

STAT2402: Week 3 Computer Laboratory

Getting Started:

Log in at one of the PCs and start up the software package RStudio. If you have problems either logging in or starting RStudio, ask the lab demonstrator for help.

When typing in R commands in the console window you can use the arrow keys to speed things up. The 'up' arrow gives you the previous command that you typed. The usual prompt sign for R is `>`. If you get a `+` prompt sign instead, it means that R is awaiting the completion of the previous command that you typed in. This can happen because you have forgotten to close parentheses, for instance. Just type in the remainder of the command and press enter.

You can ask the lab demonstrator for help at any point when you have a problem.

This laboratory session covers the following topics:

1. Revision of linear models.
 2. Probability.
 3. Random variables.
-

Exercise 1: Linear models We will work on the weight loss clinical trial discussed in lectures this week. A randomised trial will be conducted to investigate the effect of exercise and diet on adults. Each healthy subject will be randomly allocated to an exercise regime and a diet level. The following variables will be collected.

- Exercise: Exercise regime, Yes or No
 - Diet: Yes or No
 - Sex: M or F
 - Age: continuous
 - Height: in metres
 - Weight0: Initial Weight (kg), continuous
 - Weight1: Weight (kg) after twelve weeks, continuous
1. An important consideration for trials is sample size. The calculation is based on *power* of the statistical test for a difference in weight loss due to the treatments (exercise and diet). We will use the R package `pwr` and the function `pwr.f2.test` to compute a sample size. Look at the document Sample Size Calculation on LMS in the folder Computer Labs.
 - (a) You need to specify u , the numerator degrees of freedom. Note that u is simply the number of parameters in your model. You can determine this by writing your model equation.
 - (b) Compute sample size for powers 0.85 to 0.95 in steps of 0.05.
 - (c) You will also need to select an effect size f_2 . Determine the effect of this on sample size by taking values between 0.3 and 0.8 in steps of 0.1

The function gives you the value of v in the output. The required sample size is $u + v + 1$, rounded up to the next integer.

2. Take a sample size of 80, and produce a random allocation of subjects to the treatments.
3. The data for this trial is available in the file `Weight.txt`. Analyse the data and report on your findings. You should include in your analysis appropriate data exploration and model checking

Extra

On further data exploration, it appears that the minimum weight is 25.32 kg for 30 year old male in the No diet and No Exercise group. This person has a weight loss of 1.93 kg. This data record is not credible. Other similar checks for data should be conducted, not just in this example but routinely.

Exercise 2: Binomial Problems

1. On Pingelap Island, 10% of the population is colour blind. A researcher selects 50 people at random from the island. Let the random variable X denote the number of people, out of the 50, who are colour blind.
 - (a) State the distribution of the random variable X .
 - (b) Determine the probability that of these 50 people, exactly 7 are colour blind.
 - (c) Determine also the following:
 - i. $P(X = 4)$;
 - ii. $P(X \leq 6)$.
 - (d) What are the mean and variance of the random variable X ?
 2. Let X be the number of heads from 10 tosses of a fair coin. Evaluate the following probabilities:
 - (a) $P(X = 5)$;
 - (b) $P(X > 7)$;
 - (c) $P(3 \leq X \leq 8)$.
 3. A botanist researching flower bulbs knows that 90% of its bulbs will flower. They are sold in packets of 12 randomly selected bulbs with a guarantee that the packet will be replaced if 100% flowering is not achieved.
 - (a) What is the probability that it will be necessary to replace a given packet under this guarantee? Interpret this probability.
 - (b) What would be the probability of replacing a packet if the guarantee covered only at least 10 out of 12 bulbs flowering? Comment on your findings in parts (a) and (b).
-

Exercise 3: Poisson Problems

1. While John is in his office, he receives 4 phone calls per hour on average. Assume that the number of calls within any interval of time is Poisson distributed.
 - (a) What is the probability that the phone rings at least 4 times between 10am and 11am?
 - (b) If John takes a 30 minute lunch break, what is the probability that the phone does not ring during that time?
 - (c) What is the expected number of times that the phone will ring during John's lunch break? What is the variance?
 - (d) If John arrives at work at 9 am and leaves at 5 pm, what is the expected number of times that the phone will ring during the day? What is the variance?
2. Hummingbirds arrive at a flower at a rate λ per hour.
 - (a) How many visits are expected in x hours of observation?
 - (b) What is the variance of the number of visits in an hour?
 - (c) If significantly more variance is observed than expected, what might this tell you about hummingbird visits?

3. Let $Y \sim \text{Poi}(6)$. Without using R, find:
- (a) $P(Y \geq 3)$;
 - (b) $P(Y \leq 15)$;
 - (c) $P(3 \leq Y \leq 15)$.
4. Bacteria are spread across a plate at an average density of 1000 per square cm. What, therefore, is the probability of seeing at least one cell?
- (a) What is the chance of seeing no bacteria in the viewing field of a microscope if this viewing field is 4×10^{-4} square cm?
 - (b) What is therefore the probability of seeing at least one bacterium cell?
5. A firm collects large quantities of data. Occasionally, typing errors cause data to be incorrectly entered. The number of typing errors per 20 pages of data is a Poisson random variable with mean 3.
- (a) What is the probability of there being 10 or more typing errors in 40 pages of data?
 - (b) What is the probability of there being between 5 and 9 (inclusive) typing errors in 40 pages of data?
 - (c) What is the probability of there being less than 5 typing errors in 20 pages of data?
 - (d) What is the mean number of typing errors in 40 pages of data? What is the variance?

Reminder: Logging Off:

When you have finished, close down RStudio. Remember to log off from your computer before leaving.
