STATS 2107
Statistical Modelling and Inference II
Tutorial 2

Sharon Lee, Matt Ryan

Semester 2 2022

1. a. If $X \sim \chi^2_k$, show that $E(X) = k$ and $\text{Var}(X) = 2k$. **Hint: Use MGFs.**

   b. Suppose $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ independently. Find the distribution of $X_1 + X_2$.

2. A random sample of 500 hospital records shows that the length of stay in one of South Australia's hospitals had a (sample) mean 5.4 days and (population) standard deviation 3.1 days.

   a. A health agency hypothesizes that the average length of stay is 5 days. Do the data support this hypothesis? You may use $\alpha = 0.05$.

   b. For the hypothesis test in part a, and using the significance level $\alpha = 0.05$, find β for $\mu = 5.5$. **Hint: first calculate the power using the formula from the lectures.**

   c. How large should the sample size be if we require that $\alpha = 0.01$ and $\beta = 0.05$, assuming $\mu = 5.5$?

3. A study is to be conducted to investigate the amount of toxic chemicals in freshwater lakes. A common measure of toxicity for any pollutant is LC50 (lethal concentration killing 50% of test species), which is the concentration of the pollutant that will kill half of the test species in a given amount of time (usually 96 hours for fish species). In many studies, the natural logarithm of LC50 measurements, log(LC50), are normally distributed. For copper, the variance of log(LC50) measurements is around 0.4 mg/L (milligrams per litre) on fish species A and around 0.8 mg/L for fish species B.

   a. Suppose 10 samples were collected for species A. Find the probability that the sample mean of log(LC50) will differ from the population mean by no more than 0.5.

   b. If we want the sample mean (for species A) to differ from the population by no more than 0.5 with probability 0.95, how many samples do we need to collect?

   c. Assuming the population mean for both species is the same, what is the probability that the sample mean of species A will exceed the sample mean of species B by at least 1 mg/L, if we collected 10 measurements from each species?

4. Suppose $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ are independent random samples, with $X_i \sim N(\mu_1, \sigma_1^2)$ and $Y_j \sim N(\mu_2, \sigma_2^2)$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

   a. State $E(\bar{X} - \bar{Y})$.

   b. State $\text{Var}(\bar{X} - \bar{Y})$.

   c. What is the sample size needed so that $(\bar{X} - \bar{Y})$ will be within k units of $(\mu_1 - \mu_2)$ with probability $1 - \alpha$? You may assume $m = n$.