MAST20005/MAST90058: Assignment 3

Due date: 11am, Friday 18 October 2019

Instructions: Questions labelled with '**(R)**' require use of R. Please provide appropriate R commands and their output, along with sufficient explanation and interpretation of the output to demonstrate your understanding. **Such R output should be presented in an integrated form together with your explanations; do not attach them as separate sheets.** All other questions should be completed without reference to any R commands or output, except for looking up quantiles of distributions where necessary. Make sure you give enough explanation so your tutor can follow your reasoning if you happen to make a mistake. Please also try to be as succinct as possible. Each assignment will include marks for good presentation and for attempting all problems.

Problems:

1. **(R)** A study measured the weight gain after 8 weeks of two groups of mice. The first group was fed a high-protein diet, and the second group a low-protein diet. The data are given below:

High-protein:	134	146	104	119	124	161	112	83	113	129	97	123
Low-protein:	70	118	101	85	107	132	94					

   (a) Use the sign test with $\alpha = 0.05$ to test if the median weight gain in the first group is smaller than 110. Clearly state your hypotheses.

   (b) Use the Wilcoxon rank-sum test with $\alpha = 0.05$ to test if the median weight gain in the first group is larger than that in the second group. Clearly state your hypotheses.

   (c) Use a t-test with $\alpha = 0.05$ to test if the mean weight gain in the first group is larger than that in the second group. Clearly state your hypotheses.

2. **(R)** A survey was conducted that asked people of different ages how much they exercised (in hours per week). The responses received were as follows:

	0 hours	1 hour	2 hours	3 hours	4+ hours
Younger than 40 years	10	24	10	6	3
40 years or older	7	22	18	10	5

   Using a significance level of $\alpha = 0.05$, test whether:

   (a) The hours of exercise for the younger age group follows a Poisson distribution.

   (b) Age is independent of hours of exercise.

3. Let X have a shifted exponential distribution with pdf,

$$f(x) = e^{-(x-\theta)}, \quad x \geqslant \theta.$$

Suppose we have a random sample of n observations on X.

(a) Find the cdf of the sample minimum, $X_{(1)}$.

(b) Find the p quantile, π_p.

(c) Find the asymptotic variance of the sample median, $\hat{M}$.

4. **(MAST20005 students only)** Consider the one-way analysis of variance model,

$$X_{ij} = \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

where $\epsilon_{ij} \sim \mathrm{N}(0, \sigma_j^2)$ are independent but not identically distributed (σ_j^2 varies by j). Let,

$$\bar{X}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} X_{ij}.$$

(a) Find the distribution of $\bar{X}_{i\cdot}$.

(b) Find $\mathbb{E}\left\{\sum_{j=1}^{n} (X_{ij} - \bar{X}_{i\cdot})^2\right\}$.

5. **(MAST90058 students only) (R)** Retail sales are affected by store locations and the number of competitiors nearby. In a study, three different type of locations are considered: outer surburb, inner surburb, and CBD. For each type of location and number of competitiors, retail sales of three stores were reported (in thousands of dollars). The data obtained were:

	Number of competitors			
Locations	3	2	1	0
Outer surburb	270	290	446	440
	310	350	487	428
	220	305	500	530
Inner surburb	410	382	598	470
	305	320	480	415
	450	380	510	400
CBD	180	220	290	246
	290	170	283	275
	330	260	260	330

Perform a two-way analysis of variance to examine whether these data suggest that retail sales are affected by store locations. State and test appropriate hypotheses at a 5% significance level. You should report the value of the appropriate statistic, the p-value, the assumptions you have made and your conclusions. Is it possible to test for interaction? If yes, then perform the test and draw an interaction plot; otherwise, explain why it is not possible.