Hypothesis Test and Comparison

STAT-UB.0001 Statistics for Business Control

Ningshan Zhang

IOMS Department
nzhang@stern.nyu.edu

Aug 2, 2018

* Office Hours: Wed & Fri 10:00 - 11:30 AM, KMC 8-174

Review: CI for Population Mean

Setup: let $\bar{X}$ and S be the mean and standard deviation of a random sample of n observations.

When n is large ($n \geq 30$) or the population is normal, a $1-\alpha$ CI for population mean μ is

$$(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}})$$

- When n is large, $t_{\alpha/2,n-1} \approx z_{\alpha/2}$.

Review: CI for Population Proportion

Setup: let $\widehat{p}$ be the sample proportion of a random sample of n observations.

When $np \geq 15$ and $n(1-p) \geq 15$, a $1-\alpha$ CI for population proportion p is

$$\left(\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \, \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right)$$

- In practice, check for $n\widehat{p} \geq 15$ and $n(1-\widehat{p}) \geq 15$.

Hypothesis Testing

Often, someone makes a claim about the world, such as

- The revenue of a company is normally distributed with mean \$1.5 B and standard deviation \$0.1 B.
- David throws 10 heads in a roll with a fair coin.

We collect some data, and we want to evaluate the plausibility of that claim in the face of data.

Hypothesis Testing

Hypothesis:

- A statement about the numerical value of a population parameter.

Null Hypothesis (H_0):

- The statement is true.

Alternative Hypothesis (H_A):

- The statement is false.

Hypothesis Testing

Procedure of a hypothesis testing:

1. Specify the null (H_0) and alternative (H_a) hypothesis.
2. Collect a sample, find evidence against H_0 in the sample.
3. Make a conclusion:
   - reject H_0 if the evidence is strong.
   - otherwise do not reject H_0.

Hypothesis tests work like criminal courts: the defendant is presumed innocent, until proven guilty beyond a reasonable doubt.

Hypothesis Testing for Population Mean

- Setup: Population has unknown mean μ. Let $\bar{X}$ and S be the mean and standard deviation of a random sample of n observations.
- Goal: test whether or not $\mu = \mu_0$ for some given value of μ_0.

Step 1: Specify the null (H_0) and alternative (H_a) hypothesis.

Test Statistic

Step 1: $H_0 : \mu = \mu_0, \quad H_A : \mu \neq \mu_0.$

Step 2: Collect a sample, find evidence against H_0 in the sample.

Given a random sample, define the test statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

- T is a random variable. (Why?)
- If null hypothesis is true, i.e. $\mu = \mu_0$, then T is t-distributed with df $= n - 1$, when $n \geq 30$ or population is normal.

Test Statistic

Step 2: Collect a sample, find evidence against H_0 in the sample.

Given a random sample, define the test statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

The observed test statistic, t is the evidence:

- If H_0 is true, it is not very likely to observe extreme values of t, such as $t \geq 2$ or $t \leq -2$.
- Larger value of $|t|$ means stronger evidence against H_0.

* T is a random variable, t is its observed value.

p-value

Step 2: Collect a sample, find evidence against H_0 in the sample.

Evidence (test stat.): $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$, t is the observed value.

The p-value tells us how strong is the evidence against H_0:

- If H_0 is true, p-value is the probability of getting a test statistic as least as extreme as the one we observed:

$$\text{p-value} = \mathbb{P}(|T| \geq |t|) \text{ under } H_0.$$

- When the p-value is small, the evidence is strong, and vice versa.

Conclusion

Step 1: Specify the null (H_0) and alternative (H_a) hypothesis.
Step 2: Collect a sample, find evidence against H_0 in the sample.
Step 3: Make a conclusion

For significance level α (often use $\alpha = 0.05$),

- reject H_0 if the p-value $\leq \alpha$.
- do not rejct H_0 if p-value $> \alpha$.

Interpretation of Significance Level α

When H_0 is rejected, we say that the result is statistically significant at level α. A common choice is $\alpha = 0.05$.

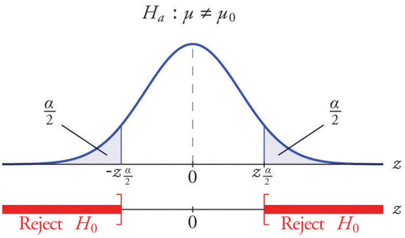
- Caution: α is not the probability that H_0 is true (there is nothing random about H_0!)
- It is possible that H_0 is true but we reject it.
- Rather, α is the probability, or proportion of the time, that a test of this kind would reject H_0 when H_0 is true.

Hypothesis Testing by Rejection Region

An alternative way of hypothesis testing: reject H_0 if the observed t falls into a pre-specified *rejection region*.

- For significance level α, reject H_0 when $|t| \geq t_{\alpha/2,n-1}$.
- Thus the rejection region is $(-\infty, -t_{\alpha/2,n-1}] \cup [t_{\alpha/2,n-1}, +\infty)$.

Figure: The rejection region (when n is large, T is Z-distributed).



Confidence Intervals and Hypothesis Testing

- Estimation: A $1-\alpha$ CI for μ is derived from

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \in (-t_{\alpha/2,n-1}, t_{\alpha/2,n-1}).$$

- Hypothesis test for $H_0 : \mu = \mu_0$ and $H_a : \mu \neq \mu_0$: Reject H_0 at significance level α when

$$\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \notin (-t_{\alpha/2,n-1}, t_{\alpha/2,n-1}).$$

Confidence Intervals and Hypothesis Testing

One can show the following two things are equivalent for a given sample:

- μ_0 is not in the $1 - \alpha$ CI for μ.
- Reject the $H_0 : \mu = \mu_0$ with significance level α;

Example: given a 95% CI for μ, and consider the hypothesis test

$$H_0 : \mu = \mu_0, \quad H_a : \mu \neq \mu_0.$$

If μ_0 is not in the CI, reject H_0 with significance 0.05; otherwise do not reject H_0.

Types of Errors

	Reject H_0	Do not reject H_0
H_0 is true	False positive; Type I error	✓
H_0 is false	✓	False negative; Type II error

* "positive" means reject null.

Types of Errors

Type I Error: Rejecting H_0 when H_0 is true.

- Probability of Type I error $=$ "significance level" $= \alpha$.

Type II Error: Not rejecting H_0 when H_0 is false.

- Probability of Type II error $= \beta$.
- Usually not specified, and can be difficult to calculate.

Type I and type II errors are conflicting:

$$\text{Decreasing } \alpha \iff \text{Less likely to reject } H_0 \iff \text{Increasing } \beta.$$

Comparisons

Many experiments involve a comparison of two distinct groups.

Example:

- Are men superior drivers?
- Who has faster 4G LTE network: AT&T or Verizon?
- Do SAT prep courses improve scores?
- Is drug A more effective than drug B (i.e. A/B testing)?

Comparisons: Setup

- There are two populations, let μ_k, σ_k indicate the population mean and standard deviation, for population $k = 1, 2$.
- There are two independent random samples from each population, let $n_k, \bar{X}_k, S_k$ indicate the sample size, sample mean, sample deviation, for population $k = 1, 2$.

	Population 1	Population 2
Population parameters	μ_1, σ_1	μ_2, σ_2
Sample statistics	$n_1, \bar{X}_1, S_1$	$n_2, \bar{X}_2, S_2$

Comparisons: Goal

Given the sample statistics, we would like to

- Estimation: construct a confidence interval for $\mu_1 - \mu_2$.
- Hypothesis Testing: test

$$H_0 : \mu_1 = \mu_2, \quad H_a : \mu_1 \neq \mu_2.$$

It is equivalent to test

$$H_0 : \mu_1 - \mu_2 = 0, \quad H_a : \mu_1 - \mu_2 \neq 0.$$

Review One Sample Case

When there is only one sample, we rely on the distribution of

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}. \tag{1}$$

- Estimation: A $1 - \alpha$ CI for μ is derived from

$$T \in (-t_{\alpha/2,n-1}, t_{\alpha/2,n-1}).$$

- Hypothesis test for $H_0 : \mu = \mu_0$ and $H_a : \mu \neq \mu_0$,

$$\text{Reject } H_0 \text{ when } T \notin (-t_{\alpha/2,n-1}, t_{\alpha/2,n-1}),$$

  with $\mu = \mu_0$ in (1).

Review One Sample Case

The key formula is

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mathbb{E}(\bar{X})}{\text{sd}(\bar{X})}.$$

Given an estimate $\bar{X}$ of μ, we need to know:

- Its mean $\mathbb{E}(\bar{X})$ and standard deviation $\text{sd}(\bar{X})$;
- The estimate is approximately t or Z distributed.

Comparison

To study $\mu_1 - \mu_2$, we need to find:

- An esitmate of $\mu_1 - \mu_2$. Solution:

$$\bar{X}_1 - \bar{X}_2.$$

- The mean of the estimate. Solution:

$$\mathbb{E}(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2.$$

- The standard deviation of the estimate. Solution:

$$\mathrm{sd}(\bar{X}_1 - \bar{X}_2) = ?$$

Standard Deviation of $\bar{X}_1 - \bar{X}_2$

Theorem

If X, Y are two independent random variables, then

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y).$$

Thus,

$$\text{var}(\bar{X}_1 - \bar{X}_2) = \text{var}(\bar{X}_1) + \text{var}(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\Rightarrow \text{sd}(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Finally, use $\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ as an estimate for $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Comparison

Assume

- Two independent random samples;
- Samples are sufficiently large ($n_1 \geq 30$ and $n_2 \geq 30$).

Then

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

is approximately Z distributed.

---

* For comparisons, we only consider the large sample case.

Comparison

- Estimaton: the $1-\alpha$ CI for $\mu_1 - \mu_2$ is

$$(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

- Hypothesis test for $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 \neq \mu_2$:

$$\text{Reject } H_0 \text{ when } \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \notin (-z_{\alpha/2}, z_{\alpha/2})$$

  at significance level α.