ECO220Y

Sampling Distributions of Sample Statistics: Sample Mean

Readings: Chapter 10, section 10.5-10.8

Fall 2011

Lecture 9 Part 2

How to Derive Sampling Distribution Analytically

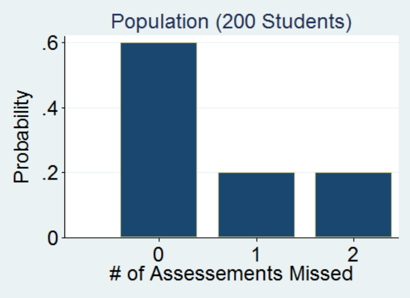
1. List <u>every</u> sample with n observations from the population of interest.
2. Find the probability of obtaining each sample (use probability rules).
3. Calculate the sample statistics for each sample.
4. Link values in 3 with probabilities in 2 (use probability rules).

Example

- In a statistics class of 200, students are allowed to be excused from assessment up to a maximum 2 times per year.
- At the end of the year professor reports the mean number of excuses as 0.6 per student.
- The University randomly selects 3 students and ask them about the number of assessments they missed.
- The mean for the sample is 2: all 3 have been excused twice.
- Why discrepancy between 2 and 0.6?

Example Cont'd

Number of Assessments Missed	Fraction of Students
0	0.6
1	0.2
2	0.2



$\mu = 0.6$
$\sigma^2 = 0.64$

Do you remember how to find these parameters?

List All Possible Samples (1)

sample
0,0,0
0,0,1
0,0,2
0,1,0
0,1,1
0,1,2
0,2,0
0,2,1
0,2,2

$\longrightarrow$

sample
1,0,0
1,0,1
1,0,2
1,1,0
1,1,1
1,1,2
1,2,0
1,2,1
1,2,2

$\longrightarrow$

sample
2,0,0
2,0,1
2,0,2
2,1,0
2,1,1
2,1,2
2,2,0
2,2,1
2,2,2

How many possible samples are there?

Probability of Each Sample (2)

sample	probability	sample	probability	sample	probability
0,0,0	$(0.6)^3 = 0.216$	1,0,0	$(0.6)^2(0.2) = 0.072$	2,0,0	$(0.6)^2(0.2) = 0.072$
0,0,1	$(0.6)^2(0.2) = 0.072$	1,0,1	$(0.6)(0.2)^2 = 0.024$	2,0,1	$(0.6)(0.2)^2 = 0.024$
0,0,2	$(0.6)^2(0.2) = 0.072$	1,0,2	$(0.6)(0.2)^2 = 0.024$	2,0,2	$(0.6)(0.2)^2 = 0.024$
0,1,0	$(0.6)^2(0.2) = 0.072$	1,1,0	$(0.6)(0.2)^2 = 0.024$	2,1,0	$(0.6)(0.2)^2 = 0.024$
0,1,1	$(0.6)(0.2)^2 = 0.024$	1,1,1	$(0.2)^3 = 0.008$	2,1,1	$(0.2)^3 = 0.008$
0,1,2	$(0.6)(0.2)^2 = 0.024$	1,1,2	$(0.2)^3 = 0.008$	2,1,2	$(0.2)^3 = 0.008$
0,2,0	$(0.6)^2(0.2) = 0.072$	1,2,0	$(0.6)(0.2)^2 = 0.024$	2,2,0	$(0.6)(0.2)^2 = 0.024$
0,2,1	$(0.6)(0.2)^2 = 0.024$	1,2,1	$(0.2)^3 = 0.008$	2,2,1	$(0.2)^3 = 0.008$
0,2,2	$(0.6)(0.2)^2 = 0.024$	1,2,2	$(0.2)^3 = 0.008$	2,2,2	$(0.2)^3 = 0.008$

Find the Mean of Each Sample (3)

sample	mean	sample	mean	sample	mean
0,0,0	0	1,0,0	0.33	2,0,0	0.67
0,0,1	0.33	1,0,1	0.67	2,0,1	1
0,0,2	0.67	1,0,2	1	2,0,2	1.33
0,1,0	0.33	1,1,0	0.67	2,1,0	1
0,1,1	0.67	1,1,1	1	2,1,1	1.33
0,1,2	1	1,1,2	1.33	2,1,2	1.67
0,2,0	0.67	1,2,0	1	2,2,0	1.33
0,2,1	1	1,2,1	1.33	2,2,1	1.67
0,2,2	1.33	1,2,2	1.67	2,2,2	2

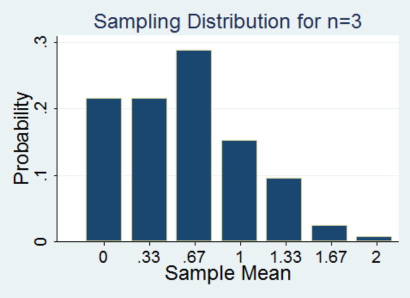
Probability of Each Mean (4)

Sample Mean	Probability
0	.216
0.33	.072+0.072+.072=.216
0.67	.072+.024+.072+.024+.024+.072=.288
1	.024+.024+.024+.008+.024+.024+.024=.152
1.33	.024+.008+.008+.024+.008+.024=.096
1.67	.008+.008+.008=.024
2	.008

Sampling Distribution of the Sample Mean

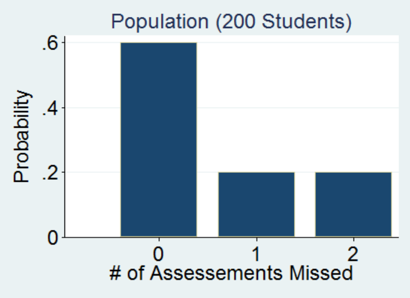
mean	probability
0.00	0.216
0.33	0.216
0.67	0.288
1	0.152
1.33	0.096
1.67	0.024
2	0.008

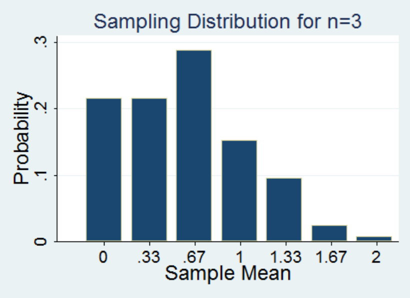
Is this a valid probability distribution?



$$E[\bar{X}] = 0.6$$
$$V[\bar{X}] = 0.214$$
$$sd[\bar{X}] = 0.462$$

Population Distribution vs Sampling Distribution





Compare variance of two distributions. What have you noticed?

Reasons for Discrepancy

- Can sampling error explain the discrepancy between population mean and sample mean? What is the chance that sample mean is equal to 2? In other words, is it statistically plausible to observe value of the sample mean equal to 2 simply by chance?
- Non-sampling errors (HW: Identify non-sampling errors)
- Population parameters are not what professor claimed

Can we find probability that all three students missed 2 assessments without constructing sampling distribution of the sample mean?

Feasibility of Analytical Method

- Given sample size n=3, the number of all possible samples is 3^3.
- What if we would like to increase sample size to 20? Number of samples 3^{20}.
- What if the number of values for the population is greater than 3? Number of samples x^3.

Sampling Distribution of Sample Mean

1. If we know that the underlying population has distribution with mean μ_X and variance σ_X^2.
2. and sample is a sum of random variables, each distributed with μ_X and σ_X^2.
3. then a sample mean, $\bar{X}$, is a sum of these random variables, divided by the sample size n, or $\bar{X} = \frac{X_1+X_2+\ldots+X_n}{n}$.

Parameters of Sampling Distribution of Sample Mean

- Use Laws of Expectation and Variance to find parameters of sampling distribution of sample mean:

$$E[\bar{X}] = E\left[\frac{X_1+X_2+\ldots+X_n}{n}\right] = \frac{1}{n}E[X_1 + X_2 + \ldots + X_n] =$$
$$\frac{1}{n}\left(E[X_1] + E[X_2] + \ldots + E[X_n]\right) = \frac{1}{n}(\mu_x + \mu_x + \ldots + \mu_x) = \frac{1}{n}n\mu_x = \mu_x$$

$$V[\bar{X}] = V\left[\frac{X_1+X_2+\ldots+X_n}{n}\right] = \frac{1}{n^2}V[X_1 + X_2 + \ldots + X_n] =$$
$$\frac{1}{n^2}\left(V[X_1] + V[X_2] + \ldots + V[X_n]\right) = \frac{1}{n^2}(\sigma_X^2+\sigma_X^2+\ldots+\sigma_X^2) = \frac{1}{n^2}n\sigma_X^2 = \frac{\sigma_X^2}{n}$$

- What must be true about the relationship between $X_i's$?

Parameters of Sampling Distribution

- Sample mean $\bar{X}$ is a random variable.
- Distributed with parameters:
  1. Mean: $\mu_{\bar{X}} = \mu_X$
  2. Variance: $\sigma^2_{\bar{X}} = \frac{\sigma^2_X}{n}$
  3. St. dev.: $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$
- Standard deviation of the sampling distribution is called standard error.

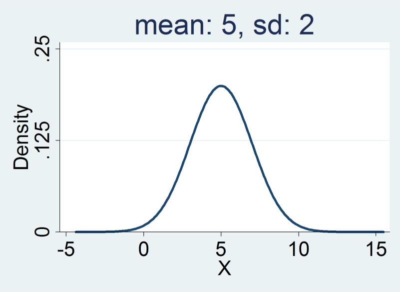
Shape of Sampling Distribution

- Use Central Limit Theorem (CLT) to learn about the shape of sampling distribution of sample mean.
- CLT implies that, no matter what the underlying distribution is, sample mean will tend to a normal distribution.
- CLT states that the sum of n independent, identically distributed random variables approaches a normal distribution as n increases.
- Recall that a sample mean is just a sum of n independent, identically distributed random variables divided by n. $\longrightarrow$ Conditions for the CLT are satisfied $\longrightarrow$ $\boxed{\bar{X} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})}$

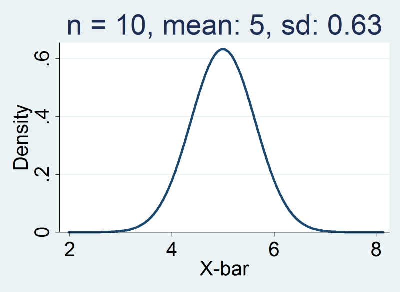
Sufficiently Large Sample Size

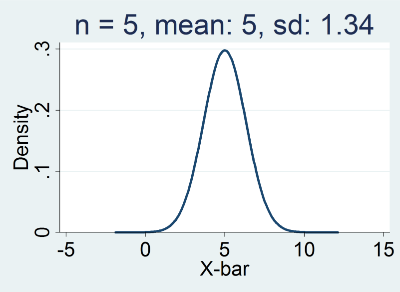
Rule of thumb:

- Sample size at least 30 ($n \geq 30$)
- Normal approximation improves as n rises
- $n < 30$ sufficient for modest departures from normal
- If population is normal, then $n \geq 1$ is sufficient



Which one is a sampling distribution?





Example

North American adults watch TV on average 6 hours per day with a standard deviation of 1.5 hours.

Notation

Example

North American adults watch TV on average 6 hours per day with a standard deviation of 1.5 hours.

A If we randomly sample 5 adults, what is the chance that on average they watch TV more than 7 hours a day?

Notation

$P(\bar{X} > 7 | \mu_x = 6, \sigma_x = 1.5, n = 5) = ?$

Example

North American adults watch TV on average 6 hours per day with a standard deviation of 1.5 hours.

A If we randomly sample 5 adults, what is the chance that on average they watch TV more than 7 hours a day?

B Is it statistically possible that the average time watching television by a random sample of 16 adults is less than 4.5 hours?

Notation

$P(\bar{X} > 7|\mu_x = 6, \sigma_x = 1.5, n = 5) =?$

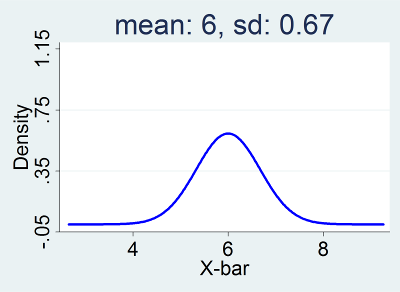
$P(\bar{X} < 4.5|\mu_x = 6, \sigma_x = 1.5, n = 16) =?$

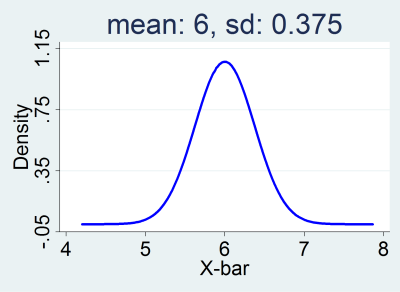
TV Example

$$\mu_{\bar{x}} = \mu_x = 6$$

(A) $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{1.5}{\sqrt{5}} = 0.67$

(B) $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{1.5}{\sqrt{16}} = 0.375$





TV Example

$$P(\bar{X} > 7) = P\left(\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{7 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$$

$$= P\left(Z > \frac{7-6}{0.67}\right) = P(Z > 1.49) = 0.5 - P(0 < Z < 1.49)$$

$$= 0.5 - 0.4319 = 0.0681$$

TV Example

$$P(\bar{X} < 4.5) = P\left(\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} < \frac{4.5 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right)$$

$$= P\left(Z < \frac{4.5 - 6}{0.375}\right) = P(Z < -4)$$

Do we need table to find $P(Z < -4)$?