Ch 9. Hypothesis Tests — Modules

36. Introduction / Terminology

37. Normal Mean Tests (variance known)

38. Normal Mean Tests (variance unknown)

39. Tests for Other Parameters

36. Introduction to Hypothesis Testing

General Approach

Easy Example

How Can You Go Wrong?

General Approach

1. State a hypothesis.

2. Select a test statistic (to test whether or not the hypothesis is true).

3. Evaluate (calculate) the test statistic.

4. Interpret results — does the test statistic allow you to reject your hypothesis?

Details.

1. **Hypotheses** are simply statements or claims about parameter values.

You perform a **hypothesis test** to prove or disprove the claim.

Set up a **null hypothesis** (H_0) and an **alternative hypothesis** (H_1) to cover the entire parameter space. The null hyp sort of represents the "status quo".

Example: We claim that the mean weight of a filled package of chicken is μ_0 ounces. (We specify μ_0.)

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

This is a **two-sided test**. We'll reject the claim if $\widehat{\mu}$ (an estimator of μ) is "too high" or "too small".

Example: We claim that a certain brand of tires lasts for at least a mean of μ_0 miles. (We specify μ_0.)

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

This is a **one-sided test**. We'll reject the claim if $\widehat{\mu}$ is "too small".

Example: We claim that emissions from a certain type of car do not exceed a mean of μ_0 ppm. (We specify μ_0.)

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

This is a **one-sided test**. We'll reject the claim if $\widehat{\mu}$ is "too large".

Idea: H_0 is the old, conservative “status quo”. H_1 is the new, radical hypothesis. Although you may not be toooo sure about the truth of H_0, you won’t reject it in favor of H_1 unless you see substantial evidence in support of H_1.

If you get substantial evidence supporting H_1, you’ll decide to reject H_0. Otherwise, you “fail to reject” H_0. (This roughly means that you grudgingly accept H_0.)

2. Select a test statistic (to test if H_0 is true).

For instance, we could compare an estimator $\widehat{\mu}$ with μ_0. The comparison is accomplished using a known sampling distribution (aka "test statistic"), e.g.,

$$z_{\text{obs}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad \text{(if } \sigma^2 \text{ is known) or}$$

$$t_{\text{obs}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad \text{(if } \sigma^2 \text{ is unknown)}$$

More details later.

3. Evaluate the test statistic. Here's the logic of hypothesis testing:

(a) Collect sample data

(b) Assume H_0 (the "status quo") is true.

(c) Determine the probability of the sample result, assuming H_0 is true.

(d) Decide from (c) if H_0 is plausible.

* If the prob from (c) is low, reject H_0 and select H_1.

* If the prob from (c) is high, fail to reject H_0.

Example: Time to metabolize a drug. Current drug takes $\mu_0 = 15$ min. Is new drug better?

Claim: Expected time for new drug is < 15 min.

$$H_0 : \mu \geq 15$$
$$H_1 : \mu < 15$$

Data: $n = 20$, $\bar{X} = 14.88$, $S = 0.333$. The test statistic is

$$t_{\text{obs}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = -1.61.$$

Now, if H_0 is actually the true state of things, then $\mu = \mu_0$, and from our discussion on CI's, we have

$$t_{\text{obs}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \sim t(19).$$

What would cause us to reject H_0?

If $\bar{X} \ll \mu_0 (= 15)$, this would indicate that H_0 is probably wrong.

Equivalently, I'd reject H_0 if t_{obs} is "significantly" $\ll 0$.

4. Interpret the Test Statistic.

So if H_0 is true, is it reasonable (or, at least, not outrageous) to have gotten $t_{\text{obs}} = -1.61$?

If yes, then we we'll *fail to reject* H_0.

If no, then we'll *reject* H_0 in favor of H_1.

Let's see. . . . From the t table, we have

$$t_{.05,19} = -1.729 \quad \text{and} \quad t_{.10,19} = -1.328.$$

I.e.,

$$\text{Pr}(t(19) < -1.729) \;=\; .05 \;\text{ and}$$

$$\text{Pr}(t(19) < -1.328) \;=\; .10.$$

This means that

$$0.05 \;<\; p \;\equiv\; \text{Pr}(t(19) < \underbrace{-1.61}_{t_{\text{obs}}}) \;<\; 0.10.$$

In English: If H_0 were true, there's a $100p\%$ chance that we'd see a value of t_{obs} that's ≤ -1.61. That's not a real high probability, but it's not toooo small.

Formally, we'd **reject** H_0 at "level" 0.10, since $t_{\text{obs}} = -1.61 < t_{.10,19} = -1.328$.

But, we'd **fail to reject** H_0 at level 0.05, since $t_{\text{obs}} = -1.61 > t_{.05,19} = -1.729$.

More on this in the next modules!

Where Can We Go Wrong?

Four things can happen:

1. If H_0 is actually true and we conclude that it's true — good.

2. If H_0 is actually false and we conclude that it's false — good

3. If H_0 is actually true and we conclude that it's false — bad. This is called **Type I error**.

Example: We conclude that a new, inferior drug is better than the drug currently on the market.

4. If H_0 is actually false and we conclude that it's true — bad. This is called **Type II error**.

Example: We conclude that a new, superior drug is worse than the drug currently on the market.

Want to keep

$$\text{Pr(Type I error)} = \text{Pr(Reject } H_0 \mid H_0 \text{ true)} \leq \alpha$$

$$\text{Pr(Type II error)} = \text{Pr(Fail to Rej } H_0 \mid H_0 \text{ false)} \leq \beta$$

We choose α and β. Of course, we need to have $\alpha + \beta < 1$.

Usually, Type I error is considered as "worse" than Type II.

Definition: The **power** of a hypothesis test is

$$1 - \beta \;=\; \text{Pr(Reject } H_0 \mid H_0 \text{ false).}$$

It's good to have high power.

Definition: The probability of Type I error, α, is called the **size** or **level of significance** of the test.