Final assessment

Final assessment: stat601/stat805/comp616

This assessment requires you to use R to demonstrate your understanding of probability and statistics. There are four questions, each of 25 marks.

- Answer each question and show your working, including your R code.
- Submit in the form of a PDF document. We cannot assess any other type of submission.
- Study tips: review the notes, labs, and course manual
- Academic Integrity: This is an individual assessment, and standard requirements around academic integrity apply. Refer to the "Academic Integrity" module for further details.
- Submission: submit as for any other assignment

Question 1

The temperature of a certain city during the year can be modelled reasonable well with a Normal distribution with mean 25 degrees Celsius and standard deviation 4 degrees Celsius.

- What is the probability that in a random day the temperature is at most 21C?
- Verify your calculation via simulation.
- What is the probability that in a random day the temperature is at least 28C?
- Verify your calculation via simulation.
- What is the probability that in a random day the probability is between 28C and 32C?
- Verify your calculation via simulation.

[excellence] We choose 14 days randomly (assume independence between the days).

- What is the probability that the temperature is greater than 30C in at least five days, chosen randomly from the 14?
- Verify the probability through simulation.
- How many days with a temperature greater than 30C would you expect? Verify your answer via simulation.

Question 2

A vetinarian is investigating a particular disease of cows. She is interested in whether English cows differ from Scottish cows in their probability of suffering from this particular disease. A sample of 15 cows is taken, of which 9 are English and 6 are Scottish. It is not known if English cows are more or less likely to have this disease. In the sample, there are 10 healthy cows, of which 8 are English. We may represent the data in R as follows:

- State a sensible null hypothesis
- State the precise definition of p-value and explain what "more extreme'' means in this context
- Is a one-sided or two-sided test needed? justify
- Perform a Fisher test using `fisher.test()` and interpret
- Estimate the probability of: (a), an English, and (b), a Scottish cow, having the disease

- give a 95% Gaussian confidence interval for the probabilities of English and Scottish cows having the disease (use the normal approximation)
- excellence: verify your p-value given by `fisher.test()` using `dhyper()` directly.

Question 3

A computer scientist in Auckland has a large number of computers in his laboratory and suspects that more failures occur when the temperature outside is higher.

He collects data on 23 days:

```
temp <- c(17.1, 19.2, 15.2, 18.1, 15, 17.8, 17.8, 15.2, 17.4, 15.7, 17.8,
16.9, 17.1, 17, 18, 15.9, 16.5, 17.3, 16.7, 15.9, 16.1, 19, 16)
fail <- c(28, 30, 27, 33, 27, 34, 31, 29, 29, 26, 33, 31, 27, 31, 32,
28, 30, 33, 32, 26, 30, 34, 27)
```

Above, `temp` is the temperature in centigrade, and `fail` the number of failures on each day.

- State a sensible null hypothesis
- State the precise definition of p-value and explain what "more extreme'' means in this context
- Is a one-sided or two-sided test needed? justify
- Perform a linear regression using R and interpret. Is there evidence that more computer failures occur during hotter weather?

Question 4

Around 37 degrees Centigrade is normal body temperature for children and adults. *Fever* is when your core body temperature is higher than 38 degrees centigrade. A certain medicine that aims to reduce body temperature has been applied to 80 patients that had a fever. The temperatures after the medicine was administrated are the following:

```
y  <-  scan( text = "
37.3 37.5 37.2 38.0 37.6 37.2 37.6 37.7 37.7 37.4 37.9 37.6 37.3
36.8 37.8 37.5 37.5 37.8 37.7 37.7 37.8 37.7 37.5 36.9 37.7 37.5
37.4 37.0 37.3 37.6 37.9 37.5 37.6 37.5 37.1 37.4 37.4 37.5 37.8
37.7 37.4 37.4 37.7 37.7 37.3 37.3 37.6 37.7 37.5 37.8 37.6 37.3
37.6 37.2 37.9 38.1 37.4 37.2 37.7 37.4 38.2 37.5 37.7 37.5 37.3
37.5 36.9 37.9 37.5 38.1 37.6 37.3 37.7 37.2 37.1 37.6 37.4 37.5
37.5 37.3")
```

We consider that the medicine is *effective* if the temperature is lower or equal to 38. Perform a Bayesian analysis assuming a skeptical position: we seek the probability that the medicine is effective.

- Define the random variable and its distribution.
- Define the prior and justify your choice.
- Define the posterior distribution.
- Plot the prior and posterior in a graph. Comment on it.
- Calculate a 95% credible interval and interpret it.