

Homework Assignment 2

Shrusti Ghela

February 03, 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.

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

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

We use the following test statistic:

$$Z = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{s_A^2/n_A + s_B^2/n_B}}.$$

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

```
##           z           p
## 60 -2.551155 0.0054
```

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

The p-value(0.0054) is less than 0.05, so we would reject the null hypothesis of true difference of means is greater than or equal to 0 at the 0.05 level of significance.

1.3. State the null hypothesis for the test.

The question clearly states that we are only interested in if there is an increase in the output when the temperature is increased. We are not interested in the scenario where increasing the temperature would decrease output. Considering this, since we are interested only in the increase in the output,

We define the null hypothesis as

$$H_0 : \mu_1 \geq \mu_2,$$

where μ_1 and μ_2 are the mean of the process output for temperatures 60 and 75, respectively.

we set the alternative hypothesis to be

$$H_1 : \mu_1 < \mu_2.$$

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.

```
with(temperature, t.test(output[temp=="60"], output[temp=="75"], var.equal=F, alternative="less"))

##
##  Welch Two Sample t-test
##
## data:  output[temp == "60"] and output[temp == "75"]
## t = -2.5512, df = 25.633, p-value = 0.008531
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -6.032275
## 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.

```
with(temperature, t.test(output[temp=="60"], output[temp=="75"], var.equal=T, alternative="less"))

##
##  Two Sample t-test
##
## data:  output[temp == "60"] and output[temp == "75"]
## t = -1.9248, df = 28, p-value = 0.03224
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -2.116827
## 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?

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

From this, we can see that the standard deviations for the two groups are not equal. Also, the sample size is not sufficiently large. Considering these two facts, I believe that Welch test would be the most valid test for this experiment.

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

Here, I have used two-tailed test instead of one-tailed test to calculate the confidence interval

```
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
upper = m[1]-m[2]+z.05*se
lower
```

```
##           60
## -32.2178
```

```
upper
```

```
##           60
## -4.222204
```

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

```
with(temperature, t.test(output[temp=="60"], output[temp=="75"], var.equal=F, alternative="less"))
```

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

For two-tailed test, the confidence interval is given by:

```
with(temperature, t.test(output[temp=="60"], output[temp=="75"], var.equal=F, alternative="two.sided"))
```

```
##
## 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
```

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

```
with(temperature, t.test(output[temp=="60"], output[temp=="75"], var.equal=T, alternative = "less"))
```

```
##
## Two Sample t-test
##
## data: output[temp == "60"] and output[temp == "75"]
## t = -1.9248, df = 28, p-value = 0.03224
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
## -Inf -2.116827
## sample estimates:
## mean of x mean of y
## 1001.24 1019.46
```

The confidence interval for two-tailed test:

```
with(temperature, t.test(output[temp=="60"], output[temp=="75"], var.equal=T, alternative = "two.sided"))
```

```
##
## 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
```

1.10. Apart from any effect on the mean output, do the results of the experiment suggest a disadvantage of the higher temperature? The variance in output has increased due to increase in temperature. High variation is a disadvantage. Another disadvantage the cost associated with the increase in temperature. But it is not a parameter in the experiment.

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

```
defects <- read.csv("defects.csv")
```

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?

```
m = with(defects, tapply(Weight,Method,mean))
s = with(defects, tapply(Weight,Method,sd))
n = with(defects, tapply(Weight,Method,length))

data.frame(m,s,n)
```

```
##           m           s  n
## A 10.002703 0.2918871 74
## B  9.982667 0.3028573 75
## C 10.082812 0.4456330 64
## D 10.167308 0.5833376 52
```

We have taken random sample of 1% of ball bearings for all 4 methods. I don't think it is sufficiently large to carry out a z-test, so conducting a one-sample two-tailed t-test would be our best option.

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

```
t.test(defects$Weight[defects$Method=="A"], mu=10, alternative="two.sided")
```

```
##
## One Sample t-test
##
## data: defects$Weight[defects$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(defects$Weight[defects$Method=="B"], mu=10, alternative="two.sided")
```

```
##
## One Sample t-test
##
## data: defects$Weight[defects$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(defects$Weight[defects$Method=="C"], mu=10, alternative="two.sided")
```

```
##
## One Sample t-test
##
## data: defects$Weight[defects$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(defects$Weight[defects$Method=="D"], mu=10, alternative="two.sided")
```

```
##
## One Sample t-test
##
## data: defects$Weight[defects$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?

The idea of the Bonferroni correction is to adjust the significance level of each test so that the overall probability of a type I error is controlled at the desired level.

Here, we want to perform 4 tests, and we want the overall type I error of 0.05, we would perform each test with significance level $0.05/4$

So, instead of comparing our p-value with 0.05, we compare the p-value obtained from the above tests with 0.0125

The p-value(0.9367) is greater than 0.0125, so we would not reject the null hypothesis of equal means at the 0.0125 level of significance.

The p-value(0.6216) is greater than 0.0125, so we would not reject the null hypothesis of equal means at the 0.0125 level of significance.

The p-value(0.1421) is greater than 0.0125, so we would not reject the null hypothesis of equal means at the 0.0125 level of significance.

The p-value(0.04371) is greater than 0.0125, so we would not reject the null hypothesis of equal means at the 0.0125 level of significance.

Overall, because the p-values of all test are greater than 0.0125, we would not reject the null hypothesis of equal means at 0.05 significance.

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?

We try to do this by testing all the pairwise hypotheses, $H_0 : \mu_a = \mu_b$, $H_0 : \mu_a = \mu_c$, $H_0 : \mu_a = \mu_d$, $H_0 : \mu_b = \mu_c$, $H_0 : \mu_b = \mu_d$, and $H_0 : \mu_c = \mu_d$, using the two-sample t-test (assuming equal variances)

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

```
A.vs.B=t.test(defects$Weight[defects$Method=="A"],
defects$Weight[defects$Method=="B"],var.equal=T)
pAB = A.vs.B$p.value
pAB
```

```
## [1] 0.6816055
```

```
A.vs.C=t.test(defects$Weight[defects$Method=="A"],
defects$Weight[defects$Method=="C"],var.equal=T)
pAC = A.vs.C$p.value
pAC
```

```
## [1] 0.2081849
```

```
A.vs.D=t.test(defects$Weight[defects$Method=="A"],
defects$Weight[defects$Method=="D"],var.equal=T)
pAD = A.vs.D$p.value
pAD
```

```
## [1] 0.0390036
```

```
B.vs.C=t.test(defects$Weight[defects$Method=="B"],
defects$Weight[defects$Method=="C"],var.equal=T)
pBC= B.vs.C$p.value
pBC
```

```
## [1] 0.1191898
```

```
B.vs.D=t.test(defects$Weight[defects$Method=="B"],
defects$Weight[defects$Method=="D"],var.equal=T)
pBD = B.vs.D$p.value
pBD
```

```
## [1] 0.02150551
```

```
C.vs.D=t.test(defects$Weight[defects$Method=="C"],
defects$Weight[defects$Method=="D"],var.equal=T)
pCD = C.vs.D$p.value
pCD
```

```
## [1] 0.3784366
```

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?

Let us first estimate the Type I Error probability using simulation

```
set.seed(5)
n=5
m=10
sigma=2
reps=2000
pvalues=data.frame(pAB=rep(NA, reps), pAC=rep(NA, reps), pAD=rep(NA, reps), pBC=rep(NA, reps), pBD=rep(NA, reps), pCD=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)
  x5 = rnorm(n,m,sigma)
  x6 = rnorm(n,m,sigma)
  pvalues$pAB[i]=t.test(x1,x2,var.equal = T)$p.value
  pvalues$pAC[i]=t.test(x1,x3,var.equal = T)$p.value
  pvalues$pAD[i]=t.test(x1,x4,var.equal = T)$p.value
  pvalues$pBC[i]=t.test(x2,x3,var.equal = T)$p.value
  pvalues$pBD[i]=t.test(x2,x4,var.equal = T)$p.value
  pvalues$pCD[i]=t.test(x3,x4,var.equal = T)$p.value
}
reject = data.frame(pvalues<0.05, any.rejection=apply(pvalues<0.05, 1, any))
apply(reject,2,mean)
```

##	pAB	pAC	pAD	pBC	pBD
##	0.0565	0.0545	0.0500	0.0520	0.0535
##	pCD	any.rejection			
##	0.0445	0.2160			

Each of the six tests has an appropriate type I error probability of close to 0.05 but the overall type I error probability is greater than 0.22

If we perform 6 tests, each of which has probability 0.05 of rejecting, then the probability of at least one of the tests rejecting cannot be greater than 0.3

In our simulation of type I error, the probability of rejecting was 0.2160, less than 0.3

The idea of the Bonferroni correction is to adjust the significance level of each test so that the overall probability of a type I error is controlled at the desired level.

Applying Bonferroni correction to the previous simulation

We perform 6 tests and we want an overall type I error of 0.05, we perform each test with significance level $0.05/6$

```
set.seed(5)
n=5
m=10
sigma=2
```



```

reps=2000
pvalues=data.frame(pAB=rep(NA, reps), pAC=rep(NA, reps), pAD=rep(NA, reps),pBC=rep(NA, reps),pBD=rep(NA,
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)
  x5 = rnorm(n,m,sigma)
  x6 = rnorm(n,m,sigma)
  pvalues$pAB[i]=t.test(x1,x2,var.equal = T)$p.value
  pvalues$pAC[i]=t.test(x1,x3,var.equal = T)$p.value
  pvalues$pAD[i]=t.test(x1,x4,var.equal = T)$p.value
  pvalues$pBC[i]=t.test(x2,x3,var.equal = T)$p.value
  pvalues$pBD[i]=t.test(x2,x4,var.equal = T)$p.value
  pvalues$pCD[i]=t.test(x3,x4,var.equal = T)$p.value
}
reject = data.frame(pvalues<0.05/6, any.rejection=apply(pvalues<0.05/6, 1, any))
apply(reject,2,mean)

```

##	pAB	pAC	pAD	pBC	pBD
##	0.0105	0.0085	0.0105	0.0110	0.0055
##	pCD	any.rejection			
##	0.0075	0.0425			

Since, all the p-values (0.6816055, 0.2081849, 0.0390036, 0.1191898, 0.02150551, 0.3784366) are greater than 0.0083, we would not reject the null hypothesis of no difference between any pair of means among the 4 methods, at significance level 0.05

The overall type I error probability after applying Bonferroni correction is 0.0425 (close to 0.05).

We can see the proof of the fact that the Bonferroni procedure is conservative, that is it leads to a test that will have strictly less than the desired type I error probability.

Also, when we deal with large number of groups, the Bonferroni procedure becomes extremely conservative. The consequence is that the procedure will have very low power to detect differences when they exist.

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.

```

xA=defects$Weight[defects$Method=="A"]
xB=defects$Weight[defects$Method=="B"]
xC=defects$Weight[defects$Method=="C"]
xD=defects$Weight[defects$Method=="D"]

nA=length(xA)
nB=length(xB)
nC=length(xC)
nD=length(xD)

xbarA = mean(xA)
xbarB = mean(xB)
xbarC = mean(xC)
xbarD = mean(xD)

```

```

xbar = mean(c(xA, xB, xC, xD))
ss.total = sum((xA-xbar)^2) + sum((xB-xbar)^2) + sum((xC-xbar)^2) + sum((xD-xbar)^2)
ss.within = sum((xA-xbarA)^2) + sum((xB-xbarB)^2) + sum((xC-xbarC)^2) + sum((xD-xbarD)^2)
ss.between = nA*(xbarA-xbar)^2+nB*(xbarB-xbar)^2+nC*(xbarC-xbar)^2 + nD*(xbarD-xbar)^2
data.frame(ss.between,ss.within,ss.total)

##      ss.between ss.within ss.total
## 1      1.289595  42.87244 44.16204

```

The ANOVA F-statistic is based on the relative sizes of the between and within sums of squares. First the sums of squares are converted to “mean squares” as follows:

$$MS_B = \frac{SS_B}{k-1}$$

where k is the number of groups, and

$$MS_W = \frac{SS_W}{N-k}$$

where $N = \sum_{i=1}^k n_i$ is the total sample size. Then the F statistic is the ratio of the mean-squares:

$$F = \frac{MS_B}{MS_W}$$

A large value of F provides evidence against the null hypothesis.

```
summary(aov(defects$Weight~as.factor(Method), data=defects))
```

```

##              Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(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? The p-value (0.0515) is (slightly) greater than 0.05, which indicates that there is not evidence against the null hypothesis of equal group means. So, we would not reject the null hypothesis of equal group means at 0.05 significance level.

2.9. How does your conclusion from ANOVA compare to the conclusion from the pairwise comparisons? We arrive at the same conclusions on the null hypothesis by performing ANOVA and pairwise comparisons (with Bonferroni correction) However, we arrive at different conclusions on the null hypothesis by performing pairwise comparisons without Bonferroni correction.