IE 431 Design of Experiments
First Midterm Exam
Duration 75 min

Date 28 Thul-Qedah 1430 (16/11/2004)

Student Name: __________________________ Student Number: __________________

Question 1 (7 points)

Vitamin D deficiency could be related to the amount of fibre in the diet. Two groups of healthy adults are randomly assigned to one of two diets: Normal or High Fibre. A measure of vitamin D is then obtained for each of the subjects:

Normal Diet	19.1	24.0	28.6	29.7	30.0	34.8	
High Fibre	12.0	13.0	13.6	20.5	22.7	23.7	24.8

Assume that the the two measures of vitamin D are normally distributed and the the variances of the two populations are equal.
(a) Is there a significance difference between the vitamin D levels of the two groups, use $\alpha = .05$? **(3 points that is 1 point for the formula and 2 points for the calculations)**
(b) Find the 95 percent confidence interval on the difference means. **(3 points that is 1 point for the formula and 2 points for the calculations)**
(c) Discuss the effect of randomization on the design of experiment? **(1 point)**

(a)

For Normal Diet , let $\bar{y}_1$ be the sample mean, S_1^2 be the sample variance; let μ_1 be the population mean, σ^2_1 be the population variance
For the High Fibre, let $\bar{y}_2$ be the sample mean, S_2^2 be the sample variance; let μ_2 be the population mean, σ^2_2 be the population variance

From the provided samples
$\bar{y}_1 = 27.7$
$S_1^2 = 29.63$
$\bar{y}_2 = 18.61$
$S_2^2 = 30.8$

Hypotheses: $H_0: \mu_1 = \mu_2$ v $H_1: \mu_1 \neq \mu_2$

Assumption: $\sigma^2_1 = \sigma^2_2$ and unknown

Statistics Test is $t_0 = \dfrac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

$$S^2{}_p = \frac{(n_1 - 1)S_1{}^2 + (n_2 - 1)S_2{}^2}{n_1 + n_2 - 2} = \frac{(6-1)29.63 + (7-1)30.8}{6+7-2} = 30.27$$

$$t_0 = \frac{\bar{y}_1 - \bar{y}_2}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{27.70 - 18.61}{\sqrt{(30.27)}\sqrt{\frac{1}{6} + \frac{1}{7}}} = \frac{9.09}{3.06} = 2.97$$

Rejection Criteria

Reject H_0 if $|t_{0.025,11}| > t_{\frac{\alpha}{2},v}$ where $v = n_1 + n_2 - 2 = 6 + 7 - 2 = 11$

Reject H_0 if $t_0 > t_{0.025,11}$ or $t_0 < -t_{0.025,11}$, i.e., $t_0 > 2.201$ or $t_0 < -2.201$ (from the *t* distribution table)

In this case we reject the null hypotheses because (t_0= 2.97) > 2.201. We conclude that there is a significant difference between the vitamin D levels of the two groups

(b) $\bar{y}_1 - \bar{y}_2 - t_{\frac{\alpha}{2},n_1+n_1-2} S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \le \mu_1 - \mu_2 \le \bar{y}_1 - \bar{y}_2 + t_{\frac{\alpha}{2},n_1+n_1-2} S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$27.7 - 18.6 - (2.201)(2.97) \le \mu_1 - \mu_2 \le 27.7 - 18.6 + (2.201)(2.97)$

$27.7 - 18.6 - (2.201)\sqrt{30.27}\sqrt{\frac{1}{6} + \frac{1}{7}} \le \mu_1 - \mu_2 \le 27.7 - 18.6 + (2.201)\sqrt{30.27}\sqrt{\frac{1}{6} + \frac{1}{7}}$

$2.35 \le \mu_1 - \mu_2 \le 15.82$

(c) Randomization is a process that assigns research participants by chance, rather than by choice. Randomization makes observations or errors independently distributed random variables which comply the assumptions in the statistical methods.

Question 2 (13 points)

Fifteen students took part in an experiment to assess the effect of study habits on the retention of material. Three different types of study habit were taken and the fifteen students were randomly assigned to one of these types (treatments)

- Treatment one was a control where the students simply read the material,
- Treatment two involved reading the material and then producing a summary.
- Treatment three involved skimming the material, thinking of key questions and then reading the passage properly.

Each student was then assessed on his knowledge of the material by means of a multiple choice exam. Results were as follows:

Study Habit	**Scores**				
Control-reading only	22	30	14	28	31
Reading and summary	32	37	42	28	21
Skimming, thinking, reading	44	37	48	35	31

(1) Does the study method affect the retention of the material? Clearly state the hypothesis and use $\alpha = .05$? **(6 Points that is 1 point for the hypothesis, 1 point for the formulas, 3 points for the calculations and 1 point for the rejection conclusion)**

(2) Test all pairs of means using Tukey's method with $\alpha = 0.05$; which study habit provides the best scores for the sample? **(4 Points that is 1 point for the formula 2 points for the calculations and 1 point for the final conclusion)**

(3) Mention the three conditions for the model adequacy; plot the residuals versus the fitted value and show which condition is valid by this plot. **(3 Points that is 1 point for the model conditions, 2 points for the plot and the conclusion)**

(1)

The Hypotheses:
$H_0 : \mu_1 = \mu_2 = \mu_3$
$H_1 : \mu_i \neq \mu_j$ for at least one pair of (i, j)

User One-Way ANOVA where $a = 3$, $n = 5$

Study Habit	**Scores**					$\bar{y}_{i.}$
Control-reading only	22	30	14	28	31	25
Reading and summary	32	37	42	28	21	32
Skimming, thinking, reading	44	37	48	35	31	39
						$\bar{y}_{..}$ =32

$$SS_{Treatments} = n\sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SS_{Total} = n\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$$

$$SS_{Error} = SS_{Total} - SS_{Treatments}$$

$$MS_{Treatments} = \frac{SS_{Treatments}}{a - 1}$$

$$MS_{Error} = \frac{SS_{Error}}{N - a}$$

Study Habit						Average	Diff	Diff^2
Control-reading only	22	30	14	28	31	25	-7	49
Reading and summary	32	37	42	28	21	32	0	0
Skimming, thinking, reading	44	37	48	35	31	39	7	49
						32		98
Residuals	-3	5	-11	3	6			
	0	5	10	-4	-11			
	5	-2	9	-4	-8			
SS Treatments								
								490
MS treatments	**245**							
SS Total								
	22	30	14	28	31			
	32	37	42	28	21			
	44	37	48	35	31			
Diff	-10	-2	-18	-4	-1			
	0	5	10	-4	-11			
	12	5	16	3	-1			
Diff ^2	100	4	324	16	1			
	0	25	100	16	121			
	144	25	256	9	1			
SS Total						**1142**		
SS error	**652**							
MS error	**54.33**							
F0	**4.51**							

One-way ANOVA: Control-reading , Reading , Skimming

Source	DF	Sum of Squares	Mean Square	F_0
Factor	2	490.0	245.0	4.51
Error	12	652.0	54.3	
Total	14	1142.0		

Reject H_o if ($F_0 = \frac{MS_{Treatments}}{MS_{Error}}$ **)** $> F_{\alpha, a-1, N-a}$

From F distribution table ; $F_{0.05,2,12} = 3.89$

The null hypotheses should be rejected and can conclude that there is a significant difference on average between the scores on the different methods

(b) Using Tukey's Method with $\alpha = 0.05$

For equal sizes Tuckey's method

$$T_\alpha = q_\alpha (a, f)\sqrt{\frac{MS_{Error}}{n}}$$

$a = 3$, $f = 12$, from Studentized Range Statistics Table

$$T_\alpha = q_\alpha (a, f)\sqrt{\frac{MS_{Error}}{n}} = 3.77 \sqrt{\frac{54.3}{5}} = 3.77(3.3) = 12.43$$

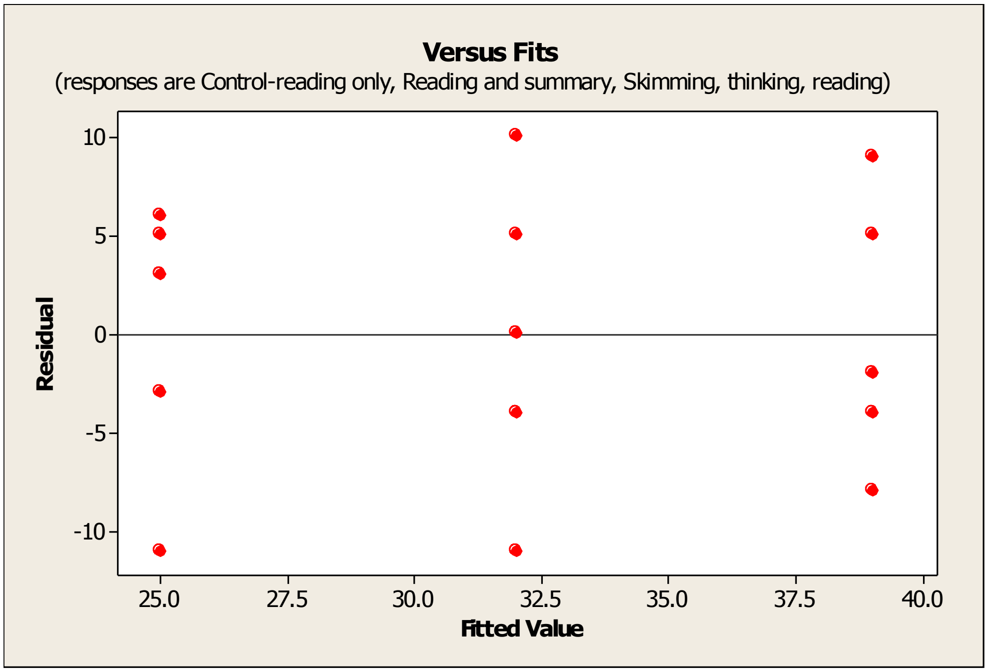
1 vs 2 = | 25-32 | = 7 $<$ 12.43
1 vs 3 = | 25-39 | = 14 $>$ 12.43
2 vs 3 = | 32-39 | = 7 $<$ 12.43

Tukey's method shows that there significant difference between sample 1 and sample 3 . But there is no significance difference between simple 1 and 2 or 2 and 3. We choose method number 3 , Skimming, thinking, reading as the best method with the highest average. We can also accept method number 2 as there is no significant difference between method 2 and 3

(c) The conditions for adequate ANOVA mode

- The observations in each group come from a normal distribution. (normal assumption)
- The population variances of each group are the same. (homogeneity of variances)
- The observations are independent of each other (independence assumption)

The plot the residuals versus the fitted value shows that the variance of each group is very close.



Versus Fits
(responses are Control-reading only, Reading and summary, Skimming, thinking, reading)
Residual
10
5
0
-5
-10
25.0
27.5
30.0
32.5
35.0
37.5
40.0
Fitted Value