University of Waterloo

Stat 372 – W16

Term Test I

Date: Thursday, Feb. 4, 2016.

Duration: 60 minutes

Family Name:________________________ First Name:________________________ I.D. #:______________

Signature:________________________

Instructor: P. Balka

Instructions:

- This exam has 6 pages including this cover page. The marks for each question are indicated (total of 30). **Show your work. Your grade will be influenced by how clearly you express your ideas, and how well you organize your solutions.**
- When using the probability tables, choose the closest degrees of freedom listed if the actual degrees of freedom are not provided.
- No questions will be permitted.

Solutions

1) In a study to compare the return of dividend paying vs non-dividend stocks in their portfolio, an investor selects a pair of dividend/non-dividend stocks from each of ten economic sectors, and calculates the return after a one-year period. (Partial) data and summary statistics appear below.

Sector	Dividend	Non-dividend
1	.2340	.2203
2	.4270	.1754
⋮	⋮	⋮
10	.0878	.1781
	$\bar{x}$: 0.22	0.17
	s: 0.179	0.123

$s_d = 0.176$ (sample standard deviation of the differences)

a) [2] Is this an observational or experimental plan? Briefly explain.

Observational. The value of the explanatory variate (stock type) was not randomly assigned - it was only observed and recorded.

b) [3] Give the treatment effects model for this study.

$Y_{ij} = \mu + \tau_i + \beta_j + R_{ij}$ $\quad R_{ij} \sim G(0, \sigma)$ $\quad i = 1, 2$ $\quad j = 1, 2, \ldots, 10$

c) **[6]** Is there a difference in mean return between the two stock types over this period? Answer this question by performing an appropriate hypothesis test. As with all hypothesis tests, always be sure to include the null hypothesis (as it relates to the model parameters), value and degrees of freedom of test statistic, p-value, and conclusion in the context of the study.

$H_0: \tau_1 - \tau_2 = 0$

$(H_a: \tau_1 - \tau_2 \neq 0)$

$$t = \frac{\hat{\tau}_1 - \hat{\tau}_2}{se(\hat{\tau}_1 - \hat{\tau}_2)} = \frac{0.22 - 0.17}{s_d / \sqrt{n_d}} \quad \text{(paired data)}$$

$$= \frac{0.05}{0.176/\sqrt{10}} = 0.90 \quad \text{(from } t_9 \text{ dist. under } H_0\text{)}$$

from t-table: $P(t_9 > 0.703) = 0.25$

$P(t_9 > 1.383) = 0.10$

$\Rightarrow \quad 0.20 < \text{p-value} < 0.50$

Do not reject Ho. There is no significant difference in the mean return of the two stock types

2) (Please note: all values given in this question are hypothetical). In a study of co-op earnings, a random sample of students is selected from each of the four Math/Bus & Accounting programs below, and data is collected on their program and the average and standard deviation of hours worked per week during their most recent work term. The summary data are shown below:

	sample size	mean	std. dev.
Math/Bus	33	44.5	4.3
DD	45	45.1	4.6
FARM	44	42.8	3.7
Math/CPA	23	47.6	5.9

a) **[2]** Is this an observational or experimental plan? Briefly explain.

Observational. The subjects were not randomly assigned to the groups (programs).

b) **[2]** Identify the assumption of the residuals addressed **by the data in the table**, and briefly indicate whether or not you feel the data supports this assumption.

$R_{ij} \sim G(0, \sigma)$ ind.

The only assumption that can be addressed by the data in the table is the assumption of constant variance.

Looking at the sample standard deviations of each group, it appears that there may be differences in the variances for each group that might invalidate this assumption, although further investigation would be required.

c) **[6]** Is there a difference in mean hours worked among the different programs? Answer this question by creating an ANOVA table and performing an appropriate hypothesis test. Show your work.

$$SS(Res) = \sum_{i,j} (y_{ij} - \bar{y}_{i+})^2 = \sum_i (n_i - 1) s_i^2$$

$$= 32(4.3)^2 + 44(4.6)^2 + 43(3.7)^2 + 22(5.9)^2$$

$$= 2877.21$$

$$SS(Treat) = \sum_i n_i (\bar{y}_{i+} - \bar{y}_{++})^2 \qquad (\bar{y}_{++} = 45.0)$$

$$= 33(44.5 - 45)^2 + \cdots + 23(47.6 - 45.0)^2$$

$$= 377.14$$

ANOVA table

Source	SS	df	MS	F
Treat	377.14	3	125.71	6.16
Res	2877.21	141	20.41	

$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$

H_a: least one $\tau_i \neq 0$, $i = 1,2,3,4$

$F = 6.16$

p-value < 0.05 (from F table: $P(F_{3,100} > 2.70) = 0.05$

100 closest available df to 143

Reject Ho.

There is a significant difference in mean hours worked among the programs

d) **[6]** Is there a difference between the effect of Math/CPA and the average effect of the Math/Bus and DD programs on hours worked? **Answer this question by providing a confidence interval** for the appropriate contrast.

$$\theta = \tau_4 - \frac{\tau_1 + \tau_2}{2}$$

$$\hat{\theta} = \bar{y}_4 - (\bar{y}_1 + \bar{y}_2)/2$$

$$= 47.6 - (44.5 + 45.1)/2 = 2.80$$

$$SE(\hat{\theta}) = \hat{\sigma}\sqrt{\tfrac{1}{23} + \tfrac{1}{4}\left(\tfrac{1}{33} + \tfrac{1}{45}\right)}$$

$$= \sqrt{20.41}\sqrt{0.0562} = 1.07$$

95% C.I. for θ;

$$\hat{\theta} \pm t_{141,.975}\, SE(\hat{\theta})$$

$$= 2.80 \pm 1.962(1.07)$$

$$= 2.80 \pm 2.10 = (0.70, 4.90)$$

Since the interval does not contain 0, we can conclude that the mean hrs for Math/CPA is greater than the average of the mean hrs worked for Math/Bus + DD.

e) **[3]** Suppose there is no difference in the treatments effects of the DD and FARM programs. Let $\theta = \tau_{DD} - \tau_{FARM}$. Find (the approximate value of) $P(|\hat{\theta}| > 2.3)$.

note this is the p-value associated with $H_0: \theta = 0$

$$\hat{\theta} \sim G(0, \sqrt{a_i^2/n_i}\,\sigma)$$

$$\Rightarrow \frac{\hat{\theta}}{SE(\hat{\theta})} = \frac{\hat{\theta}}{\hat{\sigma}\sqrt{\Sigma a_i^2/n_i}} \sim t_{141}$$

$$\Rightarrow \frac{45.1 - 42.8}{\sqrt{20.41}\sqrt{\tfrac{1}{45} + \tfrac{1}{44}}} = \frac{2.3}{0.958} \approx 2.40$$

from t-table; $P(|\hat{\theta}| > 2.3) \approx 2P(t_{141} > 2.40)$

$$\approx 2(.01) = 0.02$$