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

Estimation – Part 2 of 2
(WMS Ch. 8.5-8.8, & 8.10)
November 2, 2017

Nicole Mason
Michigan State University
Fall 2017

GAME PLAN

- Reminder: Ch. 8 HW likely due on Tuesday
- Review
- Distribute graded take-home exercise (due next Tuesday and is in lieu of a graded in-class exercise on Tuesday)
- Finish Ch. 8 – focus on interval estimates
  - Large sample confidence intervals
  - Small sample confidence interval for μ
  - Selecting the sample size

1

Review

MICHIGAN STATE UNIVERSITY

- **Target parameter:** parameter we want to estimate
- **Point estimate:** Single value given as estimate – ex) 1.7
- **Interval estimate**: Range of values given as estimate – ex) [1.2, 2.2]
- **Estimator**: rule/formula used to calculate estimate of target parameter from sample data (e.g., sample mean)

$$\boxed{\hat{\theta} \text{ is unbiased if } E(\hat{\theta}) = \theta. \quad \text{Bias: } B(\hat{\theta}) = E(\hat{\theta}) - \theta}$$

$$\boxed{\text{Error of estimation: } \varepsilon = \left|\hat{\theta} - \theta\right|}$$

$$\boxed{\text{Mean square error: } MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = V(\hat{\theta}) + \left[B(\hat{\theta})\right]^2}$$

$$\boxed{\begin{array}{l}\text{If } \hat{\theta}_1 \text{ and } \hat{\theta}_2 \text{ are both unbiased estimators of } \theta, \text{ then } \hat{\theta}_1 \\ \text{is said to be "more efficient" than } \hat{\theta}_2 \text{ if } V(\hat{\theta}_1) \le V(\hat{\theta}_2).\end{array}}$$

2

MICHIGAN STATE UNIVERSITY

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

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	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	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ *†
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$ †

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

†The two samples are assumed to be independent.

MICHIGAN STATE UNIVERSITY

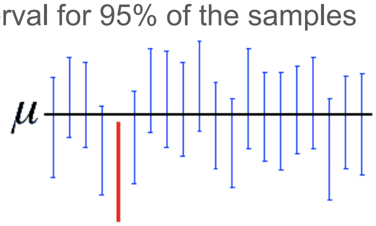
Interval estimators

- So far have focused on point estimators; now shifting focus to interval estimators
- **Interval estimator:** a rule (e.g., formula) specifying the method for using the sample measurements to calculate two numbers that form the endpoints of the interval
- Desirable properties for the interval estimate:
  1. There is a high probability that the target parameter, *θ,* falls in the interval
  2. The smaller/narrower the interval, the better. *Why?*

4

Confidence intervals

- Interval estimators often referred to as “confidence intervals”
- **Confidence interval (CI)**: a rule used to construct a random interval so that a certain percentage of all data sets yields an interval that contains the population value (target parameter)
  - EX) 95% CI: if collect repeated random samples and calculate the endpoints of the interval for each sample, then the population parameter would lie in the interval for 95% of the samples
  - *So what is a 90% CI? 99% CI?*



Source: Minitab

- The percentage we choose for the confidence interval is called the **confidence level** (**confidence coefficient** in WMS)

$$\boxed{P(\hat{\theta}_L \le \theta \le \hat{\theta}_U) = 1-\alpha}$$

- *What is the confidence level here?*

MICHIGAN STATE UNIVERSITY

Two-sided & one-sided confidence intervals

$$P(\hat{\theta}_L \le \theta \le \hat{\theta}_U) = 1-\alpha$$

What is the implied confidence interval for this probability statement?

The interval, $[\hat{\theta}_L, \hat{\theta}_U]$, is the $(1\text{-}\alpha)\%$ two-sided confidence interval for θ

$$P(\hat{\theta}_L \le \theta) = 1-\alpha$$

What is the implied confidence interval for this probability statement?

$[\hat{\theta}_L, \infty)$ is the $(1\text{-}\alpha)\%$ lower one-sided confidence interval for θ

$$P(\theta \le \hat{\theta}_U) = 1-\alpha$$

What is the implied confidence interval for this probability statement?

$(-\infty, \hat{\theta}_U]$ is the $(1\text{-}\alpha)\%$ upper one-sided confidence interval for θ

Large sample confidence intervals

- *By the CLT, what is the approximate sampling distribution for each of the estimators below for "large" samples?*
  - **Approximately normal** w/ the means and standard errors given below
  - We can use this to help us form confidence intervals

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	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$ †

Large sample confidence intervals (cont'd)

We want to find a (1-*α*)*100% confidence interval for *θ*, *i.e.;*

$$\boxed{[\hat{\theta}_L, \hat{\theta}_U] \text{ such that } P(\hat{\theta}_L \le \theta \le \hat{\theta}_U) = 1-\alpha}$$

Let $\hat{\theta}$ be a statistic that is normally distributed with mean θ and standard error $\sigma_{\hat{\theta}}$.

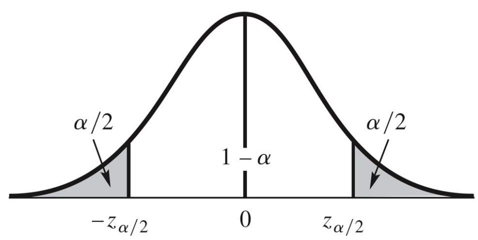
Convert $\hat{\theta}$ to a Z statistic and determine the sampling distribution of Z.

$$\boxed{Z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \sim Normal(0,1)}$$

$$\boxed{\begin{array}{l}\text{So we need to find } z_{\alpha/2} \text{ such that:}\\ P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha\end{array}}$$

Once we have $z_{\alpha/2}$,
how do we recover $\hat{\theta}_L$ and $\hat{\theta}_U$?

$$\boxed{\begin{array}{l}\hat{\theta}_L = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}\\ \hat{\theta}_U = \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\end{array}} \boxed{\Leftrightarrow \hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}}$$



$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$, $Z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}$

$P(-z_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \le z_{\alpha/2}) = 1-\alpha$

$P(-z_{\alpha/2}\sigma_{\hat{\theta}} \le \hat{\theta}-\theta \le z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$

$\boxed{P(\hat{\theta}_L \le \theta \le \hat{\theta}_u) = 1-\alpha}$

$P(-\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \le -\theta \le -\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$

$P(\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}} \ge \theta \ge \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$

$P(\underbrace{\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}}_{} \le \theta \le \underbrace{\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}}_{\hat{\theta}_u}) = 1-\alpha$

$\Rightarrow \hat{\theta}_L =$

MICHIGAN STATE UNIVERSITY

Note on substituting S for σ when constructing **large sample** CIs

- If the true value of σ (population standard deviation) is known, then use it when constructing CI
- If σ not known but N is large (30+), then there is no serious loss of accuracy is we substitute S for σ in the large sample CI formula (still Z-statistic b/c of CLT)
- After we do an example of large sample CIs, we'll study small sample CIs where the CLT doesn't apply, so we'll need to use a t-statistic instead of a Z-statistic if we don't know σ

10

$\hat{\theta}_L = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}$
$\hat{\theta}_U = \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}$

$\Leftrightarrow \hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$

MICHIGAN STATE UNIVERSITY

Large sample confidence intervals - example

EXAMPLE **8.7**

The shopping times of $n = 64$ randomly selected customers at a local supermarket were recorded. The average and variance of the 64 shopping times were 33 minutes and 256 minutes^2, respectively. Estimate μ, the true average shopping time per customer, with a confidence coefficient of $1 - \alpha = .90$. $\Rightarrow \alpha = 0.1 \Rightarrow \alpha/2 = 0.05$

level

Note: N is large (30+) here, so we can substitute the sample s.e. for the population s.e. with no serious loss of accuracy for our CI. ok to sub in S b/c N is large

$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}} \Rightarrow \bar{Y} \pm z_{\alpha/2}\sigma_{\bar{Y}} \Rightarrow \bar{Y} \pm z_{\alpha/2}\sigma/\sqrt{N}$

$33 \pm 1.645 \frac{\sqrt{256}}{8} \Rightarrow [29.71, 36.29]$ is the 90% CI for μ

11

MICHIGAN STATE UNIVERSITY

Small sample CI for μ for random samples drawn from roughly mound-shaped distributions

- 3 key things for the CIs for μ given in this section:
  1. Small sample (N < 30)
  2. Random sample
  3. Sample from roughly mound-shaped distribution (roughly normal)

12

Small sample CI for μ for random samples drawn from roughly mound-shaped distributions

Suppose we have a random sample $Y_1, Y_2, ..., Y_N$ from a normal distribution. We know $\bar{Y}$ and S^2 but not $V(Y_i) = \sigma^2$. We want to construct a CI for the population mean, μ, but the sample size is too small for the CLT to apply. We can instead use:

$$\boxed{T = \frac{\bar{Y} - \mu}{S / \sqrt{N}} \sim t \text{ with } (N - 1) \text{ d.f.}}$$

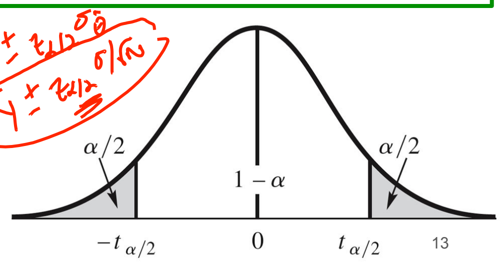
To construct a $(1 - \alpha) * 100\%$ CI for μ, we need to find $t_{\alpha/2}$ such that:
$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1 - \alpha$

Once we have $t_{\alpha/2}$, then the small-sample CI for μ is:

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

$\hat{\theta} \pm t_{\alpha/2}\, \sigma_{\hat{\theta}}$

$\bar{Y} \pm z_{\alpha/2}\, \sigma/\sqrt{n}$



13

MICHIGAN STATE UNIVERSITY

Small sample CI for μ - example

Small-sample CI for μ is:
$$\bar{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{N}}\right)$$

EXAMPLE **8.11**

A manufacturer of gunpowder has developed a new powder, which was tested in eight shells. The resulting muzzle velocities, in feet per second, were as follows:

$N = 8$

$\alpha = 0.05$

3005	2925	2935	2965
2995	3005	2937	2905

Find a 95% confidence interval for the true average velocity μ for shells of this type. Assume that muzzle velocities are approximately normally distributed.

The sample mean is $\underset{\bar{Y}}{\underline{2959}}$ and the sample standard deviation is $\underset{S}{\underline{39.1}}$.

$\bar{Y} \pm t_{\alpha/2}\, S/\sqrt{N}$

$2959 \pm 2.365\; 39.1/\sqrt{8} \Rightarrow [2926.3, 2991.7]$

is the 95% CI for μ

14

MICHIGAN STATE UNIVERSITY

Selecting the sample size (simple random sampling case)

- How large of a sample you need depends on how accurate you want your estimates to be (and your budget constraint) $67.18 \Rightarrow 68$
- E.g., say you want the sample mean to be within 5 units of the population mean with 95% probability, then you could solve for N such that:

$$P(|\bar{Y}-\mu| \le 5) = P(-5 \le \bar{Y}-\mu \le 5) = 0.95$$

$$= P(\underbrace{\frac{-5}{\sigma/\sqrt{N}}}_{-z_{\alpha/2}} \le Z \le \underbrace{\frac{5}{\sigma/\sqrt{N}}}_{z_{\alpha/2}}) = 0.95$$

$\frac{5}{S/\sqrt{N}} = z_{\alpha/2} = 1.96$

$\frac{5}{21/\sqrt{N}} = 1.96$

67.78

$\Rightarrow N = 68$

Solve for N, then plug in σ if you have it (to get exact sample size), or S if you don't (to get approximate sample size). **Round up** to the nearest N. Suppose $S = 21$ in our example.

15

MICHIGAN STATE UNIVERSITY

Selecting the sample size (simple random sampling case) - another example

$|\hat{\theta}-\theta|$

EXAMPLE 8.9 $z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} = \frac{\hat{p}-p}{\sigma_{\hat{p}}}$

The reaction of an individual to a stimulus in a psychological experiment may take one of two forms, A or B. If an experimenter wishes to estimate the probability p that a person will react in manner A, how many people must be included in the experiment? Assume that the experimenter will be satisfied if the error of estimation is less than .04 with probability equal to .90. Assume also that he expects p to lie somewhere in the neighborhood of .6. $\approx p$

$\sigma_{\hat{p}} = \sqrt{\frac{pq}{N}}$

$P(|\hat{p}-p| < 0.04) = 0.9$

$\Leftrightarrow P(-0.04 < \hat{p}-p < 0.04) = 0.9$

$P\left(\frac{-0.04}{\sqrt{\frac{pq}{N}}} < z < \frac{0.04}{\sqrt{\frac{pq}{N}}}\right) = 0.9 \Rightarrow z_{\alpha/2} = \frac{0.04}{\sqrt{\frac{pq}{N}}}$

1.645

$\Rightarrow$ solve for N

$N = 406.07 \Rightarrow$ sample size of 407

16

MICHIGAN STATE UNIVERSITY

Intuitively, how does the sample size change with increases in the standard deviation? Confidence level? Acceptable distance from the true population parameter?

For the sample mean,

$$N = \left(\frac{z_{\alpha/2}\sigma}{\text{acceptable distance from } \mu}\right)^2$$

(Rounded up to the nearest whole integer)

17

MICHIGAN STATE UNIVERSITY

Selecting the sample size – free software

- Optimal Design (out of U of M)
- Works for simple random sampling and more complex sampling schemes
- https://sites.google.com/site/optimaldesignsoftware/home
- See http://blogs.worldbank.org/impactevaluations/power-calculations-what-software-should-i-use for more on pros/cons of Optimal Design and other software options

18

Summary of Chapter 8

- Major objective of statistics is to **make inferences about population parameters based on sample data**
- Often inferences take the form of estimates – either **point estimates or interval estimates**
- We prefer **unbiased** estimators with **small variance**
- **MSE** gives us a **combined measure** of the **bias** and **variance** of an estimator
- **Confidence intervals** for many parameters can be derived from the **normal distribution b/c of the CLT**
- If **sample size is small** and we don't know the population variance, then can use the ***t* distribution** when deriving **confidence intervals for μ**

19

Homework:

- WMS Ch. 8 (part 2 of 2)
  - Large sample confidence intervals: 8.56, 8.57, 8.58, 8.60
  - Small sample confidence intervals for μ: 8.80, 8.81 (feel free to use Excel to compute the sample mean and standard deviation for 8.81)
  - Selecting the sample size: 8.70, 8.74

•**All Ch. 8 HW will likely be due on Tuesday

•**Take home graded exercise due on Tuesday

Next class **[Guest lecture by Mary Doidge]**

- Properties of point estimators (cont'd) – consistency
- Intuition on some common methods of estimation (maximum likelihood & method of moments)

Reading for next class:

- WMS Ch. 9 (sections 9.1, 9.3, 9.6-9.7, 9.9)

$\boxed{\begin{aligned}\hat{\theta}_L &= \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \\ \hat{\theta}_U &= \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\end{aligned}}$ $\boxed{\Leftrightarrow \hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}}$

Extra in-class exercise: large sample CI

EXAMPLE **8.8**

See Table 8.1 for the formulas required to answer this question.

Two brands of refrigerators, denoted A and B, are each guaranteed for 1 year. In a random sample of 50 refrigerators of brand A, 12 were observed to fail before the guarantee period ended. An independent random sample of 60 brand B refrigerators also revealed 12 failures during the guarantee period. Estimate the true difference $(p_1 - p_2)$ between proportions of failures during the guarantee period, with confidence coefficient approximately .98.

Note: N is large (30+) here, so we can substitute the sample s.e. for the population s.e. with no serious loss of accuracy for our CI.

MICHIGAN STATE UNIVERSITY

Extra practice problems

MICHIGAN STATE UNIVERSITY

Extra practice problem #1

Acme Corporation manufactures light bulbs. The CEO claims that an average Acme light bulb lasts 300 days. A researcher randomly selects 15 bulbs for testing. The sampled bulbs last an average of 282.6 days, with a sample standard deviation of 50 days. If the CEO's claim were true, what is the probability that 15 randomly selected bulbs would have an average life of no more than 282.6 days? (Hint: Is this a standard normal or *t* distribution question? Why?)

MICHIGAN STATE UNIVERSITY

Extra practice problem #2

The reading on a voltage meter is a random variable, Y, that is uniformly distributed over the interval (θ, $\theta+1$). Suppose that Y_1, Y_2, …, Y_N is a random sample of readings from the voltage meter. From our study of the uniform distribution, we know that for $i=1, 2, \ldots, N$: $E(Y_i) = \theta + \frac{1}{2}$ and $V(Y_i) = \frac{1}{12}$

a. You want to estimate θ. Find the expected value and bias of $\bar{Y}$ when it is used as an estimator of θ.

b. Find the variance of $\bar{Y}$ when it is used as an estimator of θ.

c. Find the MSE of $\bar{Y}$ when it is used as an estimator of θ.

24

25