University of Tehran,
College of Engineering

# Homework 3

## Introduction to Statistical Interference

Instructor: **Dr. Mohammadreza A. Dehaqani**

**Erfan Panahi** (*Student Number:* **810103084**)

## Problem 1. Say Hello to Neyman-Pearson

Given the random sample $X' = (X_1, \cdots, X_n)$ from the distribution with the pdf: $(X_i \sim N(\theta, 1))$

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right), \qquad -\infty < x < \infty$$

**Part 1.** The likelihood function for the sample is:

$$L(\theta; X') = \prod_{i=1}^{n} f(X_i;\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{(X_i-\theta)^2}{2}\right)$$

For the hypotheses $H_0: \theta = \theta_0$ and $H_1: \theta = \theta_1$, the likelihood ratio is:

$$\textbf{likelihood ratio }(\mathbf{X}) = \frac{L(\theta_0; X')}{L(\theta_1; X')} = \exp\left(-\sum_{i=1}^{n} \frac{(X_i-\theta_0)^2 - (X_i-\theta_1)^2}{2}\right)$$

$$= \exp\left(-\sum_{i=1}^{n} \frac{\theta_0^2 - \theta_1^2 - 2X_i(\theta_0-\theta_1)}{2}\right) = \mathbf{exp}\left(-\frac{\boldsymbol{n}}{\mathbf{2}}(\boldsymbol{\theta}_0^2 - \boldsymbol{\theta}_1^2) + (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)\sum_{i=1}^{n} \boldsymbol{X}_i\right)$$

Identifying the test statistic:

$$\textbf{Test statistic} = \mathbf{T} = \sum_{i=1}^{n} \boldsymbol{X}_i$$

---

**Part 2.** To determine the rejection region, we need to compare the likelihood ratio to a critical value $c$:

$$\text{likelihood ratio (X)} \le c$$

Taking natural logarithms and simplifying, the rejection region is:

$$\exp\left(-\frac{n}{2}(\theta_0^2 - \theta_1^2) + (\theta_0-\theta_1)\sum_{i=1}^{n} X_i\right) \le c \to (\theta_0-\theta_1)\sum_{i=1}^{n} X_i \le \frac{n}{2}(\theta_0^2 - \theta_1^2) + \log c$$

$$\to \sum_{i=1}^{n} X_i \le \frac{n}{2(\theta_0-\theta_1)}(\theta_0^2 - \theta_1^2) + \frac{\log c}{(\theta_0-\theta_1)}$$

$$\text{the rejection region: } \boldsymbol{T} \le \frac{\boldsymbol{n}}{\mathbf{2}(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)}(\boldsymbol{\theta}_0^2 - \boldsymbol{\theta}_1^2) + \frac{\mathbf{log}\,\boldsymbol{c}}{(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)}$$

---

**Part 3.** We are given:

$$\theta_0 = 0\,, \theta_1 = 1\,, \alpha = 0.05$$

- Under $H_0$ $(\theta = \theta_0 = 0)$:

$$T = \sum_{i=1}^{n} X_i \sim N(n\theta_0, n) = N(0, n) \rightarrow \boldsymbol{T \sim N(0, n)}$$

The critical value $c$ for $\alpha = 0.05$ is:

$$P(T \leq c \mid \theta = 0) = \Phi\left(\frac{c}{\sqrt{n}}\right) = 0.05 \rightarrow c = \Phi^{-1}(0.05)\sqrt{n} \cong -1.645\sqrt{n} \rightarrow \boldsymbol{c \cong -1.645\sqrt{n}}$$

- Under $H_0$ $(\theta = \theta_0 = 0)$:

$$T = \sum_{i=1}^{n} X_i \sim N(n\theta_1, n) = N(n, n)$$

The power of the test is:

$$\beta = P(T \leq c \mid \theta = 1) = P\left(\frac{T-n}{\sqrt{n}} \leq \frac{c-n}{\sqrt{n}}\right) = \Phi\left(\frac{-1.645\sqrt{n} - n}{\sqrt{n}}\right) = \Phi\left(-1.645 - \sqrt{n}\right)$$
$$= \Phi\left(-1.645 - \sqrt{n}\right) = 1 - \Phi\left(1.645 - \sqrt{n}\right)$$

$$\rightarrow \textbf{Power of the test} = 1 - \beta = \boldsymbol{\Phi\left(1.645 - \sqrt{n}\right)}$$

## **Problem 2.** Wald Test

The probability mass function (PMF) for a Poisson random variable $X_i$ is:

$$P(X_i = x_i; \lambda) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \,, \qquad x_i = 0,1,2,\ldots$$

**Part 1.** For a sample $X_1, \ldots, X_n$, the likelihood function $L(\lambda)$ is:

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Take the natural logarithm of the likelihood function to get the log-likelihood $\ell(\lambda)$:

$$\ell(\lambda) = \log L(\lambda) = \sum_{i=1}^{n} [x_i \log(\lambda) - \lambda - \log(x_i!)]$$

Differentiate $\ell(\lambda)$ with respect to $\lambda$ to find the MLE:

$$\frac{\partial \ell(\lambda)}{\partial \lambda} = \sum_{i=1}^{n} \left[\frac{x_i}{\lambda} - 1\right] = \sum_{i=1}^{n} \frac{x_i}{\lambda} - n = 0 \rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X} \rightarrow \boldsymbol{\hat{\lambda} = \bar{X}}$$

---

**Part 2.** For a Poisson distribution with parameter $\lambda$, the mean and variance are both equal to $\lambda$. That is:

$$E(X_i) = \lambda \quad \text{and} \quad var(X_i) = \lambda$$

For the sample mean $\hat{\lambda} = \bar{X}$ the variance is:

$$var(\hat{\lambda}) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

Using the property of variances for i.i.d. variables:

$$var(\hat{\lambda}) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\lambda}{n} \rightarrow \mathrm{SD}(\hat{\lambda}) = \sqrt{\frac{\lambda}{n}} = \sqrt{\frac{\bar{X}}{n}} \rightarrow \mathbf{SD}(\boldsymbol{\hat{\lambda}}) = \sqrt{\frac{\boldsymbol{\bar{X}}}{\boldsymbol{n}}}$$

---

**Part 3.** We are testing the following hypotheses:

$$H_0 : \lambda = \lambda_0$$

$$H_1 : \lambda \neq \lambda_0$$

We use the test statistic:

$$Z = \frac{\hat{\lambda} - \lambda_0}{\mathrm{SD}(\lambda_0)} = \frac{\hat{\lambda} - \lambda_0}{\sqrt{\frac{\lambda_0}{n}}}$$

Under $H_0$, $\hat{\lambda}$ (the sample mean) is approximately normal for large $n$ (by the Central Limit Theorem), with:

$$\hat{\lambda} \sim N\left(\lambda_0, \frac{\lambda_0}{n}\right)$$

Thus, the standardized test statistic $Z$ follows a standard normal distribution:

$$\boldsymbol{Z \sim N(0, 1)}$$

For a two-tailed test at significance level $\alpha$, the rejection region is:

$$|\boldsymbol{Z}| > \boldsymbol{z_{\frac{\alpha}{2}}}$$

where $z_{\frac{\alpha}{2}}$ is the critical value of the standard normal distribution corresponding to $\frac{\alpha}{2}$.

# Problem 3. $f_0(x)$ vs $f_1(x)$

## Part 1. Test Procedure to Minimize $\alpha + \beta$

Given the null and alternative hypotheses:

$H_0 : f(x) = f_0(x)$ (distribution is $f_0(x)$)

$H_1 : f(x) = f_1(x)$ (distribution is $f_1(x)$)

the objective is to minimize the sum of Type I ($\alpha$) and Type II ($\beta$) errors. To achieve this, we use the **Likelihood Ratio Test (LRT)**, which compares the likelihood ratio $\Lambda(x)$ to a threshold:

$$\Lambda(x) = \frac{f_1(x)}{f_0(x)}$$

From the given probability density functions:

$f_0(x) = \frac{3}{2}, \quad 0 \le x \le \frac{2}{3}$

$f_1(x) = \frac{9}{2}x, \quad 0 \le x \le \frac{2}{3}$

the likelihood ratio becomes:

$$\Lambda(x) = \frac{f_1(x)}{f_0(x)} = \frac{\frac{9}{2}x}{\frac{3}{2}} = 3x$$

To minimize $\alpha + \beta$, the **Neyman-Pearson Lemma** suggests rejecting $H_0$ when $\Lambda(x) > k$, where $k$ is a threshold chosen based on the desired significance level. Thus, the rejection region is:

$$\Lambda(x) = 3x > k \quad \Longrightarrow \quad x > \frac{k}{3}$$

The optimal decision boundary $c$ is the value of $x$ that satisfies this inequality:

$$x = c = \frac{k}{3}$$

The probabilities of Type I ($\alpha$) and Type II ($\beta$) errors for each decision boundary $c$:

**Type I error ($\alpha$):** Probability of rejecting $H_0$ when $H_0$ is true

$$\alpha = P(X > c \mid H_0) = \int_c^{\frac{2}{3}} f_0(x)dx = \int_c^{\frac{2}{3}} \frac{3}{2}dx = \frac{3}{2}\left(\frac{2}{3} - c\right)$$

**Type II error ($\beta$):** Probability of failing to reject $H_0$ when $H_1$ is true

$$\beta = P(X \le c \mid H_1) = \int_0^{c} f_1(x)dx = \int_0^{c} \frac{9}{2}x\, dx = \frac{9}{2}\frac{x^2}{2}\Big|_0^c = \frac{9}{4}c^2$$

Differentiate $\alpha + \beta$ with respect to $c$ and set it to zero to find the optimal decision boundary $c$:

$$\frac{\partial}{\partial c}(\alpha + \beta) = \frac{\partial}{\partial c}\left(1 - \frac{3}{2}c + \frac{9}{4}c^2\right) = -\frac{3}{2} + \frac{9}{2}c = 0 \to \boldsymbol{c = \frac{1}{3}} \ (\text{or } \boldsymbol{k = 1})$$

## Part 2. Programming Task

The **power** of the test for each $c$ is:

$$\text{Power} = 1 - \beta = 1 - \frac{9}{4}c^2$$

As it is shown if figure 3.1 we conclude that:

- **Small $c$:** High false positives ($\alpha$) but low false negatives ($\beta$), leading to high power.
- **Large $c$:** Low false positives ($\alpha$) but high false negatives ($\beta$), leading to low power.
- The plot helps visualize how $c$ affects test performance, enabling the selection of the optimal $c$ for specific goals (e.g., minimizing $\alpha + \beta$).

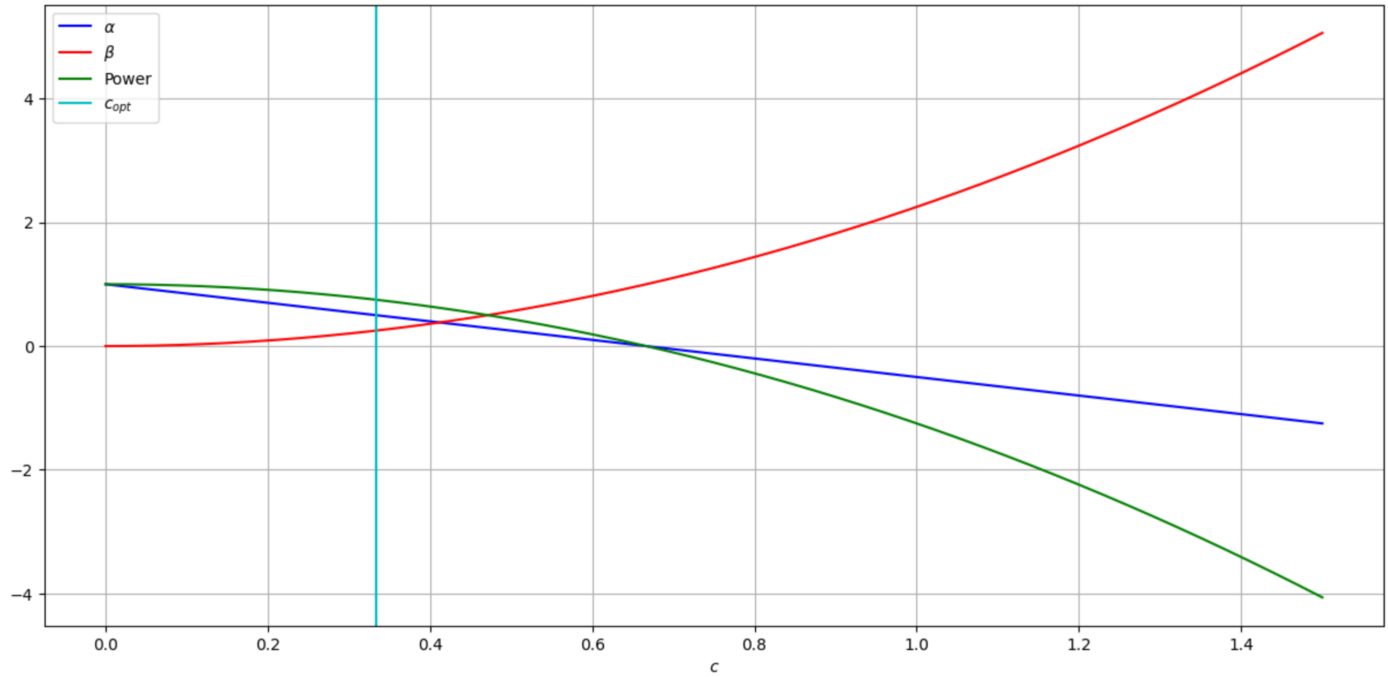


**Figure 3.1.** The power of the test and types of error respect to $c$

---

## Part 3. Minimizing Expected Risk

The problem involves assigning risks to Type I ($\alpha$) and Type II ($\beta$) errors:

$\boldsymbol{R_1}$: Risk associated with Type I error ($\alpha$)

$\boldsymbol{R_2}$: Risk associated with Type II error ($\beta$)

The **expected risk** is defined as:

$$\text{Expected Risk} = R_1\alpha + R_2\beta = \frac{3}{2}\left(\frac{2}{3} - c\right)R_1 + \frac{9}{4}c^2R_2$$

Differentiate the expected risk with respect to $c$ and set it to zero to find the optimal decision boundary $c^*$:

$$\frac{\partial}{\partial c}(\text{Expected Risk}) = \frac{\partial}{\partial c}\left(\frac{3}{2}\left(\frac{2}{3} - c\right)R_1 + \frac{9}{4}c^2R_2\right) = \frac{9}{2}cR_2 - \frac{3}{2}R1 = 0 \to \boldsymbol{c^*} = \frac{\boldsymbol{R_1}}{\boldsymbol{3R_2}}$$

## **Problem 4.** Uniformly Most Powerful

**Part 1.** The probability density function is given as:

$$f(x;\theta) = \theta x^{\theta-1}, \qquad 0 < x < 1, \qquad \theta > 0$$

For a random sample $X_1, X_2, \dots, X_n$ the likelihood function $L(\theta)$ is:

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta) = \prod_{i=1}^{n} \theta X_i^{\theta-1} = \theta^n \prod_{i=1}^{n} X_i^{\theta-1}$$

Take the natural logarithm of the likelihood function to get the log-likelihood:

$$\ell(\theta) = \log\big(L(\theta)\big) = n\log(\theta) + (\theta-1)\sum_{i=1}^{n} n\log(X_i)$$

To find the MLE $\hat{\theta}$, differentiate $\ell(\theta)$ with respect to $\theta$ and set it to zero:

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log(X_i) = 0 \to \hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log(X_i)}$$

Under $\boldsymbol{H_0: \theta = \theta_0}$**,** The likelihood ratio $\Lambda$ is defined as:

$$\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})} = \frac{\theta_0^n \prod_{i=1}^{n} X_i^{\theta_0-1}}{\hat{\theta}^n \prod_{i=1}^{n} X_i^{\hat{\theta}-1}} = \frac{\theta_0^n}{\hat{\theta}^n} \prod_{i=1}^{n} X_i^{\theta_0-\hat{\theta}}$$

Notice that $\prod_{i=1}^{n} X_i$ is the sufficient statistic for $\theta$, and the likelihood ratio $\Lambda$ depends on it. Therefore, the test statistic is:

$$T = \prod_{i=1}^{n} X_i$$

**UMP Test Using Neyman-Pearson Lemma:** The Neyman-Pearson Lemma states that the **UMP test** for $\boldsymbol{H_0: \theta = \theta_0}$ versus $\boldsymbol{H_1: \theta < \theta_0}$ rejects $\boldsymbol{H_0}$ when the likelihood ratio $\Lambda$ is sufficiently small.

The likelihood ratio depends on $\prod_{i=1}^{n} X_i$, so we reject $H_0$ when $\prod_{i=1}^{n} X_i$ is sufficiently small.

**Rejection Region:** Let $c$ be a critical value determined by the significance level $\alpha$. The rejection region is:

$$\prod_{i=1}^{n} X_i < c,$$

where $c$ is chosen such that:

$$P\left(\prod_{i=1}^{n} X_i < c \mid \theta = \theta_0\right) = \alpha$$

## **Problems 6 and 9.** Error Types

**Part 1.** A **Type I error** occurs when we reject $H_0 : \theta = 2$ even though it is true. In this case, the test rejects $H_0$ if $X \leq 0.1$ or $X \geq 1.9$. When $\theta = 2$, $X \sim U[0,2]$, so the probability density function (PDF) is:

$$f(x) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

The probability of a Type I error ($\alpha$) is the probability of being in the rejection region:

$$P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ true}) = P(X \leq 0.1 \mid \theta = 2) + P(X \geq 1.9 \mid \theta = 2)$$

Compute these probabilities:

$$P(X \leq 0.1) = \int_0^{0.1} f(x)dx = \int_0^{0.1} \frac{1}{2}dx = \frac{1}{2}(0.1) = 0.05$$

$$P(X \geq 1.9) = \int_{1.9}^{2} f(x)dx = \int_{1.9}^{2} \frac{1}{2}dx = \frac{1}{2}(0.1) = 0.05$$

Thus, the total Type I error probability is:

$$\alpha = 0.05 + 0.05 = 0.1 \rightarrow \boldsymbol{\alpha = 0.1}$$

---

**Part 2.** A **Type II error** occurs when we fail to reject $H_0$ despite the true value being $\theta \neq 2$. Let the true value be $\theta = 2.5$. In this case, $X \sim U[0,2.5]$, and the PDF is:

$$f(x) = \begin{cases} \frac{1}{2.5}, & 0 \leq x \leq 2.5 \\ 0, & \text{otherwise} \end{cases}$$

The Type II error occurs when $X \in (0.1,1.9)$, i.e., in the acceptance region. The probability of a Type II error ($\beta$) is:

$$P(\text{Type I error}) = P(\text{accept } H_0 \mid H_0 \text{ false}) = P(\,0.1 \leq X \leq 1.9 \mid \theta = 2.5\,) = \int_{0.1}^{1.9} f(x)dx = \int_{0.1}^{1.9} \frac{1}{2.5}dx$$

$$= \frac{1}{2.5}(1.8) = 0.72 \rightarrow \boldsymbol{\beta = 0.72}$$

---

**Part 3.** We want to choose $c_1$ and $c_2$ such that the Type I error probability ($\alpha$) is exactly 0.05. Recall that $\alpha$ is the sum of the rejection probabilities in the two tails when $\theta = 2$. Let $c_1$ and $c_2$ be the new thresholds:

$$\alpha = P(X \leq c_1 \mid \theta = 2) + P(X \geq c_2 \mid \theta = 2)$$

For $X \sim U[0,2]$, $P(X \leq c_1) = \frac{c_1}{2}$ and $P(X \geq c_2) = \frac{2-c_2}{2}$. Set the total Type I error to 0.05:

$$\frac{c_1}{2} + \frac{2-c_2}{2} = 0.05 \rightarrow c_1 + 2 - c_2 = 0.1 \rightarrow c_2 = 1.9 + c_1$$

However, $c_1$ and $c_2$ must remain within $[0,2]$. By symmetry, the values that balance rejection in the tails are:

$$\boldsymbol{c_1 = 0.05}\,, \qquad \boldsymbol{c_2 = 1.95}$$

**Impact on Power:** Reducing $\alpha$ narrows the rejection region, making it harder to reject $H_0$. This decreases the **power** of the test $(1-\beta)$.

---

**Part 4.** As $\theta$ increases, the PDF for $X$ spreads over a wider range $[0,\theta]$, decreasing its height ($f(x) = 1/\theta$). The fixed rejection region $[0, c_1] \cup [c_2, \theta]$ captures a smaller proportion of the total distribution, decreasing $\beta$.

**Mathematical Argument:** The Type II error probability is the area of the acceptance region $(c_1, c_2)$ under the true distribution $f(x)$:

$$P(\text{Type I error}) = P(\text{accept } H_0 \mid H_0 \text{ false}) = P(\, c_1 \leq X \leq c_2 \mid \theta\,) = \int_{c_1}^{c_2} f(x)dx = \int_{c_1}^{c_2} \frac{1}{\theta}dx = \frac{c_2 - c_1}{\theta}$$

As $\boldsymbol{\theta}$ **increases,** $\boldsymbol{\beta}$ **decreases** because $\frac{1}{\theta}$ decreases.

---

**Part 5.** For distributions with finite variance, we can use **Chebyshev's inequality** to bound the probability of observing extreme values. For the uniform distribution $X \sim U[0,\theta]$, the variance is:

$$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0 \text{ true}) = P(X \leq 0.1 \mid \theta = 2) + P(X \geq 1.9 \mid \theta = 2)$$

**Chebyshev's inequality:** $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \rightarrow P(X \geq \mu + k\sigma) + P(X \leq \mu - k\sigma) \leq \frac{1}{k^2}$

$$\rightarrow \mu + k\sigma = 1.9, \mu - k\sigma = 0.1 \rightarrow \mu = 1\,, k\sigma = 0.9$$

$$X \sim U[0,\theta] \rightarrow \sigma^2 = \frac{\theta^2}{12} = \frac{1}{3} \rightarrow k = 0.9\sqrt{3}$$

$$\rightarrow P(|X-1| \geq 0.9) \leq \frac{1}{0.81 \times 3} = \frac{1}{2.43} \approx 0.4115$$

$\boldsymbol{\alpha \leq 0.4115} \rightarrow$ **The alpha value obtained in parts 1 and 3 is confirmed.**

## **Problem 7.** Warm Up $t$-test

**Part 1.** We are performing a **two-sample t-test** to test the null hypothesis $\boldsymbol{H_0 : \mu_1 = \mu_2}$ against the alternative hypothesis $\boldsymbol{H_1 : \mu_1 \neq \mu_2}$.

$$\text{Test statistic} = t = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Under the null hypothesis $H_0 : \mu_1 = \mu_2$, we have $\mu_1 - \mu_2 = 0$. Thus, the test statistic simplifies to:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

For a two-tailed test at significance level $\alpha$, the null hypothesis $H_0$ is rejected if the test statistic $t$ falls outside the critical values of the t-distribution:

$$t_{critical} < |t|$$

$t_{critical}$ is the critical value of the t-distribution at a significance level of $\alpha$ (two-tailed).

---

**Part 2.** The **confidence interval** for the difference $\mu_1 - \mu_2$ is given by:

$$\text{CI} = \left( (\bar{x}_1 - \bar{x}_2) - t_{critical}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} , (\bar{x}_1 - \bar{x}_2) + t_{critical}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

The role of $t_{critical}$ is to determine the margin of error for the confidence interval. Larger $t_{critical}$ values (e.g., at smaller significance levels) lead to wider confidence intervals, while smaller $t_{critical}$ values (e.g., at larger significance levels) lead to narrower intervals.

---

**Part 3.** We want to prove that $H_0$ is rejected if and only if the confidence interval for $\mu_1 - \mu_2$ does not include zero.

**If the t-test rejects $\boldsymbol{H_0}$:**

From the t-test, $H_0$ is rejected when $|t| > t_{critical}$. This implies that the observed difference $\bar{x}_1 - \bar{x}_2$ is so large (in absolute value) that it lies outside the range of plausible values under $H_0$.

For the confidence interval:

$$\text{CI} = \left( (\bar{x}_1 - \bar{x}_2) - t_{critical}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} , (\bar{x}_1 - \bar{x}_2) + t_{critical}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

If $|t| > t_{critical}$, then $0$ (the hypothesized difference under $H_0$) cannot lie within the interval because the interval endpoints are farther away from zero.

Thus, the confidence interval excludes zero if the t-test rejects $H_0$.

**If the confidence interval excludes zero:**

If $0$ is not within the confidence interval, then the observed difference $\bar{x}_1 - \bar{x}_2$ is so extreme that it cannot be explained by random sampling variability under $H_0$.

This implies that the t-statistic $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ must satisfy $|t| > t_{critical}$, because the interval endpoints are determined by $\pm t_{critical}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

Thus, if the confidence interval excludes zero, the t-test must reject $H_0$.

**Result:**

The confidence interval and the t-test are two equivalent approaches to hypothesis testing:

- The **t-test** directly evaluates whether the observed difference $\bar{x}_1 - \bar{x}_2$ is statistically significant based on the critical values of the t-distribution.
- The **confidence interval** assesses whether the null hypothesis value $0$ lies within the plausible range of values for the true difference $\mu_1 - \mu_2$.

Both methods lead to the same conclusion:

- $H_0$ is rejected if the confidence interval excludes $0$.
- $H_0$ is not rejected if the confidence interval includes $0$.

## Problem 8. $t$-test

We perform a t-test for the null hypothesis $\boldsymbol{H_0 : \mu = 10}$ at significance level $\boldsymbol{\alpha = 0.05}$ by means of a dataset consisting of $\boldsymbol{n = 16}$ elements with sample mean $\boldsymbol{\bar{x} = 11}$ and sample variance $\boldsymbol{s^2 = 4}$.

**Part 1.** First, we should compute the t-statistic:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{11 - 10}{\frac{2}{4}} = 2$$

Now, we determine the critical t-values for a two-tailed test for $\alpha = 0.05$:

$$t_{critical} = \pm 2.131$$

It is clear that the test statistic $t = 2$ falls within the range $-2.131 \leq t \leq 2.131$. Therefore, we **fail to reject $\boldsymbol{H_0}$** at the 5% significance level.

---

**Part 2.** First, we determine the critical t-values for a one-tailed test for $\alpha = 0.05$:

$$t_{critical} = 1.753$$

It is clear that the test statistic $t = 2$ is greater than $t_{critical} = 1.753$. Therefore, we **reject $\boldsymbol{H_0}$** in favor of $H_A : \mu > 10$ at the 5% significance level.

---

**Part 3.**

- **Why do the results of the two tests differ?**

The results differ because the two tests (one-tailed and two-tailed) evaluate different alternative hypotheses and, consequently, have different critical regions:

**Two-tailed test:** Splits the significance level ($\alpha$) equally between the two tails of the t-distribution. For this test, both large positive and large negative deviations from the null hypothesis are considered evidence against $H_0$.

**One-tailed test:** Concentrates the entire significance level ($\alpha$) in one tail of the t-distribution, making it easier to reject $H_0$ if the test statistic falls in the specified direction (here, $\mu > 10$).

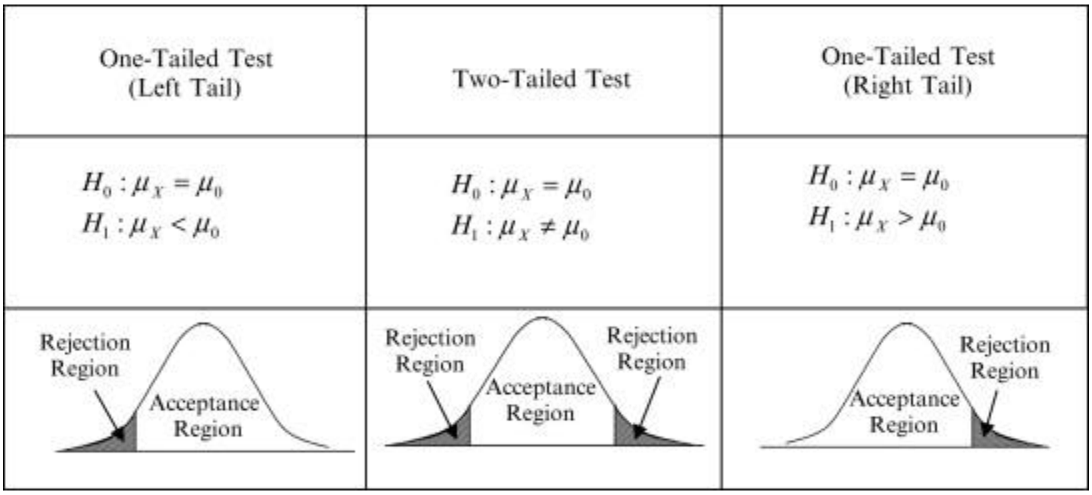


**Figure 8.1.** One and Two Tailed Tests

- **Relationship between critical regions of one-tailed and two-tailed tests (with illustration)**

The critical regions for one-tailed and two-tailed tests are related as follows:

- In a two-tailed test, the critical region is split equally between both tails of the t-distribution, so the threshold for rejection in each tail is more extreme (higher absolute value of $t$).
- In a one-tailed test, the entire critical region is concentrated in one tail, so the threshold for rejection is less extreme.

## **Problem 10.** An Apple a Day ... (Programming Question)

As it is shown in **figure 10.1**, The experiment confirmed that the confidence intervals constructed using different methods provide varying results. While the Z-distribution with known $\sigma$ and the T-distribution both give valid intervals, the T-distribution intervals are generally wider, reflecting the additional uncertainty when $\sigma$ is unknown. These findings highlight the importance of selecting appropriate methods for confidence interval estimation based on the available information about the population parameters.

```
True mean is in 99.00% of CIs in known-sigma scenario for Z-distribution
True mean is in 98.00% of CIs in unknown-sigma scenario for Z-distribution
True mean is in 99.00% of CIs in unknown-sigma scenario for T-distribution
CIs in known-sigma scenario for Z-distribution on average are: [1.70277916 2.26857203]
CIs in unknown-sigma scenario for Z-distribution on average are: [1.71586482 2.25548637]
CIs in unknown-sigma scenario for T-distribution on average are: [1.68268559 2.28866561]
Confidence intervals constructed with the Z-distribution using s instead of σ are narrower than those using the T-distribution.
```

**Figure 10.1.** Simulation Results

## Problem 11. Short Answers - Big Concepts

**Part 1.** To use a t-test, the following conditions need to be met:

- The data should be approximately normally distributed, especially for smaller sample sizes ($n < 30$).
- The sample size should be reasonably large or symmetric (for $n \geq 30$, normality is less critical due to the Central Limit Theorem).
- No significant outliers should be present.

---

**Part 2.** The manager wants to verify whether the mean burger weight matches the advertised weight of 250 grams.

**Hypothesis Testing:**

- **Null Hypothesis ($H_0$):** The mean weight of the burgers is 250 grams ($\mu = 250$).
- **Alternative Hypothesis ($H_1$):** The mean weight of the burgers is not 250 grams ($\mu \neq 250$).

Since the manager is checking for deviation in either direction (more or less than 250 grams), this is a **two-tailed test**.

---

**Part 3.** The fitness scientist wants to evaluate the program's impact on flexibility by comparing pre- and post-program flexibility scores for the same participants.

**Type of Test:** Since the same participants are measured before and after the program, this is a **paired t-test** (dependent samples).

**Hypothesis Testing:**

- **Null Hypothesis ($H_0$):** There is no difference in flexibility before and after the program ($\mu_{difference} = 0$).
- **Alternative Hypothesis ($H_1$):** The program improves flexibility ($\mu_{difference} > 0$).

This is a **one-tailed test** because the scientist is specifically interested in improvement (increase in flexibility).

---

## **Problem 12.** Speed Camera Error Rate

**Part 1.** Given that each camera has a normal measurement error of $N(0,25)$, the errors from the three cameras are independent. The mean of the three readings, $\bar{x}$, will also follow a normal distribution:

$$\bar{x} \sim N(\mu, \sigma_{\bar{x}}^2)$$

$\mu = 40$ (true speed limit under $H_0$)

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{25}{3}$

Thus:

$$\bar{x} \sim N\left(40, \frac{25}{3}\right)$$

**Hypothesis Testing Framework:**

- $H_0$: The driver is not speeding ($\mu = 40$).
- $H_A$: The driver is speeding ($\mu > 40$).

This is a **one-tailed test** where we test whether the average reading $\bar{x}$ exceeds the threshold under $H_0$.

---

**Part 2.** The threshold must be set so that no more than 4% of drivers traveling at exactly 40 mph (i.e., law-abiding, non-speeding drivers) are mistakenly issued tickets.

### a. Setting the Threshold

The rejection region for the test is where $\bar{x} > \text{Threshold}$ . For $\mu = 40$, we need:

$$P(\bar{x} > \text{Threshold} \mid H_0) = 0.04$$

Using the standard normal distribution $Z$, where:

$$Z = \frac{\bar{x} - 40}{\text{SD}(\bar{x})}$$

we want the $Z$-score such that $P(Z > z) = 0.04$. From standard normal tables:

$$z \approx 1.75$$

Thus, the threshold is:

$$\text{Threshold} = 40 + z \cdot \text{SD}(\bar{x}) = 40 + 1.75 \times \sqrt{\frac{25}{3}} = 40 + 1.75 \times 2.887 \approx 45.05$$

The police should issue tickets if $\bar{x} > 45.05$.

### b. Illustrating the Distribution and Rejection Area

The distribution of $\bar{x}$ under $H_0$ is $N(40, \frac{25}{3})$. The rejection area corresponds to values of $\bar{x} > 45.05$, representing the top 4% of the distribution.

$$if\ \bar{x} > 45.05 \rightarrow \text{reject } H_0$$

### c. Probability of Ticketing a Non-Speeder

This is the probability that the driver was not speeding, given that they received a ticket. Using Bayes' theorem:

$$P(H_0 \mid \text{ticket}) = \frac{P(\text{ticket} \mid H_0).P(H_0)}{P(\text{ticket})}$$

$$P(\text{ticket}) = \underbrace{P(\text{ticket} \mid H_0)}_{0.04}.\underbrace{P(H_0)}_{\frac{1}{2}} + \underbrace{P(\text{ticket} \mid H_1)}_{1}.\underbrace{P(H_1)}_{\frac{1}{2}} = 0.52$$

Thus:

$$P(H_0 \mid \text{ticket}) = \frac{P(\text{ticket} \mid H_0).P(H_0)}{P(\text{ticket})} = \frac{0.04 \times \frac{1}{2}}{0.52} \approx 0.038$$

### d. Probability of Tickets Given in Error When No One Speeds

If no one is speeding, all drivers are not speeding ($\mu = 40$). The probability of mistakenly issuing a ticket to a non-speeder is:

$$P(\bar{x} > 45.05 \mid H_0) = 0.04$$

Thus, if no one is speeding, **100%** of the tickets issued are in error.

---

## Part 3.

- **Power of the Test at $\mu = 45$**

To compute the power, we calculate the probability of rejecting $H_0$ when $\mu = 45$. This happens when $\bar{x} > 45.05$. Standardizing:

$$Z = \frac{\bar{x} - \mu}{\text{SD}(\bar{x})}$$

Under $H_A$, where $\mu = 45$:

$$Z = \frac{45.05 - 45}{2.887} \approx 0.0173$$

Using standard normal tables, the power is:

$$P(Z > 0.0173) \approx 0.493$$

- **Cameras Needed for Power of 0.9 and $\alpha = 0.04$**

To achieve a power of $0.9$, we need the probability of rejecting $H_0$ (when $\mu = 45$) to be at least $0.9$.

$$\alpha = 0.04 \rightarrow z_{1-\alpha} = 1.75$$

$$z_{power} = \Phi^{-1}(0.9) \approx 1.28 = \frac{\mu_A - \mu_0}{\text{SD}(\bar{x})} - z_{1-\alpha} = \frac{45 - 40}{\sqrt{\frac{25}{n}}} - 1.75 \rightarrow n \approx 9.2$$

Thus, approximately **10 cameras** are needed to achieve a power of $0.9$ with $\alpha = 0.04$.

# Problem 13. Statistically Delicious Coffee

## Part 1. Appropriate Statistical Test

**Comparison of Preparation Times Between Branch A and Branch B:**

- **Independent Samples t-Test:** Since the preparation times are independent between the two branches and we are comparing their means, an independent t-test is appropriate. We assume equal variances as specified in the question.

**Comparison of Taste Ratings Between Recipe A and Recipe B:**

- **Paired t-Test**: Since the same customers rate both recipes, the data is paired. A paired t-test is the correct test to analyze differences in taste ratings.

---

## Part 2. Preparation Times Between Branch A and Branch B

**Hypotheses**:

- $H_0 : \mu_A = \mu_B$ (No difference in preparation times between the branches)
- $H_0 : \mu_A < \mu_B$ (Branch A is faster than Branch B)

Table 13.1 shows the data related to the preparation times for Branch A and Branch B.

| Order | Branch A | Branch B |
|---|---|---|
| 1 | 6 | 7 |
| 2 | 7 | 6 |
| 3 | 7 | 7 |
| 4 | 8 | 8 |
| 5 | 6 | 9 |
| 6 | 9 | 7 |
| 7 | 5 | 10 |
| 8 | 7 | 9 |
| 9 | 8 | 9 |
| 10 | 7 | 9 |
| 11 | 5 | 7 |
| 12 | 7 | 9 |

**Table 13.1.** Preparation times for Branch A and Branch B

$$\bar{X}_A = \frac{\sum X_A}{n} = \frac{6+7+7+8+6+9+5+7+8+7+5+7}{12} = \frac{82}{12} \approx 6.833$$

$$\bar{X}_B = \frac{\sum X_B}{n} = \frac{7+6+7+8+9+7+10+9+9+9+7+9}{12} = \frac{97}{12} \approx 8.0833$$

$$s_A^2 = \frac{\sum(X_A - \overline{X_A})^2}{n_A - 1} \rightarrow (n_A - 1)s_A^2 = \sum(X_A - X_A)^2 = 15.666$$

$$s_B^2 = \frac{\sum(X_B - \overline{X_B})^2}{n_B - 1} \rightarrow (n_B - 1)s_B^2 = \sum(X_B - X_B)^2 = 16.917$$

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} = \frac{15.666 + 16.917}{12 + 12 - 2} = 1.481$$

$$SE = \sqrt{s_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} = \sqrt{1.481\left(\frac{1}{12} + \frac{1}{12}\right)} \to T = \frac{\bar{X}_A - \bar{X}_B}{SE} = \frac{6.833 - 8.0833}{\sqrt{\frac{1.481}{6}}} = -2.5159$$

$$\textbf{Test statistic} = \boldsymbol{T = -2.516}$$

For a one-tailed test at $\alpha = 0.05$, the critical value is:

$$T_{critical} = 2.201$$

**Decision**: Since $|T| = 2.516 > T_{critical} = 2.201$, we reject $H_0$.

So, there is sufficient evidence to conclude that **Branch A is faster than Branch B at the $\boldsymbol{\alpha = 0.05}$ level**.

---

## **Part 3.** Taste Ratings Between Recipe A and Recipe B

**Hypotheses**:

- $H_0 : \mu_D = 0$ (No difference in taste ratings between Recipe A and Recipe B)
- $H_0 : \mu_D > 0$ (Recipe B is rated higher than Recipe A)

Table 13.2 shows the data related to the taste ratings between Recipe A and Recipe B.

| Order | Branch A | Branch B | $D = B - A$ |
|---|---|---|---|
| **1** | 8 | 9 | 1 |
| **2** | 7 | 6 | -1 |
| **3** | 7 | 7 | 0 |
| **4** | 6 | 8 | 2 |
| **5** | 9 | 7 | -2 |
| **6** | 5 | 9 | 4 |
| **7** | 8 | 8 | 0 |
| **8** | 8 | 9 | 1 |
| **9** | 8 | 10 | 2 |
| **10** | 7 | 9 | 2 |
| **11** | 5 | 9 | 4 |
| **12** | 7 | 6 | -1 |

**Table 13.2.** Taste Ratings Between Recipe A and Recipe B

$$\bar{D} = \frac{\sum D}{n} = \frac{1 - 1 + 0 + 2 - 2 + 4 + 0 + 1 + 2 + 2 + 4 - 1}{12} = \frac{12}{12} = 1$$

$$s_D^2 = \frac{\sum(D - \bar{D})^2}{n - 1} = \frac{0 + 4 + 1 + 1 + 9 + 9 + 1 + 0 + 1 + 1 + 9 + 4}{11} \approx 3.6363 \to s_D \approx 1.91$$

$$SE = \frac{s_D}{\sqrt{n}} = \frac{1.91}{\sqrt{12}} \approx 0.552 \to T = \frac{\bar{D}}{SE} = \frac{1}{0.552} \approx 1.81$$

$$\textbf{Test statistic} = \boldsymbol{T = 1.81}$$

For a one-tailed test at $\alpha = 0.05$:

$$T_{critical} = 1.796$$

**Decision**: Since $T = 1.81 > T_{critical} = 1.796$, we reject $H_0$.

There is sufficient evidence to conclude that **Recipe B is rated higher than Recipe A at the $\boldsymbol{\alpha = 0.05}$ level.**