Introduction to Probability and Statistics

Session 2 Exercises

Israel Leyva-Mayorga

Exercise 1: Poll results

On October 14, 2003, the New York Times reported that a recent poll indicated that 52% of the population was in favor of the job performance of president Bush, with a margin of error (i.e., 95% confidence interval (CI)) of $\pm 4\%$.

a) What does this mean?
b) How many people were questioned?

Solution:

a) What does this mean?

   It means that there is a 95% confidence that the interval $(48, 56)$ captures the real opinion of the population regarding the job performance of Bush.

b) How many people were questioned?

   To calculate n, we assume that the sample is sufficiently large and calculate $Z_{\alpha/2} = 1.96$. Then, knowing that the opinion of each person is a Bernoulli random variable (RV) with parameter $\hat{p}_n$, we estimate σ^2 using the Maximum Likelihood Estimator (MLE) as

$$\hat{\sigma}_n^2 = \hat{p}_n(1 - \hat{p}_n) = 0.52\,(1 - 0.52) = 0.2496$$

   With this, we find n from

$$Z_{\alpha/2}\frac{\hat{\sigma}_n}{\sqrt{n}} \leq 0.04$$

   This gives

$$n = \left\lceil \frac{Z_{\alpha/2}^2 \hat{\sigma}_n^2}{0.04^2} \right\rceil = 600.$$

   Note that we cannot use the t-distribution for this exercise because it requires knowledge about n to calculate $T_{\alpha/2,n-1}$.

Exercise 2: Wireless communication with multiple antennas

A device transmits a signal towards a receiver. The receiver measures the signal strength, which is affected by Gaussian noise and, hence, is normally distributed with mean μ and variance $\sigma^2 = 4$. The receiver has 9 receiving antennas and each one records the value of the signal strength, which are

$$[5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5]$$

a) Calculate the 95% CI using $Z_{\alpha/2}$

b) Calculate the 95% CI using $T_{\alpha/2,n-1}$ and without knowing the variance

Solution:

a) Calculate the 95% CI using $Z_{\alpha/2}$

We easily calculate that $\hat{\mu}_n = 9$. Knowing that $\sigma^2 = 4$, we can calculate the 95% CI as

$$\begin{aligned} C_{0.95} &= \left(\hat{\mu}_n - \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}, \hat{\mu}_n + \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}\right) \\ &= \left(9 - \frac{1.96 \times 2}{3}, 9 + \frac{1.96 \times 2}{3}\right) = (7.6934, 10.3066) \end{aligned}$$

b) Calculate the 95% CI using $T_{\alpha/2,n-1}$ and without knowing the variance

Now we need to estimate the variance as

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2 = 9.5. \tag{1}$$

Then, we use the t-distribution with $n-1 = 8$ degrees of freedom to calculate the CI:

$$\begin{aligned} C'_{0.95} &= \left(\hat{\mu}_n - \frac{T_{\alpha/2,8}\,\hat{\sigma}_n}{\sqrt{n}}, \hat{\mu}_n + \frac{T_{\alpha/2,8}\,\hat{\sigma}_n}{\sqrt{n}}\right) \\ &= \left(9 - \frac{2.306 \times \sqrt{9.5}}{3}, 9 + \frac{2.306 \times \sqrt{9.5}}{3}\right) = (6.6308, 11.3692) \end{aligned}$$

The CI with the unknown variance is larger than with known variance.

Exercise 3: Execution time of an algorithm.

We measured the execution time of an image processing algorithm and obtained an average execution time of $\mu = 12.995$ ms. Assume that our measurements are subject to Gaussian noise with zero mean and variance $\sigma^2 = 1.997$

a) Get the maximum likelihood estimates for μ, denoted as $\hat{\mu}_n$ after taking a sample with $n = 10$ and $n = 1000$.

b) Plot the collected measurements along with the estimated $\hat{\mu}_n$.

c) Calculate the 95% CI for both values of n assuming that the variance σ^2 is known.

d) Calculate the 95% CI for both values of n assuming that the variance is **not known** using $Z_{\alpha/2}$

e) Calculate the 95% CI for both values of n assuming that the variance is **not known** using $T_{\alpha/2,n-1}$

f) Which of these CIs is wider?

Solution:

a) The MLE estimates are $\hat{\mu}_{10} = 13.2496$ and $\hat{\mu}_{1000} = 12.9299$.

c) With known variance, I got $\text{CI}_{10} = (12.3782, 14.1212)$ and $\text{CI}_{1000} = (12.6544, 13.2056)$

d) With unknown variance and using $Z_{\alpha/2}$, I got $\text{CI}_{10} = (12.3242, 14.1752)$ and $\text{CI}_{1000} = (12.6726, 13.1874)$

e) With unknown variance and using $T_{\alpha/2,n-1}$, I got $\text{CI}_{10} = (12.1815, 14.3179)$ and $\text{CI}_{1000} = (12.6694, 13.1906)$

f) The widest CIs are obtained with the t-distribution and the greatest difference is observed for the CIs with $n = 10$. There is no big difference between the CIs obtained with $n = 1000$.

Exercise 4: CIs with unknown parameters and small sample size.

Suppose the data 2.5, 5.5, 8.5, 11.5 was drawn from a $N(\mu, \sigma^2)$ distribution with unknown parameters. Give the 95%, 80%, and 50% CIs for μ.

Solution: We estimate $\hat{\mu}_n = 7$ and

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2 = 15. \tag{2}$$

Then, we get

$$T_{0.025,3} = 3.1824, \quad T_{0.1,3} = 1.6377, \text{ and } T_{0.25,3} = 0.7649$$

And calculate the CIs as in previous examples

$$\begin{aligned} C_{0.95} &= \left(\hat{\mu}_n - \frac{T_{\alpha/2,3}\,\hat{\sigma}_n}{\sqrt{n}}, \hat{\mu}_n + \frac{T_{\alpha/2,3}\,\hat{\sigma}_n}{\sqrt{n}}\right) \\ &= \left(7 - \frac{3.1824 \times \sqrt{15}}{2}, 7 + \frac{3.1824 \times \sqrt{15}}{2}\right) = (0.8372, 13.1628) \end{aligned}$$

$$C_{0.80} = \left(7 - \frac{1.6377 \times \sqrt{15}}{2}, 7 + \frac{1.6377 \times \sqrt{15}}{2}\right) = (3.8285, 10.1715)$$

$$C_{0.50} = \left(7 - \frac{0.7649 \times \sqrt{15}}{2}, 7 + \frac{0.7649 \times \sqrt{15}}{2}\right) = (5.5188, 8.4812)$$

Exercise 5: Chapter 9, problem 4 and 9

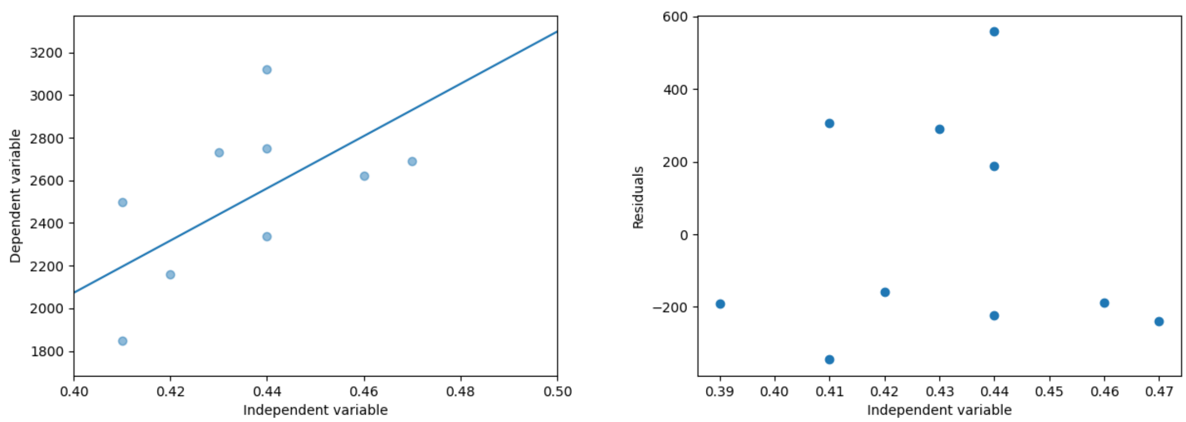
The following data indicate the relationship between x, the specific gravity of a wood sample, and Y, its maximum crushing strength in compression parallel to the grain

x_i	y_i(psi)	x_i	y_i(psi)
.41	1850	.39	1760
.46	2620	.41	2500
.44	2340	.44	2750
.47	2690	.43	2730
.42	2160	.44	3120

a) Plot a scatter diagram. Does a linear relationship seem reasonable?
b) Estimate the regression coefficients.
c) Predict the maximum crushing strength of a wood sample whose specific gravity is 0.43.
d) Estimate the variance of an individual response.

Solution:

a) Just by looking at the scatter plot, a linear model seems reasonable. This is confirmed after looking at the residuals in Fig. 1
b) The regression coefficients are $\beta_0 = -2825.9168$ and $\beta_1 = 12245.7466$
c) We estimate the response by calculating $r(0.43) = \beta_0 + \beta_1(0.43) = 2439.7542$
d) This is estimated from the residuals as $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{\epsilon}_i^2 = 105660.066$



(a) (b)

Fig. 1. Linear regression and residual analysis for the crushing strength of wood.