MAST20005/MAST90058: Assignment 2

Due date: 11am, Friday 20 September 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. Suppose that you want to know how long (in hours) it takes for a particular brand of paint to dry. Nine experiments are done and the times were measured as follows:

   ```
   6.0  5.7  5.8  6.5  7.0  6.3  5.6  6.1  5.0
   ```

   Assume these times follow a normal distribution, $\mathrm{N}(\mu, \sigma^2)$.

   (a) Assuming $\sigma = 0.6$ based on previous experience, calculate a 95% CI for μ.

   (b) Still assuming $\sigma = 0.6$, suppose we want our estimate of μ to be within 0.2 with probability near 95%. How many experiments do we need to run?

   (c) **(R)** If σ is unknown, calculate a 95% CI for μ. Compare the width of CIs in part (a) and (c).

2. An assembly line has a target of achieving an 80% success rate when making bicycles. Long experience shows that they are never more than 10% away from that target. What sample size is required for estimating the success rate:

   (a) To within 5% with probability 0.95?

   (b) To within 2% with probability 0.95?

3. **(R)** The `pressure` dataset is available in a standard installation of R. You should be able to access it directly, for example via:

   ```
   > pres <- pressure$pressure
   > temp <- pressure$temperature
   ```

   The dataset describes the relationship between temperature in degrees Celsius and the vapor pressure of mercury in millimeters (of mercury). According to the Antoine Equation, they have the relationship:

   $$\log_{10}(\text{pressure}) = \alpha + \frac{\beta}{\text{temperature} - 10},$$

   for some constants a and b, and where temperature in the formula is measured on the Kelvin scale (K). The relationship between temperature in Kelvin (T_K) and temperature in Celsius (T_C) is $T_K = T_C + 273.15$.

(a) Suppose we want to estimate the constants by fitting the simple linear regression model,

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

Define y and x appropriately for this model to work.

(b) Fit the above model and find estimates of α and β.

(c) Assess the model fit visually using some standard diagnostic plots. Does the linear model seem appropriate?

(d) Give 95% confidence intervals for the regression coefficients. According to other sources, $\alpha = 4.86$ and $\beta = -3007$. Does your model support these two claims?

(e) Give a 95% confidence interval for the mean vapor pressure when the temperature is 70 degrees Celsius.

(f) Give a 95% prediction interval for the vapor pressure when tempeature is 70 degrees Celsius.

4. Two different types of high-speed train are manufactured. We are interested in whether there is a difference in their maximum speed. The following table summarises the maximum train speed in kilometers per hour, measured over a series of tests.

Train type	Sample size	Mean	Standard deviation
Type A	10	500	1.1
Type B	20	496	1.2

Assuming the maximum speed of both types of train follows a normal distribution, determine whether a type A train is faster than a type B train when they are running at maximum speed.

5. It was claimed that 80% of users on a particular website are male. In a random sample of 200 users, 146 of them were male. Is there evidence that the proportion p of users that are male differs from 0.8?

(a) State appropriate null and alternate hypotheses.

(b) What would you conclude if the significance level is $\alpha = 0.05$?

(c) What would you conclude if the significance level is $\alpha = 0.01$?

(d) Give a 95% confidence interval for the proportion of users that are male.

6. **(R)** Consider a Poisson random variable X with pmf

$$\Pr(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

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

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

(b) What is the probability of committing a Type II error when $\lambda = 5$ in H_1?

(c) Draw a power curve for this test for alternative values of λ between 2 and 10.

(d) Find a test of these hypotheses that has an approximate significance level of 0.05. What is the *actual* significance level of your test?