STAT0008 Lecture 7
Hypothesis Testing: An Introduction

Dr. Aidan O'Keeffe

Department of Statistical Science
University College London

19th November 2018

Outline

- Neyman Pearson approach
- Types of hypotheses (simple and composite)
- Types of error (Type I and Type II)
- Significance level and power.
- General set-up for a hypothesis test.

Introduction

So far, we have concentrated mostly on inferential problems where the aim was to estimate some parameter of interest.

Another major area of statistical inference is **hypothesis testing**.

A **statistical hypothesis** is simply some assertion or conjecture about the distribution of one or more random variables.

A test of a statistical hypothesis is a rule or procedure with which we decide to reject or retain the conjecture made in a statistical hypothesis.

(Relevant?!) Example

This morning, students have taken their STAT0008 ICA test.

I (the lecturer) might be interested in whether or not there is any difference in mean test scores between undergraduate students and postgraduate students.

My feeling is that there is no difference.

I define: $\bar{X}_U$ to be the mean test score for undergraduates and $\bar{X}_P$ to be the mean test score for postgraduates.

Using the Central limit theorem, I assume that the distribution of $D = \bar{X}_U - \bar{X}_P$ is normal.

My conjecture is that the mean of D is zero. In words, this corresponds to my belief that there is no difference in mean test score between undergraduates and postgraduates.

Neyman Pearson Approach

What I have done here, is state something known as a **null hypothesis**.

In general, we follow an approach to hypothesis testing developed by Neyman and (Egon) Pearson (each of whom held academic positions at UCL in the early days of the Department of Statistical Science!)

Their approach supposes that we have a set of data $\mathbf{x} = (x_1, \ldots, x_n)^\mathsf{T}$ and we assume that these data have arisen from some density f.

We consider **two possible hypotheses** about f and then use **information from the data,** $\mathbf{x}$, to decide which hypothesis is true.

Neyman Pearson Approach

The first hypothesis is known as the **null hypothesis** (written H_0) and this is a conservative hypothesis that we shall assume true initially and then **reject only if evidence from the data, $\mathbf{x}$, is clear**.

The other hypothesis is known as the **alternative hypothesis** (written H_1) and this demonstrates a departure from the null hypothesis (i.e. a change in f) that we may be interested in establishing from the data $\mathbf{x}$.

What do we mean by 'departure from the null hypothesis'...? Essentially, this means some change in the distribution of $\mathbf{x}$ between H_0 and H_1.

Here, 'change' often means a change in a parameter value.

Neyman Pearson Approach

Often, we assume that f belongs to a particular parametric family (probability distribution), parameterised by $\theta \in \Theta$.

In other words, we would assume that our data $X_1, \ldots, X_n$ are such that $X_i \sim \mathcal{D}(\theta)$ and the pdf/pmf of X_i is $f(x_i; \theta)$.

Then we may wish to test a **hypothesis** about θ, stating our **null** and **alternative** hypotheses as

$$\mathsf{H}_0 : \theta \in \Theta_0 \quad \text{versus} \quad \mathsf{H}_1 : \theta \in \Theta_1$$

with $\Theta_0 \cap \Theta_1 = \emptyset$.

We may or may not have $\Theta_0 \cup \Theta_1 = \Theta$.

Simple and Composite Hypotheses

If Θ_0 (or Θ_1) consists of a single value, then the corresponding hypothesis is a **simple hypothesis**. With a simple hypothesis, the pdf/pmf, f, is completely specified.

For example, $\mathrm{H}_0 : \theta = 0$ is a simple hypothesis.

Otherwise, a hypothesis is a **composite hypothesis**

For example, $\mathrm{H}_1 : \theta > 0$ is a composite hypothesis.

Critical Region

In general, a hypothesis test consists of the following steps:

1. Decide on a distribution for the data (i.e. choose $X_i \sim \mathcal{D}(\theta)$).
2. State hypotheses (H_0 and H_1 – in terms of θ).
3. Assume that H_0 is true.
4. Decide whether or not the observed data, $\mathbf{x}$, are compatible with H_0.
5. Based on 4., either reject H_0 (and conclude that H_1 holds) or retain/do not reject H_0.

Steps 4. and 5. are defined by a **critical region**.

Critical Region

Critical Region (of a Hypothesis test)

The critical region of a hypothesis test is the region (or set) of values of $\mathbf{x}$ that correspond to the rejection of the null hypothesis, H_0. The critical region is written as $\mathcal{C}$ such that

$$\text{If } \mathbf{x} \in \begin{cases} \mathcal{C} & \text{then } \mathsf{H}_0 \text{ is rejected;} \\ \bar{\mathcal{C}} & \text{then } \mathsf{H}_0 \text{ is retained.} \end{cases}$$

Note that $\bar{\mathcal{C}}$ is the complement of $\mathcal{C}$.

Decisions and Errors

We have two possible decisions:

1. Retain (do not reject) H_0 and reject H_1.
2. Reject H_0 and accept H_1.

There are two possible errors that could occur:

- **Type I Error**: Reject H_0 when H_0 is true (i.e. a false positive).
- **Type II Error**: Retain H_0 when H_0 is false (i.e. a false negative).

The decision that we take (and the hypothesis test set-up) is based on the probabilities of these errors.

Type I and Type II Errors

A hypothesis test has the form

$$\mathsf{H}_0\colon \theta \in \Theta_0 \quad \text{versus} \quad \theta \in \Theta_1$$

The probability of a Type I error is

$$\begin{aligned}\alpha(\theta) &= \mathbb{P}(\text{Type I Error}) \\ &= \mathbb{P}(\mathbf{X} \in \mathcal{C} \mid \theta \in \Theta_0).\end{aligned}$$

The probability of a Type II error is

$$\begin{aligned}\beta(\theta) &= \mathbb{P}(\text{Type II Error}) \\ &= \mathbb{P}(\mathbf{X} \in \bar{\mathcal{C}} \mid \theta \in \Theta_1).\end{aligned}$$

Ideally, we should like to minimise both probabilities ($\alpha(\theta)$ and $\beta(\theta)$) simultaneously. But, in general, this is not possible.

Size/Significance Level

In general, we regard a Type I error as more serious than a Type II error.

As such, we fix an upper limit for $\alpha(\theta)$ and then aim to minimise $\beta(\theta)$, subject to the pre-specified upper limit on $\alpha(\theta)$.

The upper limit on $\alpha(\theta)$ is known as the **size** (or significance level) of the hypothesis test and we call the upper limit α where

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta) = \sup_{\theta \in \Theta_0} \mathbb{P}(\mathbf{X} \in \mathcal{C} \mid \theta).$$

Once we've chosen a value for α (e.g. $\alpha = 0.05$) then we concentrate on minimising $\beta(\theta)$.

Power of a Hypothesis Test

Minimising $\beta(\theta)$ is equivalent to maximising $1 - \beta(\theta)$.

The term $1 - \beta(\theta)$ is known as the **power** of the hypothesis test.

Hence, in maximising $1 - \beta(\theta)$ for a given size α, we are finding the **most powerful test** of size α for a given set of hypotheses.

For example, the (familiar?) two sample t-test for two independent sets of normally distributed variables is a common most powerful test.

In Lecture 8, we will examine how to find a most powerful test more closely.

Simple Null Hypothesis

Suppose that our hypotheses are

$$\mathsf{H}_0\colon \theta = \theta_0 \quad \text{versus} \quad \mathsf{H}_1\colon \theta \neq \theta_0.$$

Note: we have a simple null hypothesis.

A basic approach is

1. Assume H_0 is true.
2. Calculate a test statistic $t(\mathbf{x})$, based on an observed sample of data $\mathbf{x}$
3. Examine the value of $t(\mathbf{x})$, and decide whether or not this value is 'extreme', under the assumption of H_0.

P-value

Extreme values of t indicate a significant departure from the null hypothesis (and would imply that we should reject H_0).

Making use of the sampling distribution of the test statistic under H_0, we compute

$$p = \mathbb{P}(|T| > t \mid \mathsf{H}_0).$$

This probability is known as a **P-value**. In words, it is the probability of observing a sample $\mathbf{x}$ (or a more 'extreme' sample) under the assumption that H_0 is true.

A small P-value would imply that the probability of observing the data that you have, assuming H_0 is true, is small. In other words, the probability that our data are consistent with the null hypothesis is low – and we may wish to reject H_0.

Evidence

The natural question would be, how small ought the P-value to be for H_0 to be rejected?

Convention:

- A P-value < 0.01 might be regarded as **strong evidence** against H_0
- A P-value < 0.05 might be regarded as **sufficient evidence** against H_0.

But note: These notions depend very much on the problem being studied/sampling mechanism used (or the consequences of the errors).

Example

Suppose that we have $X_1, \ldots, X_n$ a set of iid $\mathcal{N}(\mu, 36)$ random variables and we wish to test

$$\mathsf{H}_0\colon \mu = 0 \;\text{ versus }\; \mathsf{H}_1\colon \mu = 5.$$

Our test statistic is $\bar{X}$ and we set up the test to take the form

$$\text{Reject } \mathsf{H}_0 \text{ if } \bar{X} > a$$

for some positive constant a.

The value of a will depend on the significance level and power of the test.

Example

We recall that $\alpha = \mathbb{P}(\text{Type I Error})$ and hence

$$\begin{aligned}\alpha &= \mathbb{P}(\bar{X} > a \mid \mu = 0) \\ &= \mathbb{P}\left(\frac{\bar{X}\sqrt{n}}{6} > \frac{a\sqrt{n}}{6}\right) \\ &= 1 - \Phi\left(\frac{a\sqrt{n}}{6}\right) \quad \text{since} \quad Z = \frac{\bar{X}\sqrt{n}}{6} \sim \mathcal{N}(0,1).\end{aligned}$$

Then, if z_α is the upper $100(1-\alpha)\%$ point of the standard normal distribution, we have

$$a = \frac{6z_\alpha}{\sqrt{n}}. \tag{1}$$

Example

Now, for β (the probability of a Type II error) we set

$$\begin{aligned}
\beta &= \mathbb{P}(\bar{X} \leq a \mid \mu = 5) \\
&= \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - 5)}{6} \leq \frac{\sqrt{n}(a - 5)}{6}\right) \\
&= \mathbb{P}\left(Z \leq \frac{\sqrt{n}(a - 5)}{6}\right) \quad \text{where } Z \sim \mathcal{N}(0, 1). \\
&= \Phi\left(\frac{\sqrt{n}(a - 5)}{6}\right)
\end{aligned}$$

Then we have

$$a = 5 - \frac{6 z_\beta}{\sqrt{n}}. \qquad (2)$$

Example

The two equations (1) and (2) imply that if we specify the size of the test (α) and we know the sample size n, then we can deterime both the power of the test β and the rejection region

$$\bar{X} > a.$$

The form of our test is then written as the decision rule

$$\text{Reject H}_0 \text{ if } \bar{X} > a.$$

Learning Outcomes

- Understand the Neyman Pearson approach to hypothesis testing.
- Understand and be able to define the following terms:
  - Simple and composite hypotheses.
  - Significance level and power.
  - Critical region.
  - P-value.
- Know how to set up and carry out a hypothesis test.