Hypothesis Testing

Hypothesis Testing Motivation

1. Is the chance of getting a cold different when participants 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 influenza, whereas 20% of grade-school children nationwide develop influenza. Is there evidence of an excessive number of cases in the class?

Hypothesis Testing Motivation cont'd

3. In a study of 49 HIV+ pregnant women, we find a mean arterial pressure of 79.4 mm Hg. In the general population of pregnant women, the mean arterial pressure is 78.8 mm Hg with standard deviation of 1.4 mm Hg.

- Are the data from our study consistent with the data from the general population?
- What if $\overline{X}$ = 80.1 mm Hg?
- What if $\overline{X}$ = 81.3 mm Hg?
- What if the sample was of size n=100 instead of 49?

Hypothesis Testing

Define:

μ = *population* mean arterial pressure for HIV+ pregnant women at 36 weeks gestation

Hypotheses:

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

H_0: μ = 78.8 mm Hg

2. Alternative Hypothesis: contradicts the null hypothesis.

H_A: $\mu \neq$ 78.8 mm Hg

Hypothesis Testing

α = **"size"**
$1 - \beta$ = "**power**"

Decision and Truth:

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

		Truth	
		H_0 Correct	H_A Correct
Decision	Fail to reject H_0	$1-\alpha$	β
	Reject H_0, Accept H_A	α	$1-\beta$

Hypothesis Testing

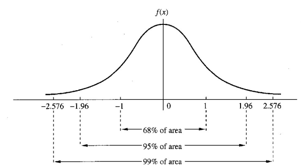
Let's fix α, for example, $\alpha = 0.05$.

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

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

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.

But note that we operate under the assumption that H_0 is true and look for evidence suggesting it is false. We can reject the null but we can't prove it is true. Therefore, we say:

- If $|Z| > 1.96$, then we **reject** H_0 and **accept** H_A.
- If $|Z| < 1.96$, then we **fail to reject** H_0.

Hypothesis Testing Example

Mean Arterial Pressure Example:

Let μ be the mean arterial pressure for HIV+ pregnant women at 36 weeks. In our sample of 49 women:

$$\overline{X} = 79.4 \text{ mm Hg}$$

Also, we are told that for the general population of pregnant women at 36 weeks...

μ_0 = mean arterial pressure = 78.8 mm Hg

σ = std. dev. of mean arterial pressure = 1.4 mm Hg

Null hypothesis (H_0): mean for HIV+ pregnant women (μ) is the same as the mean for the general population of pregnant women (μ_0).

Alternative hypothesis (H_A): mean for HIV+ pregnant women (μ) is different than the mean for the general population of pregnant women (μ_0).

$$H_0 : \mu = \mu_0 = 78.8 \text{ mm Hg}$$

$$H_A : \mu \neq \mu_0 \ \ (\mu \neq 78.8 \text{ mm Hg})$$

Hypothesis Testing Example

Mean Arterial Pressure Example:

Test H_0 with significance level $\alpha = 0.05$.

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

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

Hypothesis Testing Example cont'd

Test: **Reject** H_0 if $\overline{X} > 78.8 + 1.96\frac{1.4}{\sqrt{49}}$

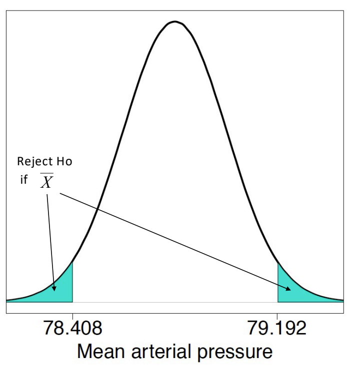
or $\overline{X} < 78.8 - 1.96\frac{1.4}{\sqrt{49}}$

or $\overline{X} > 79.192$

$\overline{X} < 78.408$

$\overline{X} = 79.4$ therefore we reject the null and accept the alternative hypothesis.

In terms of Z: $Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ We **reject** H_0 if $Z < -1.96$ or $Z > 1.96$



Hypothesis Testing and the P-value

p-value:

- a measure of how well the data fits the null hypothesis
- probability of obtaining a result as extreme or more extreme than the actual sample (when H_0 is true)
- **not** the probability that the null (or alternative) is true

Hypothesis Testing and the P-value Example

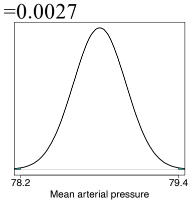
p-value: Mean arterial pressure example

$\overline{X}$ = 79.4 mm Hg n = 49σ = 1.4 mm Hg

$H_0 : \mu = 78.8$ mm Hg

$H_A : \mu \neq 78.8$ mm Hg

p-value is given by: 2 * P[$\overline{X}$ > 79.4] =0.0027



(Somewhat Outdated) Guidelines for Judging the Significance of p-value

If $.05 \leq p < .10$, than the results are *marginally significant*.

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 $p > .1$, then the results are considered *not statistically significant* (sometimes denoted by NS).

Significance is not everything!

American Statistical Association Guidelines for the Use of P-values

In March 2016, the ASA provided these guidelines:

1. P-values can indicate how incompatible the data are with a specified statistical model.

2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.

3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.

4. Proper inference requires full reporting and transparency.

5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.

6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

Hypothesis Testing and Confidence Intervals

Hypothesis Test: Fail to reject H_0 if:

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

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

Confidence Interval: Plausible values for μ are given by:

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

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

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

$$H_A : \mu \neq \mu_0 \quad (1)$$

$$H_A : \mu < \mu_0 \quad (2)$$

$$H_A : \mu > \mu_0 \quad (3)$$

(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

1-sided Hypothesis Testing

Mean Arterial Pressure Example: Instead of the two-sided alternative considered earlier we may have only been interested in the alternative that HIV+ pregnant women have higher mean arterial pressure.

$$H_0 : \mu = 78.8$$

$$H_A : \mu > 78.8$$

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

$$\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = Z > Q_Z^{(1-0.05)} = 1.64$$

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

The p-value would be half of the previous,

$$\text{p-value} = P[\overline{X} > 79.4]$$

$$= 0.0014$$

Take Home Points

Hypothesis Testing Steps

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. Calculate p-value
6. Interpret the result