Hypothesis Testing

- Testing Hypotheses (Chapter 9.1)
- The z-test (Chapter 9.1)
- T-tests (Chapter 9.5, 9.6)

Testing Hypotheses

- Let $X_1, \ldots, X_n \sim f(x; \theta)$. Suppose we want to know if $\theta = \theta_0$ or not, where θ_0 is a specific value of θ.
- If we are flipping a coin, we may want to know if the coin is fair; this corresponds to $\theta = 1/2$.
- If we are testing the effect of two drugs — whose means effects are θ_1 and θ_2 we may be interested to know if there is no difference, which corresponds to $\theta_1 - \theta_2 = 0$.
- We formalize this by stating a null hypothesis H_0 and an alternative hypothesis H_1. For example:

$$H_0 : \theta = \theta_0 \text{ versus } H_1 : \theta \neq \theta_0$$

- More generally, consider a parameter space Θ. We consider

$$H_0 : \theta \in \Theta_0 \text{ versus } H_1 : \theta \notin \Theta_0$$

Testing Hypotheses

- Let X be a random variable and let X be

$$H_0 : \theta \in \Theta_0 \text{ versus } \theta \in \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset$$

- Simple and Composite hypotheses:
  - If Θ_i contains only a single value, it is a simple hypothesis.
  - If Θ_i contains more than a single value, it is a composite hypothesis.
- One-sided vs two-sided hypotheses:
  - $H_0 : \theta \geq \theta_0, \quad H_1 : \theta < \theta_0.$
  - $H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0.$
  - If the null is simple, the alternative is usually two-sided:

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0.$$

Testing hypotheses

- You want to find a critical region, i.e., a subset of all possible data that lead to rejection of H_0.
- Define a statistic of your data $T(\mathbf{x})$.
- Define a rejection region (subset of the real line) for the statistic, e.g.

$$R = [c, \infty)$$

- c is the *critical value*.
- If $T(\mathbf{x}) \in R$ reject H_0.
- Otherwise do not reject H_0.
- Hypothesis testing is like a legal trial: Innocent unless the evidence strongly suggest they are guilty.

Example: Mean of normal distribution.

- $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, σ is known.
- We want to test:

$$H_0 : \mu \leq \mu_0 \text{ vs } H_1 : \mu > \mu_0$$

- Obvious relevant statistic: $\overline{X}_n$
- Define a rejection region for the sample mean:

$$R : \{\mathbf{x} : \overline{X}_n > \mu_0 + c, c > 0\}$$

- If $\overline{X}_n > \mu_0 + c$ reject H_0.
- Otherwise do not reject H_0.

The p-value

- We want to make a decision on whether we think H_0 or H_1 is correct.
- We want to quantify our belief.
- p-value: The probability of observing data at least as favorable to the alternative hypothesis as our current data set, **if the null hypothesis is true**.
- If the p-value is low (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence reject H_0.
- If the p-value is high (higher than the significance level, α) then it is pretty likely to observe the data even if the null hypothesis were true, so we do not reject H_0.

We never accept H_0 since we're not in the business of trying to prove it!

Example: Mean of the normal distribution:

- $X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1^2)$.
- $n = 25$
- For simplicity, assume that we want to test:

$$H_0 : \mu = 0 \text{ vs } \mu \neq 0$$

- Assume we observe $\overline{x}_n = 0.5$.
- The p-value is the probability of observing data at least as favorable to the alternative hypothesis as our current data set, **if the null hypothesis is true**.

$$P(|\overline{X}_n| > 0.5|\text{null is true}) = P(|\overline{X}_n| > 0.5|\mu = 0) = \ldots$$

Example: Mean of the normal distribution:

- If the true value of the mean is 0, then there is a 1.2% chance of observing a sample mean of 25 observations at least 0.5.
- This is a very low probability, so we can conclude that it could not have happened by chance.
- The difference between the null value of 0 and observed sample mean of 0.5 is very unlikely to be due to chance or sampling variability.
- Now assume that we have 10 observations.
- Find the p-value.

The p-value is **NOT** the probability that the null is true!

Recap:Mean of a population

- Set the hypotheses
  - H_0 : $\mu =$ null value or $\mu \geq \leq$ null value
  - H_1 : $\mu \neq$ null value or $\mu <>$ null value
- Assumptions
  - Independence of $X_1, \ldots, X_n$
  - Normality (of $\overline{X}_n$):
    - normal population
    - $n \geq 30 \rightarrow$CLT (also true for sample variance).
- Calculate a test statistic and a p-value: $Z = \frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n}$.
- Make a decision:
  - If p-value $\leq \alpha$, reject H_0.
  - If p-value $>\alpha$, do not reject H_0.

Question

- Last year, the average grade of students the first midterm of applied statistics was 68.7.
- This year, the first 50 students that were graded have an average grade of 63.7 with a standard deviation of 27.
- Is the average grade of the first midterm different this year than last year?
- Can we reject the null hypothesis?

$|P(\overline{X}_n - \mu) > |63.7 - \mu||H_0) =$

One-tailed vs two-tailed tests.

- Consider the following hypotheses
- Test 1: This year's grades are greater than last year's grades.
- Test 2: This year's grades are different than last year's grades.
- What is the difference between the two?

One sample t-test

- Assume that I only graded 10 midterm exams, so I cannot use the "large" sample approximation $s \to \sigma$.
- I assume that grades follow a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with unknown mean and standard deviation.
- What can I do?
- Use the fact: $\sqrt{n}\frac{\overline{X}_n - \mu}{s} \sim t_{n-1}$.
- This is the one-sample t-test, used for testing hypothesis about the mean of a normal distribution when the sample size is small.

  $P(|\overline{X}_n - \mu|) > |63.7 - \mu||H_0) =$

Paired t-test

- Now assume that I have the grades for the midterm X and the final Y for applied statistics.
- I want to see if students did better or worse at the finals.
- My samples are "paired": The same student i has both a X_i and a Y_i.
- I create the variable $X - Y$ and test if the mean is 0.
- This is called paired t-test.

Comparing the means of two distributions

- Assume that I want to compare the midterm grades of male and female students.
- I have graded 20 midterms from female students, and 30 midterms from male students.
- Let X denote the grades of female students, Y denote the grades of male students.
- We assume X and Y have the same unknown variance.

T-test

Theorem

Let $X_1, \ldots, X_m$ be m samples from a normal distribution with mean μ_x and variance σ^2, and $Y_1, \ldots, Y_n$ be n samples from a normal distribution with mean μ_y and variance σ^2. Let

$$S_x^2 = \sum_{i=1}^{m} (X_i - \overline{X}_m)^2, \quad S_y^2 = \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2.$$

Define the test statistic:

$$U = \frac{(m+n-2)^{1/2}(\overline{X}_m - \overline{Y}_n)}{(\frac{1}{m} + \frac{1}{n})^{1/2}(S_x^2 + S_y^2)^{1/2}}$$

Then if $\mu_x = \mu_y$, The distribution of U is the t-distribution with $m+n-2$ degrees of freedom.

T-test

- One-sided hypothesis:

$$H_0 : \mu_x \geq \mu_y \text{ vs } H_1 : \mu_x < \mu_y$$

  - p-value: $T_{m+n-2}(u)$
  - Rejection region: $U \leq T^{-1}_{m+n-2}(1\alpha)$.
- One-sided hypothesis:

$$H_0 : \mu_x \leq \mu_y \text{ vs } H_1 : \mu_x > \mu_y$$

  - p-value: $1 - T_{m+n-2}(u)$
  - Rejection region: $U \geq T^{i1}_{m+n-2}(1-\alpha)$.
- Two-sided hypothesis:

$$H_0 : \mu_x = \mu_y \text{ vs } H_1 : \mu_x \neq \mu_y$$

  - p-value: $2[1 - T_{m+n-2}(|u|)]$.
  - Rejection region: $|U| \geq T^{-1}_{m+n-2}(1-\alpha/2)$.

Unequal variances

- If the variances σ_x^2 and σ_y^2 are unknown, but $\sigma_x^2 = k\sigma_y^2$ where k is a known positive constant, then under $\mu_x = \mu_y$,

$$U = \frac{(m+n-2)^{1/2}(\overline{X}_m - \overline{Y}_n)}{(\frac{1}{m} + \frac{k}{n})^{1/2}(S_x^2 + \frac{S_y^2}{k})^{1/2}}$$

  follows a t distribution with $m + n - 2$ degrees of freedom.

Unequal variances

- If the variances σ_x^2 and σ_y^2 are unknown, the problem becomes very difficult.
- Welch's t-test:

$$V = \frac{\overline{X}_m - \overline{Y}_n}{(\frac{S_x^2}{m(m-1)} + \frac{S_y^2}{n(n-1)})^{1/2}}$$

- Let W=$\frac{S_x^2}{m(m-1)} + \frac{S_y^2}{n(n-1)})^{1/2}$, and use a Gamma distribution to approximate the distribution of W.
- If W has the approximating Gamma distribution, then V has the t-distribution with v degrees of freedom:

$$v = \frac{(\frac{s_x^2}{m(m-1)} + \frac{s_y^2}{n(n-1)})^2}{\frac{1}{(m-1)^3}(\frac{s_x^2}{m})^2 + \frac{1}{(n-1)^3}(\frac{s_y^2}{n})^2}$$

Meaning of significance.

Suppose :

- $\overline{X}_n = 50, s = 2$
- $H_0 : \mu \leq 49.5, H_1 : \mu > 49.5.$

Will the p-value be lower if n = 100 or n = 10000?

Meaning of significance.

Suppose :

- $\overline{X}_n = 50, s = 2$
- $H_0 : \mu \leq 49.9, H_1 : \mu > 49.9.$

Will the p-value be lower if n = 100 or n = 10000?

Practical vs statistical significance

- Real differences between the point estimate and null value are easier to detect with larger samples.
- However, very large samples will result in statistical significance even for tiny differences between the sample mean and the null value (effect size), even when the difference is not practically significant.
- This is especially important to research: if we conduct a study, we want to focus on finding meaningful results (we want observed differences to be real but also large enough to matter).