Ra60

Seat

Please print in pen: MC 2017

Waterloo Student ID Number:

WatIAM/Quest Login Userid:

UNIVERSITY OF WATERLOO

Examination
Midterm 2
Fall 2017
STAT 322

Times: Thursday 2017-11-23 at 10:00 to 11:10
Duration: 1 hour 10 minutes (70 minutes)
Exam ID: 3677413
Sections: STAT 322 LEC 001
Instructors: Peter Balka

Special Materials

Candidates may bring only the listed aids.

- Calculator - Non-Programmable

Solutions

STAT 322 Term Test II

Instructions and information

1. This test is out of **30** total marks. The marks for each question are indicated.
2. Be sure to show your work. Your grade will be influenced by how clearly you express your ideas, and how well you organize your solutions.
3. Use the closest available degrees of freedom from the probability tables whenever the exact degrees of freedom is not available.
4. Non-programmed calculators only.
5. No questions are permitted.
6. There are questions on both sides of the page.
7. Answer the questions in the spaces provided.
8. Only question pages will be marked.
9. Do not write on the cover page.
10. Last page is for rough work and will not be graded. **<u>DO NOT DETACH</u>** this page.

Formulas and values you may find useful:

$$n = \frac{1}{\frac{1}{N} + \frac{m^2}{z^2 \hat{\pi}(1-\hat{\pi})}} \qquad n = \frac{1}{\frac{1}{N} + \frac{m^2}{z^2 \hat{\sigma}^2}}$$

For $Z \sim G(0,1)$:
$P(Z < 1.645) = 0.950$
$P(Z < 1.96) = 0.975$
$P(Z < 2.575) = 0.995$

F-table (right tail) $\alpha = 0.05$

For each row (denominator degrees of freedom k) and column (numerator degrees of freedom j), the table entry e satisfies $P(F(j,k) \geq e) = \alpha$.

numerator degrees of freedom

denominator degrees of freedom	1	2	3	4	5	6	7	8	9	10	20	30
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	248.02	250.10
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.15	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.94	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.65	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.54	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.46	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.39	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.33	2.25
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.28	2.19
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.23	2.15
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.19	2.11
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.16	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.12	2.04
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.10	2.01
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.07	1.98
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.05	1.96
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.03	1.94
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.01	1.92
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.84	1.74
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.78	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.68	1.57

1) The table of means and variances below are from a 3 x 2 factorial study to investigate the effects of dose levels of vitamin C (0.5, 1, and 2 mg/day) and supplement type (orange juice (OJ), ascorbic acid (VC)) on tooth length (mm) in 60 guinea pigs. (Source: C. I. Bliss (1952) The Statistics of Bioassay. Academic Press. The numbers have been rounded for computational convenience).

Table 1. Treatment means

		dose			
		low	med	high	
supp	OJ	14	23	26	21
	VC	8	17	26	17
		11	20	26	$\bar{y}_{\cdots} = 19$

Table 2. Treatment variances

		dose		
		low	med	high
supp	OJ	20	16	7
	VC	8	6	23

a) **[2]** Give the response model associated with this study. Be as specific as possible.

$Y_{abj} = \mu + \tau_{ab} + R_{abj}$ $R_{abj} \sim G(0, \sigma)$

$a = 1, 2, 3$ (dose)

$b = 1, 2$ (supp)

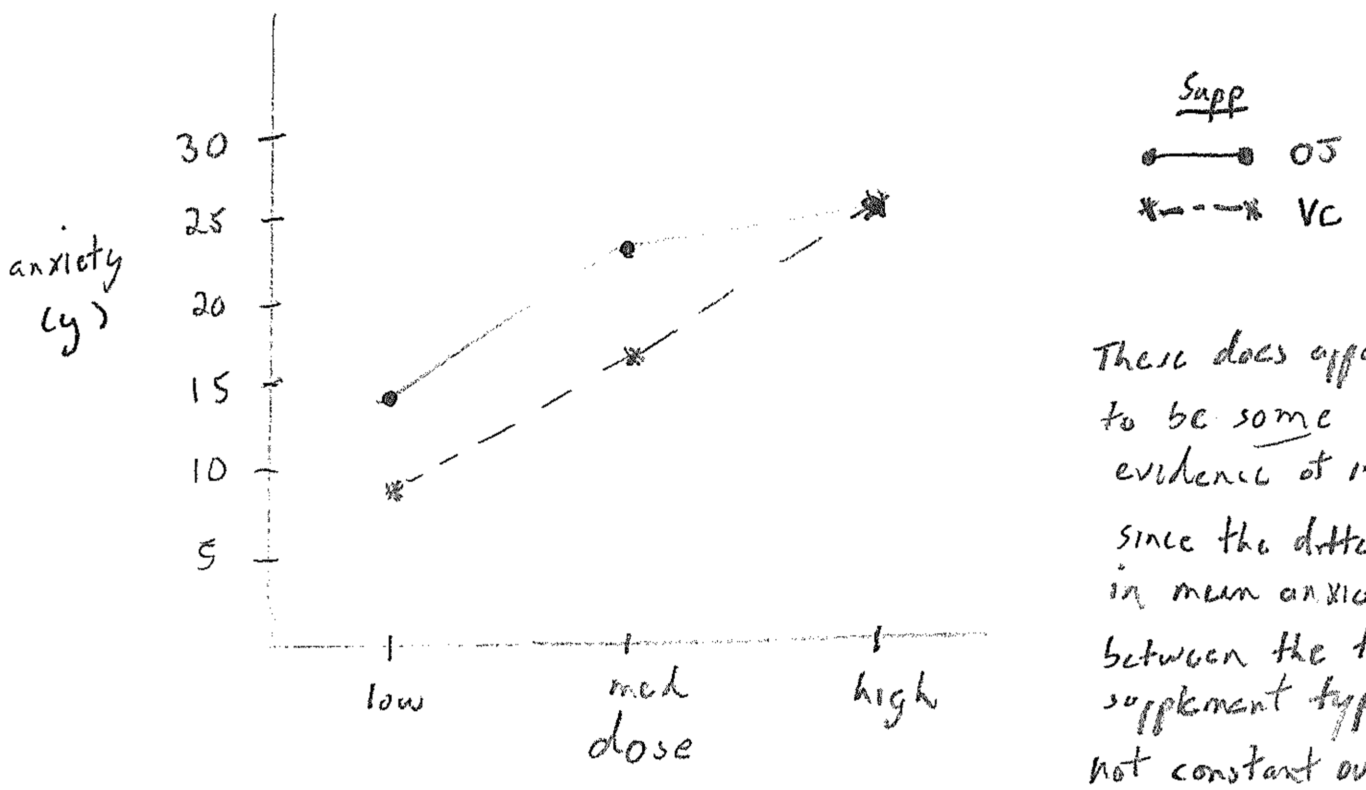
$j = 1, 2, \ldots, 10$

b) **[2]** Based on the above data, what assumption of the response model may be violated? (We will ignore this potential problem for the remaining questions)

The assumption of constant variance across groups ($R_{abj} \sim G(0, \sigma)$)

Based on Table 2, it appears from the sample variances that different treatment combinations may have different variances

c) **[3]** Create an interaction plot for this data. Comment on whether there is evidence of interaction in the plot.



There does appear to be some evidence of interaction since the difference in mean anxiety between the two supplement types is not constant over the dose levels

d) **[2]** SS(Res) for this data is 720. Show how this value was calculated from the information contained in the above tables.

$$SS(Res) = \sum_{a,b,j} (y_{abj} - \bar{y}_{ab\cdot})^2$$

$$= \sum_{ab} (r-1) s_{ab}^2$$

$$= 9(20^2 + 16^2 + \cdots + 23^2)$$

$$= 720$$

e) **[6]** Create an ANOVA table (including p-values from *F* table provided) to assess the effects of dose level, supplement type and their interaction on tooth length. Show your work, and state your conclusions in the context of the study.

$SS(Res) = 720$

$SS(Treat) = 10((14-19)^2 + (23-19)^2 + \cdots + (26-19)^2)$

$= 2640$

$SS(Dose) = 2280$

$SS(Supp) = 240$

$SS(Dose*Supp) = 2640 - 2280 - 240 = 120$

ANOVA

Source	SS	df	MS	F	p-value
Treat	2640	5	528	39.6	<.05
-Dose	2280	2	1140	85.5	<.05
-Supp	240	1	240	18.0	<.05
-Dose*Supp	120	2	60	4.5	<.05
Res	720	54	13.33		

There is significant interaction between dose & supplement type ($F=4.5 \Rightarrow$ p-value < 0.05).

Since the effect of dose on tooth length depends on supp type (and vice verse) the significance of the main effects (Dose, Supp) cannot easily be interpreted.

2) **[2]** Based on a <u>simple random sample</u> (SRS) of five numbers from the numbers {1, 2, 3, ..., 100}, which one of these two samples is more likely: {2,3,4,5,6} or {4,16, 35, 72, 91}? Explain in one sentence. A correct guess without proper justification will receive no marks.

They are both <u>equally likely</u>, since in SRS, all samples (of size 5 in this case) has the same probability of being selected.

3) You wish to undertaken a quality control study to estimate μ, the mean length of a certain automobile part produced over a single (24 hr) day, and p, the proportion of parts produced that day that are 'rejects' (e.g. outside specification limits).

The production day is comprised of three 8 hr. shifts. The parts are placed in shipping containers (10 parts per container) as part of the production process. There were a total of 2400 parts produced on the day in question. You wish to obtain a random sample of 120 of these parts.

a) **[4]** Clearly describe a reasonable sampling protocol for this study that employs simple random sampling, cluster sampling and stratified sampling. Be sure to clearly define the clusters and strata.

one way:

- Stratified sampling stage: parts produced in a day are broken up into the three shifts (stratum = shift)
- Cluster sampling stage: four boxes (clusters) are randomly selected from each stratum (shift). All parts in selected containers are sampled ($3 \times 4 \times 10 = \underline{\underline{120}}$)
- simple random sampling (SRS): Boxes from each shift selected using SRS.

b) **[2]** Is study error likely to be a problem in this study? Explain in one or two sentences.

No. study error is the difference in attribute between the target population and study population. In this study, they are identical (all 2400 parts produced that day)

c) The sample of 120 parts yields a mean length of 63 mm and standard deviation of 2 mm. Nine of the parts are rejects.

i) **[4]** Calculate a 95% confidence interval for the mean length of the 2400 parts.

$\hat{\mu} \pm 1.96\sqrt{1-\frac{n}{N}}\ \frac{s}{\sqrt{n}}$

$63 \pm 1.96\sqrt{1-\frac{120}{2400}}\ \frac{2}{\sqrt{120}}$

$= 63 \pm 0.35$

$(62.65, 63.35)$

ii) **[3]** What sample size is required to estimate the proportion of rejects produced that day to within '3 percentage points, 19 times out of 20'?

$\hat{\pi} = \frac{9}{120} = 0.075$

$n = \dfrac{1}{\frac{1}{N} + \frac{m^2}{z^2\hat{\pi}(1-\hat{\pi})}}$

$= \dfrac{1}{\frac{1}{2400} + \frac{(.03)^2}{1.96^2(.075)(.925)}}$

$= 263.6$

$\Rightarrow n \geq 264$

LEFT BLANK

FOR ROUGH WORK

WILL NOT BE MARKED