PODS Lab 9:
Parametric Significance Tests

Hamza Alshamy

Center for Data Science, NYU
ha2486@nyu.edu

03/21/25

1. Z-test
2. T-test
   - **Independent T-test – Between-group**
   - **Paired T-test – Within-group**
   - **Welch T-test**

Hypothesis Testing Framework

- **What is Hypothesis Testing? (A falsification approach)**
  - A statistical method used to make inferences about a population based on sample data.
  - It helps us determine whether an observed effect is real or due to random chance.
- **Hypothesis Testing Framework:**
  1. **Formulate Null (H_0) and Alternative (H_1) Hypotheses**
     - H_0: Assumption of no effect or no difference.
     - H_1: Assumption of an effect or a difference.
  2. **Choose a Significance Level (α)**
     - Common choices: $\alpha = 0.05, 0.01$.
  3. **Determine Test Statistic** $\implies$ Today's topic
     - Depends on data type and assumptions.
     - Examples: Z-test, T-test, KS test, Mann-Whitney U test.
  4. **Compute P-value**
     - **Reject H_0:** If $p \leq \alpha$ (statistically significant).
     - **Fail to reject H_0:** If $p > \alpha$ (not enough evidence).

Parametric Tests

- **What do we mean by "parametric?"**

Parametric Tests

- **What do we mean by "parametric?"**
  - Statistical tests that assume the data follow a specific distribution
  - Each of these parametric tests has a corresponding sampling distribution that their test statistics follow under the null hypothesis.

Parametric Tests

- **What do we mean by "parametric?"**
  - Statistical tests that assume the data follow a specific distribution
  - Each of these parametric tests has a corresponding sampling distribution that their test statistics follow under the null hypothesis.

Test	Test Statistic	Sampling Distribution (Under H_0)
Z-test	$Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	$\mathcal{N}(0,1)$ (Standard Normal)
T-test (One-sample)	$T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$	Student's t-distribution with $df = n - 1$
ANOVA (F-test)	$F = \dfrac{\text{Between-group variance}}{\text{Within-group variance}}$	F-distribution with $df_1 = k - 1$, $df_2 = N - k$

Table: Parametric tests and their distributions

Z-test

Normal Distribution ; Z-table

- **Use-case:** Use a Z-test when you want to assess whether a population mean μ is plausible given:
  - A known population standard deviation σ

Z-score

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

- **Algorithm:**
  1. Compute Z-score
  2. Look up the Z-score in a standard Z-table (Clickable link!).
  3. Determine the corresponding probability to the Z-score (P-value)

Z-test: Example 1 – One-Tailed Test (H_a smaller than)

Example: Inferring the U.S. female population height (μ) from a sample ($\bar{x}$)

- **Given:** $\sigma = 2$, $\bar{x} = 63$, $n = 50$, $\alpha = 0.01$
- **Hypotheses:**
  - $H_0 : \mu = 64$ inches (5'4")
  - $H_a : \mu < 64$ inches

Z-test: Example 1 – One-Tailed Test (H_a smaller than)

Example: Inferring the U.S. female population height (μ) from a sample ($\bar{x}$)

- **Given:** $\sigma = 2$, $\bar{x} = 63$, $n = 50$, $\alpha = 0.01$
- **Hypotheses:**
  - $H_0 : \mu = 64$ inches (5'4")
  - $H_a : \mu < 64$ inches
- **Z-score:**

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{63 - 64}{2/\sqrt{50}} \approx -3.54$$

Z-test: Example 1 – One-Tailed Test (H_a smaller than)

Example: Inferring the U.S. female population height (μ) from a sample ($\bar{x}$)

- **Given:** $\sigma = 2$, $\bar{x} = 63$, $n = 50$, $\alpha = 0.01$
- **Hypotheses:**
  - $H_0 : \mu = 64$ inches (5'4")
  - $H_a : \mu < 64$ inches
- **Z-score:**

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{63 - 64}{2/\sqrt{50}} \approx -3.54$$

- **P-value:** $P(Z \leq -3.54) = 0.00020 < \alpha = 0.01 \Rightarrow$ Reject H_0
- **Interpretation:** This provides evidence against H_0 and in favor of the claim that the population mean is less than 64 inches.

Z-test: Example 2 – One-Tailed Test (H_a greater than)

Example: Inferring the U.S. female population height (μ) from a sample ($\bar{x}$)

- **Given:** $\sigma = 2$, $\bar{x} = 63$, $n = 50$, $\alpha = 0.01$
- **Hypotheses:**
  - $H_0 : \mu = 64$ inches (5'4")
  - $\boxed{H_a : \mu > 64 \text{ inches}}$
- **Z-score:**

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{63 - 64}{2/\sqrt{50}} \approx -3.54$$

- **P-value:** $P(Z > -3.54) = 1 - 0.00020 = 0.9998 > \alpha = 0.01 \Rightarrow$ Fail to reject H_0
- **Interpretation:** We fail to reject H_0 and do **not** find statistical evidence for that mean height is greater than 64 inches.

Z-test: Example 3 – Two-Tailed Test (H_a: different than)

Example: Inferring the U.S. female population height (μ) from a sample ($\bar{x}$)

- **Given:** $\sigma = 2$, $\bar{x} = 63$, $n = 50$, $\alpha = 0.01$
- **Hypotheses:**
  - $H_0 : \mu = 64$ inches (5'4")
  - $\boxed{H_a : \mu \neq 64 \text{ inches}}$
- **Z-score:**

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{63 - 64}{2/\sqrt{50}} \approx -3.54$$

- **P-value:**

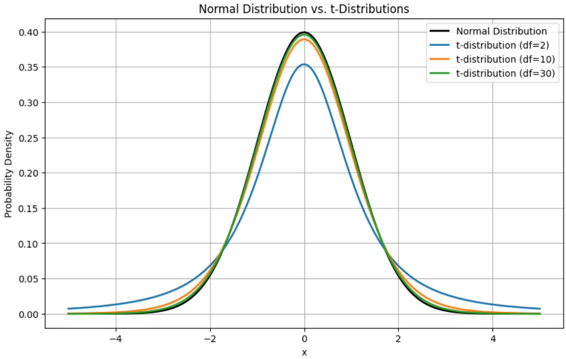
$$P = 2 \times P(Z \geq |-3.54|) = 2 \times 0.00020 = 0.00040 < \alpha = 0.01 \Rightarrow \text{Reject } H_0$$

- **Interpretation:** We reject H_0 and find statistical evidence that the population mean is **not equal to** 64 inches.

Why Do We Need Another Parametric Test? – The t-test

- **Limitation of the Z-test:** It requires knowing the population standard deviation (σ).
- **Why is the t-test helpful?**
  1. It does not require knowledge of the population standard deviation.
  2. It works well even with small sample sizes.
- **Common types of (two-sample) t-tests:**
  1. **Independent t-test (Between-group):** Comparing two separate groups.
  2. **Paired t-test (Within-group):** Comparing two related measurements on the same units.
  3. **Welch's t-test:** Adjusts for unequal variances across groups.

Student's t-distribution ; t-table



- The t-distribution has heavier ("fatter") tails compared to the normal distribution.
- It is governed by the number of **degrees of freedom (DF)**.
- As DF increases, the t-distribution approaches the normal distribution.

Degrees of Freedom (DF) – Parameter for the t-distribution

- **Degrees of Freedom (DF):** The number of independent pieces of information available for estimating a parameter.
- When we estimate population values using the sample, we often "double dip" – using the same data to estimate both the statistic and the variability around it.
- This constraint reduces how many values are truly free to vary, and that's what DF captures – the independent pieces of information

Degrees of Freedom Formula

$$DF = n - k \qquad \text{where:}$$

- n: Number of independent observations in the sample.
- k: Number of estimated statistics (e.g., sample mean).

Degrees of Freedom (DF) – Example

- **Sample:** $X = \{x_1, x_2, \ldots, x_{41}\}$, **Sample size:** $n = 41$
- Suppose you're calculating the **population standard deviation**:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

- We usually **don't know** the true population mean μ, so we estimate it with the **sample mean** $\bar{x}$:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- Because we used the sample data to estimate $\bar{x}$, we've "used up" 1 degree of freedom.
- So only $n - 1 = 40$ values are truly free to vary that's the **degrees of freedom**.

(1) Independent t-test – Between-Group

- **Use case:** Used to compare the means of **two independent groups**.
  - If we draw two samples and their sample means are far apart, it may be unlikely they came from the same underlying population.

Test statistic: Independent t-test – Between-Group

$$t = \frac{\bar{X}_1 - \bar{X}_2}{SEM_{pooled}}$$

- **Degrees of Freedom:**

$$DF = n_1 + n_2 - 2$$

- **Key Assumption:**
  - **Homogeneity of variance:** The variability within each sample is similar – only the means differ.
- **Scenario:** Group 1 receives a drug, Group 2 receives a placebo.

(2) Paired t-test – Within-group

- **Use case:** Used to compare means from the **same group** measured at two different times (e.g., before and after treatment), or under two related conditions.

  Test statistic: Paired t-test – Within-group

  $$t = \frac{\bar{D}}{SEM_D} \qquad , where$$

  - $\bar{D}$: Mean of the paired differences
  - $SEM_D = \frac{s_D}{\sqrt{n}}$: Standard error of the differences.
- **Degrees of Freedom:**

  $$DF = n - 1 \qquad \text{(where } n \text{ is the number of pairs)}$$

- **Scenario:** Measuring test scores of students **before and after** tutoring.

(3) Welch's T-test

- **Use case:** Used to compare the means of **two independent groups** when **variances across groups are not homogeneous** (most cases).

Test Statistc: Welch's T-test:

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

Welch Degrees of Freedom

$$DF = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

- Can have **fractional** DF.

T-test Summary

Test	Test Statistic	Degrees of Freedom (DF)
Independent t-test (pooled variance)	$t = \frac{\bar{X}_1 - \bar{X}_2}{SEM_{pooled}}$	$DF = n_1 + n_2 - 2$
Paired t-test (within-group)	$t = \frac{\bar{D}}{s_D/\sqrt{n}}$	$DF = n - 1$
Welch's t-test (unequal variance)	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	Adjusted (Welch-Satterthwaite)