MAST20005/MAST90058: Assignment 2

Due date: 11am, Friday 25 September 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)** Let p_1 be the proportion of residents who support more bicycle lanes in the City of Yarra and p_2 be the proportion in the City of Moreland. Respective random samples of residents from each city, of size $n_1 = 800$ and $n_2 = 1000$, gave $y_1 = 520$ and $y_2 = 600$ respondents who support more bicycle lanes. Is there evidence that the proportions differ between the two cities? Set this up as a hypothesis test.

   (a) State appropriate null and alternative hypotheses.

   (b) Carry out a test that has significance level $\alpha = 0.05$. What is your conclusion?

   (c) Give a 95% confidence interval for the difference in proportions between the two cities.

2. Tom wants to know how long (in minutes) it takes to walk from the State Library to the Peter Hall Building. Over the past ten days, Tom records the time (in minutes) needed for him to finish the trip, as follows:

   ```
   12.1  12.2  17.4  13.1  17.8  19.8  13.0  10.8  18.4  16.0
   ```

   Assuming a normal distribution $\mathrm{N}(\mu_X, \sigma_X^2)$ for these observations, please answer the following:

   (a) If σ_X is unknown, calculate a 95% confidence interval for the mean.

   (b) Assume $\sigma_X = 3$ minutes and that you want a 95% confidence interval of width 2 (i.e. ± 1). How many experiments are needed?

Tom also has a record, from a month ago, of the time for him to walk from Queen Victoria Market to the Peter Hall Bulding, as follows:

```
20.1   21.3   20.4   21.7   20.3   19.5   19.4   19.9
```

Assuming a normal distribution $\mathrm{N}(\mu_Y, \sigma_Y^2)$ for these observations, please answer the following:

(c) Calculate a 95% confidence interval for $\mu_X - \mu_Y$.

(d) Calculate a 95% confidence interval for σ_X^2/σ_Y^2.

(e) **(R)** Calculate a 95% confidence interval for σ_X^2/σ_Y^2, using the R function `var.test()`.

3. **(R)** This question refers to the data in the file `coffee.txt` from a coffee shop. It shows the sales (in dollars) and the number of customers each day for twenty days.

   (a) Fit a simple linear regression model for the sales given the number of customers. State point estimates for all parameters.

   (b) Find 95% confidence intervals for the regression coefficients.

   (c) Give a 95% confidence interval for the mean sales if the number of customers is 100.

   (d) Give a 95% prediction interval for the sales if the number of customers is 100.

   (e) Are the usual regression model assumptions appropriate?
   Answer by reporting two appropriate diagnostic plots and commenting on them.

4. Let $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$, where both μ and σ^2 are unknown. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ be the usual sample mean and sample variance. State whether the following quantities are pivots, giving your reasoning for each one.

   (a) $T_1 = \bar{X} - \mu$

   (b) $T_2 = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

   (c) $T_3 = \frac{\bar{X}-\mu}{S/\sqrt{n}}$

   (d) $T_4 = \frac{\bar{X}-\mu}{S}$

5. Consider an exponential random variable X with pdf

$$f(x \mid \theta) = \theta^{-1} e^{-x/\theta}, \quad x > 0.$$

   A single observation of such a variable is used to test $H_0\colon \theta = 2$ against $H_1\colon \theta = 5$. The null hypothesis is rejected if the observed value is greater than 4.

   (a) What is the probability of committing a Type I error?

   (b) What is the probability of committing a Type II error?

   (c) What is the power of the test?

   (d) Find a test (i.e. determine a test statistic and critical region) of these hypotheses that has significance level 0.05.