AFRE 802
Statistical Methods for Agricultural, Food, & Resource Economists

Hypothesis Testing – Part 2 of 3
(WMS Ch. 10.5-10.8)
November 14, 2017

Nicole Mason
Michigan State University
Fall 2017

Please provide feedback to Mary Doidge

- https://msu.co1.qualtrics.com/jfe/form/SV_bPMru1yogxA4Xjv
- Survey is completely anonymous
- Thanks!

1

GAME PLAN

- Return Ch. 8 HW
- Review

- Hypothesis testing – Part 2 of 3
  - Small sample hypothesis testing for μ
  - Relationship b/w hypothesis testing procedures & confidence intervals
  - Another way to report the results of a hypothesis test: p-values

REVIEW

Type I vs. Type II error

		REALITY	
		NULL HYPOTHESIS	
		TRUE	FALSE
STUDY FINDINGS	TRUE		Type II error (β) 'False negative'
	FALSE	Type I error (α) 'False positive'	

- **Type I error: reject H_0 when H_0 is true**
  - Probability: α (significance level)
- **Type II error: fail to reject H_0 when H_0 is false**
  - Probability: β
- In hypothesis testing, we choose the probability of Type I error (α) we are willing to live with, then seek to minimize the probability of Type II error (β) given α (more on this next class)

REVIEW: Hypothesis testing procedure

1. State the **null & alternative hypotheses**
2. Define an appropriate **test statistic**
3. Determine the **distribution of the test statistic under the null** hypothesis
4. **Standardize the test statistic** to something with known/tabled probabilities for its sampling distribution (e.g., *Z, t, chi-square, F*)
5. Choose a **significance level** (***α***, the *P*(Type I error)=*P*(reject the null when it is true), typically 0.01, 0.05, or 0.10) and a **rejection region** (values of standardized test statistic that lead to rejection of H_0)
6. **Reject the null hypothesis if** the standardized statistic lies **in the rejection region**; fail to reject otherwise

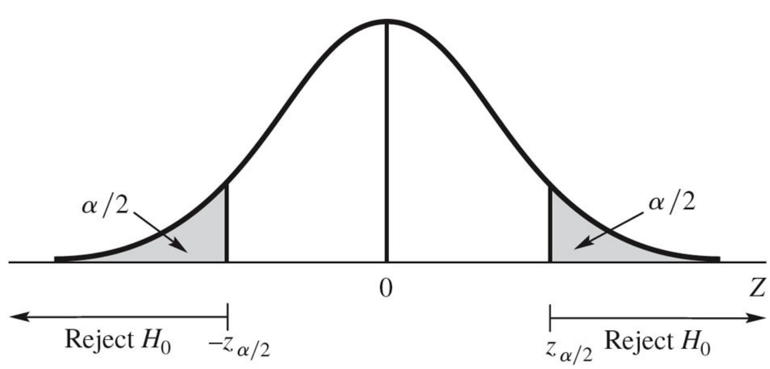
4

REVIEW

MICHIGAN STATE UNIVERSITY

Two-sided alternative & associated rejection region

$H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$ two-sided (two-tailed) alternative



5

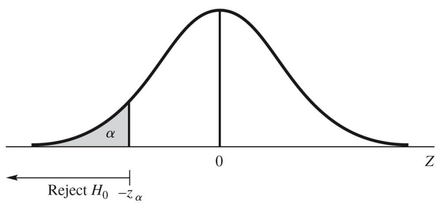
REVIEW

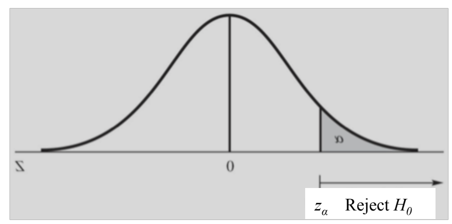
MICHIGAN STATE UNIVERSITY

One-sided alternatives & associated rejection regions

$H_0: \theta = \theta_0$ vs. $H_1: \theta < \theta_0$
one-sided (lower-tailed) alternative

$H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$
one-sided (upper-tailed) alternative





6

REVIEW

MICHIGAN STATE UNIVERSITY

Large-Sample α-Level Hypothesis Tests

$$H_0 : \theta = \theta_0.$$

$$H_a : \begin{cases} \theta > \theta_0 & \text{(upper-tail alternative).} \\ \theta < \theta_0 & \text{(lower-tail alternative).} \\ \theta \neq \theta_0 & \text{(two-tailed alternative).} \end{cases}$$

$$\text{Test statistic: } Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}.$$

$$\text{Rejection region: } \begin{cases} \{z > z_\alpha\} & \text{(upper-tail RR).} \\ \{z < -z_\alpha\} & \text{(lower-tail RR).} \\ \{|z| > z_{\alpha/2}\} & \text{(two-tailed RR).} \end{cases}$$

Small-sample hypothesis testing for μ

- The **Z test statistics** that we've used for hypothesis testing so far hinge on having a **sample size large enough** that we can **invoke the CLT** such that

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \text{ is approximately } \sim Normal(0,1)$$

- Recall our discussion of small sample confidence intervals (CIs) for μ. Today we'll discuss the closely related topic of small sample hypothesis testing for μ, then the relationship b/w hypothesis testing and Cis.
- *What test statistic should we use for small sample hypothesis testing for μ?*

8

Hypothesis testing procedure – same except that our test statistic will now be a *T* !

Steps #1-3 and #5-6 the same as before.

4. Standardize the test statistic to something with known/ tabled probabilities for its sampling distribution:

Large sample hypothesis test for μ

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} = \frac{\bar{Y} - \mu_0}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu_0}{\sigma / \sqrt{N}} \text{ approx. } \sim Normal(0,1)$$

If sample size large, don't lose much by substituting *S* for *σ;* can still use Z. (Remember *t* converges to standard normal for large *N*.)

Small sample hypothesis test for μ

$$T = \frac{\bar{Y} - \mu_0}{S / \sqrt{N}} \sim t \text{ distribution with } N - 1 \; d.f.$$

Note that here, as in small sample CI for μ discussion, we are **assuming that the random sample is from a bell-shaped (approximately normal) distribution**.

What will the rejection regions look like for two- vs. one-sided alternatives?

Example – small sample hypothesis testing for μ

EXAMPLE **10.12**

Example 8.11 gives muzzle velocities of eight shells tested with a new gunpowder, along with the sample mean and sample standard deviation, $\overline{y} = 2959$ and $s = 39.1$. The manufacturer claims that the new gunpowder produces an average velocity of not less than 3000 feet per second. Do the sample data provide sufficient evidence to contradict the manufacturer's claim at the .025 level of significance?

10

MICHIGAN STATE UNIVERSITY

Summary

A Small-Sample Test for μ

(or approximately normal)

Assumptions: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with $E(Y_i) = \mu$.

$H_0 : \mu = \mu_0$.

$$H_a : \begin{cases} \mu > \mu_0 & \text{(upper-tail alternative).} \\ \mu < \mu_0 & \text{(lower-tail alternative).} \\ \mu \neq \mu_0 & \text{(two-tailed alternative).} \end{cases}$$

Test statistic: $T = \dfrac{\overline{Y} - \mu_0}{S/\sqrt{n}}$.

$$\text{Rejection region: } \begin{cases} t > t_\alpha & \text{(upper-tail RR).} \\ t < -t_\alpha & \text{(lower-tail RR).} \\ |t| > t_{\alpha/2} & \text{(two-tailed RR).} \end{cases}$$

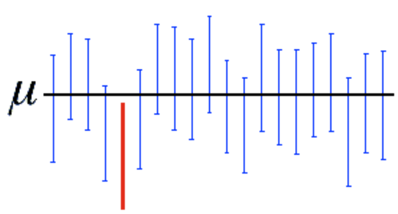
(See Table 5, Appendix 3, for values of t_α, with $\nu = n - 1$ df.)

In our previous example (10.12), why do we use H_0: μ=3000 vs. H_1: μ<3000 instead of H_0: μ≥3000 vs. H_1: μ<3000?

Gist:

- **What we really care about is the alternative hypothesis**. We specify the null and alternative hypotheses so that this is the case.
- Specifying the null as μ=3000 or μ≥3000 leads to the **exact same conclusion**
- BUT the former requires that we only compute the test statistic under the null for ONE null value. It's much more complicated for the latter.
- **Bottom line: specifying the null in terms of an equality instead of an inequality leads to the correct testing procedure and the same conclusion AND simplifies our lives**
- See WMS p. 519 for details

Relationship between hypothesis-testing procedures and confidence intervals



Source: Minitab

Recall that a 95% CI means that 95% of samples (e.g., 19 of 20 samples) drawn from the same population will produce CIs that contain the true population parameter.

If we are testing the null hypothesis that $\theta=\theta_0$ against the alternative that $\theta\neq\theta_0$ at the α=0.05 level, and θ_0 falls in the 95% confidence interval for θ, do we reject or fail to reject the null (in favor of the alternative) at the α=0.05 level?

→ **FAIL TO REJECT** because the associated ***Z*** **would NOT be in the rejection region**

Relationship between hypothesis testing procedures and confidence intervals

General rules for **TWO-SIDED** alternatives:

- **Fail to reject** H_0: $\theta=\theta_0$ in favor of H_1: $\theta\neq\theta_0$ at the α level **if θ_0 lies inside a 100(1-α)% two-sided CI**
- **Reject** H_0: $\theta=\theta_0$ in favor of H_1: $\theta\neq\theta_0$ at the α level **if θ_0 lies outside of a 100(1-α)% two-sided CI**

Relationship between hypothesis testing procedures and confidence intervals

Similar rules apply for **ONE-SIDED** alternatives:

- H_0: $\theta=\theta_0$ vs. H_1: $\boldsymbol{\theta>\theta_0}$ → fail to reject H_0 if θ_0 is in **lower CI** $[\hat{\theta}_L,\infty)$; o.w. reject H_0
  - Recall 100(1-α)% lower CI is $P(\theta \geq \hat{\theta}_L) = 1-\alpha$
- H_0: $\theta=\theta_0$ vs. H_1: $\boldsymbol{\theta<\theta_0}$ → fail to reject H_0 if θ_0 is in **upper CI** $(-\infty,\hat{\theta}_U]$; o.w. reject H_0
  - Recall 100(1-α)% upper CI is $P(\theta \leq \hat{\theta}_U) = 1-\alpha$

These one-sided rules might seem counter-intuitive so let's do an example then look at the proof

Example – CIs & hypothesis testing

Recall Example 10.12 from earlier in class.

a. Construct a 97.5% upper one-sided CI for μ.
b. Does the value of μ under H_0 (3,000) fall within this CI?
c. What would we conclude about H_0: μ=3,000 (implicitly μ≥3,000) vs. H_1: μ<3000 at the α=0.025 level based on this CI?
d. How does this compare to what we concluded in Example 10.12?

$$\boxed{\bar{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{N}}\right)}$$

Recall the formula for a 2-sided confidence interval. What would this be for an upper one-sided CI (which finds an upper bound for μ)?

16

17

18

See handout on relationship b/w CI's and hypothesis testing

19

MICHIGAN STATE UNIVERSITY

Another way to report the results of a hypothesis test: **p-values**

- Recall that **α is the probability of Type I error** (rejecting the null when it is true)
- **α** is the **"significance level"** or **"level"** of the test
- We pick α but our choice is rather arbitrary
  - Some prefer 0.10, others 0.05, others 0.01
- Instead of picking an (arbitrary) α, we can report the "p-value" a.k.a. the "attained significance level"
- **p-value** = the smallest α for which the data suggest the null hypothesis should be rejected in favor of the alternative
  - *How would be interpret a p-value of 0.04?*
- The smaller the p-value, the stronger is the evidence against the null (and in favor of the alternative)

20

MICHIGAN STATE UNIVERSITY

p-values (cont'd)

- Suppose we have chosen α=0.10 and the p-value for our statistical test is 0.04. *Do we reject or fail to reject the null?*
- *What if α=0.01?*

Finding the p-value

- Follow the usual hypothesis testing steps but rather than picking α and identifying the rejection region, **determine the significance level of your test statistic** (keeping the alternative hypothesis in mind and thus whether you are dealing with an "α" or "α/2" situation)
- EX) 2-sided alternative and Z-stat: p-value is 2*P(z > |Z-stat|)
- EX) 1-sided alternative and Z-stat: p-value is P(z > |Z-stat|)
- Similar for T (with appropriate d.f.) because also symmetric

21

MICHIGAN STATE UNIVERSITY

Example #1 – finding the p-value for a hypothesis test

- Recall Example 10.12. We were testing H_0: μ=3000 vs. H_1: μ<3000, and calculated **T = -2.966** with **d.f. = N-1 = 8-1= 7**
- *What is the p-value for this test, i.e., what is $P(T < -2.966)$?*

22

MICHIGAN STATE UNIVERSITY

Example #2 – finding the p-value for a hypothesis test

EXAMPLE **10.11** Find the p-value for the statistical test of Example 10.7.

Example 10.7 presents a test of the null hypothesis $H_0 : \mu_1 - \mu_2 = 0$ versus the alternative hypothesis $H_a : \mu_1 - \mu_2 \neq 0$. The value of the test statistic, computed from the observed data, was $z = -2.5$.

23

Homework:

- WMS Ch. 10 (cont'd):
  - CIs & hypothesis testing: 10.45, 10.49
  - p-values: 10.50, 10.51, 10.52 (WMS key is wrong), 10.53
  - Small-sample hypothesis tests for μ: 10.61, 10.62, 10.63 (a and b only), 10.64 (a only), 10.67
- **All Ch. 10 HW will be due on Tues., Nov. 28

Next class:

- Calculating Type II error probabilities and finding the sample size for Z tests
- The "power" of statistical tests
- Wrap-up of Chapter 10

Reading for next class:

- WMS Ch. 10 (sections 10.4, 10.10, 10.12)

24

MICHIGAN STATE UNIVERSITY

Table 8.1 Expected values and standard errors of some common point estimators

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\dfrac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \dfrac{Y}{n}$	p	$\sqrt{\dfrac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ *†
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}$ †

Square root of variance of estimator

*σ_1^2 and σ_2^2 are the variances of populations 1 and 2, respectively.

†The two samples are assumed to be independent.

In-class exercise on hypothesis testing

EXAMPLE **10.6**

A machine in a factory must be repaired if it produces more than 10% defectives among the large lot of items that it produces in a day. A random sample of 100 items from the day's production contains 15 defectives, and the supervisor says that the machine must be repaired. Does the sample evidence support his decision? Use a test with level .01.

***We already did this example. **Now find the p-value for your test statistic.**

In-class exercise on hypothesis testing

EXAMPLE **10.7**

A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results are shown in Table 10.2. Do the data present sufficient evidence to suggest a difference between true mean reaction times for men and women? Use $\alpha = .05$.

***We already did this example. **Now find the p-value for your test statistic.**

Table 10.2 Data for Example 10.7

Men	Women
$n_1 = 50$	$n_2 = 50$
$\overline{y}_1 = 3.6$ seconds	$\overline{y}_2 = 3.8$ seconds
$s_1^2 = .18$	$s_2^2 = .14$