# Problem 1

1. The degrees of freedom for aphid/plant combination can be found by subtracting the degrees of freedom for Error from the degrees of freedom for Total DF = 51 − 46 = 5. The sum of squares for aphid/plant combination can be found by subtracting the sum of squares for Error from the sum of squares for Total. SSGroups = 64.77 − 39.87 = 24.90. The F-value can be found by dividing the mean square for aphid/plant combination by the mean square for Error: F 4.9807/0.8667 = 5.75.
2. Since the degrees of freedom for aphid/plant combination is 5 and the degrees of freedom for treatments is the number of groups minus 1, the number of different occupations considered is 6.
3. Since the P-value is so small, we reject the null hypothesis. We have strong evidence that mean amount of honeydew produced by aphids is different for at least one aphid/plant combination.

# Problem 2

1. One-way ANOVA: meth labs versus type

Source DF SS MS F

Type 2 37.51 18.755 5.101

Error 9 33.09 3.677

Total 11 70.60

1. The MS for county type measures the amount of variability between the means of the three county types.
2. Using an F-distribution with 2 and 9 degrees of freedom, the P-value is 0.033.
3. The null hypothesis is that the three sizes of counties have the same average number of meth labs. In symbols, this is H0 : α1 = α2 = α3 = 0. The alternative hypothesis is that at least one type of county has a different average number of meth labs than the other two sizes of county. With a P-value of 0.033, we reject the null hypothesis and conclude that at least one type of county has a different average number of meth labs.

# Problem 3

The following are the boxplots of the combined scores, broken down by the race of the

individual taking the exam. It appears that there may be a difference between the scores

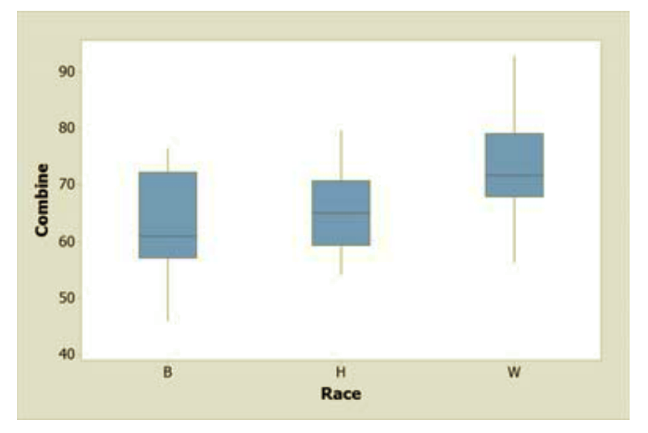
of the various races. Specifically, it appears that whites generally did a little bit better,

that Hispanics had scores in the middle, and that blacks had the lowest scores, though the

difference between Hispanics and blacks does not appear to be large and might turn out to be

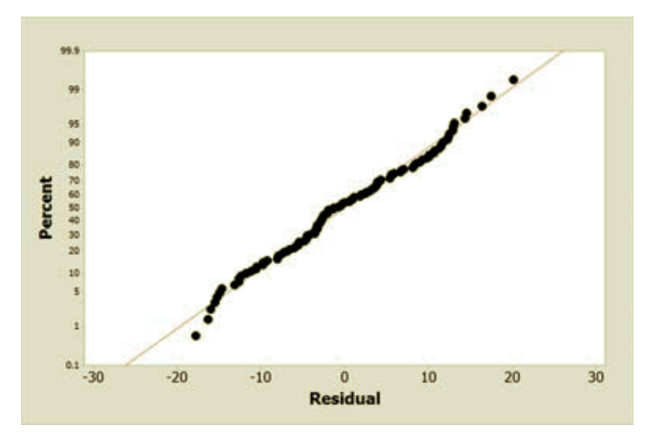
statistically insignificant. The spreads of the three samples appear to be similar so continuing

by conducting an ANOVA analysis would likely be the next step.



# Problem 4

1. To check the conditions necessary for conducting an ANOVA we begin by checking the normality of the residuals. Following is a normal plot of the residuals in this case. While there is a small amount of curvature to this plot, there is not enough to be worried about.



Next we check whether the data are consistent with the idea that the variances for the

three groups are the same. The residuals versus fitted values plot suggests that the data

are comparable with the equal variances requirement. Also the ratio of the largest standard deviation to the smallest is 8.83 - 7.14 = 1.24 which is much less than our suggested cutoff of 2.

Finally we need to evaluate both the independence of the errors and the idea that they come from a population with mean 0. In this case, there is no reason to believe that any one individual’s exam score is related to any other individual’s score, so the independence of the errors seems reasonable. We also have no reason to believe that there is any bias in these scores other than the one that we are modeling, so a mean error of 0 also seems reasonable.

1. The null hypothesis is that the mean combined exam score is the same for the popula-

tions of white, black, and Hispanic firefighters in New Haven, Connecticut. The alternative is that at least one group of firefighters has a different mean combined exam score. In symbols we should write

H0 : αB = αH = αW = 0

Ha : at least one αk is not 0

# Problem 5

The ANOVA suggests that there is a difference for the mean combined exam score for at least

one of the three groups of firefighters. The F-statistic is 13.60 (with 2 and 115 df), which

results in a P-value near 0 (Minitab reports less than 0.001). Looking at the boxplots from

Exercise 5.28 part (a) it would appear that white firefighters have a higher mean score than

the other two groups. Without further testing, we cannot tell whether there is a statistically

significant difference between blacks and Hispanics. (The output for this ANOVA follows.)

Source DF SS MS F-Value P-Value

Race 2 1972 985.83 13.60 0.000

Error 115 8339 72.51

Total 117 10311