STATISTICS 2080, Midterm Exam Solutions - October 2024

Exam is out of 50 points

1. A small study was carried out in which 6 subjects were randomized to two groups, each of size 3. The data were as follows:

Group 1:	10	20	30
Group 2:	30	40	50

Carry out a permutation test for the hypotheses $H_0 : \mu_1 = \mu_2$ vs $H_a : \mu_1 < \mu_2$.

(3) (a) How many distinct arrangements of the data are possible for creating two groups of size 3.

$$\binom{6}{3} = \frac{6!}{3!3!} = 20$$

(4) (b) Compute the p-value obtained by a permutation test. Show your work/reasoning.

Small values of $\bar{X}_1 - \bar{X}_2$ provide evidence against H_0 and in favour of H_A. The observed value of the test statistic is 20-40 = -20. There is one additional configuration which gives a value of the test statistic at least as small as the observed test statistic. This second configuration is obtained by swapping the "30's" in the two groups. Hence the p-value is $2/20 = .1$

(1) (c) What would be the p-value for the two sided alternative $H_A : \mu_1 \neq \mu_2$?

$2 \times .1 = .2$.

2. (a) For the data in problem 1, construct a 95% confidence interval interval for $\mu_1 - \mu_2$.
(6) (Note that these data are NOT paired, and that the sample variance of each sample is 100.)

You will need one of the following percentiles of the t-distribution to construct the confidence interval.

$$t_{.025,2} = 4.30 \quad t_{.025,3} = 3.18 \quad t_{.025,4} = 2.78 \quad t_{.025,5} = 2.57$$
$$t_{.05,2} = 2.92 \quad t_{.05,3} = 2.35 \quad t_{.05,4} = 2.13 \quad t_{.05,5} = 2.01$$

$n_1 = n_2 = 3$
$s_1^2 = s_2^2 = 100$
$s_p^2 = (2s_1^2 + 2s_2^2)/4 = 100, \ s_p = 10.$
$\bar{X}_1 - \bar{X}_2 = -20$
for a 95% CI, $\alpha/2 = .05/2 = .025$ and the degrees of freedom are $n_1+n_2-2 = 6-2 = 4$, so the appropriate value of t is $t_{.025,4} = 2.78$
The confidence interval is
$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2,n_1+n_2-2} s_p \sqrt{1/n_1 + 1/n_2}$
or $-20 \pm 2.78(10)\sqrt{2/3}$ or (-42.7,2.7)

(b) On the basis of the confidence interval you calculated, would you reject the null
(1) hypothesis for the two sided alternative $H_A : \mu_1 \neq \mu_2$? (Yes or No)

NO (because the CI above contains 0)

3. When carrying out the Wilcoxin Rank Sum Test (also known as the Mann-Whitney test), we need to rank the observations from smallest to largest.

(2) For the data in problem 1, fill in the ranks in the following table:

Group 1	Rank	Group 2	Rank
10	1	30	3.5
20	2	40	5
30	3.5	50	6

4. In general, if the 95% confidence interval for $\mu_1 - \mu_2$ **does not contain** 0, then the 99% confidence interval **will not contain** 0.)

(1) (c) Insufficient information to tell.

5. In general, if the 95% confidence interval for $\mu_1 - \mu_2$ **contains** 0, then the 90% confidence interval $\mu_1 - \mu_2$ **will contain 0**.

(1) (c) Insufficient information to tell.

6. A group of 32 rats were randomly assigned to each of 4 diets labelled (A,B,C,and D). The response is the liver weight as a percentage of body weight. Two rats escaped and another died, resulting in the following data

A	B	C	D
3.42	3.17	3.34	3.65
3.96	3.63	3.72	3.93
3.87	3.38	3.81	3.77
4.19	3.47	3.66	4.18
3.58	3.39	3.55	4.21
3.76	3.41	3.51	3.88
3.84	3.55		3.96
	3.44		3.91

For these data, the sum of squares for diet is 1.16, and the error sum of squares is 0.90.

(8) (a) Fill in the missing entries in the following ANOVA table.

Source	df	SS	MS	F
Diet	3	1.16	.387	10.74
Error	25	0.90	.036	—
Total	28	2.06	—	—

(1) (b) What is the estimate of σ^2? $\hat{\sigma}^2 = MSE = .036$
For this exam only, I will also accept $s^2 = SST/(n-1) = 2.06/28 = .014$.
s^2 is a good estimate of σ^2 only if there are no differences among groups, that is, when the null hypothesis holds. MSE is a good estimate under the null or alternative hypotheses.

(1) (c) What is the null hypothesis for the oneway ANOVA model?
$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$.

(1) (d) What is the alternative hypothesis for the oneway ANOVA model?
H_A : at least two of μ_A, μ_B, μ_C, μ_A are different.

(2) (e) Suppose that you reject H_0 when testing at level $\alpha = .05$, and you decide to make pairwise confidence intervals for each $\mu_i - \mu_j$, with $i \neq j$. What is the value of α^*?

$$\alpha^* = \frac{\alpha}{\binom{4}{2}} = .05/6 \approx .0083$$

7. The data from problem 6 were ranked. For each of $i = 1, 2, 3, 4$, the following table shows $\bar{R}_i$, the mean of the ranks for diet i.

diet	i	$\bar{R}_i$	n_i
A	1	18.5	7
B	2	6.44	8
C	3	11.92	6
D	4	22.81	8

(6) Calculate the value of the Kruskal-Wallis test statistic K for these data.

$n = 29$

$$K = \frac{12}{n(n+1)} \sum_{i=1}^{a} n_i \left(\bar{R}_{i.} - \frac{n+1}{2} \right)^2$$

$$= \frac{12}{29(30)} \left(7(3.5^2) + 8(8.56^2) + 6(3.08^2) + 8(7.81^2)\right)$$

$$\approx \frac{12}{29(30)} 1216.83 \approx 16.78$$

8. A two way analysis of variance with interaction was carried out. There were 32 observations in all. Factor A had 4 levels, Factor B had 4 levels, and there were 2 replicates for each combination of factors A and B. The overall estimate of variance for the data was $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 = 5$. The error sum of squares was $SSE = 64$, the sum of squares for factor A was $SSA = 4$ and the the sum of squares for factor B was $SSB = 6$.

(7) (a) Fill in the entries in the following ANOVA table.

Source	df	SS	MS	F
Factor A	3	4	4/3	1/3
Factor B	3	6	2	2/4
Interaction	9	81	9	9/4
Error	16	64	4	—
Total	31	155	—	—

$n = 32$, $SST = (n-1)s^2 = 31(5) = 155$

(1) (b) When testing the hypothesis of no interaction, what is the observed value of the test statistic?

9/4

(1) (c) When testing the hypothesis of no interaction, what is the numerator degrees of freedom?

9

(1) (d) When testing the hypothesis of no interaction, what is the denominator degrees of freedom?

16

(1) (e) If you fit a model without interaction to the same data, what this the new error sum of squares?

81+64 = 145

(1) (f) If you fit a model without interaction to the same data, what this the new error degrees of freedom?

9+16 = 25