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

Hypothesis Testing – Part 1 of 3

(WMS Ch. 10.1-10.3)

November 9, 2017

Mary Doidge & Nicole Mason
Michigan State University
Fall 2017

GAME PLAN

- Collect Ch. 8 HW; return graded HW & exercises
- Review
- Hypothesis testing – Part 1 of 3
  - Motivation / intuition on hypothesis testing
  - Type I vs. Type II error
  - The steps in the hypothesis testing procedure
  - Examples (large sample hypothesis testing)

1

REVIEW

MICHIGAN STATE UNIVERSITY

3 important properties of estimators

1. **Unbiasedness**:
2. **Efficiency**:
3. **Consistency**:
   - An unbiased estimator is consistent if:
   - Note that consistency does NOT imply unbiasedness (but unbiasedness plus zero asymptotic variance does imply consistency)
   - **Unbiasedness is nice, but consistency is essential**

3 common methods of estimation

1. **Method of moments**
2. **Maximum likelihood**
3. **Least squares**

2

REVIEW: Method of moments

MICHIGAN STATE UNIVERSITY

Method #1: The method of moments (MOM)

- The gist: replace population moments" (expected values) with their sample analogues
- *What would you propose as the MOM estimator of:*
  - *$E(Y^2)$?*
  - *$V(Y)=E(Y^2)-[E(Y)]^2$*
- **Pros**:
  - Easy & intuitive to use
  - Consistent
- **Cons**:
  - Often biased
  - Typically not very efficient

3

REVIEW: Maximum likelihood estimation

MICHIGAN STATE UNIVERSITY

Method #2: Maximum likelihood estimation (MLE)

- The gist: Finding the value of $\hat{\theta}$ that **maximizes** the likelihood function (joint distribution)
  - In practice, maximize log likelihood function
- **Pros**:
  - Usually consistent, often unbiased
  - Often most (asymptotically) efficient estimator
- **Cons**:
  - No major cons

4

REVIEW: Least squares

MICHIGAN STATE UNIVERSITY

Method #3: Least squares

- The gist: Finding the value of $\hat{\theta}$ that **minimizes** the sum of squared deviations between the observed values and the estimated values $\sum_{i=1}^{N}\left(Y_i - \hat{\mu}\right)^2$

5

MICHIGAN STATE UNIVERSITY

HYPOTHESIS TESTING

6

MICHIGAN STATE UNIVERSITY

Hypothesis testing: Motivation

- The main **objective of statistics** is to **make inferences** about unknown population parameters based on information contained in sample data
- Previous 2 sections of the course: how to estimate population parameters from sample data, and some desirable properties of estimators
- **Statistical inference** = testing hypotheses about population parameters
- Once we have our estimate of a given population parameter, can test whether it is equal to zero or to some other value, including the values of other population parameters. *Examples from your work?*

Motivation (cont'd)

- Suppose that in a recent election Candidate A got 42% of the vote, and Candidate B got 58%
- Candidate A is convinced he got more than 42% of the vote, so hires a consultant to randomly sample 100 voters and record if they voted for A or B
  - 53 of them voted for candidate A
  - → sample implies 53% voted for Candidate A, but official results were that 42% voted for Candidate A
  - *Enough to conclude that there was election fraud? How strong is the sample evidence against the official results?*
- Can set up a **hypothesis test** to determine this

Motivation (cont'd)

- Let θ be the true proportion of the population that voted for Candidate A
- The hypothesis that the official results are accurate can be stated as $\boldsymbol{H_0}$: $\theta = 0.42$ ("**null hypothesis**")
  - Null hypothesis is presumed to be true until the data strongly suggest otherwise (innocent until proven guilty)
- Candidate A believes he got more than 42% of the vote, so the "**alternative hypothesis**" of interest is $\boldsymbol{H_1}$: $\theta > 0.42$
- In order to reject H_0 in favor of H_1, we need to have evidence "beyond a reasonable doubt" against H_0
- *Is 53 out of 100 strong enough to reject H_0?*
  - Depends on how we quantify "beyond a reasonable doubt"

In hypothesis testing, we can make two kinds of mistakes:

Type I and Type II errors

		REALITY	
		NULL HYPOTHESIS	
		TRUE	FALSE
Conclusion of your hypothesis test/study: the null is…	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**
  - In medical stats: "false positive"
  - Probability: α (significance level)
  - *In our candidate A example?*

Candidate A example:
H_0: $\theta = 0.42$
H_1: $\theta > 0.42$

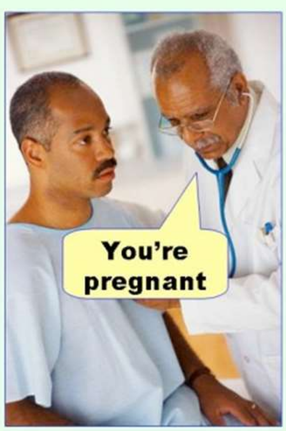
- **Type II error: fail to reject H_0 when H_0 is false**
  - In medical stats: "false negative"
  - Probability: β
  - *In our candidate A example?*

MICHIGAN STATE UNIVERSITY

Type I vs. Type II error

What are H_0 and H_1 here?

Type I error
(false positive)



Reject H_0 when H_0 is true

Type II error
(false negative)

Fail to reject H_0 when H_0 is false

Hypothesis testing rules are constructed to:

1. **Make the probability of Type I error fairly small**
   - α is the "significance level" (or simply "level") of the test
   - Commonly set at 0.01, 0.05, or 0.10
   - *What does α=0.05 mean?*

2. **Minimize the probability of Type II error (β) given the chosen significance level (α)**
   - We'll come back to this later in Chapter 10 when we talk about the "power" of a test, which is 1- β

Hypothesis testing procedure

1. State the **null & alternative hypotheses**. EX)
2. Define an appropriate **test statistic** (like an estimator; a function of the sample measurements on which the statistical decision will be based). EX)
3. Determine the **distribution of the test statistic under the null** hypothesis. EX)

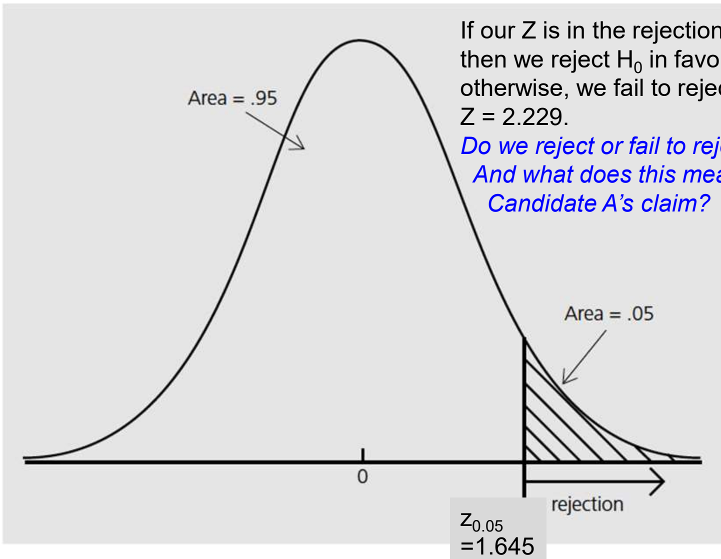
   In general, $\hat{p} \sim N\left(p, \frac{pq}{N}\right)$. Under H_0 :
4. **Standardize the test statistic** to something with known/tabled probabilities for its sampling distribution (e.g., *Z, t, chi-square, F*)

   EX) $Z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} \sim N(0, 1)$ in general, so in our example $Z =$

1. 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) & a **rejection region** (values of standardized test statistic that lead to rejection of H_0)
2. **Reject the null hypothesis if** the standardized statistic lies **in the rejection region**; fail to reject otherwise

MICHIGAN STATE UNIVERSITY

Rejection region for our example if we choose α=0.05



If our Z is in the rejection region, then we reject H_0 in favor of H_1; otherwise, we fail to reject H_0. Our Z = 2.229.

Do we reject or fail to reject H_0? And what does this mean about Candidate A's claim?

14

MICHIGAN STATE UNIVERSITY

Notes on the language of hypothesis testing

- We either **reject** or **fail to reject** a hypothesis; we never accept or prove a hypothesis
- "Reject the null hypothesis (at the α*100% level) in favor of the alternative hypothesis"
- "Fail to reject the null hypothesis (at the α*100% level) in favor of the alternative hypothesis"

15

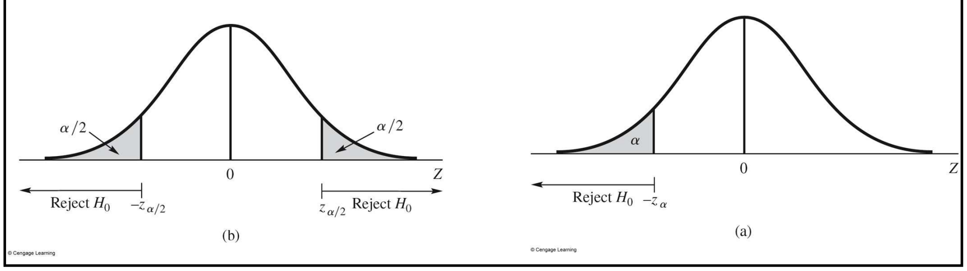
Two- vs. one-sided alternatives & associated rejection regions for *Z*-statistics (similar for *t*)

$$H_0: \ \theta = \theta_0 \quad \text{(null hypothesis)}$$

$$H_1: \begin{cases} \theta \neq \theta_0 & \text{(two-sided (two-tailed) alternative hypothesis)} \\ \theta > \theta_0 & \text{(one-sided (upper-tail) alternative hypothesis)} \\ \theta < \theta_0 & \text{(one-sided (lower-tail) alternative hypothesis)} \end{cases}$$

Rejection region for **two-sided alternative**

Rejection region for **one-sided (lower-tail) alternative**



Example: testing a hypothesis about *μ* against a two-sided alternative hypothesis

The mean lifetime of a sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a sample standard deviation of 120 hours. If μ is the mean lifetime of all the bulbs produced by the company, test the null hypothesis μ = 1600 hours against the alternative hypothesis $\mu \neq$ 1600 hours, using a level of significance of (a) 0.05 and (b) 0.01. (Note that N is large, so we can use the sample standard deviation as an estimate of σ.)

1. State the **null & alternative hypotheses**.
2. Define an appropriate **test statistic** (like an estimator; a function of the sample measurements on which the statistical decision will be based).
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) & 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

17

Continuing our previous example: Testing a hypothesis about μ against a one-sided lower tail alternative hypothesis

Now test the hypothesis μ = 1600 hours against the alternative hypothesis μ < 1600 hours, using a level of significance of (a) 0.05, (b) 0.01.

1. State the **null & alternative hypotheses**.
2. Define an appropriate **test statistic** (like an estimator; a function of the sample measurements on which the statistical decision will be based).
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) & 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

18

MICHIGAN STATE UNIVERSITY

Summary

Large-Sample α-Level Hypothesis Tests

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

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

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

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

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

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Square root of variance of estimator ← 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	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}^{\,*\dagger}$
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1q_1}{n_1} + \dfrac{p_2q_2}{n_2}}^{\,\dagger}$

$^*\sigma_1^2$ and σ_2^2 are the variances of populations 1 and 2, respectively.

†The two samples are assumed to be independent.

Homework:

- WMS Ch. 10
  - Large-sample hypothesis tests (section 10.3): 10.17-10.21 (excluding part e on 10.17)

Next class:

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

Reading for next class:

- WMS Ch. 10 (sections 10.5-10.8)

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.

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 $\alpha = .05$. Also 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
$s_1^2 = .18$	$s_2^2 = .14$