

HW2

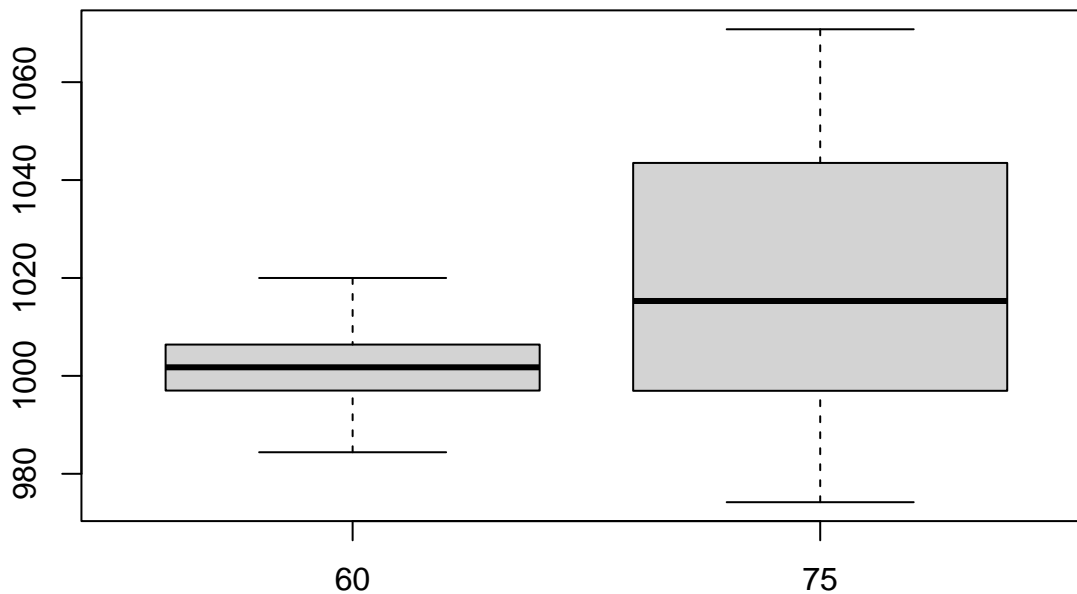
Amrit

2/3/2022

R Markdown

Question 1

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.



	mean	sd	sample_length
60	1001.24	9.569187	10
75	1019.46	28.930705	20

	z_statistic	p
60	-2.551155	0.0107

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

As $p=0.0107$ we would reject the null hypothesis of equal means at the 0.05 significance level.

1.3 State the null hypothesis for the test.

The null hypothesis is defined as

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

where μ_A and μ_B are the mean process outputs recorded from from each run (across two different temperatures)

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 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] "Test statistic: "

      t
-2.551155

[1] "p-value:"

[1] 0.008531113
```

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.

```
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] "Test statistic:"

      t
-1.924751

[1] "p-value:"

[1] 0.03223789
```

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

Descriptive stats between temperatures 60 and 75:

	mean	sd	sample_length
60	1001.24	9.569187	10
75	1019.46	28.930705	20

First, since the sample size is uncharacteristically small (less than 40), we disregard the use of the large sample Z-test. Second, the standard deviations between these two groups are very large, therefore we don't use the equal variance t-test. Therefore, the Welch T-test is most valid for this experiment.

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

```
se=sqrt(sum(sd^2/sample_length))
z.05=qnorm(0.975)
lower = mean[1]-mean[2]-z.05*se
```

```

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

      60
-32.2178
upper

      60
-4.222204
print(" 95% confidence interval for the difference between mean output using the large-sample (lower,upper): (")

[1] " 95% confidence interval for the difference between mean output using the large-sample (lower,upper): ("
print(paste(lower,",",upper,")"))

[1] "-32.2177956941121 , -4.22220430588794 )"

```

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

```

welch_t_test_results <- with(temperature_df,
  t.test(output[temp=="60"],output[temp=="75"],var.equal=F,alternative="two.sided"))
welch_t_test_results$conf.int

[1] -32.910534 -3.529466
attr(,"conf.level")
[1] 0.95

```

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

```

equal_variance_t_test_results <- with(temperature_df,
  t.test(output[temp=="60"],output[temp=="75"],var.equal=T,alternative="two.sided"))
equal_variance_t_test_results$conf.int

[1] -37.610544  1.170544
attr(,"conf.level")
[1] 0.95

```

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

Relying on results from the Welch t-test, since the t-statistic value is < 0 (-2.551155) with a p-value of 0.008531113 ($p < 0.05$), we can say that the experiment only suggests an advantage of the higher temperature conditions

Question 2

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?

As we are hypothesizing for an estimate's value within the same sample, we can use the 1 sample t-test

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

```

defects_df <- read.csv("/Users/amrit/Documents/Courses/Applied Statistics & Experimental Design/Assignments/A2/data")
t.test(defects_df[defects_df$Method == "A",]$Weight,mu=10)$p.value

[1] 0.9367316

```

```
t.test(defects_df[defects_df$Method == "B",]$Weight,mu=10)$p.value
```

```
[1] 0.6216105
```

```
t.test(defects_df[defects_df$Method == "C",]$Weight,mu=10)$p.value
```

```
[1] 0.1420953
```

```
t.test(defects_df[defects_df$Method == "D",]$Weight,mu=10)$p.value
```

```
[1] 0.04370811
```

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?

```
t.test(defects_df[defects_df$Method == "A",]$Weight,mu=10)$p.value/4
```

```
[1] 0.2341829
```

```
t.test(defects_df[defects_df$Method == "B",]$Weight,mu=10)$p.value/4
```

```
[1] 0.1554026
```

```
t.test(defects_df[defects_df$Method == "C",]$Weight,mu=10)$p.value/4
```

```
[1] 0.03552382
```

```
t.test(defects_df[defects_df$Method == "D",]$Weight,mu=10)$p.value/4
```

```
[1] 0.01092703
```

As two of the pairwise comparison between the means of method generates $p < 0.05$, i.e for method C and method D we reject the null hypothesis of no difference between any pair of means among the 4 methods

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?

As we are hypothesizing for comparison of means between 4 samples, we can use the pairwise two sample t-test or ANOVA test

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

```
def_A <- defects_df[defects_df$Method == "A",]$Weight
def_B <- defects_df[defects_df$Method == "B",]$Weight
def_C <- defects_df[defects_df$Method == "C",]$Weight
def_D <- defects_df[defects_df$Method == "D",]$Weight
```

```
t.test(def_A,def_B,var.equal = T)$p.value
```

```
[1] 0.6816055
```

```
t.test(def_A,def_C,var.equal = T)$p.value
```

```
[1] 0.2081849
```

```
t.test(def_A,def_D,var.equal = T)$p.value
```

```
[1] 0.0390036
```

```
t.test(def_B,def_D,var.equal = T)$p.value
```

```
[1] 0.02150551
t.test(def_C,def_D,var.equal = T)$p.value
```

```
[1] 0.3784366
t.test(def_B,def_C,var.equal = T)$p.value
```

```
[1] 0.1191898
```

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?

```
t.test(def_A,def_B,var.equal = T)$p.value/6
```

```
[1] 0.1136009
t.test(def_A,def_C,var.equal = T)$p.value/6
```

```
[1] 0.03469749
t.test(def_A,def_D,var.equal = T)$p.value/6
```

```
[1] 0.006500599
t.test(def_B,def_D,var.equal = T)$p.value/6
```

```
[1] 0.003584251
t.test(def_C,def_D,var.equal = T)$p.value/6
```

```
[1] 0.06307277
t.test(def_B,def_C,var.equal = T)$p.value/6
```

```
[1] 0.01986496
```

As every pairwise comparison between the means of method generates ($p < 0.05$) except the comparison between method A & method B, we reject the null hypothesis of no difference between any pair of means among the 4 methods

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.

```
summary(aov(defects_df$Weight ~ Method, data=defects_df))
```

```
          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?

As the $p\text{-value} > 0.05$ ($\Pr(>F) = 0.0515$) in the above results for the F-Test, we can conclude that the we fail to reject the null hypothesis of no difference between any pair of means among the 4 methods

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

Both the two sample t-test along with bonferroni correction differ from the conclusion of ANOVA. With ANOVA, we fail to reject the null hypothesis, whereas we reject the null hypothesis with the other two.