**1. Machine speed. The number of defective items produced by a machine () is known to be linearly related to the speed setting of the machine (). The data below were collected from recent quality control records.**

1. Fit a linear regression function by ordinary least squares, obtain the residuals, and plot the residuals against . What does the residual plot suggest?

A screenshot of a computer

Description automatically generatedA graph with numbers and dots

Description automatically generated with medium confidence

The residuals suggest that the assumption of constant variance is violated; residuals clearly increase with greater values.

1. Conduct the Breusch-Pagan test for constancy of the error variance, assuming ; use . State the alternatives, decision rule, and conclusion.

The null and alternative hypothesis are and . We reject the null if the value is less than .

A screenshot of a computer

Description automatically generated

Since , we reject the null hypothesis and conclude that there is significant evidence to suggest that variance is not constant.

1. Plot the squared residuals against . What does the plot suggest about the relation between the variance of the error term and ?

A graph with blue dots and numbers

Description automatically generated

There exists a linear relationship between and the squared residuals, which could suggest that homoscedasticity (error terms have constant variance) is being violated.

1. Estimate the variance function by regressing the squared residuals against , and then calculate the estimated weight for each case using (11.16b).

A screenshot of a computer

Description automatically generated

By 11.16b, .

1. Using the estimated weights, obtain the weighted least squares estimates of and . Are the weighted least squares estimates similar to the ones obtained with ordinary least squares in part (a)?

A screenshot of a data

Description automatically generated

The weighted least squares estimate using our previously defined weight indeed does yield similar estimates as part (a).

1. Compare the estimated standard deviations of the weighted least squares estimates and in part (e) with those for the ordinary least squares estimates in part (a). What do you find?

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

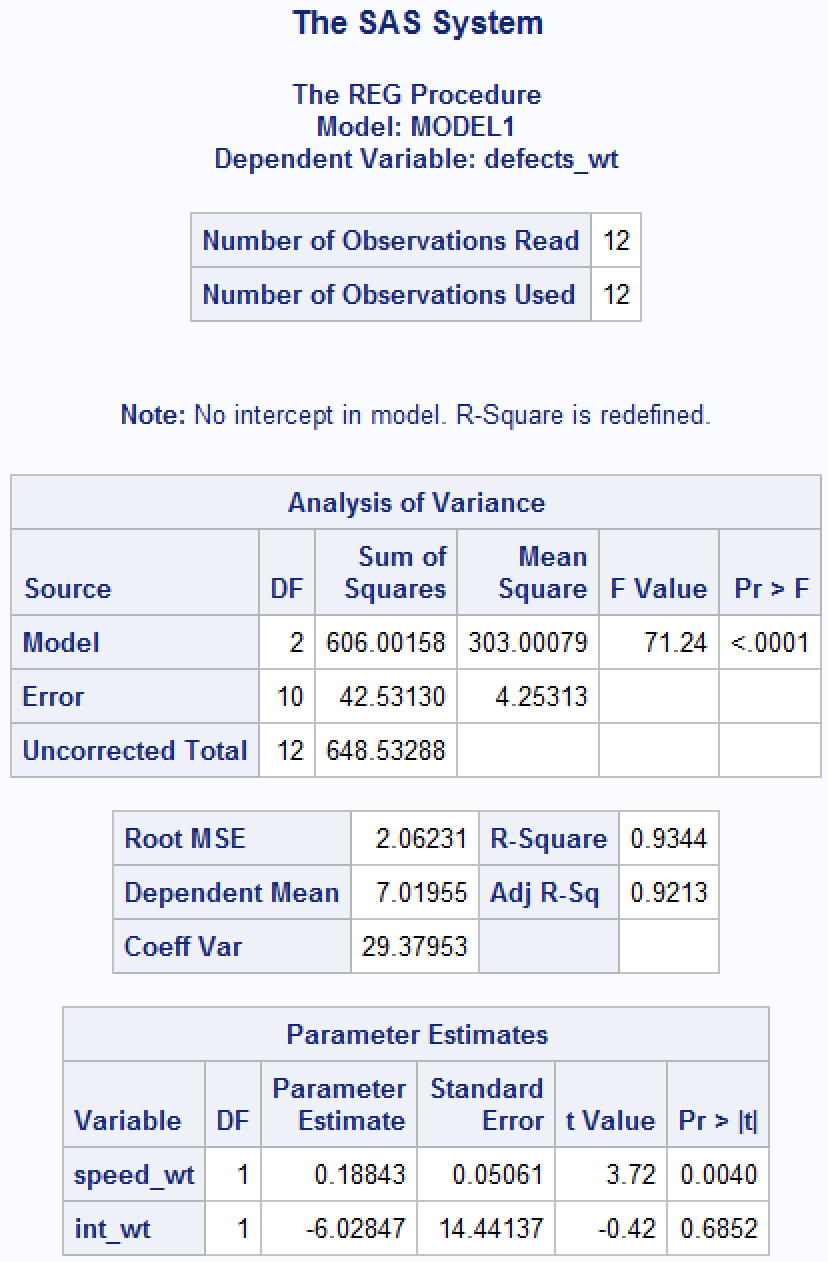
Description automatically generated

The parameter estimates for WLS and OLS are on the right and left, respectively. We find that the standard error for both the intercept and speed is greater for OLS.

1. Iterate the steps in parts (d) and (e) one more time. Is there a substantial change in the estimated regression coefficients? If so, what should you do?

After reiterating steps (d) and (e), we find no change in the estimated regression coefficients.

**2. Refer to Machine speed Problem 11.7. Demonstrate numerically that the weighted least squares estimates obtained in part (e) are identical to those obtained when using transformation (11.23) and ordinary least squares.**

**A screenshot of a data

Description automatically generated**

Transformation 11.23 was used to obtain the left while the right was obtained in part (e). They are identical, as desired. The code used to implement the transformation is given below:

**A screenshot of a computer program

Description automatically generated**

**3. A market researcher, having collected data on breakfast cereal expenditures by families with 1, 2, 3, 4, and 5 children living at home, plans to use an ordinary regression model to estimate the mean expenditures at each of these five family size levels. However, the researcher is undecided between fitting a linear or a quadratic regression model, and the data do not give clear evidence in favor of one model or the other. A colleague suggests: “For your purposes you might simply use an ANOVA model.” Is this a useful suggestion? Explain.**

This is a useful suggestion because an ANOVA model won’t assume a prior relationship before testing whether mean expenditure differs by family size, thereby bypassing the issue of choosing between linear versus quadratic regression when there is no conclusive evidence to favor one over the other. Also, since is a categorical variable (1, 2, 3, 4, or 5 children families), ANOVA will treat each family size as a separate category unlike in regression, which will treat as continuous.

**4. In a study of length of hospital stay (in number of days) of persons in four income groups, the parameters are as follows: = 5.1, = 6.3, = 7.9, = 9.5, = 2.8. Assume that ANOVA model, (16.2) is appropriate.**

1. Draw a representation of this model in the format of Figure 16.2.

A graph of a function

Description automatically generated

1. Suppose 100 persons from each income group are randomly selected for the study. Find and . Is substantially larger than here? What is the implication of this?

By 16.37a, . By 16.37b, . is substantially larger than , so is accordingly very large and so we can conclude that the average length of a hospital stay is not the same across all 4 income categories.

1. If = 5.6 and = 9.0, everything else remaining the same, what would be? Why is substantially larger here than in part (b) even though the range of the factor level means is the same?

. is substantially larger because compared to their original values, and are now farther from the grand mean.

**5. A completely randomized experiment involving treatments was carried out, based on experimental trials for each treatment. The test for equality of the treatment means is to be carried out by means of the randomization distribution of the test statistic (16.55).**

1. Determine the number of ways that the six experimental units can be divided into two groups of size three each. How many unique statistics are possible?

There are ways to divide the experimental units under the conditions listed. However, there are only 6 unique values as evidenced in part b:

1. Obtain the randomization distribution of the test statistic and the -value of the randomization test.

There are 20 ways to permute {1, 1, 1, 2, 2, 2} with 1 representing the first treatment and 2 representing the second treatment. These permutations, along with their corresponding -values, are listed below:

1. 111222; 1.64
2. 112122; 1.06
3. 112212; 0.29
4. 112221; 0.59
5. 121122; 0.29
6. 121212; 1.06
7. 121221; 0.59
8. 122112; 1.64
9. 122121; 2.74
10. 122211; 0.97
11. 211122; 0.97
12. 211212; 2.74
13. 211221; 1.64
14. 212112; 0.59
15. 212121; 1.06
16. 212211; 0.29
17. 221112; 0.59
18. 221121; 0.29
19. 221211; 1.06
20. 222111; 1.64

Hence the randomization distribution is given by:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| value | 2.74 | 1.64 | 1.06 | 0.97 | 0.59 | 0.29 |
|  | 0.10 | 0.20 | 0.20 | 0.10 | 0.20 | 0.20 |

The -value is from the observed data. By the randomization test, the associated -value is .

1. Obtain the -value of the normal-theory statistic for the sample results in part (b). How does this -value compare with the one from the randomization test in part (b)? What does this suggest about the appropriateness of the F distribution here if the error terms are far from normally distributed?

A screenshot of a computer screen

Description automatically generated

The -value is and the associated -value is 0.2690, which is smaller than the -value obtained from the randomization test. The distribution is still appropriate because it does not rely on the assumption that error terms must be normally distributred.

**6. Give a table of sample sizes, means, and standard deviations for the six different filling**

**machines.**

A table with numbers and letters

Description automatically generated

**7. Statistically examine the question of whether or not the six machines place the same amount of fill into the cartons. Write a model for this analysis, state the null and alternative hypothesis in terms of your model parameters (cell-based or factor effects), give the test statistic with degrees of freedom, the P-value, and your conclusion.**

**A screenshot of a computer

Description automatically generated**

The hypotheses are: all , , are equivalent, and at least one is different. The test statistic, degrees of freedom, and -value are all given above. With , we reject the null hypothesis and conclude that at least one of the machines has a mean fill amount that significantly deviates from the others.

**8. Examine the residuals of this analysis to make sure the assumptions are not violated. Display (and comment on) the plots and/or tests you use to do this.**

A graph of numbers and a line of dots

Description automatically generated with medium confidenceA graph of a line

Description automatically generated

We examine the residuals by way of a qqplot and a scatterplot of residuals versus machine number. The qqplot indicates that the residuals more or less normally distributed, while the scatterplot indicates there are no outlying observations and that residuals have constant variance.