Testing Hypotheses

Aside from estimating unknown parameters, testing hypotheses on those parameters is the most important aspect of an empirical study.

Sampling distributions are ABSOLUTELY FUNDAMENTAL to testing.

Four steps to hypothesis testing procedure:

- Formulate two opposing hypotheses: H_0 and H_A
- Compute the sample mean or variance or other quantity of interest
- Derive a test statistic and identify its sampling distribution when H_0 is true
- Derive a decision rule

These steps are explained in more detail below:

The **null hypothesis** (H_0) is the assumption we anticipate rejecting. The **alternative hypothesis** (H_A) describes the population if the null hypothesis is not true.

Examples of H_0 and H_A formulated on the mean of a population:

	(a)	(b)	(c)
H_0	$\mu = \mu_0$	$\mu \leq \mu_0$	$\mu \geq \mu_0$
H_A	$\mu \neq \mu_0$	$\mu > \mu_0$	$\mu < \mu_0$

Once we specify the null and alternative hypotheses, we collect our sample data. We then compute our **test statistic**. A **test statistic** is the estimator that will be used to either "**reject** the null hypothesis" or "**do not reject** the null hypothesis".

Once we have a test statistic we need to derive the its sampling distribution **under the null hypothesis**. We need this distribution in order to determine how likely our test statistic value is when the null hypothesis is true.

The final step is to derive a decision rule based on the observed value of the test statistic.

The range of values for which the test procedure recommends rejecting the null hypothesis (unlikely value of the test statistic) is called the **critical region** and the range for which it recommends not rejecting the null hypothesis (likely value of the test statistic) is called the **nonrejection region**.

Type I and Type II errors

For any test procedure three outcomes are possible:

- A correct decision
- Rejecting H_0 when it is true (**Type I error**)
- Not rejecting H_0 when it is false (**Type II error**)

Associated with each of these errors is a probability. The probability of committing a Type I error is denoted as $\boldsymbol{\alpha}$ and is also referred to as the **significance level of the test**. The probability of committing a Type II error is denoted as $\boldsymbol{\beta}$. The **power of the test**, defined as $\mathbf{1-\boldsymbol{\beta}}$, is the probability of rejecting H_0 when it is false.

Ideally, we would like to keep the probability of both these errors as low as possible. Unfortunately, an attempt to reduce the probability of a type I error automatically increases the probability of a type II error.

In practice, for hypothesis testing, we choose a maximum value for the type I error that is acceptable to us (say 5%) and then derive the decision rule for which the type II error is a minimum.

Let's consider the task of testing the mean of a normal distribution.

Consider a random variable X that is normally distributed with mean μ and variance σ^2. The most common null hypothesis is of the form H_0: $\mu = \mu_0$. The alternative H_A may be **one-sided** (or **one-tailed**) as in H_A: $\mu > \mu_0$ or **two-sided** (or **two-tailed**) as in H_A: $\mu \neq \mu_0$.

A one-sided test

Suppose we are interested in testing H_0: $\mu = \mu_0$ vs H_A: $\mu > \mu_0$.

We obtain a random sample $X_1,\ X_2, \dots, X_n$.

We know that the sample mean $\bar{X}$ is a good estimator of μ. So, if the observed $\bar{X}$ is considerably larger than μ_0, we would suspect that the true μ is probably larger than μ_0.

The sampling distribution of $\bar{X}$ tell us the probability that a sample of size *n* would give a sample average of what we observe.

The sampling distribution of $\bar{X}$ when H_0 is true (i.e. $\mu = \mu_0$): $\bar{X} \sim N(\mu_0, \sigma^2/n)$.

It is difficult however to compute this probability without standardizing the random variable $\bar{X}$.

Using $\bar{X} \sim N(\mu_0, \sigma^2/n)$, we can standardized $\bar{X}$ as

$$Z = \frac{(\bar{X} - \mu_0)}{\sigma/\sqrt{n}}$$

This variable is distributed as $N(0,1)$.

We are looking for very large value of Z ($\bar{X}$ far from μ_0) to reject the null hypothesis.

Unfortunately, we do not know σ and so we need to estimate it with the sample standard deviation (s). The actual test statistic we use is now

$$t = \frac{(\bar{X} - \mu_0)}{s/\sqrt{n}} \sim t_{n-1}$$

Why is this true?

As this sampling distribution is known, Table B-1 in the text can be used to pin down what “large” means in a statistical sense.

Look up the entry corresponding to n – 1 degrees of freedom and the given level of significance α and obtain the **critical value**, $t_{n-1,\alpha}$.

Reject H_0 if the observed t is greater than the critical value, otherwise do not reject H_0.

Example:

An insurance company needs to estimate the average amount claimed by its policyholders over one year. A random sample of 81 policyholders reveals that the sample mean is $839.98 and the sample standard deviation is $312.70. Suppose the insurance analyst wants to test the hypothesis that the average claim is more than $800. Test at the 5% level. Assume that the amount claimed is distributed normally with a mean μ and variance σ^2.

A two-sided test

Suppose we are instead interested in testing $H_0: \mu = \mu_0$ vs $H_A: \mu \neq \mu_0$.

Again, we obtain a random sample $X_1,\ X_{2,} \ldots, X_n$.

As before we use the test statistic

$$t = \frac{(\bar{X} - \mu_0)}{s/\sqrt{n}} \sim t_{n-1}$$

If the observed sample mean $\bar{X}$ deviates substantially from the null hypothesis $\mu = \mu_0$, the calculated *t* statistic will be either too large **or** too small. When this is the case, we reject H_0.

From the *t*-table B-1 look up the entry corresponding to *n* – 1 degrees of freedom and the given level of significance $\alpha/2$ and obtain the **critical value**, $t_{n-1,\alpha/2}$.

Reject H_0 if if the observed *t* is greater than the critical value or less than the negative of the critical value, otherwise do not reject H_0.

Example:

The label on a carton of light bulbs states that the bulbs are “long-life” with an average life of 935 hours. An unhappy customer does not believe this claim. She files a complaint alleging that the claim is false, in her experience the bulbs last longer or shorter than the claimed 935 hours. An analyst at the complaint’s department tested a random sample of 25 bulbs and found that the average life of the bulbs was 917 hours with a standard deviation of 54 hours.

Can the analyst reject the company’s claim? Assume that the life of a bulb is distributed normally with a mean μ and variance σ^2. Test at the 5% level.

P-values

We could have conducted our hypothesis test using a **p-value**. A **p-value** is the probability of observing a value of a test statistic as large as we did when the null hypothesis is true. It is also the largest probability of a Type I error when H_0 is true.

Interval Estimation

The estimation we have considered up to this point has given us a single estimated value for the unknown parameters of a distribution. These estimators are therefore called **point estimators**.

Rather than using a point estimator, we can use an **interval estimator** to provide a range of possible values for the population quantity with a certain probability.

Let's think about constructing a confidence interval for the mean of a normal distribution. We know that if a random variable X is distributed as $N(\mu, \sigma^2)$ then the sample mean $\bar{X}$ is distributed as $N(\mu, \sigma^2/n)$. Further we know that

$$t = \frac{(\bar{X} - \mu)}{s/\sqrt{n}} \sim t_{n-1}$$

Now if $t_{n-1,\alpha/2}$ represents the point on this t-distribution such that the area to the right of $t_{n-1,\alpha/2}$ is $\alpha/2$ and the area to the left of $-t_{n-1,\alpha/2}$ is $\alpha/2$, then

$$\Pr\left[-t_{n-1,\frac{\alpha}{2}} \leq \frac{(\bar{X}-\mu)}{s/\sqrt{n}} \leq t_{n-1,\frac{\alpha}{2}}\right] = 1-\alpha$$

Multiplying through by $s/\sqrt{n}$ and rearranging terms gives us

$$\Pr\left[\bar{X} - \left(\frac{s}{\sqrt{n}}\right) t_{n-1,\alpha/2} \leq \mu \leq \bar{X} + \left(\frac{s}{\sqrt{n}}\right) t_{n-1,\alpha/2}\right] = 1-\alpha$$

This interval is known as the $(1-\alpha)\%$ confidence interval for μ.

What does this mean?

Choice of confidence level is up to researcher.

Example:

A relationship exists between two-tailed tests and confidence intervals:

We reject the null hypothesis if the confidence interval does not include the value of the parameter in the null hypothesis. We do not reject the null hypothesis if the confidence interval includes the value of the parameter in the null hypothesis.