Statistical Inference

Detailed Solutions - TD Sheet 1
Sampling and Sampling Distributions

Abdelhakim Necir

Academic Year 2026–2027

Exercise 1: Population, Sample, Parameter and Statistic

A company manufactures light bulbs. A random sample of 50 bulbs is selected. The sample mean lifetime is 1240 hours and the sample standard deviation is 85 hours.

1. **Population.**

   The population is the set of all light bulbs produced by the company during the period under study.

2. **Sample.**

   The sample consists of the 50 selected light bulbs.

3. **Possible population parameters.**

   Two natural parameters are the population mean lifetime

   $$\mu = E(X),$$

   and the population variance

   $$\sigma^2 = \mathrm{Var}(X).$$

4. **Sample statistics.**

   The observed sample mean is

   $$\overline{X} = 1240,$$

   and the observed sample standard deviation is

   $$S = 85.$$

5. **Parameter versus statistic.**

   A parameter is a numerical characteristic of the population and is usually unknown. A statistic is computed from the observed sample.

   Thus, μ and σ are parameters, whereas $\overline{X}$ and S are statistics.

Exercise 2: Sampling Methods and Experimental Design

1. Selecting 200 students at random from the complete list of registered students corresponds to **simple random sampling**.

2. Dividing the population into four age groups and selecting a random sample from each group corresponds to **stratified sampling**.

3. Selecting every 20th name after a random starting point corresponds to **systematic sampling**.

4. Randomly selecting ten schools and surveying all students in those schools corresponds to **cluster sampling**.

5. Interviewing the first 100 people entering a shopping center is an example of **convenience sampling**.

6. The **experimental units** are the 200 patients.

7. The factor is the **treatment**. Its two levels are

$$\text{new treatment} \qquad \text{and} \qquad \text{standard treatment.}$$

8. If the purpose is to compare recovery time, then the response variable is the **recovery time** of each patient.

9. Randomization reduces systematic allocation bias and helps make the treatment groups comparable with respect to observed and unobserved characteristics. It therefore strengthens the validity of the comparison.

Exercise 3: Sampling Distribution of the Sample Mean

We are given

$$\mu = 80, \qquad \sigma = 12, \qquad n = 36.$$

The population is normal.

1. Since the parent population is normal,

$$\overline{X} \sim \mathcal{N}\left(80, \frac{12^2}{36}\right) = \mathcal{N}(80, 4).$$

   Here the second parameter is the variance. Hence the standard deviation of $\overline{X}$ is 2.

2.
$$E(\overline{X}) = \mu = 80.$$

3.
$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} = \frac{144}{36} = 4.$$

4. The standard error is
$$\mathrm{SE}(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{12}{6} = 2.$$

5. We want
$$P(78 \le \overline{X} \le 84).$$
Standardizing with
$$Z = \frac{\overline{X} - 80}{2} \sim \mathcal{N}(0,1),$$
gives
$$P(78 \le \overline{X} \le 84) = P\left(-1 \le Z \le 2\right).$$
Therefore,
$$P(78 \le \overline{X} \le 84) = \Phi(2) - \Phi(-1).$$
Using
$$\Phi(2) \approx 0.97725, \qquad \Phi(-1) \approx 0.15866,$$
we obtain
$$P(78 \le \overline{X} \le 84) \approx 0.81859.$$

Exercise 4: Sampling Distribution of a Proportion

We have
$$p = 0.40, \qquad n = 100.$$

1.
$$E(\widehat{p}) = p = 0.40.$$

2.
$$\mathrm{Var}(\widehat{p}) = \frac{p(1-p)}{n} = \frac{0.4 \times 0.6}{100} = 0.0024.$$

3.
$$\mathrm{SE}(\widehat{p}) = \sqrt{0.0024} \approx 0.04899.$$

4. The usual large-sample conditions are satisfied because
$$np = 40 \qquad \text{and} \qquad n(1-p) = 60.$$
Hence the normal approximation is appropriate.

5. Using
$$\widehat{p} \approx \mathcal{N}\left(0.4, \frac{0.4(0.6)}{100}\right),$$
we standardize:
$$Z = \frac{\widehat{p} - 0.4}{\sqrt{0.0024}}.$$
The two standardized bounds are
$$z_1 = \frac{0.35 - 0.40}{\sqrt{0.0024}} \approx -1.0206,$$
and
$$z_2 = \frac{0.45 - 0.40}{\sqrt{0.0024}} \approx 1.0206.$$
Hence
$$P(0.35 \le \widehat{p} \le 0.45) \approx P(-1.0206 \le Z \le 1.0206).$$
Thus,
$$P(0.35 \le \widehat{p} \le 0.45) \approx 2\Phi(1.0206) - 1 \approx 0.6924.$$

Exercise 5: Difference of Two Sample Means

The data are

$$\mu_1 = 50, \quad \sigma_1 = 8, \quad n_1 = 64,$$

and

$$\mu_2 = 46, \quad \sigma_2 = 6, \quad n_2 = 36.$$

The two samples are independent.

1. $$E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2 = 50 - 46 = 4.$$

2. By independence,
$$\mathrm{Var}(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$
Hence
$$\mathrm{Var}(\overline{X} - \overline{Y}) = \frac{64}{64} + \frac{36}{36} = 2.$$

3. The standard error is
$$\mathrm{SE}(\overline{X} - \overline{Y}) = \sqrt{2} \approx 1.4142.$$

4. If both populations are normal, then
$$\overline{X} - \overline{Y} \sim \mathcal{N}(4, 2),$$
where 2 is the variance.

5. We compute
$$P(\overline{X} - \overline{Y} > 6).$$
Standardization gives
$$P(\overline{X} - \overline{Y} > 6) = P\left(Z > \frac{6-4}{\sqrt{2}}\right).$$
Since
$$\frac{2}{\sqrt{2}} = \sqrt{2} \approx 1.4142,$$
we obtain
$$P(\overline{X} - \overline{Y} > 6) = 1 - \Phi(1.4142) \approx 0.07865.$$

Exercise 6: Sample Variance and Central Limit Theorem

Part A: Sample Variance

We are given

$$X_1, \ldots, X_{20} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 25),$$

so that

$$\sigma^2 = 25, \qquad n = 20.$$

1. For a normal sample,
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$
Therefore,
$$\frac{19S^2}{25} \sim \chi^2_{19}.$$

2. The number of degrees of freedom is
$$n - 1 = 19.$$

3. We want c such that
$$P(S^2 \leq c) = 0.95.$$
Since
$$\frac{19S^2}{25} \sim \chi^2_{19},$$
we have
$$P\left(\frac{19S^2}{25} \leq \frac{19c}{25}\right) = 0.95.$$
By definition of the 95th percentile,
$$\frac{19c}{25} = \chi^2_{19,0.95}.$$
Thus,
$$c = \frac{25}{19}\chi^2_{19,0.95}.$$
Using
$$\chi^2_{19,0.95} \approx 30.144,$$
we obtain
$$c \approx \frac{25}{19}(30.144) \approx 39.663.$$

Part B: Central Limit Theorem

We have
$$E(X) = 10, \qquad \mathrm{Var}(X) = 100, \qquad n = 100.$$
Thus
$$\mu = 10, \qquad \sigma = 10.$$

4. By the Central Limit Theorem,
$$\overline{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$
Hence
$$\overline{X} \approx \mathcal{N}(10, 1).$$

5. We seek

$$P(8 \leq \overline{X} \leq 12).$$

Since the approximate standard deviation of $\overline{X}$ is 1,

$$P(8 \leq \overline{X} \leq 12) \approx P(-2 \leq Z \leq 2).$$

Therefore,

$$P(8 \leq \overline{X} \leq 12) \approx \Phi(2) - \Phi(-2).$$

Using

$$\Phi(2) \approx 0.97725, \qquad \Phi(-2) \approx 0.02275,$$

we obtain

$$P(8 \leq \overline{X} \leq 12) \approx 0.95450.$$