Hypothesis Testing

The ideas in hypothesis testing are based on **deductive reasoning -** we assume that some probability model is true and then ask “What are the chances that these observations came from that probability model?”.

Hypothesis Testing

- Null Hypothesis
- Alternative Hypothesis
- Significance level
- Statistically significant
- Critical value
- Acceptance / rejection region
- p-value
- power
- Types of errors: Type I (α), Type II (β)
- One-sided (one-tailed) test
- Two-sided (two-tailed) test

Hypothesis Testing
Motivation

1. Is the chance of getting a cold different when subjects take vitamin C than when they take placebo? (Pauling 1971 data).

2. Suppose that 6 out of 15 students in a grade-school class develop influence, whereas 20% of grade-school children nationwide develop influenza. Is there evidence of an excessive number of cases in the class?

Hypothesis Testing
Motivation

3. In a study of 25 hypertensive men we find a mean serum-cholesterol level of 220 mg/ml. In the 20-74 year-old male population the mean serum cholesterol is 211 mg/ml with standard deviation of 46 mg/ml.

- Is the mean for the population of hypertensive men also 211 mg/ml?
- Is the data consistent with that model?
- What if $\bar{X}$ = 230 mg/ml?
- What if $\bar{X}$ = 250 mg/ml?
- What if the sample was of 100 instead of 25?

Hypothesis Testing

Define:

μ = population mean serum cholesterol for male hypertensives

Hypothesis:

1. Null Hypothesis: Generally, the hypothesis that the unknown parameter equals a fixed value.

   H_0: $\mu = 211$ mg/ml

2. Alternative Hypothesis: contradicts the null hypothesis.

   H_A: $\mu \neq 211$ mg/ml

Hypothesis Testing

Decision / Action:

We assume that either H_0 or H_A is true. Based on the data we will choose one of these hypotheses.

	H_0 Correct	H_A Correct
Decide H_0	$1-\alpha$	β
Decide H_A	α	$1-\beta$

α = **significance level**
$1 - \beta$ = **power**

Hypothesis Testing

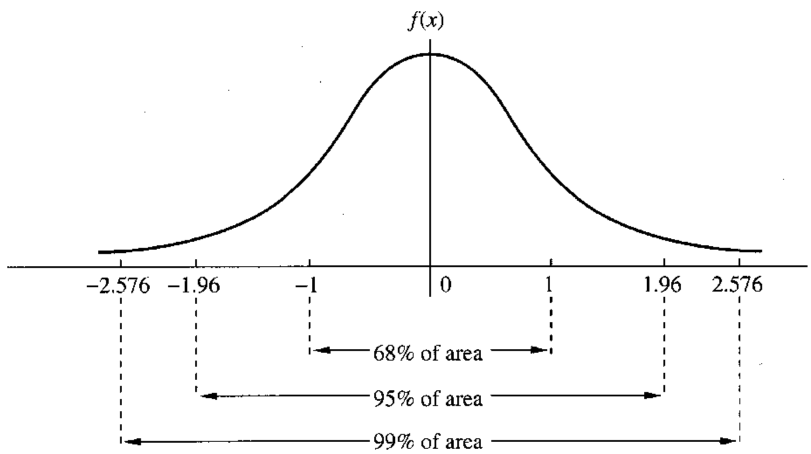
Let's fix α, for example, $\alpha = 0.05$.

$$0.05 = \alpha = P[\text{ choose } H_A \mid H_0 \text{ true }]$$

$$\alpha = P[\text{ reject } H_0 \mid H_0 \text{ true }]$$

Q: How to construct a procedure that makes this error with only 0.05 probability?

A: Suppose we assume H_0 is true and suppose that, using that assumption, the data should give us a standard normal, Z.



If $\mu = 0$ then $|Z|$ is rarely "large". A "large" $|Z|$ would make me question whether $\mu = 0$.

Hypothesis Testing

Therefore, **we reject** H_0 **if** $|Z| > 1.96$.

$\alpha = P[\text{reject } H_0 \mid H_0 \text{ true}] = 0.05$

Then if we do find a large value of $|Z|$ we can claim that:

- **Either** H_0 is true and something unusual happened (with probability α)…
- **or**, H_0 is not true.

Given α and H_0 we can construct a test of H_0 with a specified significance level. But remember, we start by assuming that H_0 is true - we haven't proved it is true. Therefore, we usually say

- $|Z| > 1.96$ then we **reject** H_0.
- $|Z| < 1.96$ then we **fail to reject** H_0.

Hypothesis Testing

Cholesterol Example:

Let μ be the mean serum cholesterol level for male hypertensives. We observe

$$\overline{X} = 220 \text{ mg/ml}$$

Also, we are told that for the general population...

μ_0 = mean serum cholesterol level for males = 211 mg/ml

σ = std. dev. of serum cholesterol for males = 46 mg/ml

NULL HYPOTHESIS: mean for male hypertensives is the same as the general male population.

ALTERNATIVE HYPOTHESIS: mean for male hypertensives is different than the mean for the general male population.

$$H_0 : \mu = \mu_0 = 211 \text{ mg/ml}$$

$$H_A : \mu \neq \mu_0 \quad (\mu \neq 211 \text{ mg/ml})$$

Hypothesis Testing

Cholesterol Example:

Test H_0 with significance level α.

Under H_0 we know:

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

Therefore,

•**Reject** H_0 if $\left| \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \right| > 1.96$ gives an $\alpha = 0.05$ test.

•**Reject** H_0 if

$$\overline{X} > \mu_0 + 1.96 \frac{\sigma}{\sqrt{n}} \quad \text{or}$$

$$\overline{X} < \mu_0 - 1.96 \frac{\sigma}{\sqrt{n}}$$

Hypothesis Testing

Cholesterol Example:

TEST: **Reject** H_0 if

$$\overline{X} > 211 + 1.96\frac{46}{\sqrt{25}} \text{ or}$$

$$\overline{X} < 211 - 1.96\frac{46}{\sqrt{25}}$$

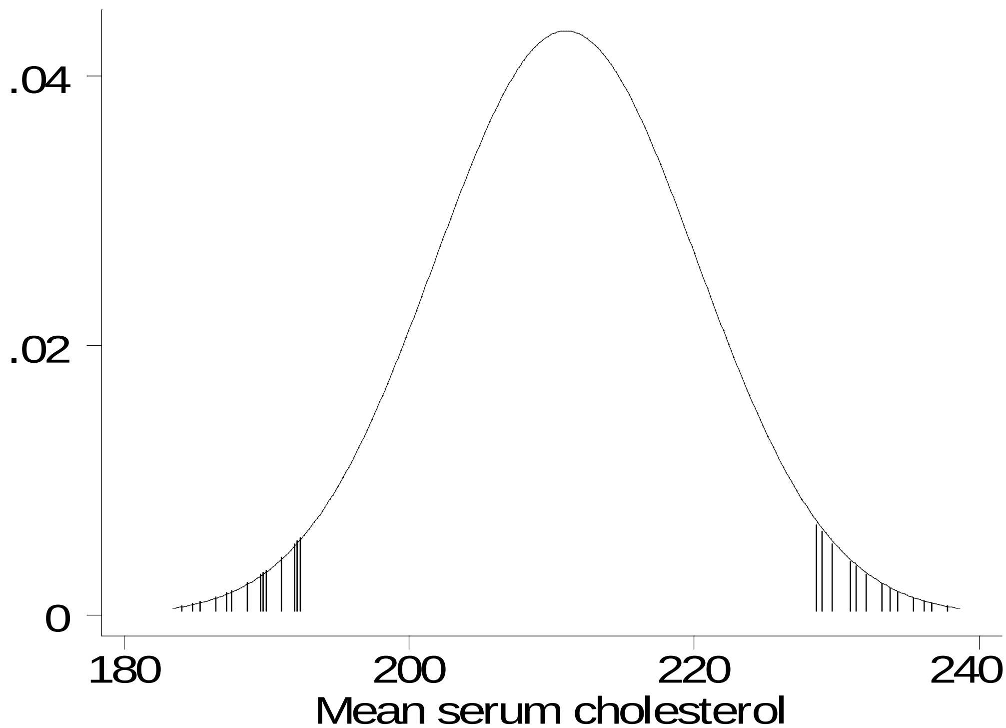
$$\overline{X} > 228.03 \text{ or}$$

$$\overline{X} < 192.97$$

In terms of Z ...

$$Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Reject H_0 if Z<-1.96 or Z> 1.96


.04
.02
0
180
200
220
240
Mean serum cholesterol

Hypothesis Testing

p-value:

- smallest possible α for which the observed sample would still reject H_0.
- probability of obtaining a result as extreme or more extreme than the actual sample (give H_0 true).

NOTE: probability calculations are <u>always</u> based on a model.

Hypothesis Testing

p-value: Cholesterol Example

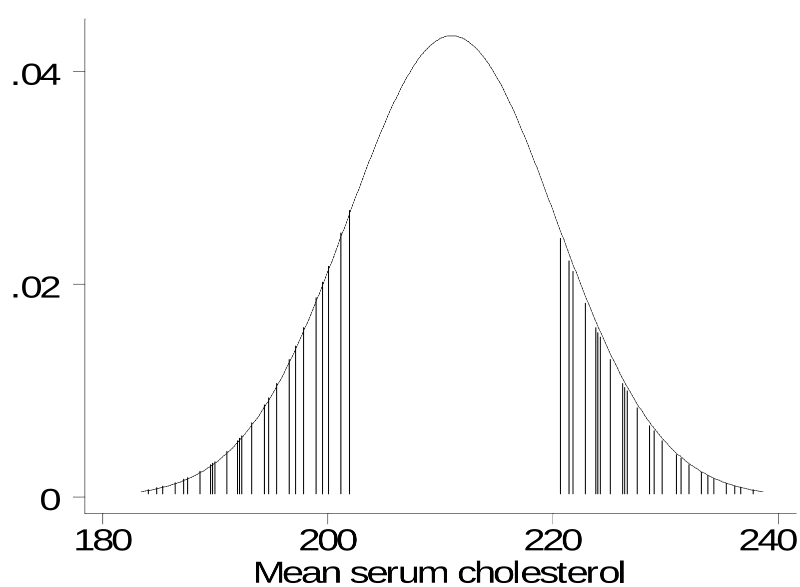
$\overline{X}$ = 220 mg/ml n = 25 σ = 46 mg/ml

$H_0 : \mu = 211$ mg/ml

$H_A : \mu \neq 211$ mg/ml

p-value is given by:

$2 * P[\overline{X} > 220] = .33$



Determination of Statistical Significance for Results from Hypothesis Tests

Either of the following methods can be used to establish whether results from hypothesis tests are statistically significant:

(1) The test statistic Z can be computed and compared with the critical value $Q_Z^{(1-\alpha/2)}$ at an α level of .05. Specifically, if H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$ are being tested and $|Z| > 1.96$, then H_0 is rejected and the results are declared *statistically significant* (i.e., $p < .05$). Otherwise, H_0 is accepted and the results are declared *not statistically significant* (i.e., $p \geq .05$). We refer to this approach as the **critical-value method**.

(2) The exact p-value can be computed, and if $p < .05$, then H_0 is rejected and the results are declared *statistically significant* . Otherwise, if $p \geq .05$ then H_0 is accepted and the results are declared *not statistically significant* . We will refer to this approach as the **p-value method** .

Guidelines for Judging the Significance of p-value (Rosner pg 200)

If $.01 \leq p < .05$, then the results are *significant*.

If $.001 \leq p < .01$, then the results are *highly significant*.

If $p < .001$, then the results are *very highly significant*.

If $p > .05$, then the results are considered *not statistically significant* (sometimes denoted by NS).

However, if $.05 \leq p < .10$, then a trend toward statistically significance is sometimes noted.

Hypothesis Testing and Confidence Intervals

Hypothesis Test: Fail to reject H_0 if

$$\overline{X} < \mu_0 + Q_Z^{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

and $$\overline{X} > \mu_0 - Q_Z^{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

Confidence Interval: Plausible values for μ are given by

$$\mu < \overline{X} + Q_Z^{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

and $$\mu > \overline{X} - Q_Z^{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

Hypothesis Testing
"how many sides?"

Depending on the alternative hypothesis a test may have a **one-sided alternative** or a **two-sided alternative**. Consider

$$H_0 \; : \; \mu = \mu_0$$

We can envision (at least) three possible alternatives

$$H_A \; : \; \mu \neq \mu_0 \qquad (1)$$

$$H_A \; : \; \mu < \mu_0 \qquad (2)$$

$$H_A \; : \; \mu > \mu_0 \qquad (3)$$

(1) is an example of a "two-sided alternative"

(2) and (3) are examples of "one-sided alternatives"

The distinction impacts

- Rejection regions
- p-value calculation

Hypothesis Testing
"how many sides?"

Cholesterol Example: Instead of the two-sided alternative considered earlier we may have only been interested in the alternative that hypertensives had a higher serum cholesterol.

$$H_0 : \mu = 211$$

$$H_A : \mu > 211$$

Given this, an $\alpha = 0.05$ test would reject when

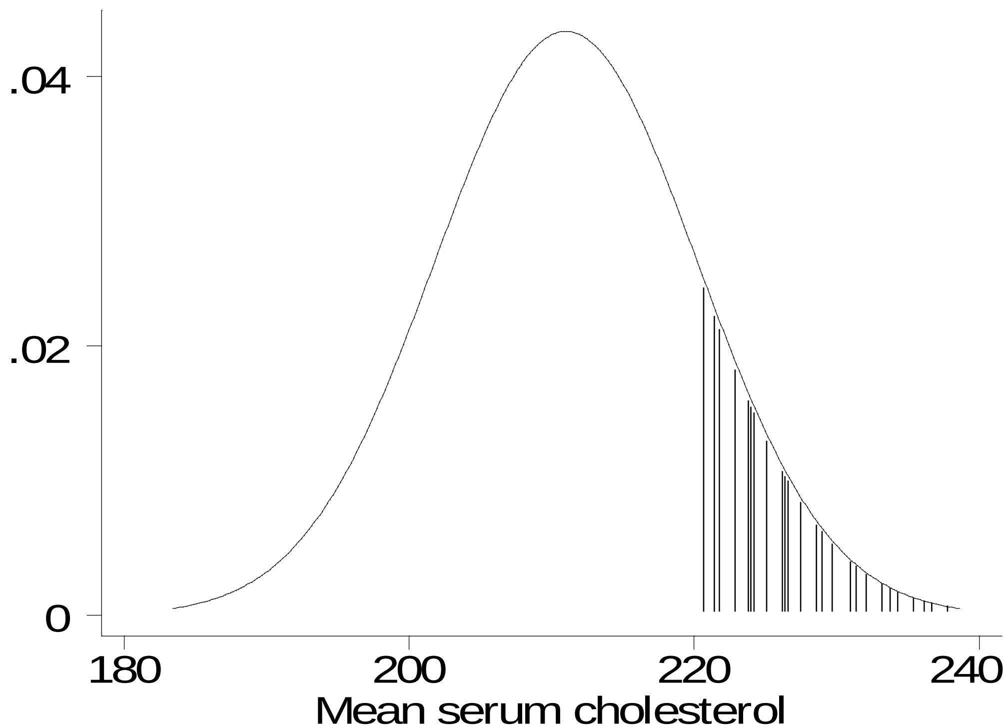
$$\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = Z > Q_Z^{(1-0.05)} = 1.65$$

We put all the probability on "one-side".

The <u>p-value</u> would be half of the previous,

$$\text{p-value} = P[\,\overline{X} > 220]$$

$$= .163$$


.04
.02
0
180
200
220
240
Mean serum cholesterol

Hypothesis Testing

Through this worked example we have seen the basic components to the statistical test of a scientific hypothesis.

Summary

1. Identify H_0 and H_A
2. Identify a test statistic
3. Determine a significance level, $\alpha = 0.05$, $\alpha = 0.01$
4. Critical value determines rejection / acceptance region
5. p-value
6. Interpret the result