STAT 2607 Solutions to Assignment # 1 Winter 2014

Last Name ——————————————- , **First** ——————————-

Student # ——————————
Lab section ($A_1, \ldots A_5$, $B_1 \ldots, B_5$, **or** $C_1 \cdots C_5$) ——————————

<u>Due in class:</u> **Sec. A, B: Mon. Feb. 3rd, Sec. C: Wed. Feb. 5th.**

Total mark=80. Marks for each question are given in [ ]

Part I. Lab questions. Use only blanks left to answer lab questions. Data are in the Excel file named 2607Lab1data. DO NOT attach any minitab output.

1. (**One-way ANOVA**) Data in columns A,B,C corresponds to Graduate record examination (GRE) scores for some students admitted at a local university. We want to run a one-way ANOVA to see if there is significant evidence to indicate a difference in mean GRE scores for these programs. One can proceed as follows: copy and paste the data into columns C1,C2,C3 and name these columns Prog1, Prog2, and Prog3 respectively. Then go to
   Stat→ANOVA→ One-way(Unstacked), then put these columns in "Responses". Click on "Comparisons" and check Tukey's family error rate (if not already checked).
   [**2**] (a) What is the value of the F test statistic?**8.55**
   [**2**] (b) What is its P-value?**.005**
   [**5**] (c) What decision would you make?**I reject the null hypothesis and conclude that there is a significant difference between these three programs.**
   [**6**] (d)What are the 3 Tukey's simultaneous paired confidence intervals Prog2-Prog1? [**14.2 , 171.8**], Prog3-Prog1?[**-101 , 56**], and Prog3-Prog2?[**-194 , -36.4**]
   [**5**] (e) Based on d), which program seems to be significantly better than the others?**Program 2 seems to be better since if one switches from 1 to 2, the average GRE will increase by between 14.2 and 171.8, and if one switches from 3 to 2 the average GRE will increase by between 36.4 and 194. Since the the other CI contains the value zero, there seems to be no statistical difference between programs 1 and 3.**

2. (**Randomized block design ANOVA**) A study was conducted to compare automobile gasoline mileage for three brands of gasoline, A,B,and C. Four automobiles, all of the same make and model, were used in the experiment, and each gasoline brand was tested in each automobile. Using each brand in the same automobile has the effect of eliminating (blocking out) automobile variability. The data (in miles per gallon) are recorded in columns F,G,H. We want to run a random block design ANOVA test to see if there is a gasoline effect (on the mileage per gallon). One can proceed as follows: copy and paste the data into columns C5,C6,C7, and name them respectively Auto, Mileage, and Brand. Then go to: Stat→ANOVA→Balanced ANOVA→ Put Mileage

in Responses, Brand and Auto in Model, and Brand in Random.

[**4**] (a) What are the F =**6.46** and P-value =**0.032** of your test?

[**5**] (b) What decision would you make? **Since the P value is below .05, I would reject H0 and conclude that there is significant evidence of a difference between these brands.**

3. (**Two-way ANOVA**) A local school board was interested in comparing test scores on a standardized reading test for fourth-grade students in their district. They selected a random sample of five male and five female fourth-grade students at each of four different elementary schools in the district and recorded the test scores. The results are stored in columns K,L,M of the Excel file. We want to run a two-way ANOVA test to see if there is a gender (male/female) effect and if there is a school effect, and also if there is an interaction (gender/school) effect. One can proceed as follows: copy and paste the data into columns c10,c11,c12 (of the minitab spreadsheet). Name these columns respectively Score, School, Gender. Then go to Stat→ANOVA→ Two-way→ Enter Score in Response, Gender in row factor, and School in Column factor. Then hit OK.

   [**6**] (a) What are the values of the F test statistics corresponding to School factor **27.75**, and gender factor **2.09**, and the interaction **1.19** ?

   [**6**] (b) What are the P values of the statistics above: **0** and **.158** and **.329** (respectively)?

   [**5**] (c) What can you conclude based on part (b)? (Give plain English answers).**I would conclude that there is significant evidence of a school factor, but no gender factor effect, and no interaction between school and gender factors.**

Part II: Long answer questions

1. Data below corresponds to sample weights of two different types (A and B) of Tuna. We want to test at significance level $\alpha = .05$ if the two tuna populations from which the samples come have the same variance.

A	.99	1.92	1.23	.85	.65	.69	.6	.53	1.41	1.12	.63	.67	.6	.66	1.9
B	2.56	1.92	1.3	1.79	1.23	.62	.66	.62	.65	.6	.67				

Figure 1: Tuna weights

   [**5**] (a) Compute sample variances S_1^2 and S_2^2 of the two types A and B.

$$S_1^2 = \frac{\sum_{i=1}^{15} x_i^2 - \frac{\left(\sum_{i=1}^{15} x_i\right)^2}{15}}{14} = \frac{16.93 - (14.45)^2/15}{14} = .215.$$

$$S_2^2 = \frac{\sum_{i=1}^{11} y_i^2 - \frac{\left(\sum_{i=1}^{11} y_i\right)^2}{11}}{10} = \frac{19.08 - (12.62)^2/11}{10} = .46.$$

   [**5**] (b) Run the appropriate statistical test and conclude if there is a difference in the two population variances.

We test H_0: $\sigma_1^2 = \sigma_2^2$ versus H_a: $\sigma_1^2 \neq \sigma_2^2$. The test statistic is $F = S_2^2/S_1^2 = .46/.215 = 2.13$. The critical value is read from F table with 10 and 14 degrees of freedom at significance level $\alpha = .05$. We get from F table the value $F_{.025} = 3.15$. Since observed F (2.13) is below table's F (3.15) we fail to reject H_0. No enough evidence of a statistical difference between the two population variances.

2. A study was conducted to investigate the effect of management training on the decision-making abilities of supervisors in a large corporation. Sixteen supervisors were selected and eight were randomly chosen to receive managerial training. Four trained and four untrained supervisors were then randomly selected to function in a situation in which a standard problem arose. The other eight supervisors were presented with an emergency situation in which standard procedures could not be used. The response was a management behaviour rating for each supervisor as assessed by a rating scheme devised by the study's author. The data are shown below.

Situation (B)	**Training (A)**		
Standard	Trained	Not trained	Total
	85	53	**519**
	91	49	
	80	38	
	78	45	
Emergency	76	40	**473**
	67	52	
	82	46	
	71	39	
Total	**630**	**362**	**992**

Figure 2: Management rating

[3] (a) What are the two factors considered in this study?
Factor 1: training. Factor 2: Situation.

[2] (b) What are the levels of each factor?
Factor 1 has two levels: Trained, and not trained. Factor 2 has two levels: Standard situation, and emergency situation.

[10] (c)Construct the ANOVA table for this study.
This is a two-way ANOVA. $a = b = 2$, and $m = 4$., the overall mean is $\bar{x} = 992/16 = 62$.

$$\bar{x}_{11} = \frac{85 + 91 + 80 + 78}{4} = 83.5, \quad \bar{x}_{21} = \frac{53 + 49 + 38 + 45}{4} = 46.25,$$

$$\bar{x}_{12} = \frac{76 + 67 + 82 + 71}{4} = 74, \quad \bar{x}_{22} = \frac{40 + 52 + 46 + 39}{4} = 44.25.$$

$\bar{x}_{1.} = 630/8 = 78.75, \quad \bar{x}_{2.} = 362/8 = 45.25, \quad \bar{x}_{.1} = 519/8 = 64.875, \quad \bar{x}_{.2} = 473/8 = 59.125.$

$$SST = (85 - 62)^2 + (53 - 62)^2 + (91 - 62)^2 + \cdots + (39 - 62)^2 = 5136,$$

$$SS(1) = 2 \times 4[(\bar{x}_{1.} - \bar{x})^2 + (\bar{x}_{2.} - \bar{x})^2] = 8[(78.75 - 62)^2 + (45.25 - 62)^2] = 4489.$$

$$SS(2) = 2 \times 4[(\bar{x}_{.1} - \bar{x})^2 + (\bar{x}_{.2} - \bar{x})^2] = 8[(64.875 - 62)^2 + (59.125 - 62)^2] = 132.25.$$

$$\begin{aligned} SS(int) &= 4\big[(\bar{x}_{11} - \bar{x}_{1.} - \bar{x}_{.1} + \bar{x})^2 + (\bar{x}_{21} - \bar{x}_{2.} - \bar{x}_{.1} + \bar{x})^2 + (\bar{x}_{12} - \bar{x}_{1.} - \bar{x}_{.2} + \bar{x})^2 \\ &\quad +(\bar{x}_{22} - \bar{x}_{2.} - \bar{x}_{.2} + \bar{x})^2] \\ &= 4[(83.5 - 78.75 - 64.875 + 62)^2 + (46.25 - 45.25 - 64.875 + 62)^2 \\ &\quad +(74 - 78.75 - 59.125 + 62)^2 + (44.25 - 45.25 - 59.125 + 62)^2] \\ &= 56.25. \end{aligned}$$

Hence $SSE = SST - SS(1) - SS(2) - SS(int) = 5136 - 4489 - 132.25 - 56.25 = 458.5$. Now we are ready to get the two-way ANOVA table:

Source of variations	Degrees of freedo m	Sum of square s	Mean square	F	F(.05) from F table with 1 and 12 df
Factor 1	1	SS(1)= 4489	MS(1)=SS(1)/1= 4489	MS(1)/MSE= 4489/38.20=117.51	4.75
Factor 2	1	SS(2)= 132.25	MS(2)=SS(2)/1= 132.25	MS(2)/MSE= 132.25/38.2=3.46	4.75
Interactio n	1	SS(int) =56.25	MS(int)=SS(int)/1 =56.25	MS(int)/MSE= 56.25/38.2=1.47	4.75
Error	12	SSE= 458.5	MSE=SSE/12= 38.20		

Figure 3: Two-way ANOVA table

[3] (d) Is there a significant interaction between the presence or absence of training and the type of decision making situation? Test at $\alpha = .05$.

Since $F(\text{int})$ is not above the table value we fail to reject H_0: we conclude that there is no evidence of an interaction between the two factors. I.e. no evidence of interaction between presence /absence of training and the type of decision making.

[3] (e) Do the data indicate a significant difference in behaviour ratings for the two types of training categories? (Test at $\alpha = .05$).?
Since $F(1) > F_{.05}$ we reject H_0: therefore there is enough evidence that training does affect the behaviour ratings.

[3] (f) Do the data indicate a significant difference in behaviour ratings for the two types of situations? (Take $\alpha = .05$).

Since $F(2) < F_{.05}$ then we fail to reject H_0: no sttistical evidence that the types of sitiations affect the behaviour ratings.