DATA557 Exercise 3

Will Wright

January 23, 2019

Problems

Applied Statistics and Design of Experiments Exercise 3 January 23, 2018

Question 1

Data: 'process.csv'

The data are from an experiment designed to assess the effect of temperature on the output of a manufacturing process. For this experiment temperatures of 50 or 100 degrees were randomly assigned to process runs and yield was recorded from each run.

1.1. Perform a large-sample Z-test to compare mean output for the two temperatures (use the sample SDs to calculate the test statistic). Is there evidence that output is affected by temperature? State the p-value and give a 95% confidence interval for the difference between mean yields.

```
pDat <- read.csv("process.csv")
xbar50 <- mean(pDat$output[which(pDat$temp==50)])
xbar100 <- mean(pDat$output[which(pDat$temp==100)])

se <- sqrt(var(pDat$output[which(pDat$temp==50)])/50+var(pDat$output[which(pDat$temp==100)]))
z <- abs((xbar50-xbar100)/se)
c <- qnorm(0.975)
z
```

```
## [1] 0.9170068
```

```
p_val <- 2*(1-pnorm(z))
c
```

```
## [1] 1.959964
```

```
conf_upper <- abs(xbar50-xbar100) + z*se
conf_lower <- abs(xbar50-xbar100) - z*se
```

The p-value is 0.36 and the confidence interval is 0 to 268.63

1.2. Perform a simulation study to estimate the type I error rate of the large-sample Z-test. Assume the population for both groups is normal with mean equal to 1000 and SD equal to 100 (mean and SD should be the same in each group). Use the same sample sizes for each group as in the data set. Use a 0.05 level of significance. What is the estimated type I error probability?

1.3. Run a simulation study assuming the common mean in the two groups is equal to 10 and the common SD is equal to 20. How do the results compare? Would you expect a different type I error rate when you change the values of the mean and SD?

1.4. Perform a simulation study to assess the performance of the large-sample Z-test for non-normal populations. Choose a distribution that is quite different from a normal distribution (e.g., t with low df, exponential, uniform or dichotomous). Is the type I error probability much different than 0.05?

1.5. Repeat the simulation studies from 1.3 and 1.4 using sample sizes for the two groups equal to 5. How do the results change?

Question 2

2.1. Perform the equal-variance t-test and the Welch t-test to compare mean output in the two temperature groups from the process experiment. Compare to the results from the large-sample Z-test. Comment on the validity of each of the three test methods for this experiment. Which one would you choose to use?

2.2. Now suppose that equipment failure had occurred and the last 15 runs of the experiment had been lost (delete rows 46-60 in the data set). Conduct each of the three tests (large-sample Z-test, equal-variance t-test, and Welch t-test), and compute the corresponding 95% confidence intervals. How do the results from the three tests compare on this new data set?

2.3. Conduct a simulation study to assess the performance of the three test procedures under the following assumptions:

Temp=50: n=30, output is distributed as N(1000,80) Temp=100: n=15, output is distributed as N(1000,20)

Estimate the type I error probabilities for the large-sample Z-test, equal-variance t-test, and the Welch t-test for each choice of group sample sizes. Use a 0.05 level of significance.

2.4. Conduct a simulation study to assess the performance of the three test procedures under these assumptions:

Temp=50: n=30, output is distributed as N(1000,20) Temp=100: n=15, output is distributed as N(1000,80)

Estimate the type I error probabilities for the large-sample Z-test, equal-variance t-test, and the Welch t-test for each choice of group sample sizes. Use a 0.05 level of significance. How do the results compare between the two sets of assumptions?

Question 3

For this question you are to work in groups. Find a group of 3 or 4 students and work together to come to a solution. One member of the group will report the results for the group at the end of the exercise.

Suppose that a new experiment is being designed to determine the effect on output of temperatures higher than 100. In particular, the aim of the new experiment is to test the null hypothesis that the mean output is

the same for temperature 100 and temperature 120. The researcher would like to have at least 90% power to detect a difference between these conditions in mean output equal to 75. Your job is to determine the sample sizes for each group and to decide which test statistic will be used to test the null hypothesis. Justify your answers.

Question 4

Data: 'fieldgoals.csv'

The data are from a random sample of field goal attempts in National Football League games from 2008.

The variables to be used here are:

distance: distance from ball placement to goal posts (yards) GOOD: indicator of whether the attempt was successful

4.1. Compare the proportions of field goals that are successful for attempts at distance 40 yards or less versus more than 40 yards using the large-sample Z-test with significance level 0.05. Report the p-value from the test. Perform the test using the chi-square test with and without continuity correction and also using Fisher's exact test. How do the results of the different tests compare?

4.2. Conduct a simulation study to estimate the type I error probability of the large-sample Z-test. Does your simulation study support the validity of the test in this situation?

4.3. Repeat the simulation study but use sample sizes of 30 in each group. For this simulation include the large-sample Z-test, the chi-square test with continuity correction and Fisher's exact test. Which test has the greatest validity in this situation based on your simulation?