101b_hw_03_Daren_Sathasivam

Daren Sathasivam

2024-04-22

Problem 1: 4.6

Consider the single-factor completely randomized experiment shown n Problem 3.7. Suppose that this experiment had been conducted in a randomized complete block design and that the sum of squares for blocks was 80.00. Modify the ANOVA for this experiment to show the correct analysis for the randomized complete block experiment.

```
##   Source DF      SS       MS        F            P
## 1 Factor  4  987.71 246.9300 33.09488 1.184819e-09
## 2  Error 25  186.53   7.4612       NA           NA
## 3  Total 29 1174.24       NA       NA           NA
```

Table for Block Design:

DF ~ Block $= DF_{total} = DF_{factor} + DF_{block} + DF_{error}$
$DF_{total} = 29$
$29 = 4 + DF_{block} + 20$
$DF_{block} = 5(alsob - 1 = 5)$
SS ~ Total $= SS_{total} = SS_{factor} + SS_{block} + SS_{error}$
$SS_{total} = 987.71 + 80.00 + 5.3265$
$SS_{total} = 1174.24$
MS ~ Block $= MS_{block} = \frac{SS_{block}}{DF_{block}}$
$MS_{block} = \frac{80}{5}$
$MS_{block} = 16$

```r
source <- c("Factor", "Block", "Error", "Total")
df <- c(4, 5, 20, 29)
ss <- c(987.71, 80, 106.53, 1174.24)
ms <- c(246.9275, 16, 5.3265, NA)  # MS for Total is typically not calculated
f <- c(46.3583, NA, NA, NA)  # F and P-values for Error and Total are not applicable
p <- c(7.669035e-10, NA, NA, NA)

# Create a data frame for the ANOVA table
block_table <- data.frame(Source = source, DF = df, SS = ss,
    MS = ms, F = f, P = p)

# Print the table
print(block_table)
```

```
##   Source DF      SS       MS       F            P
## 1 Factor  4  987.71 246.9275 46.3583 7.669035e-10
## 2  Block  5   80.00  16.0000      NA           NA
## 3  Error 20  106.53   5.3265      NA           NA
## 4  Total 29 1174.24       NA      NA           NA
```

- The experiment design effectivel controlled for specific block effects, which can be supported by the significant F-statistic value of 46.36. Additionally, the p-value is very low with a value of 7.669035e-10, indicating that the influence of the treatment on the outcome variable being independent of the block difference.

Problem 2: 4.13

```r
oil <- c(rep(1, 5), rep(2, 5), rep(3, 5))
truck <- rep(c(1:5), 3)
fuel <- c(0.5, 0.634, 0.487, 0.329, 0.512, 0.535, 0.675, 0.52,
    0.435, 0.54, 0.513, 0.595, 0.488, 0.4, 0.51)
```

a. Analyze the data from this experiment.

```r
model1 <- aov(fuel ~ factor(oil) + factor(truck))
summary(model1)
```

```
##               Df  Sum Sq  Mean Sq F value   Pr(>F)    
## factor(oil)    2 0.00671 0.003353   6.353   0.0223 *  
## factor(truck)  4 0.09210 0.023025  43.626 1.78e-05 ***
## Residuals      8 0.00422 0.000528                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Factor(oil) has an F-value of 6.353 and p-value of 0.0223. This indicates that there are statistically significant difference in fuel consumption efficiencies among different oils. Factor(truck) has an F-value of 43.626 and a p-value of 1.78e-05. This suggests that the differences in truck engines have a statistically significant impact on fuel consumption, thus supporting the choice of using a block design to control this variability.

b. Use the Fisher LSD method to make comparisons among the three lubricating oils to determine specifically which oils differ in brake-specific fuel consumption.

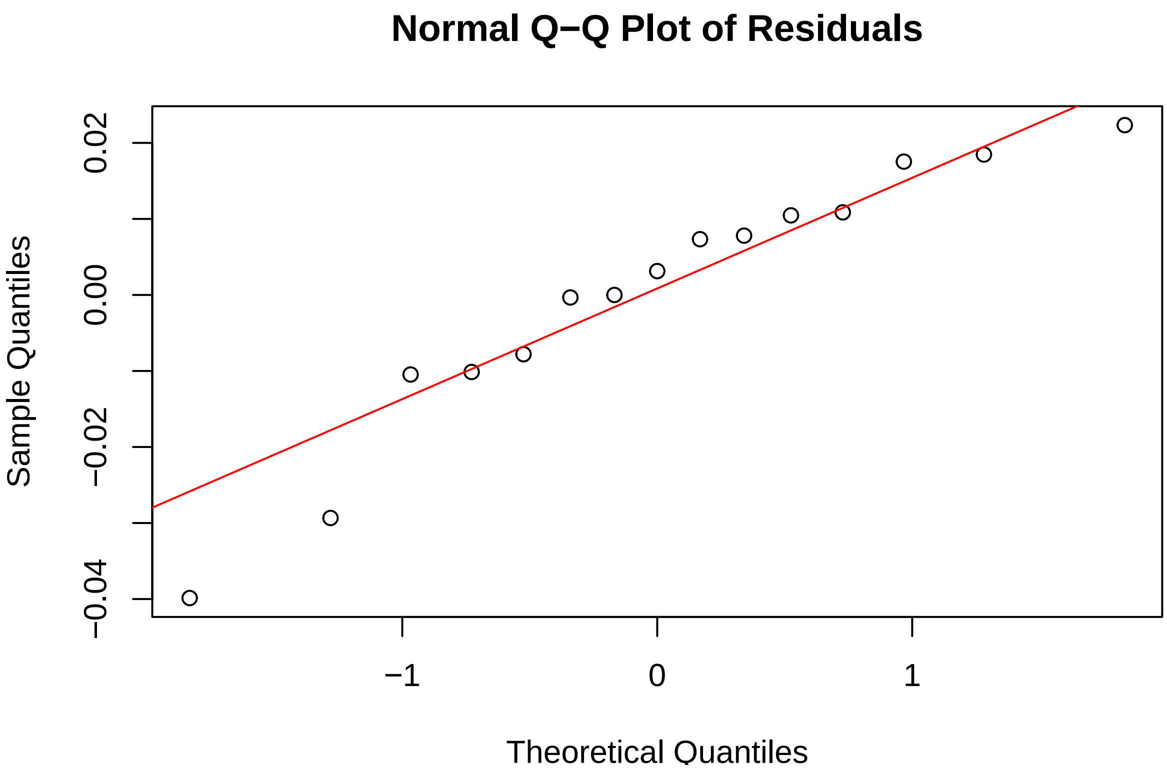
```r
library(DescTools)
PostHocTest(model1, method = "lsd")$"factor(oil)"
```

```
##        diff      lwr.ci      upr.ci       pval
## 2-1  0.0486  0.01509436  0.08210564 0.01015749
## 3-1  0.0088 -0.02470564  0.04230564 0.56152936
## 3-2 -0.0398 -0.07330564 -0.00629436 0.02548036
```

- The oils that specifically differ in brake-specific fuel consumption is oil 2 in comparison to oil 1 and oil 3 as it has the highest difference amongst the three. The difference between oil 1 and 3 is relatively small, suggesting that using oil 1 or oil 3 will result in increased fuel efficiency in comparison to oil 2.

c. Analyze the residuals from this experiment.

```r
res1 <- model1$residuals
# Normal Q-Q plot of residuals
qqnorm(res1, main = "Normal Q-Q Plot of Residuals")
qqline(res1, col = "red")
```




```r
summary(res1)
```

```
##       Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.039867 -0.008967  0.003133  0.000000  0.010667  0.022333
```

- A linear pattern can be observed in the qqplot of residuals, indicating that the residuals closely follows a normal distribution. There is no S-shapes curve therefore no indication of heavy tails or outliers within the residuals.

Problem 3: 4.24

Effect of five different ingredients on the reaction time of a chemical process is being studied.

```r
# Data
batch <- c(rep(1, 5), rep(2, 5), rep(3, 5), rep(4, 5), rep(5,
    5))
day <- rep(1:5, 5)
ingredients <- c("A", "B", "D", "C", "E", "C", "E", "A", "D",
    "B", "B", "A", "C", "E", "D", "D", "C", "E", "B", "A", "E",
    "D", "B", "A", "C")
time <- c(8, 7, 1, 7, 3, 11, 2, 7, 3, 8, 4, 9, 10, 1, 5, 6, 8,
    6, 6, 10, 4, 2, 3, 8, 8)
```

```r
# Model
model2 <- aov(time ~ factor(ingredients) + factor(batch) + factor(day))
summary(model2)
```

```
##                     Df Sum Sq Mean Sq F value   Pr(>F)    
## factor(ingredients)  4 141.44   35.36  11.309 0.000488 ***
## factor(batch)        4  15.44    3.86   1.235 0.347618    
## factor(day)          4  12.24    3.06   0.979 0.455014    
## Residuals           12  37.52    3.13                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- For factor(ingredients), it can be observed that the F-value is 11.309 and p-value is 0.000488. This suggests that the type of ingredient used in the process has a statistically significant effect on the reaction time. Both factor(batch) and factor(day) have a relatively high p-value, respectively 0.346718 and 0.455014. This indicates that both of these factors do not have a statistically significant effect on the reaction time. This suggests that future experiments should focus on the ingredients to control and optimize reaction time. Additionally, the insignificance of the batch and day effects validates the effectiveness of the Latin square design in controlling potential variability allowing for a clear assesment of the ingredients factor.

Problem 4: 4.37

Graeco-Latin square. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions

```r
# Data
batch <- c(rep(1, 5), rep(2, 5), rep(3, 5), rep(4, 5), rep(5,
    5))
acid <- rep(1:5, 5)
times <- c("A", "B", "C", "D", "E", "B", "C", "D", "E", "A",
    "C", "D", "E", "A", "B", "D", "E", "A", "B", "C", "E", "A",
    "B", "C", "D")
catalyst <- c("a", "b", "g", "d", "e", "g", "d", "e", "a", "b",
    "e", "a", "b", "g", "d", "b", "g", "d", "e", "a", "d", "e",
    "a", "b", "g")
yield <- c(26, 16, 19, 16, 13, 18, 21, 18, 11, 21, 20, 12, 16,
    25, 13, 15, 15, 22, 14, 17, 10, 24, 17, 17, 14)

# Model
model3 <- aov(yield ~ factor(times) + factor(catalyst) + factor(batch) +
    factor(acid))
summary(model3)
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)    
## factor(times)     4  342.8   85.70  14.650 0.000941 ***
## factor(catalyst)  4   12.0    3.00   0.513 0.728900    
## factor(batch)     4   10.0    2.50   0.427 0.785447    
## factor(acid)      4   24.4    6.10   1.043 0.442543    
## Residuals         8   46.8    5.85                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- From the model summary, we can observe that p-values of the catalyst(0.728900), batch(0.785447), and acid(0.442543) are larger than $\alpha = 0.05$. This suggests that the catalyst, batch, and acid factors do not have a statistically significant effect on the yield of the chemical process. The standing time has a p-value of 0.000941, indicating that the standng time factor does have a statistically significant effect on the yield of the chemical process. By having the single significant factor, this supports the effectiveness of the Graeco-Latin square design in controlling the variability from the other factors.

Problem 5: 4.44

BIBD. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

```
# Data
hardwood <- c(rep(2, 7), rep(4, 7), rep(6, 7), rep(8, 7), rep(10,
    7), rep(12, 7), rep(14, 7))
days <- rep(1:7, 7)
strength <- c(114, NA, NA, NA, 120, NA, 117, 126, 120, NA, NA,
    NA, 119, NA, NA, 137, 117, NA, NA, NA, 134, 141, NA, 129,
    149, NA, NA, NA, NA, 145, NA, 150, 143, NA, NA, NA, NA, 120,
    NA, 118, 123, NA, NA, NA, NA, 136, NA, 130, 127)

# Model
model4 <- aov(strength ~ factor(days) + factor(hardwood))
summary(model4)
```

```
##                  Df Sum Sq Mean Sq F value  Pr(>F)
## factor(days)      6 1114.3  185.71   8.814 0.00358 **
## factor(hardwood)  6 1317.4  219.57  10.420 0.00205 **
## Residuals         8  168.6   21.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 28 observations deleted due to missingness
```

- From the balanced incomplete block design experiment, it can be observed that the days factor has an F-value of 8.814 and a p-value of 0.00358. This indicates that the variation in days has a statistically significant impact on the outcome of strength of the paper produced. Additionally, the hardwood concentration factor has an F-value of 10.420 and a p-value of 0.00205, indicating that it has a statistically significant effect on the strength of paper produced. Through the BIBD, the balanced design does not affect the impact of the factors and therefore it can be concluded that the hardwood concentration and day-to-day variability have an effect on the strength of paper produced.