Final Exam – Part 2

Jasmeet Singh Saini - 0758054

2023-04-17

Cautious Caretakers of Climate Control

Note: the table and data.frame code do NOT match. Please use the data.frame code, regardless.

Trent caretakers measure the temperatures in four of the university's buildings at different times of the day and want to know what affects the temperature: the building, the time of day, or the interaction between the two. The data they collected is:

And can be loaded into a data.frame in R using:

```
btemp <- data.frame(temperature = c(21, 20.3, 20.7, 21.9, 20.7, 21.2
                                    , 20, 19.5, 20.1, 19.2, 19.3, 19.1
                                    , 20.2, 19.8, 20.1, 20.5, 19.8, 20.1
                                    , 20.1, 19.4, 19.9, 19.0, 21.2, 19.5
                                    , 18.9, 19.1, 19.3, 20.1, 20.7, 21.5
                                    , 18.7, 18.9, 19.3,  18.0, 17.9, 19.2)
                    , building = factor(rep(rep(c("DNA", "ENW", "OCA", "TSC")
                                                , each = 3), 3))
                    , time = factor(rep(c("Morning", "Afternoon", "Evening"), 12)
                                    , levels = c("Morning", "Afternoon", "Evening")))
```

a. What type of test should be used?

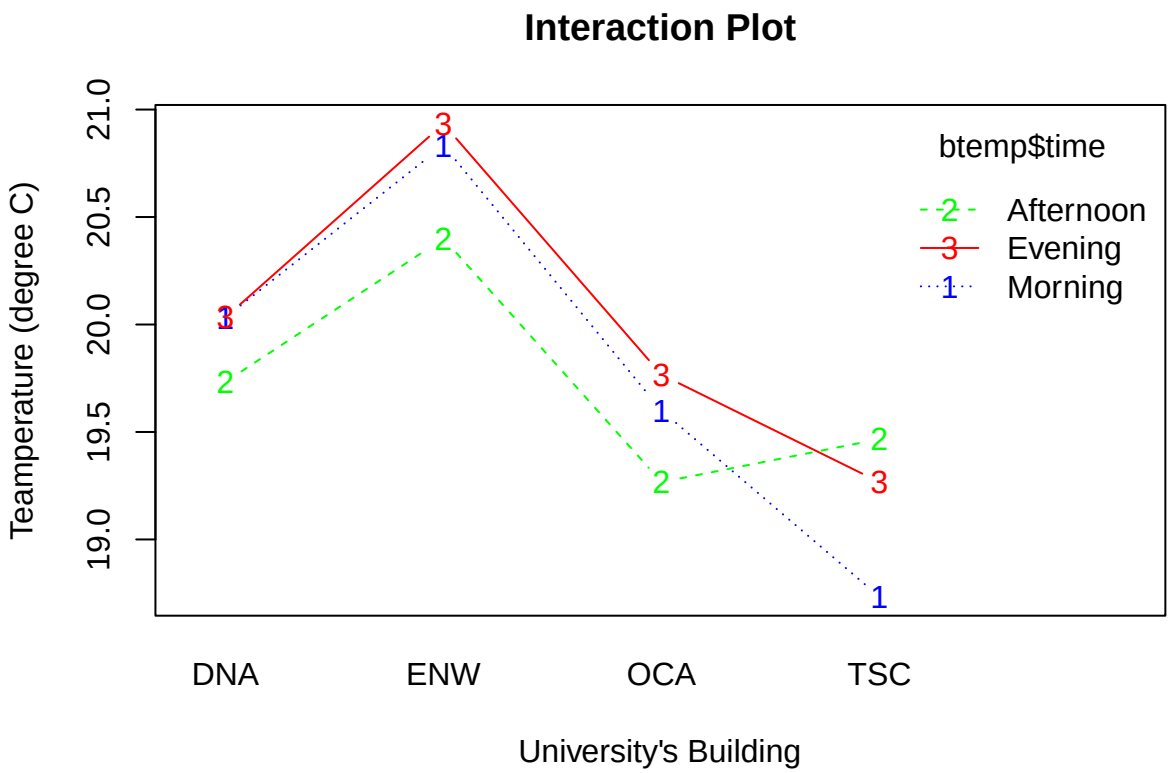
A **Two-way ANOVA (analysis of variance) test** should be used.

b. Construct an interaction plot. What does this plot indicate?

To construct an interaction plot in R, we can use the interaction.plot() function:

```
interaction.plot(x.factor     = btemp$building,
                 trace.factor = btemp$time,
                 response     = btemp$temperature,
                 main         = "Interaction Plot",
                 xlab         = "University's Building",
```

```
                ylab          = "Teamperature (degree C)",
                type          = "b",
                col           = c("blue", "green", "red"))
```



We can see from the interaction plot that a changes in the average temperature in the TSC building is observed due to an interaction between the afternoon time with evening and morning time.

c. Conduct a test to determine if there is an interaction effect on temperature due to building and time of day.

Hypothesis

The H_0 is Null hypothesis and H_A is Alternative hypothesis.

H_0 : There is no interaction effect on temperature due to building and time of day

vs.

H_A : There is no interaction effect on temperature due to building and time of day

Test Statistic and *p*-value

Preforming the *Two-way ANOVA test*:

```r
trent_uni <- aov(temperature ~ building * time, data = btemp)
# Getting the summary
summary(trent_uni)
```

```
##               Df Sum Sq Mean Sq F value  Pr(>F)
## building       3 12.086   4.029   6.219 0.00281 **
## time           2  0.509   0.254   0.393 0.67943
## building:time  6  1.404   0.234   0.361 0.89605
## Residuals     24 15.547   0.648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here, we get the test statistic as **0.361** and p-value as **0.89605**.

Statistical Decision

We **accept** the Null Hypothesis H_0 as p-value is greater than α level of significance.

Conclusion

There is enough evidence to support the null hypothesis. Therefore, we conclude that the statement, "there is no interaction effect on temperature due to building and time of day", is accurate.

d. Determine if there are any main effects: interpret these main effects (account for the interaction if necessary!). You don't need the full hypothesis testing framework here.

Determining if there are any main effects:

```r
main_effect <- lm(temperature ~ building + time, data = btemp)
summary(aov(main_effect))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## building     3 12.086   4.029    7.13 0.000939 ***
## time         2  0.509   0.254    0.45 0.641666
## Residuals   30 16.951   0.565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value for building is **0.000939**, which is less than *0.05*. And the p-value of time is **0.641666** which is greater than *0.05*. Hence, building affects the average temperature, while time does not.

e. Which building (statistically), if any, is the hottest on average?

To check which is the hottest building on average, we will use Tukey HSD - Statistical Test for Differences.

```r
TukeyHSD(trent_uni, "building")
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = temperature ~ building * time, data = btemp)
##
## $building
##               diff       lwr        upr     p adj
## ENW-DNA  0.7888889 -0.257750  1.8355278 0.1884687
## OCA-DNA -0.3888889 -1.435528  0.6577500 0.7366619
## TSC-DNA -0.7777778 -1.824417  0.2688611 0.1982811
## OCA-ENW -1.1777778 -2.224417 -0.1311389 0.0233505
## TSC-ENW -1.5666667 -2.613306 -0.5200277 0.0020143
## TSC-OCA -0.3888889 -1.435528  0.6577500 0.7366619
```

Hence, ENW building is the hottest on average.