Statistics 425: Assignment # 3 Due Date: Monday, March 2 at 9am

1. Pediatric researchers carrried out an experiment do investigate if a teaspoon on honey before bed calms a child's cough. A sample of $n =$ children who were sick with an upper respiratory tract infection and their parents participated in this study. On the first night, parents rated their child's symptoms on a scale from 0 (no problems) to 6 (extremely severe) for five different areas. On the second night, the parents were instructed to give their sick child a dosage of liquid medicine prior to the child's bedtime. Unknown to the parents, some were given a dosage of dextromethorphan (DM) - an over-the-counter cough medicine, while others were given a similar dosage of honey. A third group of parents were giving their children nothing at all. Again, the parents rated their children's cough symptoms, and the improvements in total cough symptoms score was determined for each child. These data appear in the **COUGH.csv** file.

`http://people.ucalgary.ca/~jbstall/DataFiles/COUGH.csv`

Note that these data are *not stacked*.

(a) Identify the treatment, the placebo, and the control group in this experiment.

(b) Create boxplots of these data. What do your boxplots tell you about the "'cough score"?

(c) Assuming the conditions of (i) Normality of the e_{ij} terms and (ii) the variation of the cough score is the same, do these data indicate that there exists a treatment effect? Should your null hypotheses be rejected, rarry out *one* multiple comparison method to identify where the differences. Ensure you summarize and interpret your findings.

2. Refer to Assignment 2, Question 5: Can you conclude from the data collected that on average, people recall more commercials when they are watching a "'Neutral" program when compared to television programs with content that is either "Violent" or "Sexual". Formulate the statistical hypothesis, then carry out the necessary statistical test.

3. The data appearing in the data file **GOLFRBD.csv** and found in

`people.ucalgary.ca/~jbstall/DataFiles/GOLFRBD.csv`

was the result of a random sample of four different brands of golf balls. For each brand (A, B, C, and D), a robotic golfer named "Iron Byron" was equipped with a 3-iron (a specific type of golf club) and hit a random sample of 10 balls of each of Brand A, Brand B, Brand C, and Brand D.

(a) Do these data suggest that there is variation in the distance of a golf ball (after it is hit) between the different brands of golf ball? State your statistical hypotheses, test statistic, P-value, and conclusion.

(b) Compute the 95% confidence interval for σ_{Common}, the standard deviation of the distance of a golf ball hit by Iron Byron with a 3-iron.

(c) From these data, compute the estimate for the $Var(X_{ij})$.

(d) Compute the 95% confidence interval for the intracluster correlation coefficient. Interpret the meaning of your finding(s) in the context of these data?

4. An entomologist counted the number of eggs laid by female moths on successive days in three strains of tobacco budworm from each of 15 matings. The data below are the number of eggs laid on the third day after the mating for each female in each of the strains.

Strain															
USDA	448	906	28	277	634	48	369	137	29	522	319	242	261	566	734
Field	211	276	415	787	18	118	1	151	0	253	61	0	275	0	153
Resistant	0	9	143	1	26	127	161	294	0	348	0	14	21	0	218

Enter these data into a data frame in R Studio, then answer the questions below.

(a) What do boxplots of these data tell you about the variation in the number of eggs laid between the three different strains of tobacco budworm? If it helps, carry out the relevant statistical test as well.

(b) Consider the nature of the variable. What transformation method would you suggest to stabilize the variation in the number of eggs laid?

(c) Apply the transformation you suggest in part (b). What are your finding(s)?

(d) Refer to part (c): Carry out a test of equal means to the transformed data from part (c). If you find a statistically significant difference, apply Tukey's HSD method to identify where the mean difference in your transformed data are.

(e) Consider the Box-Cox transformation method. Apply this method to these data. What transformation method does the Box-Cox method suggest? It is the same as in part (c) or different?

5. The potential of solar panels installed on the roofs build above national highways as a source of energy was investigated[1]. Compute simulation was used to estimate the monthly solar energy, in kilowatt-hours, generated from solar panels installed across 200-km stretch of highway in India. Each month, the simulation was run under each of four conditions: single-layer solar panels, double-layer solar panels that are 1 metre apart; double-layer solar panels that are 2metres apart, and double-layer solar panels that are 3 metres apart. The data collected for 12 months are found in the data file

`http://people.ucalgary.ca/~jbstall/DataFiles/SOLPAN.csv`

(a) State the statistical hypothesis that is required to investigate the mean amount of energy generated by the four solar panel configurations.

(b) Carry out the appropriate statistical test to address the inquiry in part (a). Ensure you provide the value of the test statistic, P-value, decision and conclusion.

(c) Is there a "Month" effect here? State the appropriate statistical hypotheses, then carry out the test. What can you infer?

(d) Consider your finding in part (b). Carry out the appropriate multiple comparison method to identify the "best" solar panel configuration.

6. A soft-drink manufacturer uses five agents (1, 2, 3, 4, 5) to handle premium distributions for its various products. The marketing director desired to study the timeliness with which the premiums are distributed. Twenty transactions for each agent were selected at random, and the time lapse (in days) for handling each transaction was determined. The results are given below.

Transaction	1	2	3	4	5	6	7	8	9	10
Agent 1	24	24	29	20	21	25	28	27	23	21
Agent 2	18	20	20	24	22	29	23	24	28	19
Agent 3	10	11	8	12	12	10	14	9	8	11
Agent 4	15	13	18	16	12	19	10	18	11	17
Agent 5	33	22	28	35	29	28	30	31	29	28

[1] *International Journal of Energy and Environmental Engineering*, December 2013.

Transaction	11	12	13	14	15	16	17	18	19	20
Agent 1	24	26	23	24	28	23	23	27	26	25
Agent 2	24	25	21	20	24	22	19	26	22	21
Agent 3	16	12	18	14	13	11	14	9	11	12
Agent 4	15	12	13	13	14	17	16	17	14	16
Agent 5	33	30	32	33	29	35	32	26	30	29

(a) Agents 1 and 2 distribute merchandise only, agents 3 and 4 distribute cash-value coupons only, and agent 5 distributes both merchandise and coupons. Does the data indicate that there is no difference in the time lapse of the premium distributions between a "merchandise only" approach and a "cash-value coupon" approach? Use $\alpha = 0.05$.

(b) Continued with part (a), consider the following contrasts:

$$\begin{aligned}
\mathrm{L}_1 &= \mu_1 - \mu_2 \\
\mathrm{L}_2 &= \mu_3 - \mu_4 \\
\mathrm{L}_3 &= \frac{\mu_1 + \mu_2}{2} - \mu_5 \\
\mathrm{L}_4 &= \frac{\mu_3 + \mu_4}{2} - \mu_5 \\
\mathrm{L}_5 &= \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}
\end{aligned}$$

Using Scheffe's procedure, construct a 99% family of confidence interval estimates for the above constrasts. Interpret your results.

(c) Of all the premium distributions, 25% are handled by Agent 1, 20% by Agent 2, 20% by Agent 3, 20% by Agent 4, and 15% by Agent 5. Using this new information, is there a better (faster) method for distributing premiums ? Test at $\alpha = 0.05$.