BIA 654 B: Experimental Design II

**Ephraim Schoenbrun**

CWID: 104-20-702

Date of Submission: **February 25, 2017**

Assignment:

**Assignment #5**

# **Ethical Conduct**

|  |
| --- |
| The following statement is printed in the Stevens Graduate Catalog and applies to all students taking Stevens courses, on and off campus.  “Cheating during in-class tests or take-home examinations or homework is, of course, illegal and immoral. A Graduate Academic Evaluation Board exists to investigate academic improprieties, conduct hearings, and determine any necessary actions. The term ‘academic impropriety’ is meant to include, but is not limited to, cheating on homework, during in-class or take home examinations and plagiarism.”  Consequences of academic impropriety are severe, ranging from receiving an “F” in a course, to a warning from the Dean of the Graduate School, which becomes a part of the permanent student record, to expulsion.  *Reference: The Graduate Student Handbook, Academic Year 2003-2004 Stevens*  *Institute of Technology, page 10.*  Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.  Signature: Date: February 25, 2017  Please note that assignments in this class may be submitted to [www.turnitin.com](http://www.turnitin.com/), a web-based anti-plagiarism system, for an evaluation of their originality. |

1. The plant manager wants to investigate the productivity of three groups of workers: those with little (A), average (B), and considerable work experience (C). Since the productivity depends to some degree on the day-to-day variability of the available raw materials, which affects all groups in a similar fashion, the manager suspects that the comparison should be blocked with respect to day. The results (productivity, in percent) from five production days are given in the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| (A) | 53 | 58 | 49 | 52 | 60 |
| (B) | 55 | 57 | 53 | 57 | 64 |
| (C) | 60 | 62 | 55 | 64 | 69 |

1. Assuming the underlying assumptions of ANOVA are met, are there differences in the mean productivity among the three groups? (Use α = 0:05.)

**We see from the underlying analysis that without blocking, the one-way ANOVA would result in failing to reject the null hypothesis.**

1. Has the blocking made a difference? That is, is there a difference in results between one-way ANOVA without blocks and one-way ANOVA with blocks?

**As shown, when introducing a blocking element, we find that we are forced to reject the null hypothesis**

**.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Complete Random** | | | | | |
|  | **Df** | **Sum of Sq** | **Mean Sq** | **F Value** | **Pr(>F)** |
| **experience** | **2** | **147.73** | **73.867** | **3.4625** | **0.06499** |
| **Residuals** | **12** | **256** | **21.333** |  |  |
| **Blocking** | | | | | |
|  | **Df** | **Sum of Sq** | **Mean Sq** | **F Value** | **Pr(>F)** |
| **experience** | **2** | **147.733** | **73.867** | **24.352** | **0.0003962** |
| **days** | **4** | **231.733** | **57.933** | **19.099** | **0.0003731** |
| **Residuals** | **8** | **24.267** | **3.033** |  |  |

R Code

library(readxl)

library(plyr)

data <- read\_excel("C:/Users/Stevens/Desktop/R/654\_hwk\_5.xlsx",col\_names = FALSE)

names(data)[names(data)=="X0"] <- "experience"

names(data)[names(data)=="X1"] <- "days"

names(data)[names(data)=="X2"] <- "percent"

data <- revalue(data$experience, c("A"=1, "B"=2, "C"=3))

data$experience[which(data$experience=="A")]<-1

data$experience[which(data$experience=="B")]<-2

data$experience[which(data$experience=="C")]<-3

data[, "experience"] <- as.factor(data[, "experience"])

data[, "days"] <- as.factor(data[, "days"])

str(data)

model <- lm(percent ~ experience, data=data)

anova(model)

summary(model)

modelblock <- lm(percent ~ experience + days, data=data)

anova(modelblock)

summary(modelblock)

1. Suppose you want to determine whether the brand of laundry detergent used and the temperature affects the amount of dirt removed from your laundry. To this end, you buy two different brand of detergent (\Super" and \Best") and choose three different temperature levels (\cold", \warm", and \hot"). Then you divide your laundry randomly into 6 r piles of equal size and assign each r piles into the combination of (\Super" and \Best") and \cold", \warm", and \hot").

We are interested in testing Null Hypotheses

H0D: The amount of dirt removed does not depend on the type of detergent

H0T: The amount of dirt removed does not depend on the temperature

H0DT: There is no interaction effect between the type of detergent and the temperature.

This experiment has two factors (Factor Detergent, Factor Temperature) at a = 2 (Super and Best)

and b = 3 (cold, warm, and hot) levels. Thus there are ab = 32 = 6 different combinations of detergent

and temperature. With each combination you wash r = 4 loads and r is called the number of replicates.

This sums up to n = abr = 24 loads in total. The amounts Xijk of dirt removed when washing sub pile

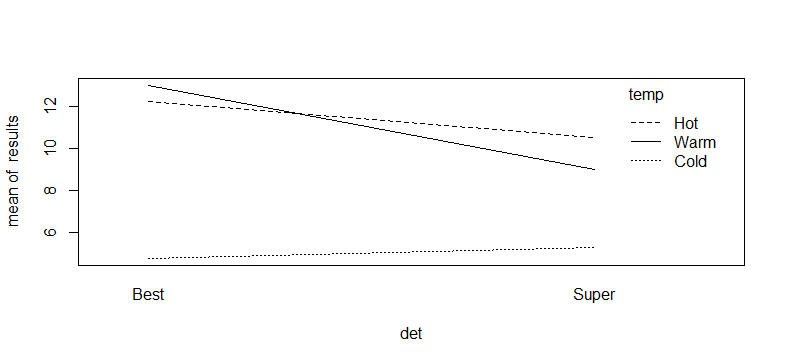
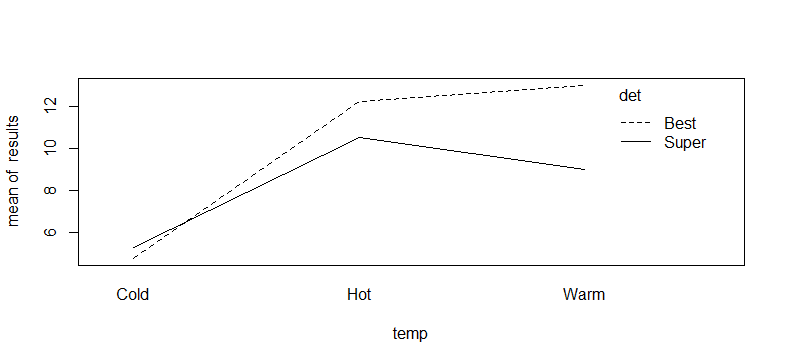
k (k = 1; 2; 3; 4) with detergent i (i = 1; 2) at temperature j (j = 1; 2; 3) are recorded in the following

Table. Perform appropriate analysis to test the above three hypotheses with α = 0:05.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cold | Warm | Hot |
| Super | 4, 5, 7, 5 | 7, 9, 8, 12 | 10, 12, 11, 9 |
| Best | 6, 5, 4, 4 | 13, 15, 12, 12 | 12, 13, 11, 13 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Factorial Analysis** | | | | | |
|  | **Df** | **Sum of Sq** | **Mean Sq** | **F Value** | **Pr(>F)** |
| **det** | **1** | **18.375** | **18.375** | **9.383** | **0.006697** |
| **temp** | **2** | **204.75** | **102.375** | **52.2766** | **3.18E-08** |
| **det:temp** | **2** | **20.25** | **10.125** | **5.1702** | **0.016819** |
| **Residuals** | **18** | **35.25** | **1.958** |  |  |

**We see that for all three metrics, the difference in means due to the type of detergent, the difference in means due to the temperature of the water, and the interaction between the factors, we must reject the null hypothesis that these factors do not impact the cleanliness of the clothes and that there is no interaction amongst the factors. This is apparent from the interaction plot below:**

R Code

data <- read.csv('data.csv', header=TRUE)

results <- data$result

temp <- data$temp

det <- data$det

tapply(results, temp, mean)

tapply(results, det, mean)

tapply(results, det:temp, mean)

model <- lm(results ~ det\*temp)

anova(model)

interaction.plot(det, temp, results)

interaction.plot(temp, det, results)