STAT7203: Applied Probability and Statistics
Week 9 Exercises

1. Let $X_1, X_2, \ldots, X_n$ be a collection of n independent random variables with $\mathbb{E}X_i = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Define the random variables

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

   (a) What is the expected value and variance of $\bar{X}$?

   (b) Compute the covariance of X_1 and $\bar{X}$.

   (c) What is the expected value of S^2?

2. Let $X_1, \ldots, X_5$ be a simple random sample from a $\mathsf{N}(\mu, \sigma^2)$ distribution. The probability that the interval

$$\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{5}}, \, \bar{X} + 1.96\frac{\sigma}{\sqrt{5}}\right)$$

contains μ is 0.95. If we do not know σ^2, then we might replace it with the estimator S^2. Simulate 10 000 simple random samples from a $\mathsf{N}(0,1)$ distribution. For each sample construct the interval

$$\left(\bar{x} - 1.96\frac{s}{\sqrt{5}}, \, \bar{x} + 1.96\frac{s}{\sqrt{5}}\right)$$

and evaluate how often this interval contains 0 . What effect does replacing σ^2 with S^2 have on the coverage probability.

Some useful R commands for this problem:

- `rnorm` is the built in function to simulate normal random variables. `rnorm(n,mean,sd)` simulates `n` random variables from the normal distribution with mean `mean` and standard deviation `sd`. If the mean and standard deviation are not specified, then the standard normal distribution is used.
- `mean` computes the average (sample mean) of the values in the input vector.
- `sd` computes the sample standard deviation of the values in the input vector.
- `matrix` takes an vector input and forms a matrix. The number of rows and columns can be specified with the `nrow` and `ncol` arguments. Note that the length of the input vector should be equal to `nrow * ncol`.

- `apply` Returns a vector of values obtained by applying a function to margins of a matrix. For example, if we have a matrix `A` typing `apply(A,1,mean)` returns the row means of `A`.

3. A quality control worker in a factory wishes to estimate the mean weight for boxes of strawberries being packed at the factory so as to be within 50 grams. The distribution of the weight of boxes of strawberries is known to have a standard deviation of 120 grams from past experience. How many boxes of strawberries would the quality control worker be required to sample for the width of the 95% confidence interval to be less than 100 grams.

   The critical values from the normal distribution can be obtained from tables or using the R function `qnorm`. e.g. `qnorm(0.975)` gives the 0.975 quantile of the standard normal distribution.

4. *Attempt after the lecture on Wednesday* 7^{th} *October.* A 2010 study examined students' perceptions about the use of video games in the classroom. 50 students were surveyed (23 females, 27 males) from flemish secondary schools.

   For the following he critical values from the t-distribution can be obtained from tables or using the R function `qt`. e.g. `qt(0.975,8)` gives the 0.975 quantile of the t-distribution with 8 degrees of freedom (also written as the t_8-distribution).

   (a) Male students spent an average 6.96 hours per week playing video games with a sample standard deviation of 6.42 hours per week. Construct a 95% confidence interval for the mean number of hours per week a male student spends playing video games.

   (b) Female students spent an average 2.16 hours per week playing video games with a sample standard deviation of 4.55 hours per week. Construct a 90% confidence interval for the mean number of hours per week a male student spends playing video games.