Central Limit Theorem

Math 230: Introductory Statistics

Central Limit Theorem

Math 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.

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 $\mu = 7 lbs$, 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?

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 $\mu = 7 lbs$, 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?

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

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 $\mu = 7lbs$, 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:

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 $\mu = 7lbs$, 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:

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

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 $\mu = 7lbs$, 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 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 $5.5 lbs$ 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 5.5lbs 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 6-5: Problems 5 through 13 odd.
- All answers are in the back.
- The even problems will make for useful practice when studying for a future exam.