DATA 557: HW Assignment 2

Hriday Baghar

February 4, 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

```
##       run           temp         output
##  Min.   : 1.00   Min.   :60   Min.   : 974.2
##  1st Qu.: 8.25   1st Qu.:60   1st Qu.: 997.0
##  Median :15.50   Median :75   Median :1008.3
##  Mean   :15.50   Mean   :70   Mean   :1013.4
##  3rd Qu.:22.75   3rd Qu.:75   3rd Qu.:1031.2
##  Max.   :30.00   Max.   :75   Max.   :1070.8
```

```
library(dplyr)
temp.exp.60 <- (temp.exp %>% filter(temp == 60))$output
temp.exp.75 <- (temp.exp %>% filter(temp == 75))$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.

We create a Z-statistic as follows. The subscript denotes the temperature of the run.

$$Z = \frac{\mu_{75} - \mu_{60}}{\sqrt{\frac{s_{60}^2}{n_{60}} + \frac{s_{75}^2}{n_{75}}}}$$

```
mu.60 <- mean(temp.exp.60); sd.60 <- sd(temp.exp.60); n.60 <- length(temp.exp.60)
mu.75 <- mean(temp.exp.75); sd.75 <- sd(temp.exp.75); n.75 <- length(temp.exp.75)

Z.1 <- (mu.75 - mu.60)/sqrt((sd.60^2/n.60)+(sd.75^2/n.75))
P.1 <- 1-pnorm(Z.1)

data.frame(Z.statistic = abs(round(Z.1,3)), p.value = round(P.1,3))
```

```
##   Z.statistic p.value
## 1      2.551   0.005
```

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

Yes, at a significance level of 0.05 we **reject the null hypothesis** (since $p < \alpha$)

1.3. State the null hypothesis for the test.

We define a test of null hypothesis as follows:

$$H_0 : \mu_{75} = \mu_{60}$$

which means the average output at 60 deg C and 75 deg C are not different.

We can also define this as,

$$H_0 : \mu_{75} - \mu_{60} = 0$$

The alternate hypothesis is as follows:

$$H_1 : \mu_{75} > \mu_{60}$$

Also defined as,

$$H_1 : \mu_{75} - \mu_{60} > 0$$

Note: We use a one-tailed alternate hypothesis since the problem states the following:

> It seems unlikely that increasing the temperature would decrease output and, even if it did, there would be no value
> in having that information.

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.

```
welch.df.1 <- (sd.60^2/n.60 + sd.75^2/n.75)^2 /
              (sd.60^4/(n.60^2*(n.60 - 1)) + (sd.75^4/(n.75^2*(n.75 - 1)))))

data.frame(Z.statistic = abs(round(Z.1,3)), p.value = round(1-pt(Z.1, df = welch.df.1),3))
```

```
##   Z.statistic p.value
## 1      2.551   0.009
```

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.

```
pooled.var.1 <- ((n.60-1)*sd.60^2 + (n.75-1)*sd.75^2)/(n.60 - 1 + n.75 - 1)
Z.1.5 <- (mu.75 - mu.60)/sqrt(pooled.var.1*(1/n.60 + 1/n.75))
data.frame(Z.statistic = abs(round(Z.1.5,3)), p.value = round(1-pt(Z.1.5, df = (n.60 + n.75 - 2)),3))
```

```
##   Z.statistic p.value
## 1      1.925   0.032
```

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

Based on our observations, the **Welch t-test would be most appropriate** for the sample of data. Here are the reason why:

1. **Small sample of data** - which is why we should not rely on the large sample z-test

2. **Unequal variances of the observed samples** - In the observed samples, the difference in variances seems far too high for us to assume equal variances ($s^2_{60} = 91.56$ and $s^2_{75} = 836.98$). Which is why equal variance t-test should not be relied on.

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

```
z.05 <- qnorm(0.950)
se <- sqrt(sd.60^2/n.60 + sd.75^2/n.75)
data.frame(lower.ci  = mu.75 - mu.60 - z.05*se, upper.ci =  "Inf")
```

```
##    lower.ci upper.ci
## 1  6.47268      Inf
```

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

```
w.05 <- qt(0.950, df = welch.df.1)
data.frame(lower.ci  = mu.75 - mu.60 - w.05*se, upper.ci ="Inf")
```

```
##    lower.ci upper.ci
## 1 6.032275      Inf
```

```
#Sanity check
t.test(temp.exp.75, temp.exp.60, alternative = "greater", paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  temp.exp.75 and temp.exp.60
## t = 2.5512, df = 25.633, p-value = 0.008531
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.032275      Inf
## sample estimates:
## mean of x mean of y
##   1019.46   1001.24
```

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

```
t.05 <- qt(0.950, df = n.60+n.75-2)
se.t <- sqrt(pooled.var.1*(1/n.60+1/n.75))
data.frame(lower.ci  = mu.75 - mu.60 - t.05*se.t, upper.ci = "Inf")
```

```
##    lower.ci upper.ci
## 1 2.116827      Inf
```

```
#Sanity check
t.test(temp.exp.75, temp.exp.60, alternative = "greater", paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  temp.exp.75 and temp.exp.60
## t = 1.9248, df = 28, p-value = 0.03224
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.116827      Inf
## sample estimates:
## mean of x mean of y
```

```
##    1019.46    1001.24
```

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

While the tests conclude that the there is an increase in output however, there is **very high variance in the sample**. It is possible that the increment in output might not be worth the additional cost that would be incurred.

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?

We can use a **one sample t-test** to test the above hypothesis.

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

```r
#Separating into 4 vectors for each method A-D
library(dplyr)
wgt.subsets <- lapply(unique(defects$Method),
                function(m) {
                        (defects %>% filter(Method == m))$Weight
                })
names(wgt.subsets) <- paste0("method.", unique(defects$Method))

#Defining one sample t-test function
my.t.test <- function(x, mu.bar = 10){
  mu <- mean(x); s <- sd(x); n <- length(x)

  z <- (mu - mu.bar)/(s/sqrt(n))
  p <- round(2*(1 - pt(abs(z), df = n-1)),3)
```

```
    return(p)
}
print("P-value for t-test by Method")
```

```
## [1] "P-value for t-test by Method"
```

```
sapply(wgt.subsets, my.t.test)
```

```
## method.A method.B method.C method.D
##    0.937    0.622    0.142    0.044
```

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?

To apply a Bonferroni correction, we must divide the significance level $\alpha = 0.05$ by the number of groups we want to compare. The new significance level is $\alpha_{corrected} = 0.05/4 = 0.0125$

Applying this correction results in **not having sufficient evidence to reject the null hypothesis for any of the methods** at a 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?

We can use a **two-sample t-test** to perform a pairwise comparison of each method.

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

```
combs <- combn(names(wgt.subsets),2)
df <- data.frame(method1 = combs[1,], method2 = combs[2,],
                 p.val = rep(NA,length(combs[1,])), stringsAsFactors = FALSE)

for(row in 1:nrow(df)){
  x <- as.vector(unlist(wgt.subsets[df[row,1]]))
  y <- as.vector(unlist(wgt.subsets[df[row,2]]))
  df[row, "p.val"] <- round(t.test(x,y, var.equal = FALSE, paired = FALSE)$p.value,3)
}
df
```

```
##     method1  method2 p.val
## 1 method.A method.B 0.682
## 2 method.A method.C 0.222
## 3 method.A method.D 0.065
## 4 method.B method.C 0.131
## 5 method.B method.D 0.040
## 6 method.C method.D 0.392
```

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?

To apply a Bonferroni correction, we must divide the significance level $\alpha = 0.05$ by the number of tests we want to conduct. The new significance level is $\alpha_{corrected} = 0.05/6 = 0.0083$

Applying this correction results in **not having sufficient evidence to reject the null hypothesis for any of the methods** at a 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.

We specify an ANOVA test with the null hypothesis for weight for each Method (A, B, C, D) as follows:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$$

```r
ssw <- sum(sapply(unique(defects$Method),
             function(mthd){
               sum(
                 (defects[defects$Method == mthd, "Weight"]
                  - mean(defects[defects$Method == mthd, "Weight"]))^2
               )}))

ssb <- sum(sapply(unique(defects$Method),
             function(mthd){
               nrow(defects[defects$Method == mthd,])*
                 (mean(defects[defects$Method == mthd, "Weight"])
                  - mean(defects[,"Weight"]))^2
             }))

tot <- sum((defects$Weight - mean(defects$Weight))^2)
cat("Checking ANOVA decomposition equality:", all.equal(tot, ssb+ssw))
```

```
## Checking ANOVA decomposition equality: TRUE
```

```r
k <- length(unique(defects$Method))
N <- nrow(defects)

F.score <- (ssb/(k-1))/(ssw/(N-k))
anova.p.val <- 1-pf(F.score, df1 = k-1, df2 = N-k)
cat("F-score = ", round(F.score,3), "\nP-value = ", round(anova.p.val, 4))
```

```
## F-score =  2.617
## P-value =  0.0515
```

```r
#sanity check
summary(aov(Weight ~ Method, data = defects))
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Method        3   1.29  0.4299   2.617 0.0515 .
## Residuals   261  42.87  0.1643
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.8. What do you conclude from the ANOVA?

We **fail to reject the null hypothesis** since the p-value is 0.0515.

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

We arrive at the same conclusion from ANOVA although we have a single p-value that makes our interpretation easier.