MAST20005/MAST90058: Assignment 3

Due date: 11am, Sunday 25 October 2020

Instructions: Please submit your assignment via the LMS, ensuring that you follow the submission instructions provided online. **Remember to submit ON TIME, since late submission will receive zero points. Do not wait until the last minute!** We suggest that you submit your assignment promptly once you finish all questions. You can always re-submit your assignment before the deadline if you make any changes.

If for any reason you think you will not be able to submit on time, you need to notify the Subject coordinator Tingjin Chu in a timely manner (as soon as you become aware of any issue and preferably prior to the deadline). In general, a medical certificate is required. Note that extensions are only granted in exceptional circumstances and only for a very limited time period.

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.** 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.

Problems:

1. **(R)** The daily new coronavirus cases in Victoria between Sep 7th and Sep 27th are recorded as

48	70	47	40	35	41	30
39	41	25	44	20	13	11
28	14	11	13	12	16	5

where 48 is the number of new coronavirus cases on Sep 7th, 70 is the number of new coronavirus cases on Sep 8th, and so on.

   (a) Use the sign test with $\alpha = 0.05$ to test if the median of daily new coronavirus cases is below 15 between Sep 7th and Sep 27th. Clearly state your hypotheses.
   (For the purpose of this question, treat these data as a random sample from a static population, which is an unrealistic assumption in practice.)

   (b) Use the Wilcoxon rank-sum test with $\alpha = 0.05$ to test if the median number of daily new coronavirus cases in the second week (the second row) is higher than the median number of daily new coronavirus cases in the third week (the third row). Clearly state your hypotheses.
   (For the purpose of this question, treat the data from each week as a random sample from a static population, which is an unrealistic assumption in practice.)

2. Let X be an exponential distribution with pdf,

$$f(x) = \lambda e^{-\lambda x}, \quad x \geqslant 0.$$

Suppose we observe the following random sample of $n = 30$:

0.11	0.21	0.75	1.14	1.35	1.63	1.63	1.83	1.93	2.04
2.16	2.25	2.41	2.52	2.65	2.83	2.92	2.92	4.83	7.23
8.80	9.80	11.54	12.16	12.91	13.93	19.68	20.94	21.73	24.09

   (a) Find the p quantile, π_p.

   (b) Calculate the 'Type 7' sample quantile $\hat{\pi}_{0.25}$.

   (c) Find the asymptotic distribution of $\hat{\pi}_{0.25}$.

   (d) Calculate a standard error for $\hat{\pi}_{0.25}$.

3. Let $X_1, \ldots, X_n$ be a random sample with pdf

$$f(x) = \beta^2 x e^{-\beta x}, \quad x \geqslant 0,$$

and for β use the prior distribution $f(\beta) = e^{-\beta}$ with $\beta \geqslant 0$.

   (a) Derive the posterior distribution of β.

   (b) Derive the posterior mean and the posterior standard deviation of β.

4. Consider a random sample $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$.

   (a) If μ is unknown and σ^2 is known, find a sufficient statistic for μ.

   (b) If μ is known and σ^2 is unknown, find a sufficient statistic for σ^2.

   (c) If μ is known and σ^2 is unknown, find a sufficient statistic for σ.

5. **(MAST20005 students only)** Let $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$ be a random sample from an exponential distribution with mean $1/\lambda$. We are interested in testing $H_0\colon \lambda = \lambda_0$ versus $H_1\colon \lambda \neq \lambda_0$.

   (a) Derive the likelihood ratio test and show it is based on the statistic $Y = \sum_{i=1}^{n} X_i$.

   (b) What is the distribution of Y when H_0 is true?

   (c) For $n = 50$ and $\lambda_0 = 1$, find a test based on Y with significance level 0.05.

6. **(MAST90058 students only) (R)** Pedestrian numbers are recorded over four different locations and four different time slots over three days, as shown in the table below.

	Locations			
Time slot	Flagstaff Station	Melbourne Central	Town Hall	Bourke Street Mall
1pm – 2pm	1706	2387	3715	3715
	1636	2284	3541	3689
	1339	2116	3369	2884
2pm – 3pm	1063	2062	3209	2940
	1065	1885	2907	2753
	977	1819	3077	2525
3pm – 4pm	1380	2108	3030	2751
	1306	1896	2837	2508
	1261	1893	2978	2288
4pm – 5pm	2539	1980	2964	2687
	2544	2025	2824	2423
	2297	2064	2987	2429

Perform a two-way analysis of variance to examine whether these data suggest that pedestrian numbers vary by time. 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.