Name: *Caleb McWhorter — Solutions*
MAT 222
Spring 2019
Minitab Lab 3

"There is Dana, only Zuul"
–Dana Barrett/Zuul, Ghostbusters

The following exercises make use of the `abalone.mpj` file. This data is from a study by Dr. Sam Waugh's Ph.D. thesis at the University of Tasmania. The goal of the study was to study whether one could predict the age of abalone using physical measurements. The variables in the study were sex (male, female, and infant), length (mm), diameter (mm), height (mm), whole weight (g), shucked weight (g), viscera weight (g), shell weight (g), and rings (each ring is 1.5 years). Use Minitab to complete the following exercises.

Problem 1: Find a 5-number summary for the variable 'Whole Weight'. Fill in the data below.

Min	Q_1	Median	Q_3	Max
0.00200	*0.44150*	*0.79950*	*1.15350*	*2.82550*

We will now sort the abalone into categories called 'small', 'medium', and 'large'. Assume that the 5-number summary above is a good representative of the weight distribution of all abalone. We will call an abalone small if its weight is less than Q_1, medium if its weight is between Q_1 and Q_3, and large if its weight is larger than Q_3. One can count individuals meeting certain conditions in Minitab. Under `Calc` → `Calculator`, choose `Logical` under `Functions`. Entering `IF('Sex' = "M" And 'Whole Weight' >= 5 And 'Whole Weight' <= 6,1,0)`, Minitab enters a '1' if the abalone is male and has whole weight ≥ 5 and ≤ 6. One can then enter these results into a blank column, and then using `Calc` → `Column Statistics`, find the sum of this new column. The sum will be the number of male abalone with weight between 5 and 6.

Problem 2: Use Minitab to complete the following chart of abalone broken down by categories male, female, or infant and small, medium, or large.

Size \ Sex	Male	Female	Infant	Totals
Small	*187*	*97*	*759*	*1043*
Medium	*815*	*719*	*556*	*2090*
Large	*526*	*491*	*27*	*1044*
Totals	*1528*	*1307*	*1342*	*4077*

C L Blake, C J Merz. UCI repository of machine learning databases University of California, Irvine, Department of Information and Computer Sciences. 1998, Sam Waugh (1995) "Extending and benchmarking Cascade-Correlation", PhD thesis, Computer Science Department, University of Tasmania.

Problem 3: Enter the table from Problem 2 into Minitab and use Minitab to perform a χ^2-test. [Use $\alpha = 0.05$.] For this test, fill in the expected values and the χ^2-squared contribution in the tables below. Then state the null and alternative hypothesis for the test along with its degrees of freedom and p-value. Be sure to state the conclusion for this test.

Expected Values

Size \ Sex	Male	Female	Infant
Small	*381.5*	*326.4*	*335.1*
Medium	*764.5*	*645.0*	*671.5*
Large	*381.9*	*326.7*	*335.4*

χ^2-squared Contributions

Size \ Sex	Male	Female	Infant
Small	*99.19*	*161.19*	*536.24*
Medium	*3.33*	*6.47*	*19.86*
Large	*54.36*	*82.66*	*283.59*

$$\begin{cases} H_0 : \textit{there is no association between sex and size} \\ H_a : \textit{there is an association between sex and size} \end{cases}$$

$X^2 = \mathit{1246.899}$

degrees of freedom= *4*

p-value= *0.000*

Therefore, we reject the null hypothesis that there is no association between sex and size.

Problem 4: Use `Stat` → `Regression` → `Fit Regression Model` to create a multilinear regression to predict the variable 'Whole Weight' using the variables 'Length', 'Diameter', 'Height', and 'Rings'. Provide a print of the regression analysis. Be sure to also give the regression equation.

For the solution, see the table provided at the end.

Problem 5: Is the model statistically significant? Explain.

Yes, the F-value has associated p-value 0.000. Therefore, the model is statistically significant.

Problem 6: For this regression, which variables are statistically significant? Which variables are not statistically significant? For the variables which are *not* statistically significant, how might you tell from their coefficient that it is not statistically significant?

Examining p-values, the variables 'Length', 'Diameter', and 'Height' are statistically significant while the variable 'Rings' is not. One must suspect that the variable 'Rings' is not statistically significant because its coefficient, 0.00031, is very close to 0. However, a coefficient being close to 0 does not always imply that it is not statistically significant. One must examine the p-value for the t-test of the coefficient.

```
Analysis of Variance

          Source           DF   Adj SS   Adj MS  F-Value  P-Value
          Regression        4   872.46  218.115  6904.77    0.000
             Length         1     4.85    4.850   153.53    0.000
             Diameter       1     3.47    3.468   109.79    0.000
             Height         1     6.56    6.559   207.64    0.000
             Rings          1     0.00    0.003     0.09    0.770
          Error          4172   131.79    0.032
             Lack-of-Fit 3968   130.37    0.033     4.71    0.000
             Pure Error   204     1.42    0.007
          Total          4176  1004.25

Model Summary

                       S    R-sq  R-sq (adj)  R-sq (pred)
                0.177733  86.88%      86.86%       86.04%

Coefficients

             Term         Coef  SE Coef  T-Value  P-Value    VIF
             Constant  -1.0957   0.0131   -83.48    0.000
             Length      1.763    0.142    12.39    0.000  38.58
             Diameter    1.849    0.176    10.48    0.000  40.53
             Height      1.746    0.121    14.41    0.000   3.40
             Rings     0.00031  0.00106     0.29    0.770   1.55
```

```
The regression equation is

Whole Weight =-1.0957 + 1.763 Length + 1.849 Diamter + 1.746 Height + 0.00031 Rings
```