1. Mothers of gifted children and IQ

   (a) State the null and alternative hypothesis in statistical notation, and in words.

   H_0: Mothers of gifted children, on average, have the same IQ as the population at large.

   $\mu = 100$

   H_A: Mothers of gifted children, on average, have a higher IQ than the population at large.

   $\mu > 100$

   (b) Give the formula for the test statistic you will use, and calculate it.

   I will use the Z-statistic, given the sample mean $\bar{x}$ and sample standard deviation s.

   $Z = (\bar{x} - \mu) / (s/ \sqrt{n})$

   (c) Give the p-value for the test, and the line of code you used to calculate it.

   p-value = 2.538738e-63

   P <- pnorm(abs(Z), mean = 0, sd = 1, lower.tail = FALSE)

   (d) Calculate a point estimate and a 95% confidence interval for the mean IQ of mothers of gifted children.

   Point estimate: $\bar{x} = 118.1667$

   Confidence interval: (116.0418, 120.2916)

   (e) Give a summary of your findings.

   The evidence suggests that the mean IQ of mothers of gifted children is higher than the population at large (Z-test, n = 36, p-value = 2.54e-63).

   Given such a low p-value we reject the null hypothesis H_0 in favor of the alternative hypothesis.

   Furthermore, we are 95% confident that the mean IQ of mothers of gifted children falls in the range (116.0418, 120.2916).

2. By understanding the sampling distribution, we can use sample summary statistics to approximate population summary statistics, and then quantify the uncertainty in our approximation using the variability in our sampling distribution. Given a significance level, a confidence interval, or range of plausible values, can be determined for our parameter of interest.

3. Game show

   a. In this case, we're interested in creating a hypothesis test for the population proportion of wins by southern participants.

      i. State the hypothesis in plain language, and then in mathematical formulae. Do the same for an alternative hypothesis we might be hoping to accept. In this case, the null hypothesis should be that the proportion of southerners winning the game show is equal to their participation level, or 25%. The alternative hypothesis would be that the proportion of southern winners is greater than 25%.

         H_0: $\hat{p}_s = 0.25$

         H_A: $\hat{p}_s > 0.25$

      ii. Identify the point estimate of the parameter of interest. In our case, we'd use the sample proportion to estimate the population proportion.

iii. The conditions for $\bar{x}$ to be considered nearly normal with an accurate standard error must be verified.
    1. Are the sample observations independent? In this case participants are chosen randomly and are much less than 10% of the population.
    2. Is the sample size reasonably large? For this case, yes, the sample size is 250.
    3. Do we have evidence that the population distribution is strongly skewed? There is no clear evidence of skew or significant outliers for our example.

iv. Compute the standard error, SE, draw a distribution of the estimate assuming H_0 is true, and shade the area under the curve that corresponds to the p-value.

v. Compute the Z-score and identify the p-value to evaluate the hypotheses. Write the conclusion in plain language.

b. For the estimate to be unbiased, our sampling distribution must be centered around the population distribution. We need to confirm that $\bar{x}$ is nearly normal, and our standard error is accurate.
   As outlined in 3.a, this entails ensuring that sample observations are independent, the sample size is reasonably large, and the population distribution is not strongly skewed.

c. The uncertainty of the estimate will decrease as the sample size increases. This is intuitive, as extreme values ("outliers") would have less and less impact as more observations were considered.

4. College grads playing Sudoku
    a. We should use the sample mean of 5 minutes to estimate the population mean time it takes to solve the puzzle. This point estimate will, on average, estimate the true population mean, given a nearly normal sampling distribution.
    b. Uncertainty in our estimate arises from sampling variation: the mean time will change from one sample to another. In this particular case, a sample size of 10 will result in a large standard error, and hence our estimate will exhibit significant variability.
    c. The uncertainty will decrease as we increase the sample size. SE would decrease tenfold in this case.
5. Idahoan farmer: $\bar{x} = 100$, CI = (90, 110)
    a. The confidence interval represents a range of plausible values given a certain significance level; in other words, it represents how confident we are that the population mean is contained within the range. It says nothing of the likelihood of individual observations.
    b. "We are 95% confident that the mean time to grow a potato is between 90 and 110 days."
    c. Confidence interval indicates a range of plausible values and the probability that the mean time it takes to grow a potato for the population is in this range.
    d. In general, increasing the confidence level would broaden the confidence interval: in order to provide a stronger confidence level we have to increase the range. Conversely, decreasing the confidence level would allow narrower ranges to be specified.