Mathematics for Policy and Planning Science

Stephen Turnbull

Graduate School of Systems and Information

Homework 4: Due May 25, 2020 at 12:00 noon

Read and understand the following instructions on submission of homework. **If you do not follow them, you will not receive credit.**

Submit this assignment by *email*. Give the mail the subject "01CN101 Homework #<number> by <your name>" in *hankaku romaji* and send it to `turnbull@sk.tsukuba.ac.jp`. (This subject is necessary for automatically sorting incoming mail.) It should look like this:

```
Subject: 01CN101 Homework #1 by Stephen Turnbull
```

for Homework #1.

Make sure that the body of the email contains your *name* and *student ID number*.

If you are late, submit the assignment for partial credit. The later, the less credit you will receive. If you believe that the late submission is in part due to lack of care by the instructor, or some event (such as hospitalization) required your full attention for two full days or more, you may explain for additional credit. Otherwise, I don't care why it was late.

Submit this assignment as a *plain text* e-mail. This homework requires no special symbols or wordprocessor features. (*Do not* attach a wordprocessor file such as Microsoft Word document or a Excel spreadsheet. *You will lose some credit!*)

Homework 4

Due Monday, May 25, 2020 at 12:00 noon.

1. Explain each of the following in 1–2 lines.

   (a) What is the purpose of a statistic like the mean when used as a *descriptive statistic*?

   (b) What is the purpose of a statistic like the mean when used as an *inferential statistic*?

   (c) When conducting statistical inference, what is the purpose of the *domain model*?

   (d) When conducting statistical inference, what is the purpose of the *statistical model*?

   (e) Why do statistician prefer *randomized assignment* of individuals to treatment groups in experiments?

   (f) Why do we need a *null hypothesis* in statistical tests?

   (g) What is a *false positive* (Type I error) in statistical testing?

   (h) What is a *false negative* (Type II error) in statistical testing?

   (i) In most research using a statistical hypothesis test, does the research want to *accept* or *reject* the null hypothesis? Explain why.

2. Briefly explain how and when *regression models* can be used in studying a cause-and-effect relationship.

3. In the Anscombe data, we can see "problems" of the linear model by looking at the *x-y plot* of the data compared to the *graph* of the regression line. But suppose instead of one explanatory variable x you had 10 explanatory variables $x_1, \ldots, x_{10}$. Explain briefly how you might detect some problems using the residuals instead of an (impossible!) 11-dimensional plot.

4. In the Anscombe data, it's obvious how to order the x-y or residual plot from left to right: use the x variable in increasing order. What orders could you use if there are 10 explanatory variables as in Problem 3? Explain why.

5. Explain how the *base rate* or *prior distribution* of taxicab colors affects the court's estimate of the probability that the taxicab in the accident was a Blue taxicab. Briefly explain what the *base rate* is.

6. In Bayesian analysis, explain each of

   (a) *informative prior* distribution

   (b) *weakly informative prior* distribution

   (c) *uninformative prior* distribution

   in one or two lines.

7. As a manager in a fairly large corporation, you are interviewing a student with a master degree in Policy and Planning Science. This position requires a master degree in a business field. Before interviewing, you consider the likelihood that this candidate will be in the bottom third, middle third, or top third of candidates you will interview. In Bayesian analysis, what do you call this set of likelihoods?

8. Now consider your prior distribution about the performance of the candidate from Problem 7 as a member of your team. Would you say your prior distribution is *informative*, *weakly informative*, or *uninformative*? Explain briefly.

9. Explain the significance of the *Bayes factor* in evaluating the contribution of an experiment to knowledge about the theory the experiment is trying to test.