8112

THE UNIVERSITY OF SYDNEY

FACULTIES OF ARTS AND SCIENCE

STAT 2012: Statistical Tests

Semester 2, 2012 | Time allowed: Two hours

Part B: Extended answer questions (50 marks)

Answer ALL questions.

1. (a) A pharmaceutical company develops two formulas to provide quicker relief of minor pains. Twelve participants with minor pain were recruited and randomly assigned to take the pills using each of the two formulas. They were then asked to report the length of time (in minutes) until some reliefs were felt after taking the pills. One participant dropped out and the results from other participants are reported below:

Pain relief	Time until some reliefs felt						Mean $\bar{y}_i$	Var. s_i^2
Formula 1	11	8	12	16	13	14	12.333	7.467
Formula 2	8	6	5	11	9		7.800	5.700

(i) (4 marks) Calculate a 95% confidence interval for the difference in time until some reliefs were felt for the two formulas. At the 5% significance level, does the time until some reliefs were felt differ between the two formulas?

(ii) (5 marks) Calculate the test statistic and p-value to test if the time until some reliefs were felt differ between the two formulas using the *Wilcoxon rank sum* test.

(b) An experiment is designed to study the germination rate under different temperatures and soil moistures. Results on the number of beans germinated out of 120 beans under 3 levels of temperature and 3 levels of soil moisture are reported below:

	Temperature		
Moisture	$15°C$	$25°C$	$35°C$
50%	63	74	81
70%	72	80	105
90%	57	65	61

(i) (4 marks) Test at the 5% significance level if the germination rate differs across temperature levels using the *Friedman* test.

State the two hypotheses, the test statistic, the p-value and the conclusion. If an exact p-value cannot be found from the table, state the range of values from the table.

(ii) (1 mark) Explain why the Friedman test is used.

2. A company examines three methods of irradiation of food to reduce bacteria for food preservation on four types of food: beef, chicken, egg and milk. The bacteria counts of three specimens of the same size for each type of food after irradiation for 10 minutes using each method are given in the following table:

Counts	Method			Block
	A	B	C	mean $\bar{y}_{i\cdot\cdot}$
Beef	47	52	36	
	50	56	41	
	53	57	46	48.667
Mean $\bar{y}_{1j\cdot}$	50	55	41	
Var. s_{ij}^2	9	7	25	
Chicken	53	61	48	
	57	67	50	
	49	73	46	56.000
Mean $\bar{y}_{2j\cdot}$	53	67	48	
Var. s_{ij}^2	16	36	4	
Egg	25	20	25	
	28	17	29	
	22	26	30	24.667
Mean $\bar{y}_{3j\cdot}$	25	21	28	
Var. s_{ij}^2	9	21	7	
Milk	45	48	39	
	40	51	33	
	38	45	30	41.000
Mean $\bar{y}_{3j\cdot}$	41	48	34	
Var. s_{ij}^2	13	9	21	$\bar{y}_{\cdots}$
Tr. mean $\bar{y}_{\cdot j\cdot}$	42.25	47.75	37.75	42.583

The following summary statistics are given:

$$\sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{k=1}^{3} y_{ijk}^2 = 71751, \quad \bar{y}_{\cdots} = \frac{1}{36}\sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{k=1}^{3} y_{ijk} = 42.583 \text{ and } \sum_{i=1}^{4}\sum_{j=1}^{3} s_{ij}^2 = 177$$

(a) (7 marks) Complete the *two-way ANOVA* table under the factorial design.

(b) (3 marks) Test at the 5% significance level if the bacteria counts for the three methods of irradiation are the same using the *two-way ANOVA* test under the factorial design.

State the two hypotheses, the test statistic, the test assumptions if any, the p-value and the conclusion. If an exact p-value cannot be found from the table, state the range of values from the table.

(c) (2 marks) If the interaction effect is dropped from the model, calculated the revised test statistic for the test in (b).

3. A movie producer has observed that, in general, movies with popular actors/actresses seem to generate more revenue than those movies with less well-known actors/actresses. To examine his belief, he records the gross revenue and the payment (in million $) given to the two highest-paid actors/actresses in the movie for ten recently released movies.

Movie	**Payment to two highest paid actors/actresses** x_i	**Gross revenue** y_i
1	5.3	48
2	7.2	65
3	1.3	18
4	1.8	20
5	3.5	31
6	2.6	26
7	8.0	73
8	2.4	23
9	4.5	39
10	6.7	58

The following summary statistics are given:

$$\sum_{i=1}^{10} y_i = 401;\ \sum_{i=1}^{10} x_i = 43.3;\ \sum_{i=1}^{10} y_i^2 = 19633;\ \sum_{i=1}^{10} x_i y_i = 2161.2; \sum_{i=1}^{10} x_i^2 = 238.77.$$

(a) (4 marks) Determine the least squares regression line with Y as the dependent variable.

(b) (4 marks) Test the significance of the regression model in (a) at the 5% significance level.

State the two hypotheses, the test statistic, the test assumptions if any, the p-value and the conclusion. If an exact p-value cannot be found from the table, state the range of values from the table.

(c) (5 marks) Determine a 95% *prediction* interval for the gross revenue of a movie if 8 millions were paid to the two highest-paid actors/actresses in the movie.

(d) (2 marks) Show that the sum of residuals $\sum_{i=1}^{n} \hat{e}_i = \sum_{i=1}^{n}(y_i - \hat{y}_i)$ is always 0.

4. In a certain district, 100 families each with 4 members are selected. The numbers of families with $i = 0$ to 4 members having blood type A are counted and are denoted by O_i. The results are summarized in the following table:

No. of members having blood type A (i)	0	1	2	3	4
No. of families (O_i)	22	32	20	16	10

(a) (1 mark) Assume that the 100 counts i follow a binomial distribution $B(4, \pi)$. Find a sample estimate of π using the following formula:

$$\hat{\pi} = \frac{1}{400} \sum_{i=1}^{4} i \times O_i.$$

(b) (5 marks) Complete the following table *in your answer book* to calculate the expected frequencies and hence the chi-square test statistic under the condition that $E_i \geq 5$.

i	O_i	$c_i = \sum_{j=0}^{i} p_j$	$p_i = c_i - c_{i-1}$	$E_i = 100\, p_i$
0	22			
1	32			
2	20			
3	16			
4	10	1.0000		
Sum	100		1.0000	100

i	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
0			
1			
2			
≥ 3			
Sum	100	100	

The cumulative probabilities $c_i = \sum_{j=0}^{i} p_j$ where $p_i = C_i^4\, \hat{\pi}^i (1 - \hat{\pi})^{4-i},\ i = 0, \ldots, 4$ for the count i can be obtained from a suitable column of $\hat{\pi}$ in the binomial table with $n = 4$.

(c) (3 marks) Test at the 5% significance level if the 100 counts i follow a binomial distribution.

State the two hypotheses, the test statistic, the p-value and the conclusion. If an exact p-value cannot be found from the table, state the range of values from the table.

************ This is the Last Page ************