

STA 312f22 Assignment Four¹

Please bring hard copy of your complete R input and output from Question 9a to the quiz. The non-computer questions are practice for the quiz on Friday Oct. 21st, and are not to be handed in.

1. This is a better version of a question from Assignment 3. Based on a random sample of size n from a Bernoulli, you want to test $H_0 : \pi = \pi_0$.
 - (a) Write the formula for the likelihood ratio statistic G^2 as a function of n , p and π_0 .
 - (b) Write the formula for the Pearson chi-squared statistic X^2 as a function of n , p and π_0 . Keep simplifying until it equals Z_1^2 , where Z_1 is given on the formula sheet.
2. What do you call a study in which the values of the X variable are assigned or selected, and then the values of the Y variable are observed later?
 - (a) Retrospective
 - (b) Prospective
 - (c) Introspective
 - (d) Cross-sectional
 - (e) Multi-sectional
3. What do you call a study in which the X and Y values are just observed, with no assignment or selection?
 - (a) Retrospective
 - (b) Prospective
 - (c) Introspective
 - (d) Cross-sectional
 - (e) Multi-sectional
4. What do you call a study in which cases (there are n cases) are selected based on their Y values, and then X is observed?
 - (a) Retrospective
 - (b) Prospective
 - (c) Introspective
 - (d) Cross-sectional
 - (e) Multi-sectional

¹Copyright information is at the end of the last page.

5. For a prospective design, which marginal totals are fixed by the design? Assume rows correspond to X and columns correspond to Y .
 - (a) Row totals
 - (b) Column totals
 - (c) Both row and column totals
 - (d) Neither row nor column totals
6. For a retrospective design, which marginal totals are fixed by the design? Assume rows correspond to X and columns correspond to Y .
 - (a) Row totals
 - (b) Column totals
 - (c) Both row and column totals
 - (d) Neither row nor column totals
7. For a cross-sectional design, which marginal totals are fixed by the design? Assume rows correspond to X and columns correspond to Y .
 - (a) Row totals
 - (b) Column totals
 - (c) Both row and column totals
 - (d) Neither row nor column totals
8. Do Problem 2.14 on p. 58 of the text. What homework problem from Assignment 3 shows you how to test the difference between those 2 percentages? What one number is missing?
9. Do Problem 2.16, parts a through c . Look at how many smokers there were in those days! Part b is asking whether the study is prospective, retrospective or cross-sectional.
 - (a) Carry out a likelihood ratio test for association of smoking and lung cancer. Use R, and do it the easiest way you can. Display the expected frequencies. Your printout should also include G^2 , the degrees of freedom, and the p -value. Do you reject H_0 at $\alpha = 0.05$? In plain, non-statistical language, what do you conclude?
 - (b) Suppose that in this population the overall probability of lung cancer is one percent. Using this information and the data in the table, estimate the probability of lung cancer given that the person is a smoker.
10. Do Problems 2.17, 2.18 and 2.21. For each question, state whether the study is prospective, retrospective or cross-sectional.

11. For each of five Instagram models, a random sample of 200 followers is selected, and they are classified as Female, Male or Other.
- (a) What kind of design is this?
 - (b) What is I ? What is J ?
 - (c) How many free parameters are there in the unrestricted model?
 - (d) Make a 3 by 5 table and write probabilities (parameters or functions of the parameters) in the cells. Call the models A, B, C, D and E.
 - (e) We want to test whether the models attract different gender profiles. What is the null hypothesis?
 - (f) Does the number of equals signs in your null hypothesis equal $(I - 1)(J - 1)$?
 - (g) How many free parameters are there in the restricted model?
 - (h) Write the restricted likelihood. Do not hesitate to use the notation for marginal totals.
 - (i) Based on your restricted likelihood, give the estimated expected number of male followers of Instagram model B. Is your answer on the formula sheet?

Please bring hard copy of your complete R input and output from Question 9a to the quiz.

This assignment was prepared by [Jerry Brunner](#), Department of Statistics, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/312f22>