STAT8007

Exercise Sheet 2

1. What exactly does a p-value tell you?
2. What is a type II error?
3. As the sample size increases does the probability of making a type II error increase or decrease?
4. Define the bias of an estimator. Explain what it means for an estimator to be *biased*.
5. The Central Limit Theorem states that as the sample size increases, the sampling distribution of the sample mean, , can be approximated by a normal distribution with mean μ and standard deviation σ/√n where:

μ is the population mean,

σ is the population standard deviation,

n is the sample size.

In other words, if we repeatedly take independent random samples of size n from any population, then when n is large, the distribution of the sample means will approach a normal distribution and the variance will reduce as n increases.

To demonstrate this we will simulate rolling a 6 sixed die.

The distribution of a single roll of a die is shown below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 |
| P(X = *x*) | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

The mean or expected value, µ, of this distribution is:

The aim of this question is to create a series of histograms showing the distribution of the sample mean ( for different sample sizes. We can sample from the uniform distribution to simulate rolling a die using R.

x <- as.integer(runif(1,1,7))

The uniform distribution is a continuous distribution so it is necessary to use the as.integer() function to round down to the nearest integer. Any value in the interval [1,2) will become 1, any value in the interval [2,3) will become 2 etc.

To get an accurate representation of the population distribution, let’s roll the die 10,000 times. Can you recreate the histograms below to show the distribution of the sample mean for samples of size, 1, 2, 5, 10, 100, 1000?

