STAT 5000 HOMEWORK #7
FALL 2024 DUE FRI, OCT 25TH @ 11:59 PM NAME: _____

Directions: Type or clearly handwrite your solutions to each of the following exercises. Partial credit cannot be given unless all work is shown. You may work in groups provided that each person takes responsibility for understanding and writing out the solutions. Additionally, you must give proper credit to your collaborators by providing their names on the line below (if you worked alone, write "No Collaborators"):

---

1. **[+24]:** 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.

(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.

(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.

2. **[+40]:** 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) and appear in the table below.

Subject	Neutral (N)	Back (B)	Staggered (S)
1	26.3	23.0	21.3
2	22.7	19.7	19.8
3	21.1	20.0	19.5
4	25.9	23.3	22.4
5	25.6	21.1	18.5
6	22.5	19.9	19.6
7	26.1	23.4	22.5
8	21.4	18.8	19.0
9	17.4	17.7	16.6
10	22.8	21.9	20.0
11	23.0	22.2	22.1
12	18.4	17.3	18.6
13	29.6	25.5	21.8
14	27.4	23.2	22.3

Use R to complete the following exercises:

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

(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.

(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)?

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

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

3. **[+36]:** 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) and appear in the table below.

Yield (Spacing) Values for Latin Squares Millet Experiment

		Column				
		1	2	3	4	5
Row	1	250 (4)	230 (10)	279 (2)	287 (6)	202 (8)
	2	235 (8)	283 (2)	245 (10)	280 (4)	260 (6)
	3	180 (10)	252 (4)	280 (6)	246 (8)	250 (2)
	4	243 (2)	244 (6)	227 (8)	193 (10)	259 (4)
	5	231 (6)	271 (8)	266 (4)	334 (2)	338 (10)

(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.

(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.

(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.

(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.

Total: 100 points **# correct:** _____ **%:** _____