# Hypothesis Testing

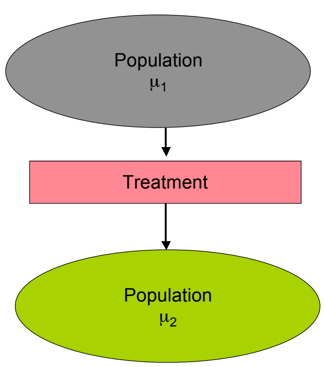
I. Introduction
II. Two Hypotheses
III. Selection of alpha
IV. Two Types of Errors
V. Directionality
VI. Four Steps
VII. Examples
VIII. Power

# I. Introduction

Based on statements of probability we make conclusions concerning populations.

For example, suppose I hit a coin with my "magic" hammer. I then toss the coin 4 times and get four heads in a row.

p(4 heads in 4 tosses of a coin) = .0625
We might want to conclude that the population of coin tosses with this new coin does not have a mean = np, where p = .5.




If $\mu_1 \neq \mu_2$ then we would conclude that there was a treatment effect

## II. Two Hypotheses

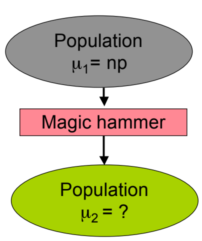
$H_O$ = null hypothesis
No treatment effect
$\mu_1 = \mu_2$

$H_1$ = alternative hypothesis
There is a treatment effect
$\mu_1 \neq \mu_2$

## II. Two Hypotheses

A statistical decision is reached by choosing between these two hypotheses.

We do this by calculating the probability that the null hypothesis is true. If this probability is small, we reject the null hypothesis, and accept the alternative hypothesis.




Example:

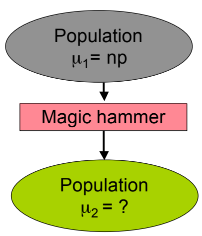
Toss my coin 4 times, I expect 2 heads (4*.5). But I get 4. The probability that I get 4 if $\mu_2$= np is small, p = .0625. Do we reject the null hypothesis?

# Selection of alpha

Definition: Alpha ($\alpha$) is the probability that we select beyond which we will reject $H_0$.

We choose a small level of $\alpha$ so as to minimize our chances of incorrectly rejecting $H_0$.

In social sciences, we usually set $\alpha$ = .05 or smaller. In medical sciences, we choose $\alpha$ = .01 or smaller.




Back to my magic coin. p = .0625. Do we reject the null hypothesis?

Not if we select $\alpha$ = .05. That is because .0625 > .05.

As a result, we do not reject $H_0$. We conclude that I do not have a magic hammer.

# Two Types of Errors

Conclusions

| | | Reject $H_0$ (treatment effect) | Accept $H_0$ (no treatment effect) |
|---|---|---|---|
| Actual State of the real world. | There is a treatment effect. | Correct Decision | Type II error $\beta$ |
| | There is not a treatment effect. | Type I error $\alpha$ | Correct Decision |

## Two Types of Errors

Type I errors (concluding there is a treatment effect when there is not) are usually consider the more costly.

- expense of useless treatments
- possible negative side effects of treatments.

As a result we directly control type I errors.

$\alpha$ = the probability of a Type I error.

## Two Types of Errors

Conclusions

| | | Magic coin | Fair coin |
|---|---|---|---|
| Actual State of the real world. | Magic coin. | Correct Decision | Type II error |
| | Fair coin (no treatment effect). | Probability of a Type I error, p = .0625 | Correct Decision |

Because the p(Type I error) > .05, we conclude that the coin is fair.

## Legal system as an analogy.

Conclusions

| | | Guilty | Not Guilty |
|---|---|---|---|
| Actual State of the real world. | Did it. | Correct Decision | Type II error |
| | Did not do it. | Type I error | Correct Decision |

Type I error (sentencing an innocent person to jail) is usually considered the worse error, as a result we have a presumption of innocence.

## Directionality

When evaluating the probability of a type I error, one has to decide between directional (one tailed) and non-directional (two tailed) tests.

If you do not know whether the treatment should raise or lower the scores, use a two tailed (non-directional) test.

For example, if I hit my coin with a hammer, I don't know if that will increase or decrease the probability of a head.

## Two Tailed tests:

$H_O$ = null hypothesis

No treatment effect

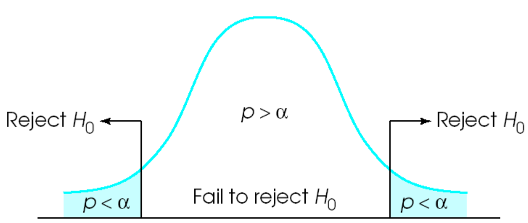
$\mu_1 = \mu_2$

$H_1$ = alternative hypothesis

There is a treatment effect

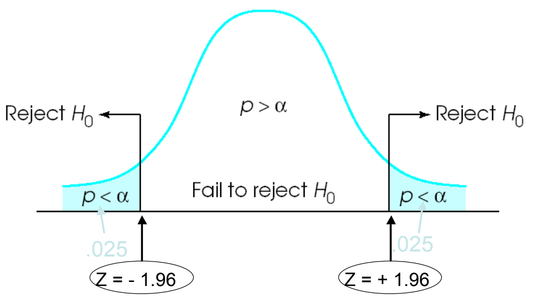
$\mu_1 \neq \mu_2$

## Two Tailed tests:




Split $\alpha$ *in half, and put 1/2 in each tail.*

## Two Tailed tests:




## Directionality

If I am confident as to the direction of the treatment effect (e.g., I know it will raise the scores), the I should use a 1 tailed test.

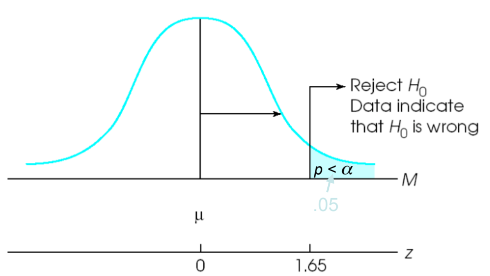
$H_O : \mu_1 = \mu_2$

$H_{1:}\ \mu_1 > \mu_2$

or

$H_{1:}\ \mu_1 < \mu_2$

## One tailed test.



## Four Steps in Hypothesis Testing

1. State $H_0$, $H_1$, and choose $\alpha$
2. Determine what type of observation it would take to reject $H_0$
   a) What is the appropriate test statistic?
   b) What is the critical value?
3. Evaluate the sample data
4. Reach a conclusion

## Examples

The simplest case: Known $\mu$, and $\sigma$

*Example 1: The effect of Project Help on academic performance.*

Population of disadvantaged children:

$\mu$ = 2 years below grade level

$\sigma$ = .5 years

Sample of n = 1 child, after project help, who is only 1 year below grade level.

Is there a treatment effect? Use $\alpha = .05$

## Examples

Rework the problem, with $n$ = 25, and a sample mean $M$ = 1.5.

Example 2:

Adult males have a life expectancy, $\mu = 74$ *years,* with $\sigma = 15$.

A sample, $n = 16$, is encouraged to exercise regularly. This sample lived on average $M$ = 79 years.

Test the hypothesis that exercise increases longevity. Use the .01 level of significance.

## Examples

Rework this example, with the same sample mean, but assume n = 100.

What do these examples demonstrate about hypothesis testing and sample size?

## Power

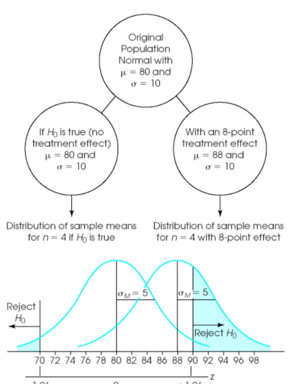
Power = the ability to detect a treatment effect when there is a treatment effect.

Power increases with:

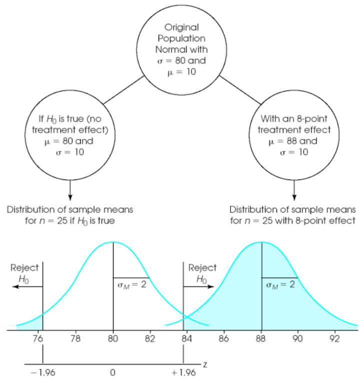
sample size

the magnitude of the effect.

## Power and sample size:



## Power and sample size:




## Power

Steps for calculating power

1) Locate critical region for the hypothesis testing, given a value for $\mu$, the standard error and the effect size.

$$M = Z\,\sigma_M + \mu_1$$

2) Calculate Z score for this value assuming $H_1$ were true.

$$Z = (M - \mu_2)/\,\sigma_M$$

3) Determine power from the Z table.

## Power Example

A researcher is examining a treatment effect. The population has the following properties: $\mu = 74$, with $\sigma = 15$. He expects a treatment effect of 5. To test this effect he has a sample of 25 participants, and tests the effect at the .05 level. What is the power of the test?