MAST20005/MAST90058: Assignment 2

Due date: 11am, Friday 17 September 2021

Instructions: See the LMS for the full instructions, including the submission policy and how to submit your assignment. Remember to submit early and often: multiple submission are allowed, we will only mark your final one. Late submissions will receive **zero** marks.

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:

   ```
   8.2  7.6  7.8  8.5  9.1  8.3  7.6  8.1  6.8
   ```

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

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

   (b) Still assuming $\sigma = 0.6$, suppose we want a 95% confidence interval (for μ) that has width at most 0.2. How many experiments do we need to run?

   (c) If σ is unknown, calculate a 95% confidence interval for μ. Comment on how this compares with the confidence interval from part (a).

2. An assembly line has a target of achieving an 90% 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 using each of the following:

   (a) A 95% confidence interval that is $\pm 5\%$?

   (b) A 95% confidence interval that is $\pm 2\%$?

3. **(R)** Enter the following command in R to access the `Animals` dataset:

   ```
   > data(Animals, package = "MASS")
   ```

   (Note that this requires the `MASS` package; if you don't have it yet, install it first.) You will now have a variable called `Animals` in your R session which is a data frame with measurements of the average body weight (kg) and average brain weight (g) of several animals. We wish to fit a simple linear regression model to relate these two measurements, with brain weight as the response variable and body weight as the predictor variable.

   (a) The raw data are unsuitable for fitting this model. Why?

   (b) Take the logarithm of all of the measurements and fit the model. Show an appropriate summary of the model fit.

   (c) Explore the model fit via relevant diagnostic plots. What do you notice?

   (d) Omit three animals that are clearly different from the rest and refit the regression model. Show an appropriate summary of the model fit.

   (e) Show a plot of the data together with the new regression line.

   (f) Give a 95% confidence interval for the average brain weight of camels that weigh on average 500 kg.

4. Plants convert CO_2 in the atmosphere, along with water and energy from sunlight, into the energy they need for growth and reproduction. Experiments were performed with normal air atmospheric conditions and those with enriched CO_2 concentrations to determine the effect on plant growth. The plants were given the same amount of water and light for a four-week period. The following table summarises the data for the plant growth, in grams. You may assume a normal distribution for each group of observations.

Condition	**Sample size**	**Mean**	**Standard deviation**
Enriched air	8	8.21	1.610
Normal air	12	7.36	0.956

Based on these data, does a CO_2-enriched atmosphere increase plant growth?
Justify your answer by calcluating relevant 95% confidence interval(s).

5. **(R)** Let p_1 be the proportion of babies with low birth weight (below 2.5 kg) in Africa and p_2 be the proportion in the Americas. Respective random samples from each continent, of size $n_1 = 800$ and $n_2 = 600$, gave $y_1 = 120$ and $y_2 = 60$ babies with a low birth weight. Is there evidence that the rates differ between the two continents? Set this up as a hypothesis test.

   (a) State appropriate null and alternate hypotheses.

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

   (c) What would be your decision if $\alpha = 0.01$?

   (d) Give a 95% confidence interval for the difference in rates.

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

$$\Pr(X = x \mid p) = p(1-p)^x, \quad x = 0, 1, 2, \ldots$$

   A single observation of such a variable is used to test $H_0\colon p = 0.4$ against $H_1\colon p = 0.2$. 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?

   (c) Draw a power curve for this test for all possible alternative values of p (not just 0.2).

   (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?

   Hint: You'll need to find out how to work with the geometric distribution in R. The lab notes from week 3 should be helpful.