50008 - Probability and Statistics - Lecture 7

Oliver Killane

07/03/22

Hypothesis Testing

Lecture Recording

Lecture recording is available here

Definition: Hypothesis Test

Given two samples, determine if the difference is significant enough to suggest the parameters are different.

- **Null Hypothesis** No statistical relation, there is no evidence for a claim. (H_0)
- **Alternative Hypothesis** There is a statistical relation. (H_1)

We can partition the parameter space Θ into two disjoint sets Θ_0 and Θ_1 for the null and alternative hypotheses, which can be expressed as:

$$H_0 \ : \ \theta \in \Theta_0 \ \text{ and } \ H_1 \ : \ \theta \in \Theta_1$$

(We are testing if based on a given sample, based onm the estimated parameter, if it is plausible the sample distribution is from another distribution)

- **Simple Hypothesis** Test that $\theta = \theta_0$
- **Composite Hypothesis** Test that $\theta > \theta_0$ or $\theta < \theta_0$

Typically a test is of the form:

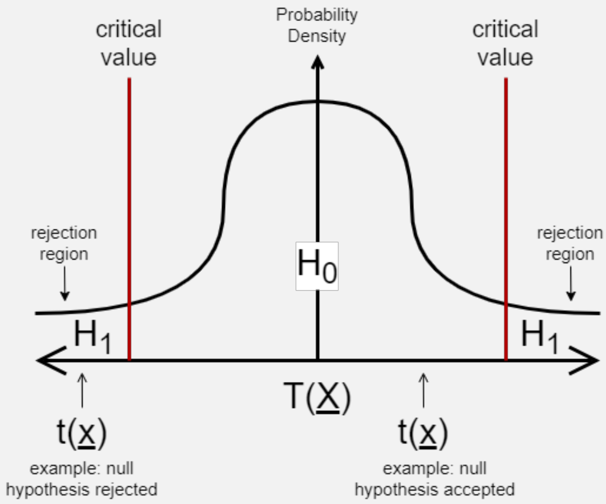
$$H_0 \ : \ \theta = \theta_0 \ \text{ versus } H_1 \ : \ \theta \neq \theta_0$$

Some tests are one-sided, for example:

$$H_0 \ : \ \theta > \theta_0 \ \text{ versus } H_1 \ : \ \theta < \theta_0$$

To test the validity of H_0:

1. Choose a **test statistic** $T(\underline{X})$ to use on the data.
2. Find a distribution P_T under H_0 from the **test statistic**.
3. Determine the rejection region (the region in which a result would invalidate H_0).
4. Calculate the observed **test statistics** $t(\underline{x})$.
5. If $t(\underline{x})$ is in the rejection region, reject H_0 and accept H_1, else retain H_0.



The **significance level/Type 1 Error Rate** $\alpha \in (0, 1)$ of as hypothesis test determines the size of the rejection regions.

- $\alpha \to 0$ Less and less likely to reject H_0, rejection region samller, confidence in our result is lower - easier test.
- $\alpha \to 1$ More and more likely to reject H_0, rejection region larger, confidence higher - stricter test.

The **p-value** of a test is the significance level threshold between rejection/acceptance of H_0 for a given test.

Definition: Test Errors

- **Type 1** Reject H_0 when it is actually true. $\alpha = P(T \in R|H_0)$
  (significance is the probability of incorrectly rejecting the null hypothesis)
- **Type 2** Accepting H_0 when H_1 is true. $\beta = P(T \notin R|H_1)$
  Probability a test statistic is not in the rejecting region, when H_1 is true.

Definition: Test Power

The probability of correctly rejecting the null hypothesis

$$Power = 1 - \beta = 1 - P(T \notin R|H_1) = P(T \in R|H_1)$$

For a given significance level:

$$\alpha = P(T \in R|H_0)$$

A good **test statistic** T and **rejection region** R will have a high **power**, the highest **power** test under H_1 is called the **most powerful**.

Example: Drug Effects

Given a control group (placebo) and a test group (given some pharmaceutical), we can test the hypotheis that the drug has an effect on survival rates.

$$H_0 : \text{The drug has no effect - survival rates are the same.}$$
$$H_1 : \text{The drug has an effect - survival rates are different.}$$

Testing For Population Mean

Lecture Recording

Lecture recording is available here

Sample mean belongs to a normal distribution (**Central Limit Theorem**):

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

We have our two hypotheses:

$$H_0 \ : \ \mu = \mu_0 \quad \text{versus} \quad H_1 \ : \ \mu \neq \mu_0$$

We can derive a new distribution in terms of the standard normal:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Hence for significance α (or confidence interval $1 - \alpha$) we can get the rejection/acceptance regions.

$$\Phi(1 - \alpha) = threshold \quad \text{results in acceptance region: } [-threshold, threshold]$$

Hence we can calculate z for a given sample, and then determine if it is in the region, if it is then accept H_0, else rejected H_0 and accept H_1.

Example: Weight of Crisp Packets (Known Variance)

A crisp manufacturer sells packets listed as having weight $454g$. From a sample size of 50, we get the mean weight of a bag as $451.22g$.

Assume the variance of bag weights is 70. Is the observed sample consistent with the claim made by the company at the 5% significance.

$$H_0: \ \mu = 454g$$
$$H_1: \ \mu \neq 454g$$

We have the following information:

$$\overline{x} = 451.22g \quad \sigma^2 = 70 \quad n = 50 \quad \alpha = 0.05$$

Hence we can state the hypothesized distribution of the sample mean:

$$\overline{X} \sim N\left(454g, \frac{70}{50}\right)$$

We can get this in terms of the standard normal distribution:

$$Z = \frac{\overline{X} - 454}{\sqrt{35}/5} \sim N(0,1)$$

At the 5% significance, we have 2.5% are each tail. Hence we get our critical value as $z(critical) = 0.975$, where 1.96.

Hence the rejection region is:

$$\frac{\overline{X} - 454}{\sqrt{35}/5} < -1.96$$
$$\frac{\overline{X} - 454}{\sqrt{35}/5} > 1.96$$

Hence in order to accept H_0, $\overline{X}$ must be in the interval:

$$451.6809 < \overline{X} < 456.3191$$

As $\overline{x} = 451.22$ it is in the rejection region, hence at the 95% significance there is sufficient evidence to reject the company's claim.

Example: Weight of Crisp Packets (UnKnown Variance)

crisp manufacturer sells packets listed as having weight $454g$. From a sample size of 50, we get the mean weight of a bag as $451.22g$.

Assume the variance of bag weights is 70. Is the observed sample consistent with the claim made by the company at the 5% significance.

$$\begin{aligned} H_0 &: \ \mu = 454g \\ H_1 &: \ \mu \neq 454g \end{aligned}$$

We have the following information:

$$\overline{x} = 451.22g \quad n = 50 \quad \alpha = 0.05$$

We first calculate the bias corrected sample variance:

$$\begin{aligned} s_{n-1} &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \\ &= \sqrt{70.502} \text{ (Need to calculate from each observation in the sample)} \end{aligned}$$

Hence we can now use the **student's t distribution** with degrees of freedom $n - 1 = 49$.

$$\frac{\overline{x} - \mu_0}{s_{n-1}/\sqrt{n}} \sim t_{49}$$

For $\alpha = 5\%$ we take the tails as 0.025, so use $t_{49,\ 0.975} \approx 2.01$. Hence will reject the regions:

$$\begin{aligned} \frac{\overline{X} - 454}{\sqrt{70.502}/5\sqrt{2}} &< -2.01 \\ \frac{\overline{X} - 454}{\sqrt{70.502}/5\sqrt{2}} &> 2.01 \end{aligned}$$

Hence to accept H_0, $\overline{X}$ must be:

$$451.6123 < \overline{x} < 456.3868$$

Hence at the 5% significance there is sufficient evidence to reject H_0 and accept H_1.

Example: Optimising Code

The previous code had a mean run time of $6s$. Following an optimisation a sample of runs is taken, with sample of size 16, mean $5.8s$ and bias corrected sample standard deviation of $1.2s$. Is the new code faster?

Our test is as follows:

$$H_0 \; : \; \mu \geq 6s \text{ (mean time is same or larger)} \quad \text{versus} \quad H_1 \; : \; \mu < 6s \text{ (mean time is lower)}$$

We have the following information:

$$\overline{x} = 5.8 \quad s_{n-1} = 1.2s \quad n = 16$$

Hence we have the distribution:

$$\frac{\overline{X} - \mu}{s_{n-1}/\sqrt{n}} \sim t_{15}$$

Hence we can use the significance (one ended/top tail) of 5% to find $t_{15,0.95} \approx 1.75$.

Hence will reject the regions:

$$\frac{\overline{X} - 6}{1.2/4} < -1.75$$

$$\frac{\overline{X} - 6}{1.2/4} > 1.75$$

Hence to accept H_0, $\overline{X}$ must be:

$$5.475 < \overline{X} < 6.525$$

Hence as $\overline{x} = 5.8$ this is within the acceptable region, so at the 95% significance we have insufficient evidence to reject H_0.

Samples from Two Populations

Lecture Recording

Lecture recording is available here

When given two random samples:

$$\underline{X} = (X_1, \ldots, X_n) \text{ from } P_X \quad \text{and} \quad \underline{Y} = (Y_1, \ldots, Y_n) \text{ from } P_Y$$

We may want to determine the similarity of the distributions of P_X and P_Y.

Typically this involves testing to see if the means of the populations are equal:

$$H_0 \; : \; \mu_X = \mu_Y \quad \text{versus} \quad H_1 \; : \; \mu_X \neq \mu_Y$$

Definition: Paired Data

A special case when $\underline{X}$ and $\underline{Y}$ are pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ (each X_i and Y_i are possibly dependent on each-other).

For example, where for a person i, X_i is the heart rate before exercise, and Y_i the rate afterwards.

We can consider a sample of the differences, if this has mean 0:

$$Z_i = X_i - Y_i \quad \text{testing } H_0 \ : \ \mu_Z = 0 \quad \text{versus} \quad H_1 \ : \ \mu_Z \neq 0$$

Known Variance, X and Y are Independent

Given that:

$$\underline{X} = (X_1, \dots, X_{n_1}) \quad X_i \sim N(\mu_X, \sigma_X^2) \quad \overline{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n_1}\right)$$
$$\underline{Y} = (Y_1, \dots, Y_{n_2}) \quad Y_i \sim N(\mu_Y, \sigma_Y^2) \quad \overline{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n_2}\right)$$

We can therefore get the distribution of the difference in sample means:

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}\right)$$

And hence:

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \sim N(0, 1)$$

As we assume for H_0 that $\mu_x = \mu_Y$ we have:

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \sim N(0, 1)$$

So we can calculate the **test statistic**:

$$z = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}}$$

And use this to determine if H_0 is rejected.

Unknown Variance, X and Y are Independent, Variances Equal

Lecture Recording

Lecture recording is available here

Definition: Bias-Corrected Pooled Sample Variance

If the variance of two samples is the same, given:

$$\underline{X} = (X_1, \ldots, X_{n_1}) \quad \text{and} \quad \underline{Y} = (Y_1, \ldots, Y_{n_2})$$

We can get an unbiased estimator of the variance as:

$$S^2_{N_1+n_2-2} \frac{(n_1-1)S^2_{n_1-1,\ X} + (n_2-1)S^2_{n_2-1,\ Y}}{(n_1-1)+(n_2-1)}$$

Which is equivalent to:

$$S^2_{n_1+n_2-2} = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2}{n_1+n_2-2}$$

If σ^2_X and σ^2_Y are unknown, but it is know that $\sigma^2 = \sigma^2_X = \sigma^2_Y$ we can use an estimator to get an estimate of the variance σ^2 using the samples from the two populations.

$$\frac{(\overline{X} - \overline{Y}) - (\mu_x - \mu_Y)}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

Hence if the $H_0 \ : \ \mu_X = \mu_Y$ then:

$$\frac{\overline{X} - \overline{Y}}{\sigma\sqrt{1/n_1 + 1/n_2}} \sim N(0,1)$$

To get an estimate for the variance we can use the **Bias-Corrected Pooled Sample Variance**

Example: Compiler Comparison

Given two compilers, attempt to determine if compiler 2 produces is faster code (to 5% significance).

Compiler 1	**Compiler 2**
$n_1 = 15$	$n_2 = 15$
$\overline{x} = 114s$	$\overline{y} = 94s$
$s_{14}^2 = 310$	$s_{14}^2 = 290$
μ_1	μ_2

$$H_0 \ : \ \mu_1 \leq \mu_2 \ \text{ versus } \ H_1 \ : \ \mu_1 > \mu_2$$

We assume that the variances of the population variances are the same for both compilers.

We can get the **Bias-Corrected Pooled Sample Variance**:

$$S_{28} = \frac{14 \times 310 + 14 \times 290}{14 + 14} = 300$$

Hence our **test statistic** is:

$$\frac{\overline{x} - \overline{y}}{\sigma\sqrt{1/n_1 + 1/n_2}} = \frac{20}{\sqrt{300}\sqrt{\frac{2}{15}}} = \sqrt{10} \approx 3.162$$

We can now use the **student's t distribution** to get the rejection region (one-sided):

$$t_{28,0.95} = 1.701$$

Hence as $3.162 > 1.701$ we have sufficient evidence at the 5% significance to reject H_0 and accept H_1. The second compiler produces faster code.

Welch's t-test

If the variances are unknwon, and not equal, we can use **Welch's t test**.

The **test statistic** is:

$$\frac{(\overline{x} - \overline{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S^2_{n_1,X}}{n_1} + \frac{S^2_{n_1,Y}}{n_1}}}$$

We then use a t distribution t_ν with the ν degrees of freedom determined by rounding the following to the nearest whole number:

$$\nu = \frac{\left(\left(\frac{S^2_{n_1,\ X}}{n_1}\right) + \left(\frac{S^2_{n_1,\ X}}{n_1}\right)\right)^2}{\left(\frac{1}{n_1 - 1}\right)\left(\frac{S^2_{n_1,\ X}}{n_1}\right)^2 + \left(\frac{1}{n_2 - 1}\right)\left(\frac{S^2_{n_2,\ Y}}{n_2}\right)^2}$$

The we proceed as normal, checking the test statistic is within the rejection regions.