HW7

2024-10-22

STAT 5000 HOMEWORK #7
FALL 2024 DUE FRI, OCTOBER 25TH @ 11:59 PM NAME: SAM OLSON
COLLABORATORS: **The Hatman**

Q1

For each of the following experiments, determine its

(i) design: completely randomized, randomized complete block (RCBD), or neither;

(ii) experimental units, treatments, and blocks (if any blocks are used).

(iii) Outline an ANOVA table listing sources of variation and degrees of freedom.

(a)

An agronomist is interested in the effects of plant density on the yield of a certain crop. There are six densities of interest (7, 8, 9, 10, 11, and 12 plants/m2) that are interesting. The agronomist has five fields and each field is divided into 6 plots of equal size. She suspects that there is non-homogeneity of soil fertility across fields. She chooses a design such that each density is randomly assigned to one plot in each of the five fields. A separate random assignment is done within each field.

i. Design

- Randomized Complete Block Design (RCBD): The design uses blocks (fields) to account for potential variability in soil fertility across fields, and within each block (field), the treatments (densities) are randomly assigned to the plots.

ii. Experimental units, treatments, and blocks:

- Experimental units: Plots within fields (each plot within a field where the density treatment is applied)
- Treatments: Six plant densities (7, 8, 9, 10, 11, and 12 plants/m²)
- Blocks: Five fields (each field serves as a block to account for soil fertility differences)

Source of Variation	Degrees of Freedom (DF)

iii. ANOVA Table Outline:

Source of Variation	Degrees of Freedom (DF)
Blocks (Fields)	4
Treatments (Densities)	5
Error	$(5-1)(6-1) = 20$
Total	$30 - 1 = 29$

- Blocks: $k - 1$, where $k = 5$ fields (blocks)
- Treatments: $t - 1$, where $t = 6$ densities (treatments)
- Error: $(k-1)(t-1)$, where $k = 5$ fields and $t = 6$ treatments
- Total: Total number of plots minus one $(30 - 1 = 29)$

(b)

A router is used to cut notches in a printed circuit board. The vibration of the board as it is cut is considered a major source of variation in the diameter of the notches. Three cutting speeds, 50, 70, and 90 rpm, were used in an experiment to assess the impact of cutting speed on vibration. For each cutting speed, notches were cut in five different circuit boards. Once a notch is cut into a board it cannot be used again, so 15 different boards were used. The assignment of boards to cutting speed was done at random.

i.

Design:

- Completely Randomized Design (CRD): Since the assignment of boards to the three cutting speeds was done randomly and no blocking is mentioned to account for other sources of variation, this is a completely randomized design.

ii.Experimental units, treatments, and blocks:

- Experimental units: Individual circuit boards (15 different boards used in the experiment)
- Treatments: Three cutting speeds (50, 70, and 90 rpm)
- Blocks: No blocks are used in this experiment, as the design is completely randomized.

iii. ANOVA Table Outline:

Source of Variation	Degrees of Freedom (DF)
Treatments (Cutting Speeds)	2
Error	$15 - 3 = 12$
Total	$15 - 1 = 14$

- Treatments: $t - 1$, where $t = 3$ cutting speeds (treatments)
- Error: Total experimental units minus number of treatments, $n - t$, where $n = 15$ boards and $t = 3$ treatments
- Total: Total number of experimental units minus one $(15 - 1 = 14)$

(c)

Industrial psychologists wish to investigate the effect of music in the factory of the productivity of workers. Four distinct music programs and no music make up the five treatments. The experiment is run in 8 plants. Each music program is used for one week. Within each plant music programs are randomly assigned to weeks (1,2,3,4,5) so that all five music programs are used in each plant. Production at each plant is recorded for each of the five weeks.

i. Design:

Randomized Complete Block Design (RCBD): Since each plant serves as a block and all five music treatments are applied in a randomized order within each plant (across the weeks), this is a randomized complete block design.

ii. Experimental units, treatments, and blocks:

- Experimental units: Week of production within each plant (the productivity during a week under a specific music program in each plant)
- Treatments: Five music programs (four distinct music programs and no music)
- Blocks: Eight plants (each plant is a block)

iii.

ANOVA Table Outline:

Source of Variation	Degrees of Freedom (DF)
Blocks (Plants)	7
Treatments (Music Programs)	4
Error	$(8 - 1)(5 - 1) = 28$
Total	$40 - 1 = 39$

Explanation of Degrees of Freedom: - Blocks: $k - 1$, where $k = 8$ plants (blocks) - Treatments: $t - 1$, where $t = 5$ music programs (treatments) - Error: $(k - 1)(t - 1)$, where $k = 8$ plants and $t = 5$ treatments - Total: Total number of observations (plants $\times$ weeks) minus one $(40 - 1 = 39)$

Q2

Standing from a seated position generates stress on the knee joint. Finding ways to minimize this stress is important in individuals who have had knees replaced (Total Knee Arthroplasty, TKA). An experiment is conducted to test whether different feet placement (neutral, back, staggered) affects the amount of torque on the knee (measured in Newton meters, Nm) of 14 older men with TKA. Each of the men will experience all three treatments, in randomly selected order, during the experiment. For all trials, the height of the chair will be the same and participants will all wear comfortable clothing and tennis shoes. The data are located in the file knee.txt (posted on Canvas).

```
dat <- read.table("C:/Users/samue/OneDrive/Desktop/Iowa_State_PS/STAT 5000/PS/PS7/knee.txt") |>
  data.frame(
  )
names(dat) <- c("row", "placement", "torque")
names(dat)
```

```
## [1] "row"       "placement" "torque"
```

(a)

Construct an ANOVA table for the model with fixed placement and subject effects. Include the full table below.

```
# Ensure placement is treated as a factor
dat$placement <- as.factor(dat$placement)

# Fit the model with placement as a factor
anova_model <- aov(torque ~ placement + row, data = dat)

# Display the ANOVA table
summary(anova_model)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## placement    2  81.09   40.54   5.719 0.00674 **
## row          1   1.26    1.26   0.177 0.67596
## Residuals   38 269.39    7.09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b)

Are there significant differences in the mean torque between the three feet placements? Conduct an F-test and properly state your conclusion in the context of the study.

(c)

Apply Tukey's HSD method with an experiment-wise Type I error level of $\alpha = 0.05$. Properly state your conclusions about the pairwise comparison of the mean torque between the three placement positions in the context of the study.

```
# Apply Tukey's HSD method
tukey_results <- TukeyHSD(anova_model, "placement", conf.level = 0.95)
```

```
## Warning in replications(paste("~", xx), data = mf): non-factors ignored: row
```

```
# Display the results
print(tukey_results)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = torque ~ placement + row, data = dat)
##
## $placement
##           diff         lwr        upr       p adj
## N-B   2.3714286 -0.08290519  4.8257623 0.0600707
## S-B -0.9285714 -3.38290519  1.5257623 0.6294866
## S-N -3.3000000 -5.75433376 -0.8456662 0.0061796
```

(d)

Specify and test two orthogonal contrasts that compare (i) the mean torque of the feet neutral group to the mean torque of the other two groups, and (ii) the mean torque of the feet back group to the mean torque of the feet staggered group. How do these results support your conclusions from part (c)?

```
unique(dat$placement)
```

```
## [1] N B S
## Levels: B N S
```

```
# Convert placement to a factor with the desired levels
dat$placement <- factor(dat$placement, levels = c("N", "B", "S"))

# Check the levels of placement to confirm
levels(dat$placement)
```

```
## [1] "N" "B" "S"
```

```
contrasts(dat$placement) <- matrix(c(
  1, -0.5, -0.5,   # Contrast 1: Compare 'neutral' to 'back' and 'staggered'
  0, 1, -1         # Contrast 2: Compare 'back' to 'staggered'
), ncol = 2)

# Fit the ANOVA model with contrasts
anova_model_contrasts <- aov(torque ~ placement + row, data = dat)

# Display the summary of the model with contrasts
summary.lm(anova_model_contrasts)
```

```
##
## Call:
## aov(formula = torque ~ placement + row, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2501 -1.2999 -0.3213  2.1074  5.7782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.37326    0.86780  24.629  < 2e-16 ***
## placement1   1.89048    0.58102   3.254  0.00239 **
## placement2   0.46429    0.50318   0.923  0.36198
## row          0.04293    0.10192   0.421  0.67596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.663 on 38 degrees of freedom
## Multiple R-squared:  0.2341, Adjusted R-squared:  0.1736
## F-statistic: 3.872 on 3 and 38 DF,  p-value: 0.01645
```

(e)

Examine the normal Q-Q plot of the residuals. What does this plot suggest?

```
residuals_anova <- residuals(anova_model)

# Create a Q-Q plot of the residuals
qqnorm(residuals_anova)
qqline(residuals_anova)
```

Normal Q–Q Plot
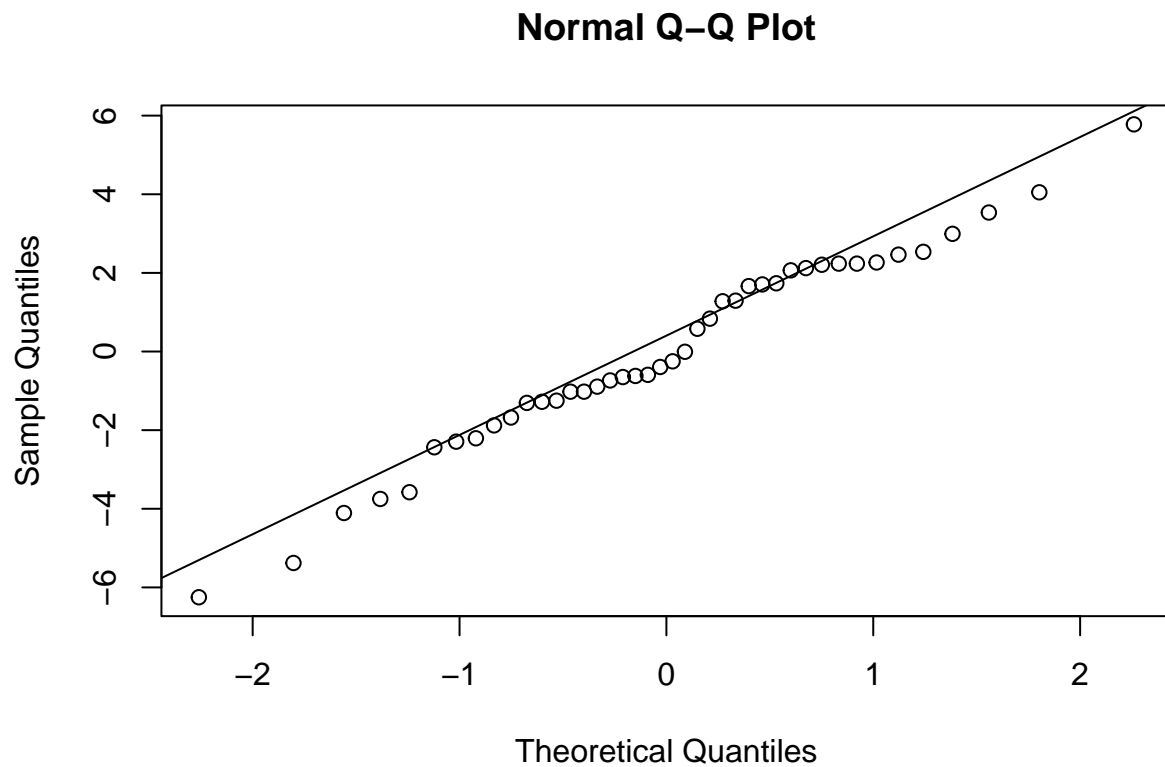
(f)

Plot the residuals versus the estimated means (fitted values). What does this plot suggest?

```r
residuals_anova <- residuals(anova_model)

# Extract the fitted values from the ANOVA model
fitted_values <- fitted(anova_model)

plot(fitted_values, residuals_anova,
     xlab = "Fitted Values",
     ylab = "Residuals",
     main = "Residuals vs Fitted Values",
     pch = 19)  # Adds solid dots for clarity

# Add a horizontal reference line at 0
abline(h = 0, col = "red")
```

Residuals vs Fitted Values
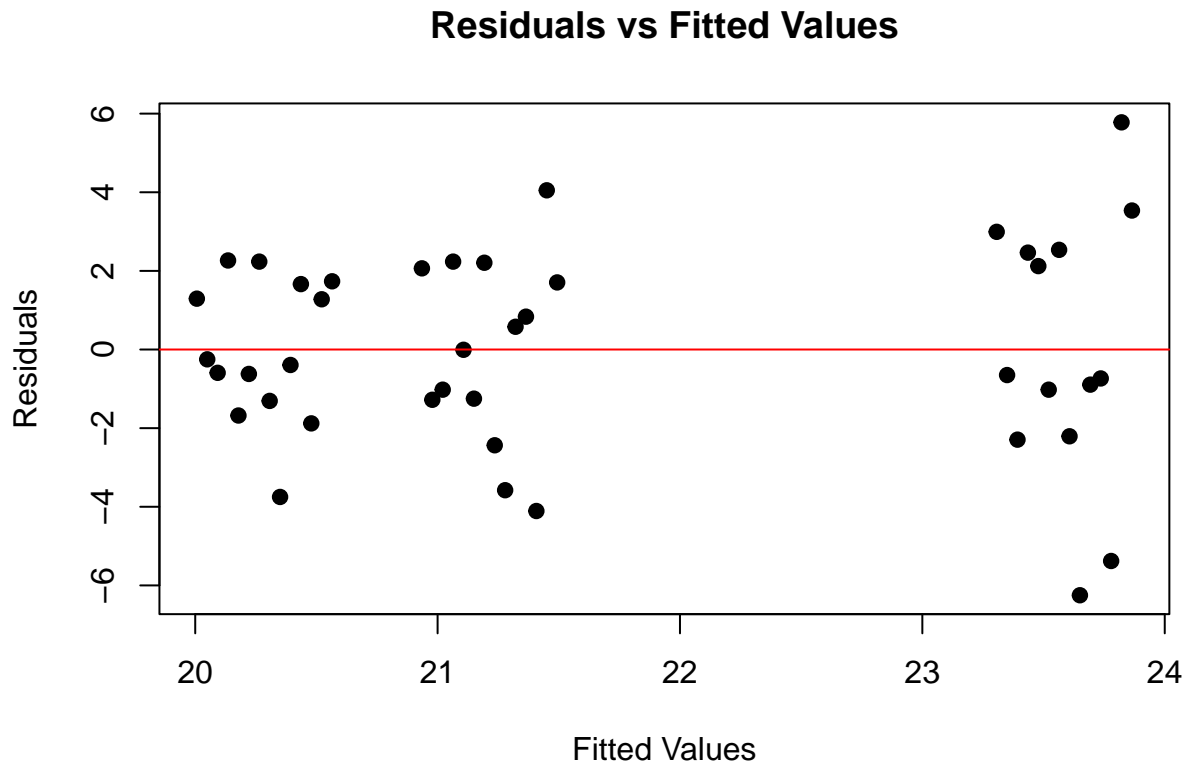
Q3

A group of investigators are interested in the effect of planting density on the yield of millet, a small grain. The investigators are comparing five planting densities, labeled 2, 4, 6, 8, and 10. A field of land was divided into 25 plots (5 rows and 5 columns) to account for known differences in soil quality in the field. In their first study, the investigators randomly assigned density treatments to plots in a 5 x 5 Latin Square. The data are in the millet.txt file (posted on Canvas).

```
millet <- read.table("C:/Users/samue/OneDrive/Desktop/Iowa_State_PS/STAT 5000/PS/PS7/millet.txt",
                     header=TRUE) |>
  data.frame(
  )
```

(a)

Compute the ANOVA table by-hand for the model with fixed row, column, and spacing effects. Include all sources of variation, degrees of freedom, and sums of squares.

(b)

Construct an ANOVA table using R for the model with fixed row, column, and spacing effects to verify your calculations in part (a). Include the full table below.

```
# Fit the ANOVA model with row, col, and spacing as fixed effects
anova_model_millet <- aov(yield ~ row + col + spacing, data = millet)

# Display the ANOVA table
summary(anova_model_millet)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## row          1   1220    1220   1.151 0.2955
## col          1   3200    3200   3.018 0.0970 .
## spacing      1   5660    5660   5.339 0.0311 *
## Residuals   21  22265    1060
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c)

Are there significant differences in the mean millet yield between the 5 planting densities? Conduct an F-test and properly state your conclusion in the context of the study

```
# Ensure spacing is treated as a factor
millet$spacing <- as.factor(millet$spacing)

# Fit the ANOVA model with spacing (planting densities) as a factor
anova_model_spacing <- aov(yield ~ spacing, data = millet)

# Display the ANOVA table
summary(anova_model_spacing)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## spacing       4   6297    1574   1.209  0.338
## Residuals    20  26049    1302
```

Null Hypothesis: There is no difference in mean millet yield between the 5 planting densities. Alternative Hypothesis:: There is a difference in mean millet yield between the 5 planting densities.

(d)

Apply Tukey's HSD method with an experiment-wise Type I error level of $\alpha = 0.05$. Properly state your conclusions about the pairwise comparison of the mean millet yield between the five planting densities in the context of the study.

```
# Ensure spacing is treated as a factor
millet$spacing <- as.factor(millet$spacing)

# Fit the ANOVA model with spacing as a factor
anova_model_spacing <- aov(yield ~ spacing, data = millet)

# Apply Tukey's HSD method
tukey_results <- TukeyHSD(anova_model_spacing, "spacing", conf.level = 0.95)

# Display the Tukey HSD results
print(tukey_results)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = yield ~ spacing, data = millet)
##
## $spacing
##        diff       lwr       upr      p adj
## 4-2   -16.4  -84.70065  51.90065 0.9497427
## 6-2   -17.4  -85.70065  50.90065 0.9384253
## 8-2   -41.6 -109.90065  26.70065 0.3888165
## 10-2  -40.6 -108.90065  27.70065 0.4122834
## 6-4    -1.0  -69.30065  67.30065 0.9999991
## 8-4   -25.2  -93.50065  43.10065 0.8024362
## 10-4  -24.2  -92.50065  44.10065 0.8241995
## 8-6   -24.2  -92.50065  44.10065 0.8241995
## 10-6  -23.2  -91.50065  45.10065 0.8448325
## 10-8    1.0  -67.30065  69.30065 0.9999991
```

None of the comparisons between the spacing levels (e.g., Spacing 4 vs. Spacing 2, Spacing 8 vs. Spacing 2) have a p adj (adjusted p-value) less than 0.05.

This indicates that there are no statistically significant differences in mean millet yield between any of the planting densities at the 5% significance level. Summary of Pairwise Comparisons:

All confidence intervals include 0 (e.g., -84.70 to 51.90 for Spacing 4 vs. Spacing 2), which also suggests that the differences in mean millet yield between these spacing levels could be 0, implying no detectable difference in yield.

There are no significant differences in mean millet yield between the five planting densities based on this analysis. Even though there are numerical differences in yield between the spacing levels, these differences are not statistically significant at the 0.05 significance level in this experiment.

(e)

If the study is repeated next year, would you recommend that it still be a Latin Squares design? Or, should the investigators just use row blocks, just use column blocks, or just use a CRD? Justify your answer with appropriate calculations and explanations.