STAT 452_Hw 8

Julia Andrus

10.22) Use the F test at level a $\alpha = .05$ to test for any differences in true average yield due to the different salinity levels.

x_1=1.6 = 59.5, 53.3, 56.8, 63.1, 58.7; n=5

x_2=3.8 = 55.2, 59.1, 52.8, 54.5; n=4

x_3=6.0 = 51.7, 48.8, 53.9, 49.0; n=4

x_4=10.2 = 44.6, 48.5, 41.0, 47.3, 46.1; n=5

n=18

means:
$\overline{x}_1$=58.28
$\overline{x}_1$=55.40
$\overline{x}_1$=50.85
$\overline{x}_1$=45.50

Sum of samples:
$x_{1.}$=291.4
$x_{2.}$=221.6
$x_{3.}$=203.4
$x_{4.}$=227.5
$x_{..}$=943.9
$\sum x_1^2 = 291.4^2 = 84914$
$\sum x_2^2 = 221.6^2 = 49106.6$
$\sum x_3^2 = 203.4^2 = 41371.6$
$\sum x_4^2 = 227.5^2 = 51756.3$

Now,

SSTr=$\sum \frac{x_{ij}^2}{n_i} - \frac{x_{..}^2}{n}$

$$= \frac{291.4^2}{5} + \frac{221.6^2}{4} + \frac{203.4^2}{4} + \frac{227.5^2}{5} - \frac{943.9^2}{18}$$
$$= \frac{84914}{5} + \frac{49106.6}{4} + \frac{41371.6}{4} + \frac{51756.3}{5} - 49497.06722$$
$$= 49953.6 - 49497.06722$$

$$456.543$$

SST=$\sum\sum x_{ij}^2 - \frac{x_{..}^2}{n}$

$$= 59.5^2 + 53.3^2..... + 41.0^2 + 47.3^2 - \frac{943.9^2}{18}$$

Squaring each of the 18 observations and adding it up

$$= 50078.07 - 49497.06722$$

$$= 581.003$$

With the Fundamental identity: SST=SSTr + SSE, we can SSE=SST-SStr:

$$581.003 - 456.543$$

$$= 124.46$$

I=4, n=18
Now $F = \frac{MSTr}{MSE}$:
$MSTr = \frac{SSTr}{I-1}$

$$= \frac{456.543}{4-1}$$

$$= 152.181$$

MSE=$\frac{SSE}{n-I}$

$$= \frac{124.46}{18-4}$$

$$= 8.89$$

$F = \frac{MSTr}{MSE}$:

$$\frac{152.181}{8.89}$$

$$= 17.1182$$

"Statistical theory says that the test statistic has an F distribution with numerator df I-1 and denominator df n-I when H_0 is true.As in the case of equal sample sizes, the larger the value of F, the stronger is the evidence against H_0. Therefore the test is upper-tailed; the P-value is the area under the $F_{I-1,n-}$ curve to the right of f."

$F_{0.05,I-1,n-I} = F_{0.05,3,14} = 3.70$ Looked at Table A.10.

Since $F_{0.05,3,14} = 3.70 < f = 17.1182$, we can reject null; there is significant evidence for concluding that a true average the yield of tomatoes depend on the EC levels.

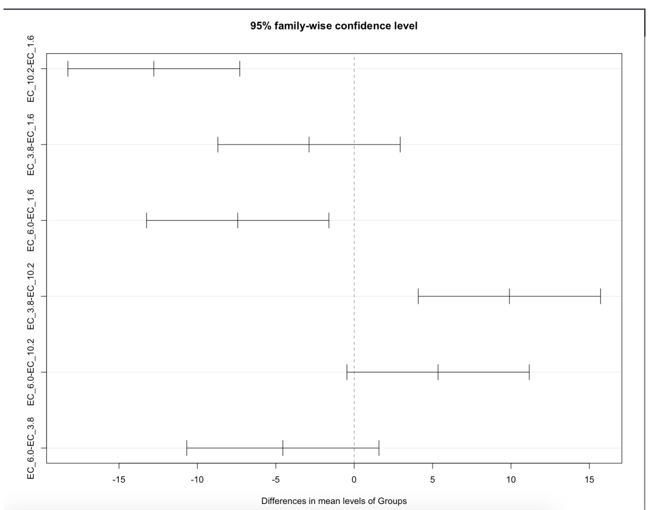
Done in R:

```
> EC_ANOVA <- aov(Y ~ Groups, data = EC_data)
> summary((EC_ANOVA))
            Df Sum Sq Mean Sq F value   Pr(>F)    
Groups       3  456.5  152.17   17.11 5.87e-05 ***
Residuals   14  124.5    8.89                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Tukey’s Method in R:

```
Fit: aov(formula = Y ~ Groups, data = EC_data)

$Groups
                  diff         lwr       upr     p adj
EC_10.2-EC_1.6 -12.78 -18.2618543 -7.298146 0.0000472
EC_3.8-EC_1.6   -2.88  -8.6943846  2.934385 0.4969233
EC_6.0-EC_1.6   -7.43 -13.2443846 -1.615615 0.0109691
EC_3.8-EC_10.2   9.90   4.0856154 15.714385 0.0010845
EC_6.0-EC_10.2   5.35  -0.4643846 11.164385 0.0760885
EC_6.0-EC_3.8   -4.55 -10.6788995  1.578899 0.1830783
```



10.25a) What assumptions must be made about the four total polyunsaturated fat distributions before carrying out a single-factor ANOVA to decide whether there are any differences in true average fat content?

We need to first assume that the variances equal and that there is normality.

10.25b) Carry out the test suggested in (a). What can we say about P-value.

-Breast milk, n=8, $\overline{x_1}$=43.0, SD=1.5
-CO , n=13, $\overline{x_2}$=42.4, SD=1.3
-SO, n=17, $\overline{x_3}$=43.1, SD=1.2
-SMO, n=14, $\overline{x_4}$= 43.5, SD=1.2

1) SSTr=$\sum \frac{x_{ij}^2}{n_i} - \frac{x_{..}^2}{n}$

-$\overline{x} = \frac{n_1(\overline{x_1})+n_2(\overline{x_2})....+n_i(\overline{x_i})}{n_1+n_2...n_i}$

$$= \frac{(8)(43.0)+(13)(42.4)+(17)(43.1)+(14)(43.5)}{8+13...+14} = \frac{2236.9}{52} = 43.017$$

$$SSTr = \sum \frac{x_{ij}^2}{n_i} - \frac{x_{..}^2}{n} =$$

$$8(43.0-43.017)^2 + ... + 14(43.5-43.017)^2 = 8.33$$

2)$SSE = (n_1-1)S_1^2 + ...(n_i-1)S_i^2$

$$= (8-1)(1.5)^2 + (13-1)(1.3)^2 + (17-1)(1.2)^2 + (14-1)(1.2)^2 = 77.79$$

3) SST=SSTr+SSE=8.334428+ 77.79=86.124428

n=52, I=4

1) $MSTr = \frac{J}{I-1}[(\overline{X}_1 - \overline{X}_{...})^2 + (\overline{X}_2 - \overline{X}_{..})^2 + ...\overline{X}_I - \overline{X}_{..})^2]$
or $MSTr = \frac{SSTr}{I-1} = \frac{8.33}{3} = 2.78$

2) $MSE = \frac{S_1^2+S_1^2+......+S_I^2}{I}$
or $MSE = \frac{SSE}{n-I} = 77.7948 = 1.621$

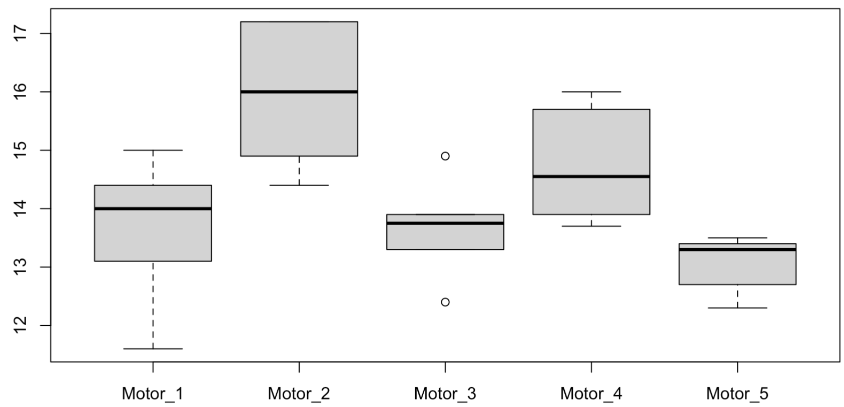
$F = f = \frac{MSTr}{MSE} = 1.713$

$F_{I-1,n-I} = F_{3,48} \approx 0.163.$

Answer: Looking at the book: $F_{0.1,4,48} = 2.2$, where the area to right right of $F_{0.1,4,50}$ under the $F_{0.1,4,48}$ curve is 0.01. Since $f = 1.713 \leq 2.2 = F_{0.1,4,48}$, the we do not reject H_0

10.37) State and test the relevant hypotheses at significance level .05, and then carry out a multiple comparisons analysis if appropriate.

Boxplot in R:



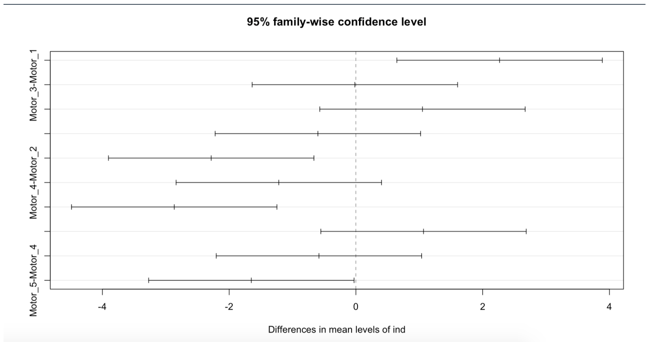

```
> summary(ANOVA_Motors)
            Df Sum Sq Mean Sq F value   Pr(>F)    
ind          4  30.86   7.714   8.444 0.000187 ***
Residuals   25  22.84   0.914                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 
```

Answer: Since our p-value is 0.000187 $<$ 0.05=α, then we can reject H_0 concluding that there is sufficient evidence to support the claim that population means are different.

Now with the formula $w = Q_{\alpha,I,I(J-1)}(\sqrt{\frac{MSE}{J}})$

$= Q_{\alpha,I,I(J-1)}(\sqrt{\frac{MSE}{J}}) = 4.15(\sqrt{\frac{4.15}{6}}) = 1.61974$

Done in R:




```
> TukeyHSD(ANOVA_Motors)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = values ~ ind, data = All_Motors)

$ind
                       diff        lwr        upr     p adj
Motor_2-Motor_1  2.26666667  0.6460270  3.8873064 0.0031588
Motor_3-Motor_1 -0.01666667 -1.6373064  1.6039730 0.9999998
Motor_4-Motor_1  1.05000000 -0.5706397  2.6706397 0.3418272
Motor_5-Motor_1 -0.60000000 -2.2206397  1.0206397 0.8112981
Motor_3-Motor_2 -2.28333333 -3.9039730 -0.6626936 0.0029299
Motor_4-Motor_2 -1.21666667 -2.8373064  0.4039730 0.2106883
Motor_5-Motor_2 -2.86666667 -4.4873064 -1.2460270 0.0002024
Motor_4-Motor_3  1.06666667 -0.5539730  2.6873064 0.3268245
Motor_5-Motor_3 -0.58333333 -2.2039730  1.0373064 0.8262091
Motor_5-Motor_4 -1.65000000 -3.2706397 -0.0293603 0.0445279
```

Answer: If the results of the differences of the means is $<$ then w=1.6194, then that the means the means belong in the same group.

10.42a) State and test the relevant hypotheses at significance level .05.

Unequal sample size in R:

```
> summary(Iris_Anova)
            Df Sum Sq Mean Sq F value Pr(>F)
Groups       2  23.00  11.499   4.802 0.0232 *
Residuals   16  38.31   2.394
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

10.42b) Investigate differences between iris colors with respect to mean cff.

Differences of mean in R:

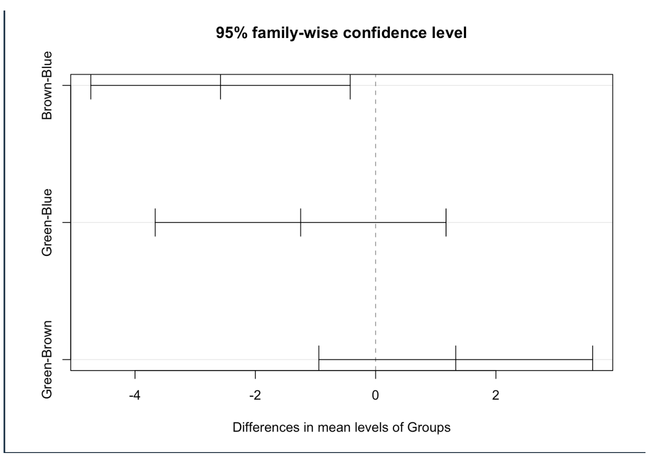
```
> Tuckey_Iris[["Groups"]]
                 diff        lwr       upr      p adj
Brown-Blue  -2.579167 -4.7354973 -0.422836 0.01835789
Green-Blue  -1.246667 -3.6643959  1.171063 0.39943189
Green-Brown  1.332500 -0.9437168  3.608717 0.31242245
>
```

Difference of means done by hand:

$\overline{x}_2 - \overline{x}_1 = 26.92 - 25.59 = 1.33 < 2.19$

$\overline{x}_3 - \overline{x}_1 = 28.17 - 25.59 = 2.58 > 2.19$

$\overline{x}_3 - \overline{x}_2 = 28.17 - 26.92 = 1.25 < 2.19$



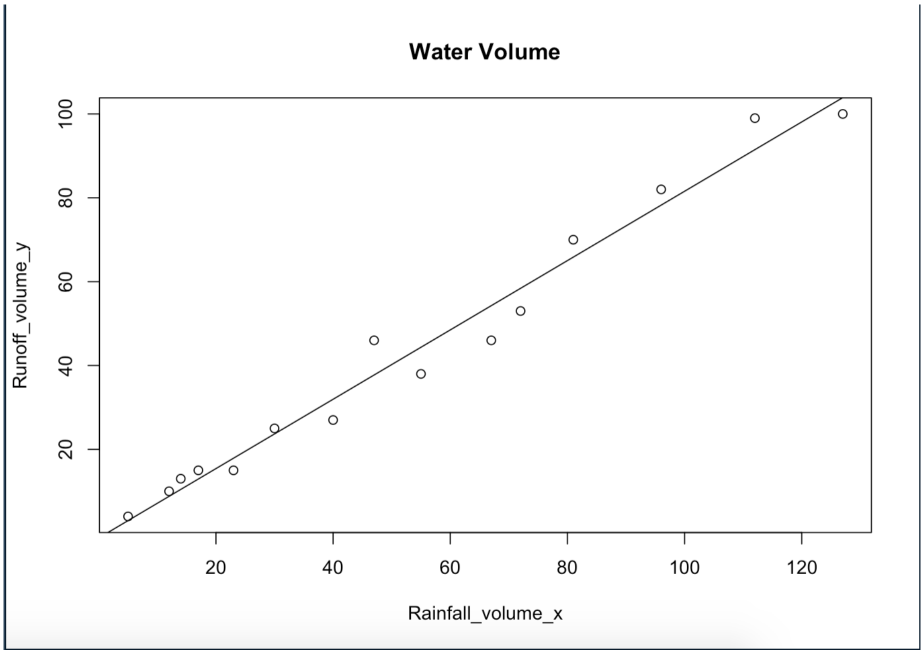
Answer: As you can see, μ_3 (Blue) and μ_1(Green) appear to be significantly different.

12.16a) Does a scatterplot of the data support the use of the simple linear regression model?

Data: x: 5, 12, 14, 17, 23, 30, 40, 47, 55, 67, 72, 81, 96, 112, 127

y: 4, 10, 13, 15, 15, 25, 27, 46, 38, 46, 53, 70, 82, 99, 100

Done in R:



The scatter plot of data does support the use of the simple linear regression model.

12.16b) Calculate point estimates of the slope and intercept of the population regression line.

Point estimator of slope:
$b_1 = \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

$$-1)S_{xy} = \sum(x_i y_i) - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$= (5)(4) + (12)(10) + ... + (127)(100) - \frac{(5 + 12 + ...127)(4 + 10 + ...100)}{15}$$

$$= 51232 - (798)(643)$$

$$= 17024.4$$

$$-2)S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$= (5^2 + 12^2 ... + 127^2) - \frac{(5 + 12 + ... + 127)^2}{15}$$

$$= 63040 - \frac{798^2}{15}$$

$$20586.5$$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{17024.4}{20586.5} = 0.826973$

Answer: $\hat{\beta}_1 = 0.826973$

Point Estimator of Intercept:

$b_0 = \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$

$$\overline{y} - \hat{\beta}_1 \overline{x}$$

$$\frac{4 + 10...100}{15} - (0.826973)\frac{(5 + 12 + ...127}{15})$$

$$= -1.1283$$

Answer: $\hat{\beta}_0 = -1.1283$

With $\hat{\beta}_1 = 0.826973$ and $\hat{\beta}_0 = -1.1283$, then $\hat{y} = -1.1283 + 0.826973x$.

12.16c) Calculate a point estimate of the true average runoff volume when rainfall volume is 50.

Answer: $\hat{y} = -1.1283 + 0.826973x = -1.1283 + 0.826973(50) = 40.22$

12.16d) Calculate a point estimate of the standard deviation σ^2.

$S^2 = \frac{SSE}{n-2}$

1) $SSE = S_{yy} - \hat{\beta}_1 S_{xy} = 14435.73 - (0.826973)17024.4 = 357.01086$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 41999 - \frac{(643)^2}{15} = 14435.73$$

$$S_{xy} = \sum (x_i y_i) - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$= (5)(4) + (12)(10) + ... + (127)(100) - \frac{(5 + 12 + ...127)(4 + 10 + ...100)}{15}$$

$$= 51232 - (798)(643)$$

$$= 17024.4$$

$$S^2 = \frac{SSE}{n-2} = \frac{357.01086}{15-2} = 27.462$$

$$\sqrt{27.462} = 5.24$$

Answer: S=5.24

12.16e) What **proportion of the observed variation** in runoff volume can be attributed to the simple linear regression relationship between runoff and rainfall?

$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{SSE}{S_{yy}} = 1 - \frac{357.01086}{14435.73} = 1 - 0.0247 = 0.975$

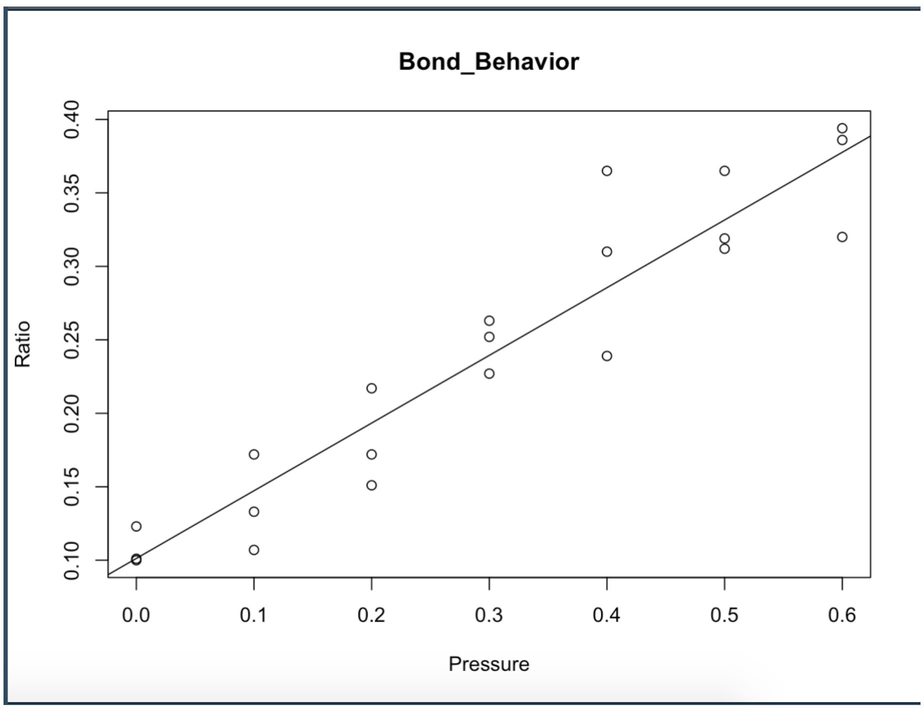
Answer: 97.5%

12.20a) Does a scatterplot of the data support the use of the simple linear regression model?

Data:

x: 0, 0, 0, .1, .1, .1, .2, .2, .2, .3, .3, .3, .4, .4, .4, .5, .5, .5, .6, .6, .6

y: 0.123, 0.100, 0.101, 0.172, 0.133, 0.107, 0.217, 0.172, 0.151, 0.263, 0.227, 0.252, 0.310, 0.365, 0.239, 0.365, 0.319, 0.312, 0.394, 0.386, 0.320



The scatter plot of data does support the use of the simple linear regression model.

12.20b) Use the accompanying Minitab output to give point estimates of the slope and intercept of the population regression line.

Answer: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + = 0.10121 + 0.46071x$

12.20c) Calculate a point estimate of the true average bond capacity when lateral pressure is $.45 f_{cu}$.

Answer: Estimate of the true ratio is $\hat{y} = 0.10121 + 0.46071x = 0.10121 + 0.46071(.45) = .3085$

As mentioned in the book: the ratio of bond strength (MPa) to $\sqrt{f_{cu}}$, so

Answer: Estimate of the true ratio is:

$$\textbf{Bond strength} = estimateratio \times \sqrt{f_{cu}}$$

,

$$= .3085 \times \sqrt{.45} = 0.2069$$

12.20d) What is a point estimate of the error standard deviation σ, and how would you interpret it?

$S^2 = \frac{SSE}{n-2}$

1) $SSE = S_{yy} - \hat{\beta}_1 S_{xy} = 0.199 - 0.46071(.387) = 0.199 - 0.178 = 0.0207$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 1.403136 - \frac{(5.028)^2}{21} = 1.403136 - 1.2038 = 0.199$$

$$S_{xy} = \sum (x_i y_i) - \frac{(\sum x_i)(\sum y_i)}{n}$$
$$= 1.8954 - \frac{(6.3)(5.028)}{21} = 1.8954 - 1.5084 = .387$$

$$S^2 = \frac{SSE}{n-2} = \frac{0.021}{21-2} = 0.0011$$
$$S = \sqrt{0.0011} = .0332$$

Answer: .0332

12.20e) What is the **value of total variation**, and what **proportion of it can be explained by the model relationship**?

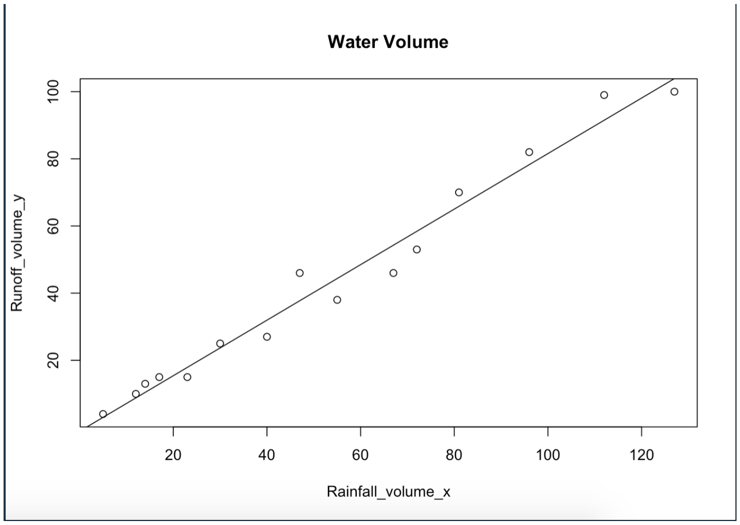
Total Variation: $SST = 0.199$

$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{SSE}{S_{yy}} = 1 - \frac{0.021}{0.199} = 1 - .1055 = .8945$

Answer: 89.5%

12.32) Use the accompanying Minitab output to decide whether there is a useful linear relationship between rainfall and runoff, and then calculate a confidence interval for the true average change in runoff volume associated with a $1m^3$ increase in rainfall volume.

From Exercise 16:



CI inter for slope $\hat{\beta}_1$ of the true regression line is:

$\hat{\beta}_1 \pm t_{\frac{\alpha}{2},n-2}(S_{\hat{\beta}_1})$

1) $\hat{\beta}_1 = 0.827$

$t_{\frac{\alpha}{2},n-2} = 1 - 0.95 = \frac{0.05}{2} = 0.025, 15 - 2 = t_{0.025,13} => TableA.5 = 2.160$

2) $(S_{\hat{\beta}_1}) = \frac{S}{\sqrt{S_{xx}}} = \frac{5.24}{\sqrt{20586.5}} = 0.03652$

$\hat{\beta}_1 \pm t_{\frac{\alpha}{2},n-2}(S_{\hat{\beta}_1}) = 0.827 \pm 2.160(.03652) = (0.748, 0.906)$

Answer: (0.748,0.906)

12.34a) Obtain the equation of the least squares line and interpret its slope.

Point estimator of slope:

$b_1 = \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

$$1) S_{xy} = \sum(x_i y_i) - \frac{(\sum x_i)(\sum y_i)}{n}$$
$$= 472.8149 - \frac{(107.35)(79.44)}{18}$$
$$472.8149 - 473.77$$
$$-0.95643$$

$$2) S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$
$$= (653.61313) - \frac{(107.35)^2}{18}$$
$$= 653.61313 - 640.223$$
$$12.81$$

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-0.95643}{12.81} = -0.07466$

Answer: $\hat{\beta}_1 = -0.07466$

Point Estimator of Intercept:

$b_0 = \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$

$$\overline{y} - \hat{\beta}_1 \overline{x}$$
$$4.41 - (-0.07466)(5.96)$$
$$= 4.85$$

Answer: $\hat{\beta}_0 = 4.85$

With $\hat{\beta}_1 = -0.07466$ and $\hat{\beta}_0 = 4.85$, then $\hat{y} = 4.85 - 0.07466x$.

12.34b) What **proportion of observed variation in dielectric constant can be attributed to the approximate linear relationship** between dielectric constant and air void.

$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{SSE}{S_{yy}} = 1 - \frac{357.01086}{14435.73} = 1 - 0.0247 = 0.975$

1) $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 350.6868 - \frac{(79.44)^2}{18} = 0.0916$$

$$S_{xy} = \sum(x_i y_i) - \frac{(\sum x_i)(\sum y_i)}{n} = 472.8149 - \frac{(107.35)(79.44)}{18} = -0.9564433$$

$SSE = S_{yy} - \hat{\beta}_1 S_{xy} = 0.0916 - (-0.07466)(-0.9564433) = 0.0916 - 0.07140 = 0.0202$

$$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{SSE}{S_{yy}} = 1 - \frac{0.0202}{0.0916} = 0.779$$

Answer: 77.9%

12.34c) Does there appear to be a useful linear relationship between dielectric constant and air void? State and test the appropriate hypotheses.

$H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0$

$t = \frac{\hat{\beta}_1 - \beta_{10}}{\frac{S}{\sqrt{S_{xx}}}}$

$$= \frac{-0.07466}{\frac{0.0335}{\sqrt{12.81}}} = \frac{-0.07466}{0.0099195} = -7.52657$$

$$t = -7.52657$$

$t_{\frac{\alpha}{2}, n-2} = t_{\frac{0.01}{2}, 18-2} = t_{.005,16} => Table A.5 => 2.921$

Answer: Since $t = -7.52657 < 2.921 = t_{.005,16}$, we can reject H_0. This concludes that there is a useful linear relationship between the variables.

12.34d) Suppose it had previously been believed that when air void increased by 1 percent, the associated true average change in dielectric constant would be at least -.05. Does the sample data contradict this belief? Carry out a test of appropriate hypotheses using a significance level of .01.

$H_0 : \beta_1 = -0.05, H_a : \beta_1 > -0.05$

$t = \frac{\hat{\beta}_1 - \beta_{10}}{\frac{S}{\sqrt{S_{xx}}}}$

$$= \frac{-0.07466 - (-.05)}{\frac{0.0335}{\sqrt{12.81}}} = \frac{-02466}{0.0099195} = -2.486$$

Now for our p-value: $P(T > -2.49) = 0.99$

Answer: Since $p - value = 0.99 > 0.01 = \alpha$, we can not reject H_0.

12.52a) Does the simple linear regression model specify a useful relationship between chlorine flow and etch rate?

The simple linear regression model does specify a useful relation between chlorine flow and etch rate.

Data:

x: 1.5, 1.5, 2.0, 2.5, 2.5, 3.0, 3.5, 3.5, 4.0

y: 23.0, 24.5, 25.0, 30.0, 33.5, 40.0, 40.5, 47.0, 49.0

The model utility test is $H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0$ We can use the the model test:

$t = \frac{\hat{\beta}_1 - \beta_0}{S_{\hat{\beta}_1}} =$

Couple things we need:

$S_{\hat{\beta}_1} = \frac{S}{S_{\sqrt{S_{xx}}}} = \frac{2.5}{6.5} = 0.9985$

$SSE = 45.4$

$S = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{45.4}{9-2}} = 2.5$

$S_{xx} = 6.5$

$\hat{\beta}_1 = 10.602564$

$t = \frac{\hat{\beta}_1 - \beta_0}{S_{\hat{\beta}_1}} = \frac{10.602564-0}{0.9985} = 10.6186$

P-value is $P = 2P(T > |10.6186|) = 0.00$, where v=n-2=7. Since $P = 0.00 < 0.05 = \alpha$, we can reject H_0

12.52b) Estimate the true average change in etch rate associated with a 1-SCCM increase in flow rate using a 95% confidence interval, and interpret the interval.

$\hat{\beta}_1 \pm t_{\frac{\alpha}{2},n-2}(S_{\hat{\beta}_1})$

$t_{\frac{\alpha}{2},n-2}(S_{\hat{\beta}_1)=t_{0.025,7}} = 2.365$

$(S_{\hat{\beta}_1}) = 0.9985$

Now,

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2},n-2}(S_{\hat{\beta}_1})$$

$$10.602564 \pm 2.365(0.9985) = (8.24, 12.96)$$

Answer: (8.24, 12.96)

12.52c) Calculate a 95% CI for mY? 3.0, the true average etch rate when flow 5 3.0. Has this average been precisely estimated?

$x^* = 3.0$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^* = 6.448718 + 10.602564(3.0) = 38.256$

$t_{\frac{\alpha}{2},n-2}(S_{\hat{\beta}_1}) = t_{0.025,7} = 2.365$

An estimate of the standard deviation of $\hat{Y}$ is:

$S_{\hat{y}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}}$

$S_{\hat{y}} = 2.5456(\sqrt{\frac{1}{9} + \frac{(3.0-2.67)^2}{6.5}}$
$= 0.35805$
95% CI:

$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\frac{\alpha}{2},n-2}(S_{\hat{\beta}_0) + \hat{\beta}_1 x^*} = \hat{y} \pm t_{\frac{\alpha}{2},n-2}(S_{\hat{y}})$
$= 38.256 \pm 2.365(0.35805)$

Answer: $(36.10, 40.41)$

12.52d) Calculate a 95% PI for a single future observation on etch rate to be made when flow 3.0. Is the prediction likely to be accurate?

$\hat{y} \pm t_{\frac{\alpha}{2},n-2}(\sqrt{S^2 S^2_{\hat{Y}}}) = 38.256 \pm 2.365(\sqrt{2.5456^2 + 0.35805^2})$

Answer: $(31.859, 44.655)$

You can be 95% confident that the future value of Y when x=3.0 is between the two values. Prediction is likely to be accurate.

12.52e) Would the 95% CI and PI when flow 5 2.5 be wider or narrower than the corresponding intervals of parts (c) and (d)? Answer without actually computing the intervals.

The confidence interval and prediction interval in (c) and (d) we be narrower when $x^* = 2.5$ because it is closer to $\overline{x}$ than when $x^* = 3.0$.

$\overline{x} = 2.667$

$x^* = 2.5$

Answer: narrower

12.52f) Would you recommend calculating a 95% PI for a flow of 6.0? Explain.

The value x=6.0 is not in the range of observed values and therefore you should not calculate the prediction interval for such value.

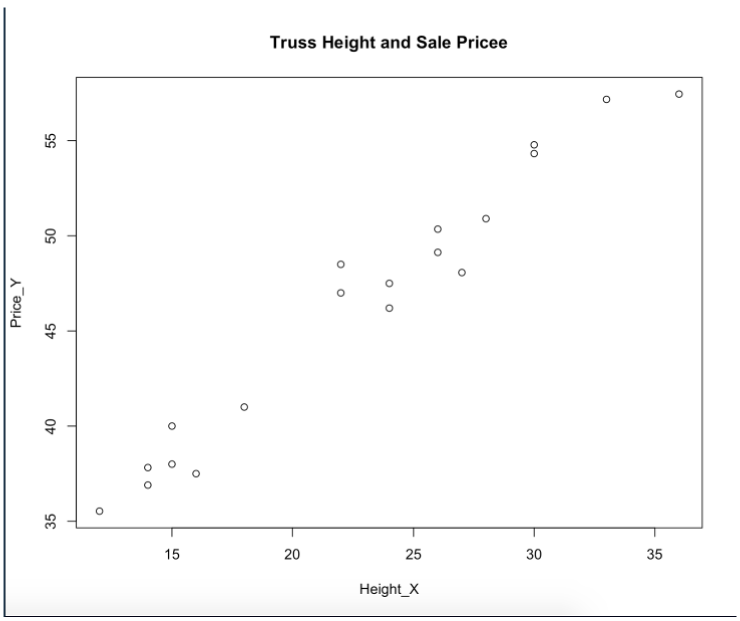
12.68a) Is it the case that truss height and sale price are "deterministically" related—i.e., that sale price is deter- mined completely and uniquely by truss height?

x: 12, 14, 14, 15, 15, 16, 18, 22, 22, 24,24, 26, 26, 27, 28, 30, 30, 33, 36

y: 35.53, 37.82, 36.90, 40.00, 38.00, 37.50, 41.00, 48.50, 47.00, 47.50, 46.20, 50.35, 49.13, 48.07, 50.90, 54.78, 54.32, 57.17, 57.45

Answer: It is not the case that the sale price is not uniquely determined by height. If you look closely at the data, we should have same price if our heights are the same (ex: height 14). Since the two same values of heights gets us different prices, then sale price is not uniquely determined by height.

12.68b) Construct a scatterplot of the data. What does it suggest?
Done in R:



Answer: Looking at the scatterplot it seems that the relationship between height and sale price is strong/perfect correlation between them (points are close to the line) and liner.

12.68c) Determine the equation of the least squares line.

I used R to do this problem: $\hat{y} = 23.77 + 0.9872x$

R:

```
Call:
lm(formula = Price_Y ~ Height_X)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.35522 -0.63584 -0.08796  0.92263  3.01053 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 23.77215    1.11347   21.35 1.03e-13 ***
Height_X     0.98715    0.04684   21.07 1.27e-13 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.416 on 17 degrees of freedom
Multiple R-squared:  0.9631,	Adjusted R-squared:  0.961 
F-statistic: 444.1 on 1 and 17 DF,  p-value: 1.271e-13

> |
```

Answer: $\hat{y} = 23.77 + 0.9872x$

12.68d) Give a point prediction of price when truss height is 27 ft, and calculate the corresponding residual.

Answer: Point prediction of price when height is 27ft is: $\hat{y} = 23.77 + 0.9872(27) = 50.42$

Corresponding Residual: residual=$y - \hat{y}$

Where y=48.07, when height is 27ft from looking at the table.

$\hat{y} = 50.42$
Answer: residual=$y - \hat{y} = 48.07 - 50.42 = -2.35$

As you can see the residual is negative and prediction is higher than the actual sale price at height of 27ft.

12.68e) What percentage of observed variation in sale price can be attributed to the approximate linear relationship between truss height and price?

$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{34.0814}{924.4364} = 1 - .036867 = 0.963132$
I used R and got that percentage of variation in the observed vales of the sale price that is explained by regression is 96.31%, which indicates that 96.31% of the variability in the sale price is explained by variability in the height.

Done in R:

```
Call:
lm(formula = Price_Y ~ Height_X)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.35522 -0.63584 -0.08796  0.92263  3.01053 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 23.77215    1.11347   21.35 1.03e-13 ***
Height_X     0.98715    0.04684   21.07 1.27e-13 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.416 on 17 degrees of freedom
Multiple R-squared:  0.9631,	Adjusted R-squared:  0.961 
F-statistic: 444.1 on 1 and 17 DF,  p-value: 1.271e-13

> 
```