MICHIGAN STATE UNIVERSITY

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

Estimation – Part 1 of 2

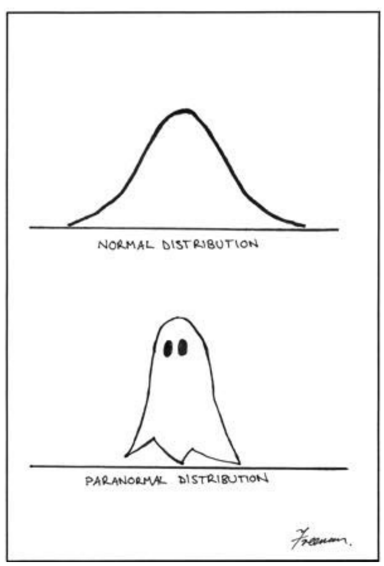
(WMS Ch. 8.1-8.4)

October 31, 2017

Nicole Mason
Michigan State University
Fall 2017

MICHIGAN STATE UNIVERSITY

A little Halloween stats humor



1

MICHIGAN STATE UNIVERSITY

GAME PLAN

-Housekeeping issue: collect Ch. 7 HW

-Graded in-class exercise on sampling distributions

-Review: CLT, LLN, and normal approximation to binomial

-Ch. 8 (Estimation – yay!)

a. Definitions
b. The bias & mean square error of an estimator
c. Some common unbiased estimators
d. The standard error of an estimator
e. The error of estimation

2

Graded in-class exercises - sampling distributions

If $Y_1, Y_2, \ldots, Y_N$ is a random sample of sizes N from a normal distribution with mean, μ, and variance, σ^2, then:

$$\boxed{Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{N}} \sim N(0,1)}$$

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

$$\boxed{\frac{(N-1)S^2}{\sigma^2} \sim \chi^2 \text{ with } (N - 1) \text{ d.f.}}$$

If we have two independent random samples from normal populations with variances σ_1^2 and σ_2^2, then:

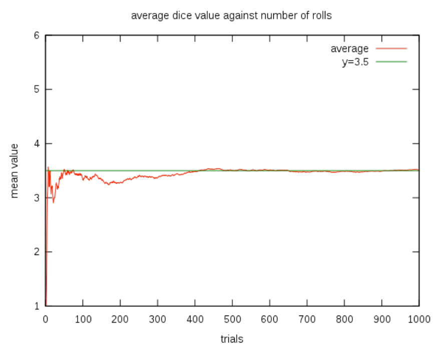
$$\boxed{F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F \text{ with } (N_1 - 1) \text{ numerator d.f. \& } (N_2 - 1) \text{ denominator d.f.}}$$

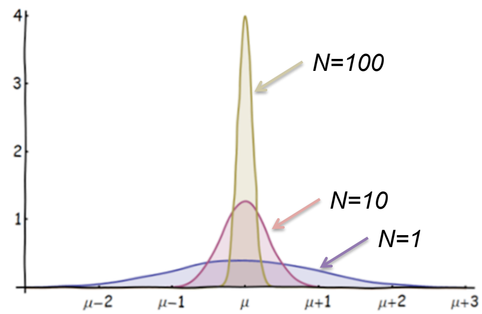
3

Review: The Law of Large Numbers

- **As $N \to \infty$, the sample mean converges (in probability) to the population mean**

$$P(|\bar{Y}_N - \mu| > \varepsilon) \to 0 \quad \text{as } N \to \infty \quad \text{for any } \varepsilon > 0$$
$$\Leftrightarrow$$
$$P(|\bar{Y}_N - \mu| < \varepsilon) \to 1 \quad \text{as } N \to \infty \quad \text{for any } \varepsilon > 0$$



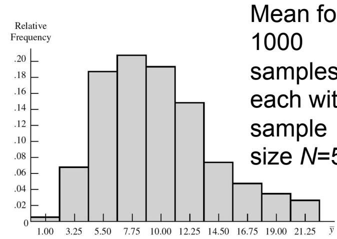


Review: The Central Limit Theorem (CLT)

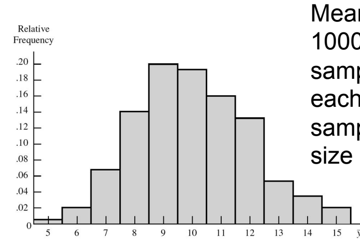
- **As $N \to \infty$**, the **sampling distribution of the sample mean will be approximately normal** regardless of the distribution of Y_i

Let $Y_1, Y_2, \ldots, Y_N$ be i.i.d. distributed RVs with $E(Y_i) = \mu$, $V(Y_i) = \sigma^2 < \infty$, then the distribution of $\dfrac{\bar{Y} - \mu}{\sigma / \sqrt{N}}$ converges to the standard normal as $N \to \infty$

- "Large" sample size: roughly ***N*≥*30***
- Note: CLT applies to a random sample from **ANY distribution** with finite mean & variance & large N



Mean for 1000 samples, each with sample size *N*=5



Mean for 1000 samples, each with sample size *N*=25

5

Review: Normal approx. to binomial distribution

- Recall that a binomial RV, *Y*, is the # of successes in *n* trials, where the *P(success)* on one trial is *p*

$$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

- Can think of as *Y* as the sum of *n* binary variables

$$Y = \sum_{i=1}^{n} X_i, \qquad X_i = \begin{cases} 1, & \text{if the } i\text{th trial results in success,} \\ 0, & \text{otherwise.} \end{cases}$$

- Divide both sides by *n*: $\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

- As *n* gets large, by the CLT:

$$\boxed{\frac{Y}{n} = \overline{X} \sim Normal(p, \frac{pq}{n})}$$

Note: This approximation works well if:

$$n > 9\left(\frac{\text{larger of } p \text{ and } q}{\text{smaller of } p \text{ and } q}\right)$$

MICHIGAN STATE UNIVERSITY

ESTIMATION (FINALLY!)

7

MICHIGAN STATE UNIVERSITY

Motivation

- *Recall from Day 1: what are the two major objectives of statistics?*
  1. To make an inference about a population based on info in a sample from that population
  2. To provide a measure of the 'goodness' of that inference
- This section of the course is about estimation. *What might we want to estimate?*
  - In quantitative work, we are usually interested in some numerical descriptive measure of the population – e.g., the population mean (μ), variance (σ^2), prob. of "success" (p), etc.
  - *Examples that may be of interest in your research?*
  - These are called (population) **parameters**
  - En route to making inferences, we'll often need to use our sample info to come up with an estimate of the (population) parameter(s)

8

MICHIGAN STATE UNIVERSITY

Terminology

- **Target parameter** = the parameter that we are trying to estimate
- **Point estimate** vs. **interval estimate** (e.g., for the population mean, μ). *What's the difference?*
  - **Point**: Single value given as estimate – EX) 0.5
  - **Interval**: Range of values given as estimate – EX) (0.3, 0.7)
  - First focus on point estimates, then interval estimates
- **Estimator** = rule (e.g., formula) used to calculate estimate of target parameter from sample data
- *Estimator for the population mean?*

$$\boxed{\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i}$$

What makes this a "good" estimator?

9

One measure of “goodness”: unbiasedness

- *What is an unbiased estimator?*

Notation: Let $\hat{\theta}$ denote the point estimator of θ.

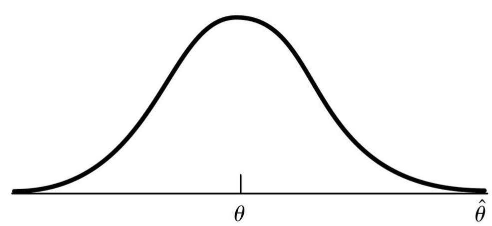
What do we want $E(\hat{\theta})$ to equal?

We want $E(\hat{\theta}) = \theta$.
If true, then $\hat{\theta}$ is an "**unbiased estimator**" of θ.

- *What is a biased estimator?* $E(\hat{\theta}) \neq \theta$
- *How could we measure the bias in our estimator?*

Bias: $B(\hat{\theta}) = E(\hat{\theta}) - \theta$
Positive bias if $B(\hat{\theta}) > 0$, i.e., $E(\hat{\theta}) > \theta$
Negative bias if $B(\hat{\theta}) < 0$, i.e., $E(\hat{\theta}) < \theta$

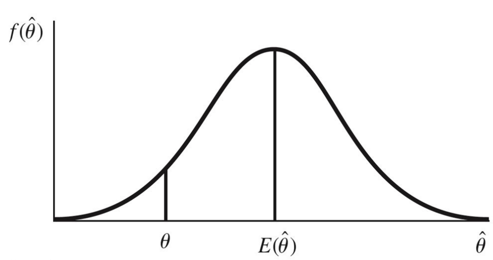
10



$E(\hat{\theta}) = \theta$

Biased or unbiased? And if biased, positive or negative bias?

Sampling distribution for an unbiased estimator of θ



$E(\hat{\theta}) \neq \theta$, with $E(\hat{\theta}) > \theta$ (positively biased)

Biased or unbiased? And if biased, positive or negative bias?

Sampling distribution for a positively biased estimator of θ

11

MICHIGAN STATE UNIVERSITY

8.2 **a** If $\hat{\theta}$ is an unbiased estimator for θ, what is $B(\hat{\theta})$? 0

b If $B(\hat{\theta}) = 5$, what is $E(\hat{\theta})$?

c Is the estimator in (b) positively or negatively biased?

$B(\hat{\theta}) = E(\hat{\theta}) - \theta = 5$

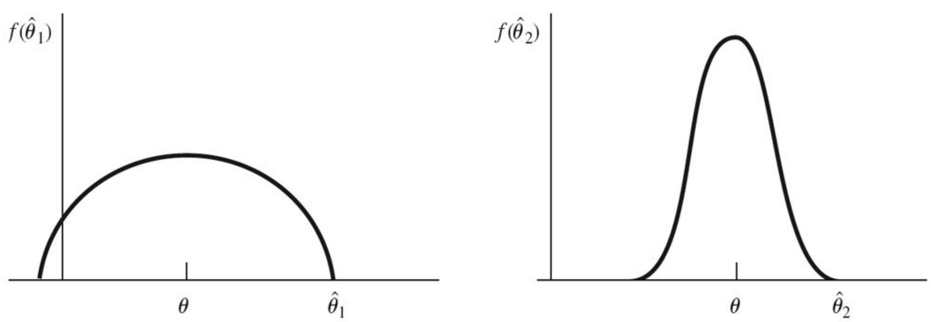
$\Rightarrow E(\hat{\theta}) = 5 + \theta$

Bias: $B(\hat{\theta}) = E(\hat{\theta}) - \theta$
Positive bias if $B(\hat{\theta}) > 0$, i.e., $E(\hat{\theta}) > \theta$
Negative bias if $B(\hat{\theta}) < 0$, i.e., $E(\hat{\theta}) < \theta$

12

Another desirable property of a point estimator: greater "efficiency"

- *Given 2 unbiased estimators with different variances, which would you prefer and why?*
- The one with the **smaller variance**! Also referred to as the "**more efficient**" estimator
- *Which of the estimators below is more efficient?*



(a) (b)

Mean square error (MSE): a combined measure of the variance & bias of an estimator

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

where $V(\hat{\theta}) = E\left[\left(\hat{\theta}-E(\hat{\theta})\right)^2\right]$

See https://www.youtube.com/watch?v=KtNwjbWbnh8 for proof

- *What is the MSE if the estimator is unbiased?*
- *What happens to the magnitude of the MSE as:*
  - *the bias increases?*
  - *the variance increases?*
- *If two estimators have the same mean but different variances, which has the smaller MSE?*
- *What's better: a big MSE or a small MSE?*

14

MICHIGAN STATE UNIVERSITY

Examples – bias and MSE

$MSE(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$

Suppose $B(\hat{\theta}) = 5$ and $V(\hat{\theta}) = 2$.

a. What is $MSE(\hat{\theta})$? 27

b. If another estimator, $\breve{\theta}$, has $B(\breve{\theta}) = 5$ and $V(\breve{\theta}) = 1$, which estimator do you prefer, $\breve{\theta}$ or $\hat{\theta}$?

$MSE(\breve{\theta}) = 26 \Rightarrow$ prefer $\breve{\theta}$

c. If another estimator, $\tilde{\theta}$, has $B(\tilde{\theta}) = 0$ and $V(\tilde{\theta}) = 4$, which estimator do you prefer, $\tilde{\theta}$ or $\hat{\theta}$? $MSE(\tilde{\theta}) = 4$

15

Point estimator for the population mean

Recall that if $Y_1, Y_2, \ldots, Y_N$ is a random sample from a population with $E(Y)=\mu$, and $V(Y)=\sigma^2$, then the sample mean,

$$\boxed{\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \qquad \text{and} \qquad V(\bar{Y}) = \frac{\sigma^2}{N}}$$

- *Is the sample mean an unbiased estimator? Why or why not?*

- *What is the MSE of the sample mean?*

Point estimator for the binomial parameter, *p* (probability of success)

If you have a series of N independent and identical Bernoulli trials, where Y is the # of successes in N trials (i.e., Y is a binomial RV), and p is the probability of success in a single trial, how would you estimate p?

$$\boxed{\hat{p} = \frac{Y}{N}}$$

- *Find the expected value and variance of p-hat*

$$\boxed{\begin{aligned} E(\hat{p}) &= p \\ V(\hat{p}) &= \frac{pq}{N} \end{aligned}}$$

- *Is it an unbiased estimator?*
- *What is the MSE of this estimator?*

MICHIGAN STATE UNIVERSITY

How would you estimate:

a. $\mu_1 - \mu_2$ (i.e., the difference of means from 2 independent populations given random samples of size N_1 and N_2 for these populations)?

$$\boxed{\text{Unbiased estimator for } \mu_1 - \mu_2: \ \bar{Y}_1 - \bar{Y}_2}$$

b. $p_1 - p_2$ (i.e., the difference of binomial parameters for 2 different binomial RVs, Y_1 and Y_2 given N_1 and N_2 independent trials)?

Unbiased estimator for $p_1 - p_2$:

$$\hat{p}_1 - \hat{p}_2 = \frac{Y_1}{N_1} - \frac{Y_2}{N_2}$$

18

MICHIGAN STATE UNIVERSITY

The standard error of an estimator

- A fancy name for the standard deviation of an estimator
- The square root of the variance of an estimator
- A measure of the variability of the estimator

19

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	$\frac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \frac{Y}{n}$	p	$\sqrt{\frac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ *†
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$ †

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

†The two samples are assumed to be independent.

Why we divide by *N-1* instead of *N* in the sample variance formula: to get an unbiased estimator of σ^2

$$\boxed{S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})^2}$$

- Full proof is in the book (pp. 398-399) but gist is that:

Do not use this formula!!!

$$E(S'^2) = E\left[\frac{1}{N}\sum_{i=1}^{N}(Y_i - \overline{Y})^2\right] = \frac{N-1}{N}\sigma^2$$

Is S'^2 an unbiased estimator of σ^2?

$$\boxed{E(S^2) = E\left[\frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})^2\right] = \frac{N-1}{N-1}\sigma^2 = \sigma^2}$$

21

The error of estimation

- *Intuitively, if you wanted to measure how far your estimate was from the true value of the population parameter, θ, what difference would you consider?*

$$\varepsilon = \left|\hat{\theta} - \theta\right| \qquad \varepsilon \text{ is called the "error of estimation"}$$

- We want the error of estimation to be a small as possible
- Less commonly used

22

Homework:

- WMS Ch. 8 (part 1 of 2)
  - Section 8.2: 8.3 (part a only), 8.4, 8.6 (part a only), 8.8 (but ignore θ_4)
- **All Ch. 8 HW will most likely be due on Tuesday

Next class:

- Estimation (Part 2 of 2)

Reading for next class:

- WMS Ch. 8 (sections 8.5-8.8, 8.10)

23

$MSE(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2 = E[(\hat{\theta}-\theta)^2]$

Recall, $V(X) = E(X^2) - [E(X)]^2$

Let $X = \hat{\theta} - \theta$

① $\underbrace{V(\hat{\theta}-\theta)}_{ⓐ} = \underbrace{E[(\hat{\theta}-\theta)^2]}_{ⓑ} - \underbrace{[E(\hat{\theta}-\theta)]^2}_{ⓒ}$

a) $V(\hat{\theta}-\theta) = V(\hat{\theta})$

b) $E[(\hat{\theta}-\theta)^2] = MSE$ per definition

c) $E(\hat{\theta}-\theta) = E(\hat{\theta}) - \theta = B(\hat{\theta})$

$(E(\hat{\theta}-\theta))^2 = (B(\hat{\theta}))^2$

$V(\hat{\theta}) = MSE(\hat{\theta}) - [B(\hat{\theta})]^2$

$\Rightarrow MSE(\hat{\theta}) = V(\hat{\theta}) + (B(\hat{\theta}))^2$

24