STA305/1004 (L0101/L0201) Winter 2017 - Homework 1 - Solutions

Due: Electronic submission on UofT Learning Portal course page by Friday, January 27, 2017 at 22:00. NB: e-mail submissions will NOT be accepted.

STA305 students are not required to complete the bonus question. If the bonus question is completed then it will be worth extra marks.

If you work with other students on this assignment then:

- indicate the names of the students on your solutions;
- your solutions must be written up independently (i.e., your solutions should not be the same as another students solutions).

1. A scientist has two light objects to weigh. She decides to use an old fashioned pan balance scale in the lab since she heard about a design from a colleague that is supposed to increase the accuracy of her measurements and take less time.

   The scientist decides to obtain weight measurements using the following design (**DESIGN I**):

   - weigh the two objects together in one pan;
   - weigh one object in one pan, and the other object in the other pan;
   - weigh one object alone.

   In a pan balance scale when one object is in one pan and another object is in another pan the measurement obtained is the difference in weight between the two objects.

   Let y_1, y_2, y_3 be the readings from the scale, and β_1 the weight of the object that has been on the scale through all three weighings and β_2 the other object. The standard deviation of each weighing is denoted by σ.

   Answer the following questions.

   (a) Write three equations relating the observed weights y_1, y_2, y_3 to the unknown weights β_1, β_2. Make sure to include an appropriate error term and any necessary assumptions about the error term.

   $$\begin{aligned} y_1 &= \beta_1 + \beta_2 + \epsilon_1 \\ y_2 &= \beta_1 - \beta_2 + \epsilon_1 \\ y_3 &= \beta_1 + \epsilon_1, \end{aligned}$$

   where $Var(\epsilon_i) = \sigma^2, i = 1, 2, 3$.

   (b) Find the least-squares estimates of β_1, β_2.

   (c) Find the standard error of the least-squares estimates of β_1, β_2.

   (d) If the scientist measured each object three times could she achieve the same precision (standard error) as this design? Explain.

   (e) (**DESIGN II**) Suppose that instead of the design above the scientist uses the following design.

   - weigh both objects in one pan together twice;
   - weigh the objects in opposite pans.

   Question: Find the least-squares estimates of the weights and standard error of the weights using this design.

   (f) Does **DESIGN II** determine the weights of the objects with equal precison compared to **DESIGN I**? Explain your reasoning.

2. Ten thousand users of an e-commerce web site were randomly allocated to see if a new version of the website (B) would lead to increased sales compared to the existing version of the web page (A). The data is in the file `ABtest.csv`. The column labeled *page* indicates which website the user viewed and the column labelled *sales* indicates sales in Canadian dollars.

   Answer the following questions using R.

   (a) How many values does the randomization distribution of the difference in median sales between the two web pages contain?

   (b) Create a histogram of this randomization distribution. What is the randomization P-value of your test? **(Hand in your R code and output for this part)**

   (c) Is there evidence of a significant difference in sales? Explain your answer including how you define 'significant'. **(Hand in your R code and output for this part)**

   (d) In order to fully implement the new version of the web page it will cost approximately $5000. In your opinion should the company implement the new version of the web page? Explain your reasoning.

3. Six equal size plots of land will be divided into two subplots. Two fertilizers, F and G, will be randomly assigned to each subplot by flipping a fair coin. The table below shows the yield in bushels per acre after the fertilizers were administered.

Plot	Subplot	Yield	Fertilizer
1	2	78	F
2	1	82	F
3	2	82	F
4	1	65	F
5	1	51	F
6	2	75	F
1	1	72	G
2	2	70	G
3	1	55	G
4	2	85	G
5	2	59	G
6	1	80	G

   (a) What type of design was used in this study? Explain.

   (b) What is the probability that a subplot will receive Fertilizer G? Explain.

   (c) What is the probability of this treatment allocation? Would you have been surprised if all the subplots 1 were assigned to, say, fertilizer F? Explain.

   (d) Describe the randomization distribution for this comparison. How many values does this distribution contain?

   (e) Use the randomization test to determine if there is evidence of a difference in yield between the two fertilizers. **(Hand in your R code and output for this part)**

4. Use the data in the previous question to conduct an appropriate t-test. **(Hand in your R code and output for this part)**

   (a) Are the assumptions behind the t-test satisfied?

   (b) Do the results of the t-test agree with the results of the randomization test? Explain.