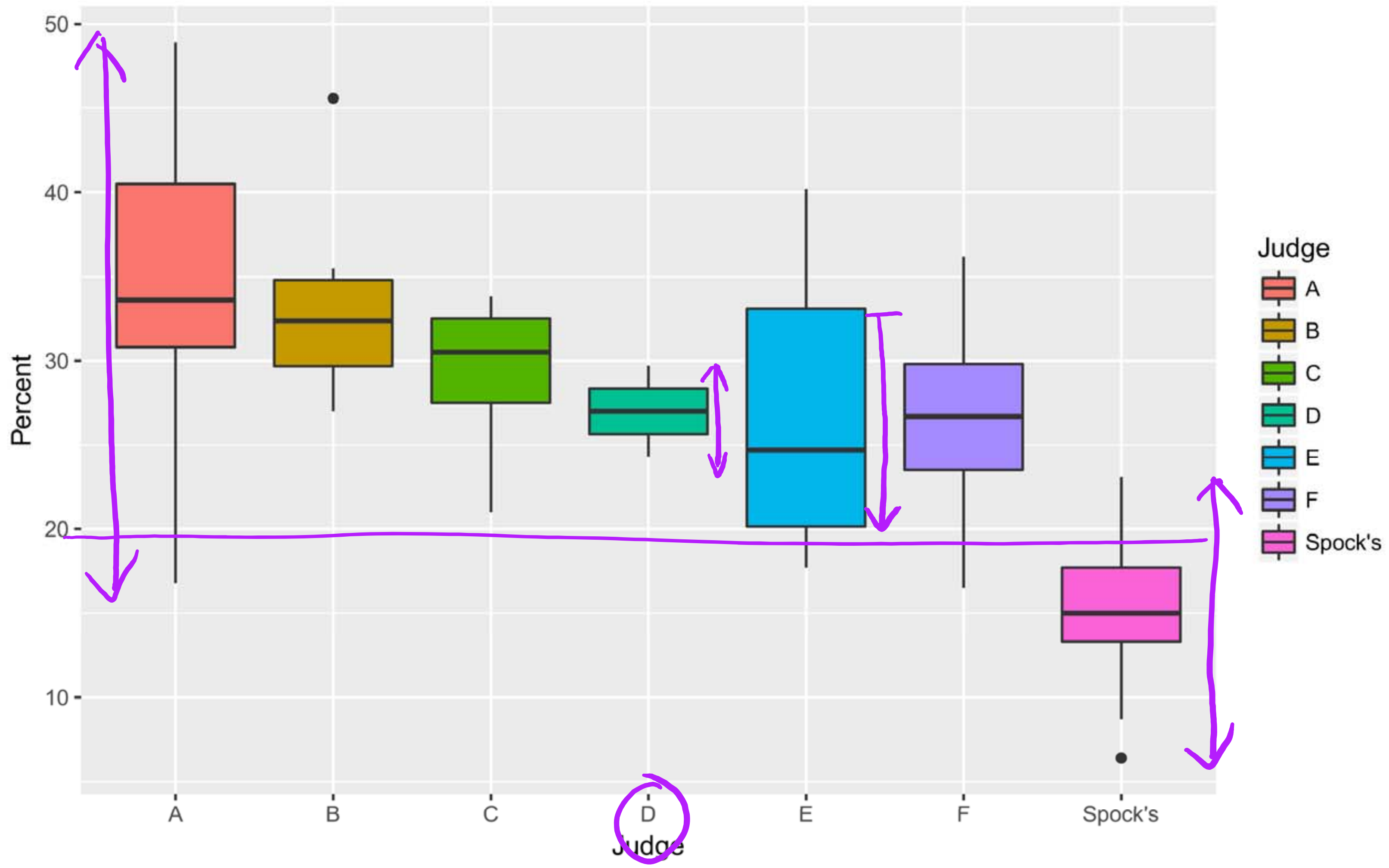
Note: Group sizes are uneven and some are very small

Evidence of equal variance.

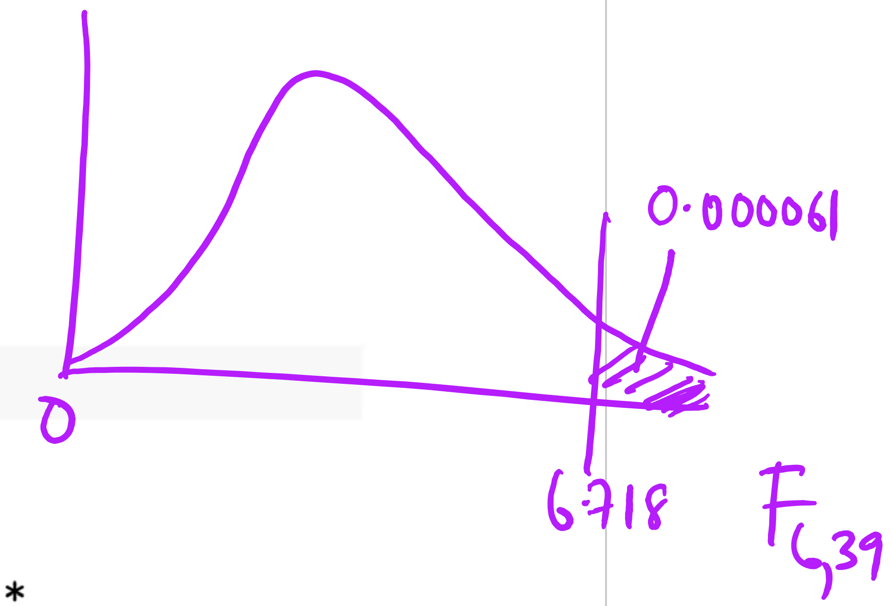
Q3

Compare all 7 judges

```
#boxplot(Percent~Judge)
library(ggplot2)
ggplot(jury, aes(x=Judge,y=Percent, fill=Judge))+geom_boxplot()
```



Compare means of all 7 judges: One-way ANOVA




```
summary(aov(Percent~Judge))
```

7-1

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Judge        6   1927   321.2   6.718 6.1e-05 ***
## Residuals   39   1864    47.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Small p-value

46-7

Compare means of all 7 judges: Gen Linear Model

```
summary(lm(Percent~Judge))
```

```
##
## Call:
## lm(formula = Percent ~ Judge)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.320   -4.367   -0.250    3.319   14.780
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      34.1200     3.0921  11.034 1.47e-13 ***
## JudgeB           -0.5033     4.1868  -0.120   0.9049
## JudgeC           -5.0200     3.8566  -1.302   0.2007
## JudgeD           -7.1200     5.7848  -1.231   0.2258
## JudgeE           -7.1533     4.1868  -1.709   0.0955 .
## JudgeF           -7.3200     3.8566  -1.898   0.0651 .
## JudgeSpock's    -19.4978     3.8566  -5.056 1.05e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.914 on 39 degrees of freedom
```

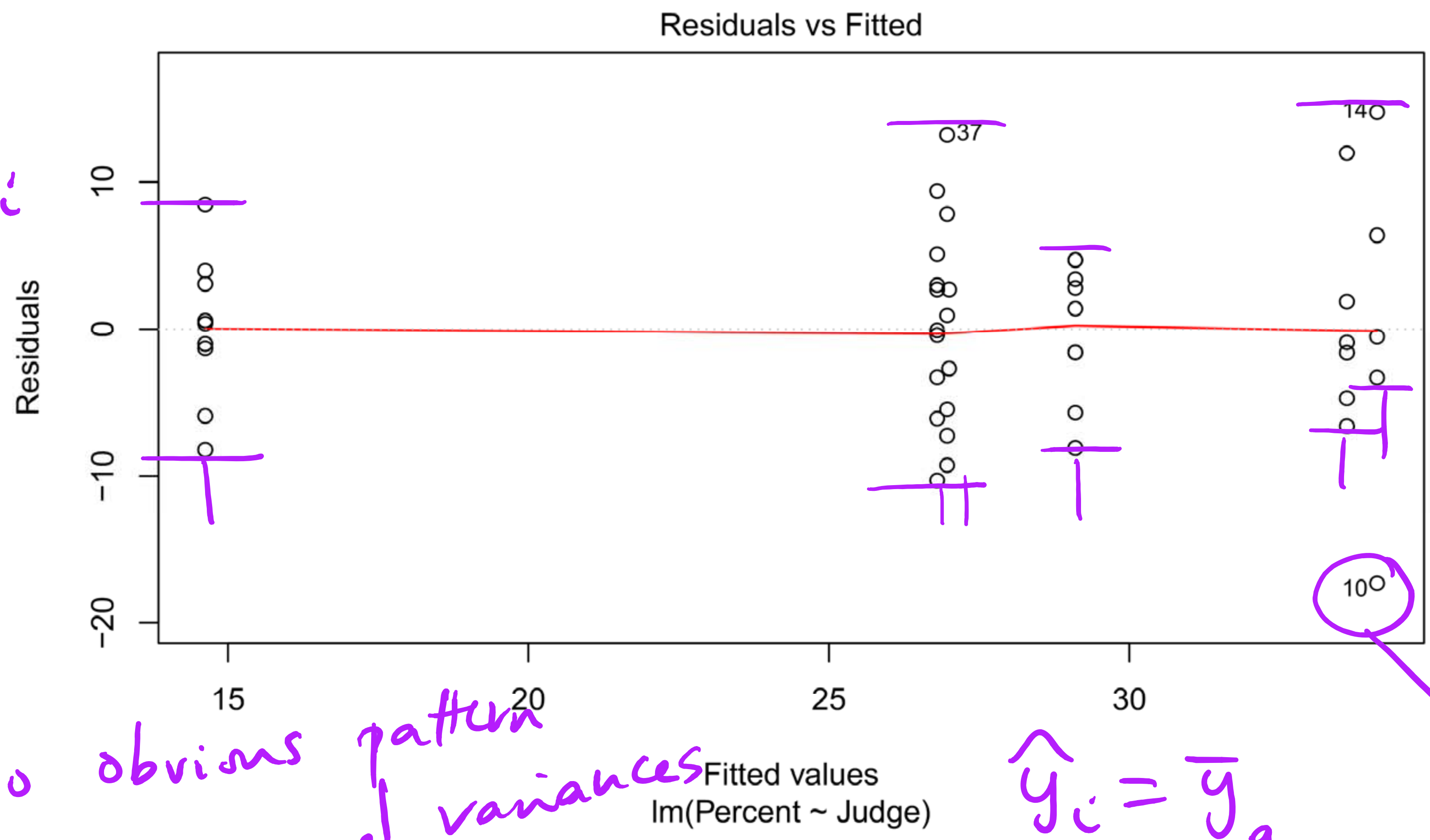
$\bar{y}_A$

$\bar{y}_g - \bar{y}_A$

${}^7C_2 = \frac{7(6)}{2} = 21.$

Check Normality: Linear Model

```
plot(lm(Percent~Judge), which=1)
```



$\hat{e}_i$

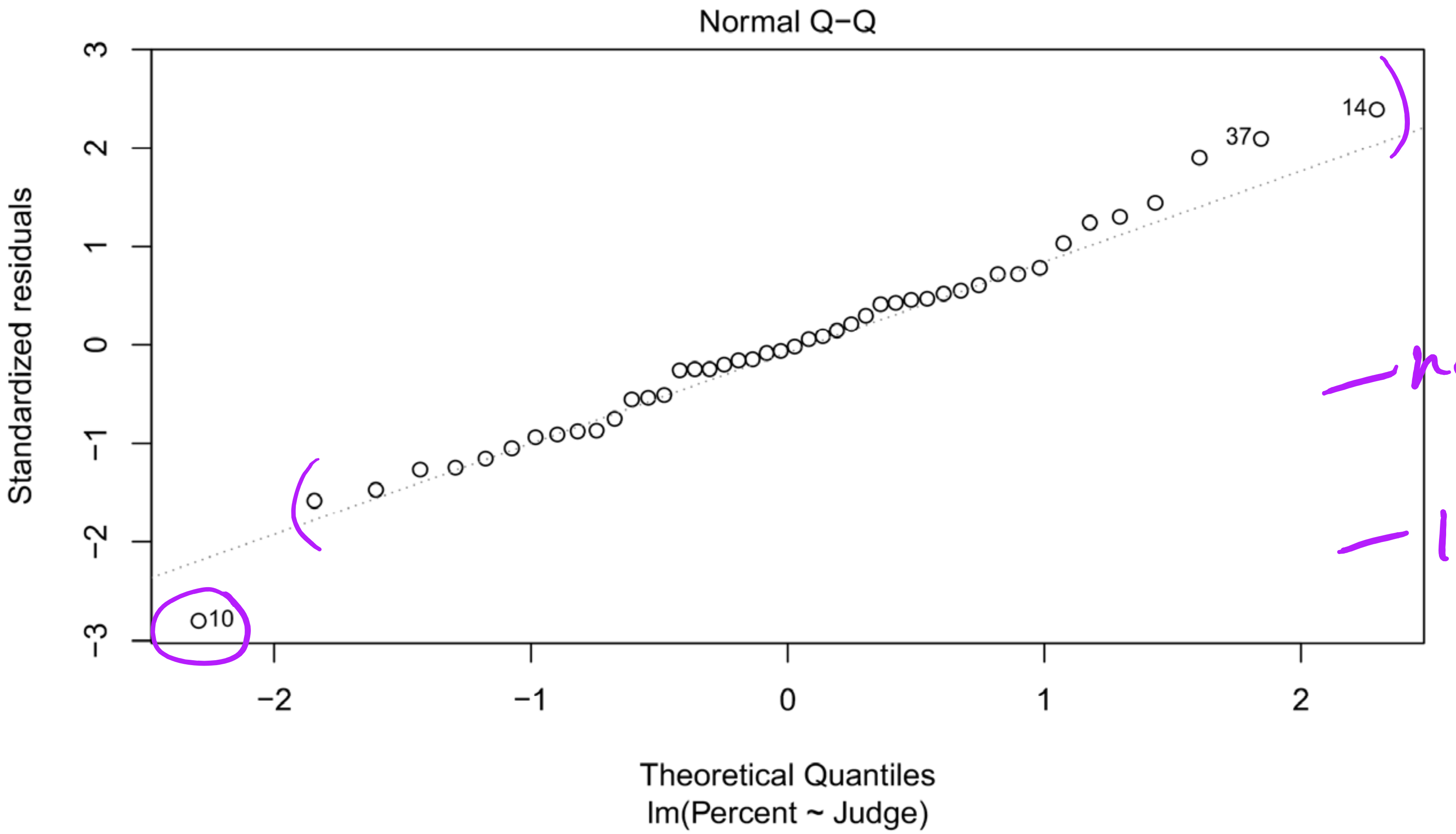
- no obvious pattern
- assume equal variances

$\hat{y}_i = \bar{y}_g$

large -ve residual

Check Normality: Linear Model

```
plot(lm(Percent~Judge), which=2)
```



— normality satisfied

— 1 suspect outlier

Compare variances of all 7 judges: RoT

$n_A = 5$ $n_D = 2$ $n_S = 9$

```
ssa=with(jury, tapply(Percent, Judge, sd))
ssa
```

```
##         A        B        C        D        E        F   Spock'
## 11.941817 6.582224 4.592929 3.818377 9.010142 5.968878 5.03879
```

```
isTRUE((max(ssa, na.rm=T)/min(ssa, na.rm=T))>2)
```

```
## [1] TRUE
```

$$\frac{11.94}{3.81} \simeq 3 > 2.$$

Compare variances of all 7 judges: Bartlett's

```r
bartlett.test(Percent~Judge, data=jury)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  Percent by Judge
## Bartlett's K-squared = 7.7582, df = 6, p-value = 0.2564
```

Assume equal variance

$S_p = \sqrt{MSE}$

Case Study 1: Bonferroni's

```r
Judge=relevel(Judge, ref="Spock's")
pairwise.t.test(Percent, Judge, p.adj="bonf")
```

y   factor

$k = {}_6C_2$

$k = \binom{7}{2}$

$N = 46$

$G = 7$

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  Percent and Judge
##
##   Spock's A       B       C       D       E
## A 0.00022 -       -       -       -       -
## B 0.00013 1.00000 -       -       -       -
## C 0.00150 1.00000 1.00000 -       -       -
## D 0.57777 1.00000 1.00000 1.00000 -       -
## E 0.03408 1.00000 1.00000 1.00000 1.00000 -
## F 0.01254 1.00000 1.00000 1.00000 1.00000 1.00000
##
## P value adjustment method: bonferroni
```

AB

① $H_0: \mu_i - \mu_j = 0$

$H_a: \mu_i - \mu_j \neq 0$ (2-sided)

② $t = \dfrac{\bar{x}_i - \bar{x}_j}{S_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}$

③ P-value $= 2P(T_{N-G} > |t|)$

④ Bonf. (Adj. P-value) K(P-value)

α/k vs P-value

α vs $\min(k \cdot \text{Pvalue}, 1)$

Case Study 1: Bonferroni's CIs

```
lmod=lm(Percent~Judge)
nlevels(jury$Judge)
```

```
## [1] 7
```

```
confint(lmod, level=1-0.05/nlevels(jury$Judge))
```

```
##                  0.357 % 99.643 %
## (Intercept)   8.078085 21.16636
## JudgeA        8.547341 30.44821
## JudgeB        8.647255 29.34163
## JudgeC        5.222970 23.73259
## JudgeD       -2.969585 27.72514
## JudgeE        1.997255 22.69163
## JudgeF        2.922970 21.43259
```

$|\bar{y}_i - \bar{y}_j| \pm t_{\alpha/2k}\, s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

C.I for μ_{spocks}

Inc. 0

$\mu_i - \mu_{spock} = \beta_i = 0$

Case Study 1: Bonferroni's CIs

```r
summary(lmod)
```

```
##
## Call:
## lm(formula = Percent ~ Judge)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.320  -4.367  -0.250   3.319  14.780
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   14.622      2.305   6.344 1.72e-07 ***
## JudgeA        19.498      3.857   5.056 1.05e-05 ***
## JudgeB        18.994      3.644   5.212 6.39e-06 ***
## JudgeC        14.478      3.259   4.442 7.15e-05 ***
## JudgeD        12.378      5.405   2.290 0.027513 *
## JudgeE        12.344      3.644   3.388 0.001623 **
## JudgeF        12.178      3.259   3.736 0.000597 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.914 on 39 degrees of freedom
```

Case Study 1: Tukey's CIs

```r
amod=aov(Percent~Judge)
TukeyHSD(amod,"Judge")
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Percent ~ Judge)
##
## $Judge
##                   diff        lwr       upr     p adj
## A-Spock's 19.49777778   7.514686 31.480870 0.0001992
## B-Spock's 18.99444444   7.671487 30.317402 0.0001224
## C-Spock's 14.47777778   4.350216 24.605339 0.0012936
## D-Spock's 12.37777778  -4.416883 29.172438 0.2744263
## E-Spock's 12.34444444   1.021487 23.667402 0.0248789
## F-Spock's 12.17777778   2.050216 22.305339 0.0098340
## B-A       -0.50333333 -13.512422 12.505755 0.9999997
## C-A       -5.02000000 -17.003092  6.963092 0.8470097
## D-A       -7.12000000 -25.094638 10.854638 0.8777485
## E-A       -7.15333333 -20.162422  5.855755 0.6146238
## F-A       -7.32000000 -19.303092  4.663092 0.4936379
## C-B       -4.51666667 -15.839625  6.806291 0.8742030
## D-B       -6.61666667 -24.158118 10.924784 0.9003280
```

LL UL P-value

- C.I includes 0
- P-value > (α = 0.05)

Case Study 1: Tukey's CI's

```
plot(TukeyHSD(amod,"Judge"))
```