STAT 210
Applied Statistics and Data Analysis
Hypothesis Tests

Joaquín Ortega
KAUST

Fall 2020

Hypothesis Tests

Hypothesis Tests

A **hypothesis** is an assumption about the value or values of a parameter or parameters of the population.

Hypotheses are formulated in mutually exclusive pairs. If possible, hypotheses are chosen so that they are exhaustive (i.e., their union covers all possible outcomes of an experiment). This forces us to make a choice.

If θ is the parameter in question and Θ is the parameter space, then the **null hypothesis** defines the region $\{\theta \in \Theta_0\}$ while the **alternative hypothesis** describes $\{\theta \in \Theta_1\}$.

These regions are mutually exclusive.

Hypothesis Tests

When a hypothesis completely specifies the distribution of the population, we say that the hypothesis is **simple**.

Any hypothesis that is not simple is **composite**.

We will only consider simple null hypotheses, while the alternatives will usually be composite.

Null hypothesis	Alternative hypothesis	Type of alternative
	$H_1 : \theta < \theta_0$	lower one-sided
$H_0 : \theta = \theta_0$	$H_1 : \theta > \theta_0$	upper one-sided
	$H_1 : \theta \neq \theta_0$	two-sided

Hypothesis Tests

In the Neyman-Pearson approach, one must choose one of the two alternatives using information from a sample, from which a test statistic is computed.

The sample space is split into two regions,

- $\mathcal{R}$, known as the **rejection** region, and
- $\mathcal{R}^c$, known as the **acceptance** region.

The value of θ that separates these two regions is known as the **critical** value.

The test statistic is computed from the sample, and if it falls in the acceptance region, the null hypothesis is accepted; otherwise, the alternative hypothesis is accepted.

Hypothesis Tests

R. Fisher proposed a different approach.

In this approach, one determines how much evidence there is in the sample against the null hypothesis.

The null hypothesis is not accepted, it is just ‘not rejected’.

The test will determine if the sample collected can be due to chance alone under the null hypothesis. If this is not likely, the researcher has evidence to reject the null hypothesis.

A test with these characteristics is called a **significance test**.

Level and Power of a Test

Types of error

Our decision is always subject to error. There are two types of error:

Type I: Reject H_0 when it is true

Type II: Fail to reject H_0 when it is false.

State of Nature	H_0 not rejected	H_0 rejected
H_0 is true	No error	**Type I error**
H_0 is false	**Type II error**	No error

Types of error

Each error has a probability associated with it.

We denote by α the probability of a Type I error and by β the probability of a Type II error.

α is known as the **level** or **significance level** of the test.

For a fixed sample size, α and β go in opposite ways: the smaller α, the larger β, so we cannot reduce both at the same time unless we change the sample size.

p-value

The p-value for a hypothesis test is defined as the probability of observing a difference (or value) as extreme or more extreme than the observed difference (or value) under the assumption that the null hypothesis is true.

The smaller the p-value, the stronger the evidence against the null hypothesis.

Calculation of p-values for continuous distributions.

	p-value
$H_1 : \theta < \theta_0$	$P(t \leq t_{obs} \vert H_0)$
$H_1 : \theta > \theta_0$	$P(t \geq t_{obs} \vert H_0)$
$H_1 : \theta \neq \theta_0$	$2 \min\{P(t \leq t_{obs} \vert H_0), P(t \geq t_{obs} \vert H_0)\}$

Power

Consider

$$H_0 : \theta = \theta_0 \qquad vs \qquad H_1 : \theta = \theta_1$$

where both hypotheses are simple.

The **power** of this test is the probability of rejecting H_0 when H_1 is true.

$$P(\text{Reject } H_0 | \theta = \theta_1) = 1 - P(\text{Accept } H_0 | \theta = \theta_1) = 1 - \beta(\theta_1)$$

The power reflects the capacity of the test to detect the alternative hypothesis when it is true.

Observe that we need to know the sampling distribution for the test statistic **under the alternative hypothesis**.

If instead of a simple alternative we have a composite hypothesis $H_1 : \theta \in \Theta_1$, we consider the power as a function of θ for values $\theta \in \Theta_1$.

Shiny App

Shiny app for hypothesis tests:

https://casertamarco.shinyapps.io/power/

Example (from Ugarte et al.)

Test the null hypothesis that for a certain age group, the mean score on an achievement test is equal to 40 against the alternative that it is not equal to 40. Scores follow a normal distribution with $\sigma = 6$.

(a) Find the probability of type I error for $n = 9$ if the null hypothesis is rejected when the sample mean is less than 36 or greater than 44.

(b) Find the probability of type I error for $n = 36$ if the null hypothesis is rejected when the sample mean is less than 38 or greater than 42.

(c) Plot the power functions for $n = 9$ and $n = 36$ for values of μ between 30 and 50.

Example

Let $\bar{X}_n$ denote the sample mean when the sample size is n. We know that

$$\bar{X}_n \sim N(\mu, \sigma^2/n) = N(\mu, 36/n)$$

For n given, this distribution depends only on μ.

1. In the first question, we reject H_0 when

$$\bar{X}_9 < 36 \quad \text{or} \quad \bar{X}_9 > 44.$$

Example

Therefore, the probability of a type I error, α is

$$\begin{aligned}
\alpha &= P(\{\bar{X}_9 < 36\} \cup \{\bar{X}_9 > 44\} | \mu = 40) \\
&= P(\{\bar{X}_9 < 36\} | \mu = 40) + P(\{\bar{X}_9 > 44\} | \mu = 40) \\
&= P\Big(\frac{\bar{X}_9 - 40}{6/\sqrt{9}} < \frac{36 - 40}{6/\sqrt{9}}\Big) + P\Big(\frac{\bar{X}_9 - 40}{6/\sqrt{9}} > \frac{44 - 40}{6/\sqrt{9}}\Big) \\
&= P(N(0,1) < -2) + P(N(0,1) > 2) \\
&= 2P(N(0,1) < -2)
\end{aligned}$$

```r
2*pnorm(-2)
```

```
## [1] 0.04550026
```

Example

2. In the second question, we reject H_0 when

$$\bar{X}_{36} < 38 \quad \text{or} \quad \bar{X}_{36} > 42.$$

A similar calculation as before shows that, in this case, we get the same value for α.

3. The power function for $n = 9$ is

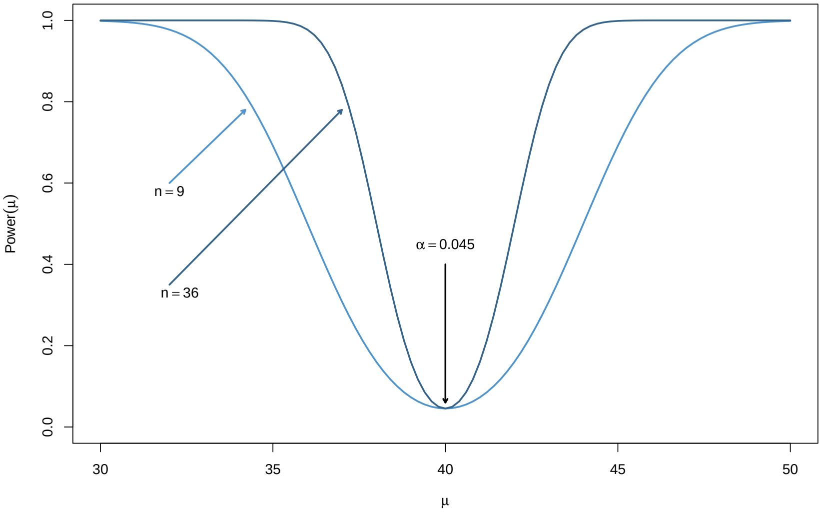
$$\begin{aligned} Power(\mu) &= P(\bar{X}_9 < 36 | N(\mu, \frac{36}{9})) + P(\bar{X}_9 > 44 | N(\mu, \frac{36}{9})) \\ &= P(\frac{\bar{X}_9 - \mu}{2} < \frac{36 - \mu}{2}) + P(\frac{\bar{X}_9 - \mu}{2} > \frac{44 - \mu}{2}) \\ &= P(N < \frac{36 - \mu}{2}) + P(N > \frac{44 - \mu}{2}) \end{aligned}$$

Example

while the power function for $n = 36$ is

$$\begin{aligned} Power(\mu) &= P(\{\bar{X}_{36} < 38\} | N(\mu, \frac{36}{36})) + P(\{\bar{X}_{36} > 42\} | N(\mu, \frac{36}{36}) \\ &= P(\{\frac{\bar{X}_{36} - \mu}{1} < \frac{38 - \mu}{1}\}) + P(\{\frac{\bar{X}_{36} - \mu}{1} > \frac{42 - \mu}{1}\}) \\ &= P(N < 38 - \mu) + P(N > 42 - \mu) \end{aligned}$$

Example



Example

The following code produces the previous graph.

```
power.ex <- function(x,n,a){
  1-pnorm(40+a, x,6/sqrt(n)) + pnorm(40-a, x,6/sqrt(n))}
curve(power.ex(x,9,4),30,50, ylab=expression(Power(mu)),
      xlab=expression(mu), ylim=c(0,1), lwd=2,
      col='steelblue3')
curve(power.ex(x,36,2),30,50,add = TRUE, lwd=2,
      col='steelblue4')
arrows(32, 0.6 , 34.2, .78, lwd=2, length=0.05,
       col='steelblue3')
arrows(32, 0.35 , 37, .78, lwd=2, length=0.05,
       col='steelblue4')
arrows(40, 0.4 , 40, 0.06, lwd=2, length=0.05)
text(32,0.58, expression(n==9))
text(32.3,0.33, expression(n==36))
text(40,0.45, expression(alpha==0.045))
```