DATA 557: HW Assignment 2

Hriday Baghar

January 20, 2022

Question 1

Data: 'temperature_experiment.csv'

A manufacturing process is run at a temperature of 60 deg C. The manufacturer would like to know if increasing the temperature would yield an increase in output.

Increasing the temperature would be more expensive, so an increase would only be used in future if it increased output. It seems unlikely that increasing the temperature would decrease output and, even if it did, there would be no value in having that information.

An experiment was performed to assess the effect of temperature on the output of a manufacturing process. For this experiment, temperatures of 60 or 75 degrees C were randomly assigned to process runs.

It was desired to gather more information about output at the new temperature so temperatures were randomly assigned to process runs at a ratio of 2 to 1 (2 runs at temperature 75 for every 1 at temperature 60). The process output was recorded from each run.

The variables in the data set are:

- *run:* Run number
- *temp:* Temperature
- *output:* Process output

1.1. Perform the large-sample Z-test to compare mean output for the two temperatures. Give the value of the test statistic and the p-value for the test.

1.2. Do you reject the null hypothesis at a significance level of ### 0.05?

1.3. State the null hypothesis for the test.

1.4. Perform the unequal-variance (Welch) t-test to compare mean output in the two temperature groups. Report the test statistic and the p-value for the test.

1.5. Perform the equal-variance t-test to compare mean output in the two temperature groups. Report the test statistic and the p-value for the test.

1.6. Which of the three tests do you think is most valid for this experiment? Why?

1.7. Calculate a 95% confidence interval for the difference between mean output using the large-sample method.

1.8. Calculate a 95% confidence interval for the difference between mean output using a method that corresponds to the Welch test.

1.9. Calculate a 95% confidence interval for the difference between mean output using a method that corresponds to the equal-variance t-test.

1.10. Apart from any effect on the mean output, do the results of the experiment suggest a disadvantage of the higher temperature?

Question 2

Data set: 'defects.csv'

The data are from an experiment to compare 4 processing methods for manufacturing steel ball bearings.

The 4 process methods were run for one day and a random sample of 1% of the ball bearings from the day was taken from each of the 4 methods. Because the processes produce ball bearings at different rates the sample sizes were not the same for the 4 methods.

Each sampled ball bearing had its weight measured to the nearest ### 0.1 g and the number of surface defects was counted.

The variables in the data set are:

- *Sample:* sample number
- *Method:* A, B, C, or D
- *Defects:* number of defects
- *Weight:* weight in g

```
##      Sample     Method    Defects           Weight
##  Min.   :  1   A:74    Min.   :0.0000   Min.   : 9.00
##  1st Qu.: 67   B:75    1st Qu.:0.0000   1st Qu.: 9.80
##  Median :133   C:64    Median :0.0000   Median :10.00
##  Mean   :133   D:52    Mean   :0.1623   Mean   :10.05
##  3rd Qu.:199           3rd Qu.:0.0000   3rd Qu.:10.30
##  Max.   :265           Max.   :3.0000   Max.   :11.90
```

2.1. The target weight for the ball bearings is 10 g. For each of the 4 methods it is desired to test the null hypothesis that the mean weight is equal to 10. What test should be used?

2.2. Give the p-values for the tests for each method. Include your R code for this question.

2.3. Apply a Bonferroni correction to your results from the previous question to account for inflation of type I error rate due to multiple testing. How does the Bonferroni correction change your conclusions? In particular, do you have evidence to reject the null hypothesis that the mean weight for all 4 methods is equal to 10, at significance level ### 0.05?

2.4. It is is desired to compare mean weights of the 4 methods. This is to be done first by performing pairwise comparisons of mean weight for the different methods. What test should be used for these comparisons?

2.5. Report the p-values from all pairwise comparisons. Include your R code for this question.

2.6. Apply a Bonferroni correction to your results of the previous question to account for inflation of type I error rate due to multiple testing. What conclusion would you draw from these results? Would you reject the null hypothesis of no difference between any pair of means among the 4 methods, at significance level ### 0.05?

2.7. Compare the mean weights for the 4 methods using ANOVA. State the F-statistic and the p-value for the F-test. Include your R code for this question.

2.8. What do you conclude from the ANOVA?

2.9. How does your conclusion from ANOVA compare to the conclusion from the pairwise comparisons?