

Data557 Homework Assignment 2

Anuhya B S

February 3, 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.

In this case since we want to compare the population means of two samples of outputs at different temperatures, the hypothesis test will be based on the two sample means $\bar{X}_A - \bar{X}_B$.

The means are defined as

$$\bar{X}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} X_{Ai}$$

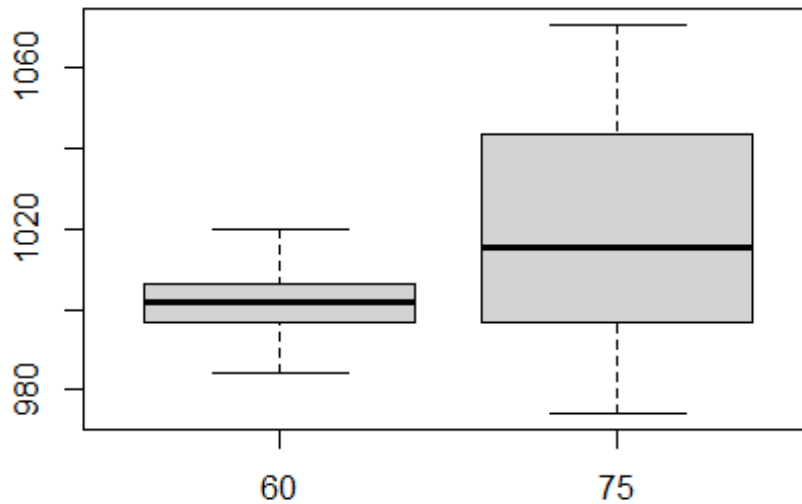
$$\bar{X}_B = \frac{1}{n_B} \sum_{i=1}^{n_B} X_{Bi}$$

To calculate the test statistic,

$$\text{Z Test Statistic} = \frac{\text{Sample Estimate} - \text{Hypothesized Value}}{\text{SE of the Sample Estimate}}$$

Here the $\bar{X}_A = 1001.24$ and $\bar{X}_B = 1019.46$ and the calculated values of **Z test statistic = -2.551 and p-value = 0.0107.**

```
data <- read.csv('temperature_experiment.csv')
boxplot(split(data$output, data$temp))
```



```
m = with(data, tapply(output, temp, mean))
s = with(data, tapply(output, temp, sd))
n = with(data, tapply(output, temp, length))
data.frame(m, s, n)
```

```
##           m           s  n
## 60 1001.24  9.569187 10
## 75 1019.46 28.930705 20
```

```
z = (m[1] - m[2]) / sqrt(sum(s^2 / n))
p = round(2 * (pnorm(z)), 4)
data.frame(z, p)
```

```
##           z           p
## 60 -2.551155 0.0107
```

```
# z = -2.551155
# p = 0.0107
```

However, since the question states that there is no value in knowing whether there is a decrease in output when temperature is increased it makes sense to only consider one half of the p-value which is 0.0054.

```
p = round((pnorm(z)),4)
p
##      60
## 0.0054
```

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

We **reject the null hypothesis** of equal means at the 0.05 level of significance since the p-value (0.01) is lesser than the alpha value (0.05).

1.3. State the null hypothesis for the test.

We define the null hypothesis as

$$H_0: \mu_A = \mu_B,$$

where μ_A and μ_B are the mean outputs at temperatures 60 and 75 degrees C respectively.

The alternative hypothesis is

$$H_1: \mu_A \neq \mu_B.$$

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.

The calculated values of **test statistic = -2.5512** and **p-value = 0.01706**.

```
with(data, t.test(output[temp==60], output[temp==75], var.equal = F))
##
## Welch Two Sample t-test
##
## data: output[temp == 60] and output[temp == 75]
## t = -2.5512, df = 25.633, p-value = 0.01706
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -32.910534 -3.529466
## sample estimates:
## mean of x mean of y
## 1001.24 1019.46
```

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.

The equal-variance test statistic uses a pooled estimate of the variance which is the weighted average of the two sample variances. Thus the test statistics has been calculated as follows:

$$Z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{s^2(1/n_A + 1/n_B)}}$$

The calculated values of **test statistic = -1.925 and p-value = 0.065**.

```
pooled_sample_var = sum((n-1)*s^2)/sum(n-1)
z_t = (m[1]-m[2])/sqrt(pooled_sample_var*sum(1/n))
p_t = round(2*(pt(z_t,sum(n)-2)),4)
data.frame(z_t,p_t)
```

```
##           z_t      p_t
## 60 -1.924751 0.0645
```

```
#z_t = -1.924751
#p_t = 1.9355
```

Cross checking the values using the t.test function:

```
with(data,t.test(output[temp==60],output[temp==75],var.equal = T))

##
## Two Sample t-test
##
## data: output[temp == 60] and output[temp == 75]
## t = -1.9248, df = 28, p-value = 0.06448
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -37.610544 1.170544
## sample estimates:
## mean of x mean of y
## 1001.24 1019.46
```

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

Firstly, the large sample test must be ruled out as the groups have a significantly small data size(10 and 20). The appropriate test here would be a t-test. The variance of both the groups are far from equal as shown below:

```
s^2

##           60           75
## 91.56933 836.98568
```

This rules out the equal-variance T-Test. Thus the most valid test for this experiment would be the **Welch's T-Test of unequal variances**.

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

On calculating the 95% confidence interval for the difference between the mean outputs, we get the lower bound = -32.218 and higher bound = -4.222. Thus the confidence interval is

$$\text{ConfidenceInterval} = [-32.218, -4.222]$$

```
se = sqrt(s[1]^2/n[1] + s[2]^2/n[2])
z.05 = qnorm(0.975)
lower = m[1]-m[2] - z.05 * se
higher = m[1]-m[2] + z.05 * se
lower #-32.2178

##          60
## -32.2178

higher #-4.22204

##          60
## -4.22204
```

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

On calculating the 95% confidence interval for the difference between the mean outputs, we get the lower bound = -32.911 and higher bound = -3.529. Thus the confidence interval is

$$\text{ConfidenceInterval} = [-32.911, -3.529]$$

```
se_w = sqrt(sum(s^2/n))
t.05=qt(0.975,25.633)
lower_w = m[1]-m[2]-t.05*se_w
upper_w = m[1]-m[2]+t.05*se_w
lower_w #-32.91055

##          60
## -32.91055

upper_w #-3.529452

##          60
## -3.529452
```

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

On calculating the 95% confidence interval for the difference between the mean outputs, we get the lower bound = -37.611 and higher bound = 1.171. Thus the confidence interval is

$$\text{ConfidenceInterval} = [-37.611, 1.171]$$

```

se_t=sqrt(pooled_sample_var*sum(1/n))
t.05=qt(0.975,df=sum(n)-2)
lower_t = m[1]-m[2]-t.05*se_t
upper_t = m[1]-m[2]+t.05*se_t
lower_t #-37.61054

##          60
## -37.61054

upper_t #1.170544

##          60
## 1.170544

```

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

From the above conducted tests, we reject the null hypothesis of equal mean outputs at 0.05 level of significance in the large samples test as well as the Welch's T-test. We do not have enough evidence to reject the null hypothesis of equal mean outputs as per the equal variance T-Test. Apart from the effect of the mean output, there would be an increase in the cost of the manufacturing process as given in the question however the results of the experiment do not give any indication of this disadvantage.

It can be seen from the results of this particular experiment that with an increase in temperature there is an increase in the variance of the weights which may 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

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?

I believe that the **1-sample T-test** must be used to test the null hypothesis that mean weight is equal to 10. The conditions to satisfy a valid 1-Sample T-Test are:

1. The population distribution is normal.
2. The sample size is sufficiently large.

There is no suggestion that the population is normally distributed in the question. However, The sample size of the methods lie in the range of 50-75. Thus, the sample size is sufficiently large for the distribution of t-test and normal distribution to not be distinguishable.

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

```
data_defect = read.csv('defects.csv')
mean_methods = with(data_defect, tapply(data_defect$Weight,
data_defect$Method, mean))
sd_methods = with(data_defect, tapply(data_defect$Weight, data_defect$Method,
sd))
len_methods = with(data_defect, tapply(data_defect$Weight, data_defect$Method,
length))
data.frame(mean_methods, sd_methods, len_methods)

##   mean_methods sd_methods len_methods
## A    10.002703  0.2918871          74
## B     9.982667  0.3028573          75
## C    10.082812  0.4456330          64
## D    10.167308  0.5833376          52

var_methods = sd_methods^2
var_methods

##           A           B           C           D
## 0.08519807 0.09172252 0.19858879 0.34028281
```

To perform 1 sample t-test, the null hypothesis is

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D = 10,$$

The alternative hypothesis is that the means are not all equal. At level of significance, $\alpha=0.05$ the p-values for the methods are:

Method A: 0.9367

Method B: 0.6216

Method C: 0.1421

Method D: 0.04371

```
t.test(data_defect$Weight[data_defect$Method=='A'], mu = 10, alternative =
"two.sided")

##
## One Sample t-test
```

```
##
## data: data_defect$Weight[data_defect$Method == "A"]
## t = 0.079652, df = 73, p-value = 0.9367
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
## 9.935078 10.070327
## sample estimates:
## mean of x
## 10.0027

t.test(data_defect$Weight[data_defect$Method=="B"],mu = 10, alternative =
"two.sided")

##
## One Sample t-test
##
## data: data_defect$Weight[data_defect$Method == "B"]
## t = -0.49565, df = 74, p-value = 0.6216
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
## 9.912986 10.052348
## sample estimates:
## mean of x
## 9.982667

t.test(data_defect$Weight[data_defect$Method=="C"],mu = 10, alternative =
"two.sided")

##
## One Sample t-test
##
## data: data_defect$Weight[data_defect$Method == "C"]
## t = 1.4866, df = 63, p-value = 0.1421
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
## 9.971497 10.194128
## sample estimates:
## mean of x
## 10.08281

t.test(data_defect$Weight[data_defect$Method=="D"],mu = 10, alternative =
"two.sided")

##
## One Sample t-test
##
## data: data_defect$Weight[data_defect$Method == "D"]
## t = 2.0682, df = 51, p-value = 0.04371
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
## 10.00491 10.32971
```



```
## sample estimates:  
## mean of x  
## 10.16731
```

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?

Applying Bonferroni correction to the results of the previous question, Since we performed 4 tests the new level of significance $\alpha = 0.05/4 = 0.0125$.

Since the p-value for all methods is more than the new α value, we **do not have evidence to reject the null hypothesis** that the mean weight for all 4 methods is equal to 10.

2.4. It 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?

Assuming that all the methods have an equal variance, the **Analysis of Variance (ANOVA) Test** would be an appropriate test to perform pairwise comparisons of mean weights for the different methods.

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

The p-values are as follows:

Method AB:0.6816

Method AC:0.2082

Method AD:0.039

Method BC:0.1192

Method BD:0.02151

Method CD:0.3784

```
pAB =  
t.test(data_defect$Weight[data_defect$Method=='A'],data_defect$Weight[data_de  
fect$Method=='B'],var.equal = T)  
pAB  
  
##  
## Two Sample t-test  
##  
## data: data_defect$Weight[data_defect$Method == "A"] and  
data_defect$Weight[data_defect$Method == "B"]  
## t = 0.41109, df = 147, p-value = 0.6816
```

```

## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.07628331 0.11635538
## sample estimates:
## mean of x mean of y
## 10.002703 9.982667

pAC =
t.test(data_defect$Weight[data_defect$Method=='A'],data_defect$Weight[data_defect$Method=='C'],var.equal = T)
pAC

##
## Two Sample t-test
##
## data: data_defect$Weight[data_defect$Method == "A"] and
data_defect$Weight[data_defect$Method == "C"]
## t = -1.2646, df = 136, p-value = 0.2082
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.20538623 0.04516664
## sample estimates:
## mean of x mean of y
## 10.00270 10.08281

pAD =
t.test(data_defect$Weight[data_defect$Method=='A'],data_defect$Weight[data_defect$Method=='D'],var.equal = T)
pAD

##
## Two Sample t-test
##
## data: data_defect$Weight[data_defect$Method == "A"] and
data_defect$Weight[data_defect$Method == "D"]
## t = -2.0863, df = 124, p-value = 0.039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.320768556 -0.008441424
## sample estimates:
## mean of x mean of y
## 10.00270 10.16731

pBC =
t.test(data_defect$Weight[data_defect$Method=='B'],data_defect$Weight[data_defect$Method=='C'],var.equal = T)
pBC

##
## Two Sample t-test
##

```

```

## data: data_defect$Weight[data_defect$Method == "B"] and
data_defect$Weight[data_defect$Method == "C"]
## t = -1.568, df = 137, p-value = 0.1192
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.22644205 0.02615038
## sample estimates:
## mean of x mean of y
## 9.982667 10.082812

pBD =
t.test(data_defect$Weight[data_defect$Method=="B"],data_defect$Weight[data_de
fect$Method=="D"],var.equal = T)
pBD

##
## Two Sample t-test
##
## data: data_defect$Weight[data_defect$Method == "B"] and
data_defect$Weight[data_defect$Method == "D"]
## t = -2.3282, df = 125, p-value = 0.02151
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.34159532 -0.02768673
## sample estimates:
## mean of x mean of y
## 9.982667 10.167308

pCD =
t.test(data_defect$Weight[data_defect$Method=="C"],data_defect$Weight[data_de
fect$Method=="D"],var.equal = T)
pCD

##
## Two Sample t-test
##
## data: data_defect$Weight[data_defect$Method == "C"] and
data_defect$Weight[data_defect$Method == "D"]
## t = -0.88422, df = 114, p-value = 0.3784
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2737961 0.1048057
## sample estimates:
## mean of x mean of y
## 10.08281 10.16731

```

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?

```

set.seed(123)
n=5
m=10
sigma=2
reps=2000
pvalues=data.frame(p_AB=rep(NA, reps), p_AC=rep(NA, reps), p_AD=rep(NA, reps), p_BC
=rep(NA, reps), p_BD=rep(NA, reps), p_CD=rep(NA, reps))
for(i in 1:reps){
  x1=rnorm(n,m,sigma)
  x2=rnorm(n,m,sigma)
  x3=rnorm(n,m,sigma)
  x4=rnorm(n,m,sigma)
  pvalues$p_AB[i]=t.test(x1,x2,var.equal=T)$p.value
  pvalues$p_AC[i]=t.test(x1,x3,var.equal=T)$p.value
  pvalues$p_AD[i]=t.test(x1,x4,var.equal=T)$p.value
  pvalues$p_BC[i]=t.test(x2,x3,var.equal=T)$p.value
  pvalues$p_BD[i]=t.test(x2,x4,var.equal=T)$p.value
  pvalues$p_CD[i]=t.test(x3,x4,var.equal=T)$p.value
}
reject = data.frame(pvalues < 0.05, any.rejection=apply(pvalues<0.05, 1,
any))
apply(reject,2,mean)

##           p_AB           p_AC           p_AD           p_BC           p_BD
##          0.0560          0.0545          0.0450          0.0530          0.0535
##           p_CD any.rejection
##          0.0470          0.2230

```

Since we perform 6 tests, each of which has a probability 0.05 of rejecting, then the probability of at least one of the tests rejecting cannot be greater than $0.05 \times 6 = 0.90$. In the simulation the probability of rejecting is 0.223 which is less than 0.90.

After applying Bonferroni correction to the results of the simulation, the overall type I error probability is 0.04. Since the Bonferroni test is a conservative one, the type I error probability is strictly below 0.05 which is its disadvantage.

```

reject.Bonf=data.frame(pvalues < 0.05/6, any.rejection=apply(pvalues<0.05/6,
1, any))
apply(reject.Bonf,2,mean)

##           p_AB           p_AC           p_AD           p_BC           p_BD
##          0.0080          0.0125          0.0055          0.0045          0.0075
##           p_CD any.rejection
##          0.0070          0.0400

```

On adjusting the level of significance using Bonferroni's correction, we would get the new value to be $0.05/6 = 0.008$. On the basis of the new level of significance, we **do not have evidence to reject the null hypothesis** of no difference between the pairs.

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.

```

mean_total = mean(data_defect$Weight)
method_A = data_defect$Weight[data_defect$Method=='A']
method_B = data_defect$Weight[data_defect$Method=='B']
method_C = data_defect$Weight[data_defect$Method=='C']
method_D = data_defect$Weight[data_defect$Method=='D']
ss_total = sum((method_A-mean_total)^2) + sum((method_B-mean_total)^2) +
  sum((method_C-mean_total)^2) + sum((method_D-mean_total)^2)
ss_within = sum((method_A-mean_methods[1])^2) + sum((method_B-
mean_methods[2])^2) +
  sum((method_C-mean_methods[3])^2) + sum((method_D-mean_methods[4])^2)
ss_between = sum((len_methods[1]*(mean_methods[1]-mean_total)^2) +
  (len_methods[2]*(mean_methods[2]-mean_total)^2) +
  (len_methods[3]*(mean_methods[3]-mean_total)^2) +
  (len_methods[4]*(mean_methods[4]-mean_total)^2))
data.frame(ss_between,ss_within,ss_total)

##   ss_between ss_within ss_total
## 1    1.289595  42.87244 44.16204

method <- as.factor(data_defect$Method)
summary(aov(data_defect$Weight~method))

##              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

```

The F-statistic value is 2.617. The p-value is 0.0515.

2.8. What do you conclude from the ANOVA?

Through ANOVA test, we can see that the F-statistic value is 2.617. A large value of F provides evidence against the null hypothesis. However, in this case we can not say for sure that F-statistic value is large enough. The p-value is 0.0515 which indicates there is **not enough evidence against the null hypothesis** of equal group means.

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

The conclusion from ANOVA is the same as the conclusion from the pairwise comparison. In ANOVA, we do not reject the null hypothesis.

In pairwise comparison, before the Bonferroni's correction we reject the null hypothesis. However after applying Bonferroni's correction, we do not reject the null hypothesis of equal group means in the pairwise comparison.