# Problem set 7: CI for the Mean and Variance.

*During the seminars, we will work on problems across three different categories of difficulty, with Category 1 being the easiest and Category 3 the most challenging.*

## CATEGORY 1:

**Confidence Intervals for Mean and Variance of a Normal Distribution**
**Model:**
Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution:

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad i = 1, \ldots, n.$$

**Theorem 1 (Confidence Interval for the Mean when Variance is Known).**
If $\sigma^2$ is known, then:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Hence, a $(1 - \alpha) \times 100\%$ confidence interval for $\mu$ is:

$$\overline{X} \ \pm \ z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

**Theorem 2 (Confidence Interval for the Mean when Variance is Unknown).**
If $\sigma^2$ is unknown, then:

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Therefore, a $(1 - \alpha) \times 100\%$ confidence interval for $\mu$ is:

$$\overline{X} \ \pm \ t_{\alpha/2,\, n-1} \cdot \frac{S}{\sqrt{n}},$$

where $t_{\alpha/2,\, n-1}$ is the upper $\alpha/2$ quantile of the $t$-distribution with $n - 1$ degrees of freedom.

**Theorem 3 (Confidence Interval for the Variance).**
Under normality:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Hence, a $(1 - \alpha) \times 100\%$ confidence interval for $\sigma^2$ is:

$$\left( \frac{(n-1)S^2}{\chi^2_{\alpha/2,\, n-1}}, \ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\, n-1}} \right),$$

where $\chi^2_{p,\, n-1}$ is the $p$-quantile of the chi-squared distribution with $n - 1$ degrees of freedom.

**Remark.**
Confidence intervals for $\sigma$ can be obtained by taking square roots of the limits of the interval for $\sigma^2$.

1. You are interested in the average emergency room (ER) wait time at your local hospital. You take a random sample of 50 patients who visit the ER over the past week. From this sample, the mean wait time was 30 minutes and the standard deviation was 20 minutes. Find a 95% confidence interval for the average ER wait time for the hospital.

2. The yearly salary for mathematics assistant professors are normally distributed. A random sample of 8 math assistant professor's salaries are listed below in thousands of dollars. Estimate the population mean salary with a 99% confidence interval.

   66.0 75.8 70.9 73.9 63.4 68.5 73.3 65.9

3. The weight of the world's smallest mammal is the bumblebee bat (also known as Kitti's hog-nosed bat or *Craseonycteris thonglongyai*) and is approximately normally distributed with a mean of 1.9 grams. Such bats are roughly the size of a large bumblebee.

   A chiropterologist believes that the Kitti's hog-nosed bats in a new geographical region under study have a different average weight than 1.9 grams. A sample of 10 bats weighed in grams in the new region are shown below:

| **Weight (g)** | 1.9 | 2.24 | 2.13 | 2.00 | 1.54 | 1.96 | 1.79 | 2.18 | 1.81 | 2.30 |
|---|---|---|---|---|---|---|---|---|---|---|

   Use the **confidence interval** method to test the claim that the **mean weight** for all bumblebee bats is not 1.9 g using a **10% level of significance**.

4. A researcher is interested in estimating the average salary of teachers. She wants to be 95% confident that her estimate is correct. In a previous study, she found the population standard deviation was \$1,175. How large a sample is needed to be accurate within \$100?

5. A random sample of 20 nominally measured 2 mm diameter steel ball bearings is taken and the diameters are measured precisely. The measurements, in mm, are as follows:

|  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| 2.02 | 1.94 | 2.09 | 1.95 | 1.98 | 2.00 | 2.03 | 2.04 | 2.08 | 2.07 |
| 1.99 | 1.96 | 1.99 | 1.95 | 1.99 | 1.99 | 2.03 | 2.05 | 2.01 | 2.03 |

   Assuming that the diameters are normally distributed with unknown mean $\mu$ and unknown variance $\sigma^2$:

   (a) Find a two-sided 95% confidence interval for the variance, $\sigma^2$.

   (b) Find a two-sided confidence interval for the standard deviation, $\sigma$.

6. The feeding habits of two species of net-casting spiders are studied. The species, the *deinopis* and *menneus*, coexist in eastern Australia. The following data were obtained on the size, in millimeters, of the prey of random samples of the two species:

   **Size of Random Prey Samples of the Deinopis Spider in Millimeters**

| **sample 1** | **sample 2** | **sample 3** | **sample 4** | **sample 5** | **sample 6** | **sample 7** | **sample 8** | **sample 9** | **sample 10** |
|---|---|---|---|---|---|---|---|---|---|
| 12.9 | 10.2 | 7.4 | 7.0 | 10.5 | 11.9 | 7.1 | 9.9 | 14.4 | 11.3 |

   **Size of Random Prey Samples of the Menneus Spider in Millimeters**

| **sample 1** | **sample 2** | **sample 3** | **sample 4** | **sample 5** | **sample 6** | **sample 7** | **sample 8** | **sample 9** | **sample 10** |
|---|---|---|---|---|---|---|---|---|---|
| 10.2 | 6.9 | 10.9 | 11.0 | 10.1 | 5.3 | 7.5 | 10.3 | 9.2 | 8.8 |

   What is the difference, if any, in the mean size of the prey (of the entire populations) of the two species?

7. Let's return to the example, in which the feeding habits of two species of net-casting spiders are studied. The species, the *deinopis* and *menneus*, coexist in eastern Australia. The following summary statistics were obtained on the size, in millimeters, of the prey of the two species:

| **Adult DEINOPIS** | **Adult MENNEUS** |
|---|---|
| $n = 10$ | $m = 10$ |
| $\overline{x} = 10.26$ mm | $\overline{y} = 9.02$ mm |
| $s_X^2 = (2.51)^2$ | $s_Y^2 = (1.90)^2$ |

   What is the difference in the mean sizes of the prey (of the entire populations) of the two species?

**Confidence Intervals for the Difference of Means**
**Model:**
Suppose we have two independent samples:

$$X_1, \dots, X_{n_1} \sim \mathcal{N}(\mu_1, \sigma_1^2),$$
$$Y_1, \dots, Y_{n_2} \sim \mathcal{N}(\mu_2, \sigma_2^2).$$

**Theorem 4 (Difference of Means, Equal Variances).**
If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, define the pooled variance:

$$S_p^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}.$$

Then:

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}.$$

Hence, a $(1 - \alpha) \times 100\%$ confidence interval for $\mu_1 - \mu_2$ is:

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2,\, n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right).$$

**Theorem 5 (Difference of Means, Unequal Variances – Welch's Approximation).**
If variances are not assumed equal, then:

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}} \sim t_\nu,$$

where the approximate degrees of freedom $\nu$ are given by:

$$\nu = \frac{\left(\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}\right)^2}{\frac{\left(\frac{S_X^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_Y^2}{n_2}\right)^2}{n_2-1}}.$$

Thus, a $(1 - \alpha) \times 100\%$ confidence interval for $\mu_1 - \mu_2$ is:

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2,\, \nu} \cdot \sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}\right).$$