Central Limit Theorem

Mgmt 230: Introductory Statistics

Goals of this class meeting

- Learn about the behavior of statistics obtained from samples.
- Learn about the particular behavior of the sample mean, and the importance of the normal distribution.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, mode, standard deviation, variance, range, a weighted mean, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
  - Is this the same thing as the probability distribution of the population?
  - The re-sampling experiment is assumed to be done *with replacement* to assure the samples are independent.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, mode, standard deviation, variance, range, a weighted mean, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
  - Is this the same thing as the probability distribution of the population?
  - The re-sampling experiment is assumed to be done *with replacement* to assure the samples are independent.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, mode, standard deviation, variance, range, a weighted mean, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
  - Is this the same thing as the probability distribution of the population?
  - The re-sampling experiment is assumed to be done *with replacement* to assure the samples are independent.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, mode, standard deviation, variance, range, a weighted mean, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
  - Is this the same thing as the probability distribution of the population?
  - The re-sampling experiment is assumed to be done *with replacement* to assure the samples are independent.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, mode, standard deviation, variance, range, a weighted mean, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
  - Is this the same thing as the probability distribution of the population?
  - The re-sampling experiment is assumed to be done *with replacement* to assure the samples are independent.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, mode, standard deviation, variance, range, a weighted mean, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
  - Is this the same thing as the probability distribution of the population? NO! They may coincidentally be the same though.
  - The re-sampling experiment is assumed to be done *with replacement* to assure the samples are independent.

Sampling distribution

- Imagine taking a sample of size 100 from a population and computing some kind of statistic.
- The statistic you compute can be anything, such as: mean, median, mode, standard deviation, variance, range, a weighted mean, or anything else you might imagine.
- Suppose you repeated this experiment over and over: take a sample of 100 and compute and record the statistic.
- A **sampling distribution** is the probability distribution *of the statistic*
  - Is this the same thing as the probability distribution of the population? NO! They may coincidentally be the same though.
  - The re-sampling experiment is assumed to be done *with replacement* to assure the samples are independent.

Example

- Suppose the following data is an entire *population*:
  5 7 10 13 14 15 19 22 25 29 30 33 34 48 48 49 51 52 55 57 59 60 62 64 66
  The population mean is $\mu = 37.08$. The population standard deviation is $\sigma = 19.484$.
- Let's take 14 random samples of size $n = 5$:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
51	30	64	19	64	10	49	59	62	33	55	52	30	59
29	55	34	52	60	14	59	7	60	55	10	62	14	48
62	52	49	13	62	52	10	49	52	34	48	34	52	22
34	22	59	48	64	5	7	57	48	48	30	66	7	34
33	25	34	29	19	10	25	52	19	22	29	19	14	5

Example

- Suppose the following data is an entire *population*:
  5 7 10 13 14 15 19 22 25 29 30 33 34 48 48 49 51 52 55 57 59 60 62 64 66
  The population mean is $\mu = 37.08$. The population standard deviation is $\sigma = 19.484$.
- Let's take 14 random samples of size $n = 5$:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
51	30	64	19	64	10	49	59	62	33	55	52	30	59
29	55	34	52	60	14	59	7	60	55	10	62	14	48
62	52	49	13	62	52	10	49	52	34	48	34	52	22
34	22	59	48	64	5	7	57	48	48	30	66	7	34
33	25	34	29	19	10	25	52	19	22	29	19	14	5

Example

- Suppose the following data is an entire *population*:
  5 7 10 13 14 15 19 22 25 29 30 33 34 48 48 49 51 52 55 57 59 60 62 64 66
  The population mean is $\mu = 37.08$. The population standard deviation is $\sigma = 19.484$.
- Let's take 14 random samples of size $n = 5$:

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
51	30	64	19	64	10	49	59	62	33	55	52	30	59
29	55	34	52	60	14	59	7	60	55	10	62	14	48
62	52	49	13	62	52	10	49	52	34	48	34	52	22
34	22	59	48	64	5	7	57	48	48	30	66	7	34
33	25	34	29	19	10	25	52	19	22	29	19	14	5

Example (continued)

- Take the mean of every sample.

1	**2**	**3**	**4**	**5**	**6**	**7**
51	30	64	19	64	10	49
29	55	34	52	60	14	59
62	52	49	13	62	52	10
34	22	59	48	64	5	7
33	25	34	29	19	10	25
41.8	36.8	48	32.2	53.8	18.2	30

8	**9**	**10**	**11**	**12**	**13**	**14**
59	62	33	55	52	30	59
7	60	55	10	62	14	48
49	52	34	48	34	52	22
57	48	48	30	66	7	34
52	19	22	29	19	14	5
44.8	48.2	38.4	34.4	46.6	23.4	33.6

Sampling distribution of $\bar{x}$

The mean of the sampling distribution:

$$\mu_{\bar{x}} = \frac{1}{14}\sum_{s=1}^{14} \bar{x}_s = 37.871$$

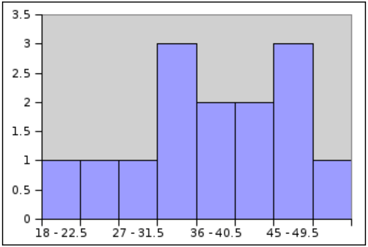
Variance of the sampling distribution:

$$\sigma^2_{\bar{x}} = \frac{1}{14}\sum_{s=1}^{14} (\bar{x}_s - \mu_{\bar{x}})^2 = 94.632$$

Standard deviation of the sampling distribution:

$$\sigma_{\bar{x}} = \sqrt{\sigma^2_{\bar{x}}} = 9.7279$$

Histogram of the sampling distribution:



Desirable qualities

What are some qualities you would like to see in a sampling distribution?

- The average of the sample statistics is equal to the true population parameter.

$$\mu_{\bar{x}} = \mu$$

- Want the variance *of the sampling distribution* to be as small as possible. Why?
- Want the *sampling distribution* to be normal, regardless of the distribution of the population.

Desirable qualities

What are some qualities you would like to see in a sampling distribution?

- The average of the sample statistics is equal to the true population parameter.

$$\mu_{\bar{x}} = \mu$$

- Want the variance *of the sampling distribution* to be as small as possible. Why?
- Want the *sampling distribution* to be normal, regardless of the distribution of the population.

Desirable qualities

What are some qualities you would like to see in a sampling distribution?

- The average of the sample statistics is equal to the true population parameter.

$$\mu_{\bar{x}} = \mu$$

- Want the variance *of the sampling distribution* to be as small as possible. Why?
- Want the *sampling distribution* to be normal, regardless of the distribution of the population.

Desirable qualities

What are some qualities you would like to see in a sampling distribution?

- The average of the sample statistics is equal to the true population parameter.

$$\mu_{\bar{x}} = \mu$$

- Want the variance *of the sampling distribution* to be as small as possible. Why?
- Want the *sampling distribution* to be normal, regardless of the distribution of the population.

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a sample mean is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (consistent):

$$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a sample mean is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (consistent):

$$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a sample mean is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (consistent):

$$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a sample mean is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (consistent):

$$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a sample mean is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (consistent):

$$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a sample mean is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (consistent):

$$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

- Given:
  - Suppose a RV x has a distribution (it need not be normal) with mean μ and standard deviation σ.
  - Suppose a sample mean is computed from a sample of size n.
- Then, if n is sufficiently large, the sampling distribution of $\bar{x}$ will have the following properties:
  - The sampling distribution of $\bar{x}$ will be normal.
  - The mean of the sampling distribution will equal the mean of the population (consistent):

$$\mu_{\bar{x}} = \mu$$

  - The standard deviation of the sampling distribution will decrease with larger sample sizes, and is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem: Small samples

If n is small (rule of thumb: $n < 30$)

- The sample mean is still consistent.
- Sampling distribution will be normal if the distribution of the population is normal.

Central Limit Theorem: Small samples

If n is small (rule of thumb: $n < 30$)

- The sample mean is still consistent.
- Sampling distribution will be normal if the distribution of the population is normal.

Example 1

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a sample of size $n = 30$ will have a mean of $7.3lbs$ or greater?

Example 1

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a sample of size $n = 30$ will have a mean of $7.3lbs$ or greater?

Example 1

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a sample of size $n = 30$ will have a mean of $7.3lbs$ or greater?

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Example 1

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a sample of size $n = 30$ will have a mean of $7.3lbs$ or greater?

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{7.3 - 7}{0.8/\sqrt{30}} = 2.05$$

Example 1

Suppose average birth weight is $\mu = 7 lbs$, and the standard deviation is $\sigma = 0.8 lbs$.

What is the probability that a sample of size $n = 30$ will have a mean of $7.3 lbs$ or greater?

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$z = \frac{7.3 - 7}{0.8/\sqrt{30}} = 2.05$$

The probability the sample mean is greater than $7.3 lbs$ is:

Example 1

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a sample of size $n = 30$ will have a mean of $7.3lbs$ or greater?

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$z = \frac{7.3 - 7}{0.8/\sqrt{30}} = 2.05$$

The probability the sample mean is greater than $7.3lbs$ is:

$$P(\bar{x} > 7.3) = P(z > 2.05) = 0.0202$$

Example 2

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a randomly selected baby will have a weight of 7.3lbs? What do you need to assume to answer this question?

Example 2

Suppose average birth weight is $\mu = 7 lbs$, and the standard deviation is $\sigma = 0.8 lbs$.

What is the probability that a randomly selected baby will have a weight of 7.3lbs? What do you need to assume to answer this question?

Example 2

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a randomly selected baby will have a weight of 7.3lbs? What do you need to assume to answer this question?

Must assume the population is normally distributed. Why?

Example 2

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a randomly selected baby will have a weight of 7.3lbs? What do you need to assume to answer this question?

Must assume the population is normally distributed. Why?

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{7.3 - 7}{0.8} = 0.38$$

Example 2

Suppose average birth weight is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.

What is the probability that a randomly selected baby will have a weight of 7.3lbs? What do you need to assume to answer this question?

Must assume the population is normally distributed. Why?

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{7.3 - 7}{0.8} = 0.38$$

The probability that a baby is greater than 7.3lbs is:

Example 2

Suppose average birth weight is $\mu = 7 lbs$, and the standard deviation is $\sigma = 0.8 lbs$.

What is the probability that a randomly selected baby will have a weight of 7.3lbs? What do you need to assume to answer this question?

Must assume the population is normally distributed. Why?

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{7.3 - 7}{0.8} = 0.38$$

The probability that a baby is greater than 7.3lbs is:

$$P(x > 7.3) = P(z > 0.38) = 0.352$$

Example 3

- Suppose average birth weight of all babies is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.
- Suppose you collect a sample of 30 newborn babies whose mothers used illegal drugs during pregnancy.
- Suppose you obtained a sample mean $\bar{x} = 5.5lbs$. If you assume the mean birth weight of babies whose mothers used illegal drugs has the same sampling distribution as the rest of the population, what is the probability of getting a sample mean this low or lower?

Example 3

- Suppose average birth weight of all babies is $\mu = 7 lbs$, and the standard deviation is $\sigma = 0.8 lbs$.
- Suppose you collect a sample of 30 newborn babies whose mothers used illegal drugs during pregnancy.
- Suppose you obtained a sample mean $\bar{x} = 5.5 lbs$. If you assume the mean birth weight of babies whose mothers used illegal drugs has the same sampling distribution as the rest of the population, what is the probability of getting a sample mean this low or lower?

Example 3

- Suppose average birth weight of all babies is $\mu = 7lbs$, and the standard deviation is $\sigma = 0.8lbs$.
- Suppose you collect a sample of 30 newborn babies whose mothers used illegal drugs during pregnancy.
- Suppose you obtained a sample mean $\bar{x} = 5.5lbs$. If you assume the mean birth weight of babies whose mothers used illegal drugs has the same sampling distribution as the rest of the population, what is the probability of getting a sample mean this low or lower?

Example 3 continued

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

Example 3 continued

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{5.5 - 7}{0.8/\sqrt{30}} = -10.27$$

Example 3 continued

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$z = \frac{5.5 - 7}{0.8/\sqrt{30}} = -10.27$$

The probability the sample mean is less than or equal to $5.5 lbs$ is:

Example 3 continued

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$z = \frac{5.5 - 7}{0.8/\sqrt{30}} = -10.27$$

The probability the sample mean is less than or equal to 5.5lbs is:

$$P(\bar{x} < 7) = P(z < -10.27) = 0.0000$$

Example 3 continued

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z = \frac{5.5 - 7}{0.8/\sqrt{30}} = -10.27$$

The probability the sample mean is less than or equal to 5.5*lbs* is:

$$P(\bar{x} < 7) = P(z < -10.27) = 0.0000$$

This is an extremely unlikely event if the assumption is true. Therefore it is likely the assumption is not true.

Homework

- Section 7.4, page 271-272: 7.17, 7.18, 7.23 - 7.26.