7.5 The Central Limit Theorem

Essential Questions:

Why does the central limit theorem make all distributions approximately normal?

Focus Points:

- For a normal distribution, use μ and σ to construct the theoretical sampling distribution for the statistics $\bar{x}$.
- For large samples, use sample estimates to construct a good approximate sampling distribution for the statistic $\bar{x}$.
- Learn the statement and underlying meaning of the central limit theorem well enough to explain it to a friend who is intelligent, but (unfortunately) does not know much about statistics.

Jan 3-11:17 AM

Theorem: For a Normal Probability Distribution

x = random variable for a normal distribution

μ = population mean

σ = standard deviation parameter

$\bar{x}$ = sample mean corresponding to random samples of size n from the x distribution.

The following MUST be true:

1) The $\bar{x}$ distribution is a *normal distribution*

2) The mean of the $\bar{x}$ distribution is μ.

3) The standard deviation of the $\bar{x}$ distribution is $\boxed{\frac{\sigma}{\sqrt{n}}}$

Jan 3-11:22 AM

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

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

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

where n is the sample size,

μ is the mean of the x distribution, and

σ is the standard deviation of the x distribution.

Jan 3-11:27 AM

Example 1: Fishing Pond in Pinedale, Wisconsin

Suppose a team of biologists has been studying the Pinedale children's fishing pond. Let x represent the length of a single trout taken at random from the pond. This group of biologists has determined that x has a normal distribution with mean $\mu = 10.2$ inches and standard deviation $\sigma = 1.4$ inches.

a) What is the probability that a *single trout* taken at random from the pond is between 8 and 12 inches long?

(Hint: Use P(8 < x < 12) to find z scores to find the probability.)

$z = \frac{x-\mu}{\sigma}$

$x=8 \quad x=12$

$P(8<x<12) = \frac{8-10.2}{1.4} < z < \frac{12-10.2}{1.4}$

$P(1.57<z<1.29) = 0.9015 - 0.0582 = \boxed{0.8433}$

Jan 3-11:31 AM

b) What is the probability that the mean length $\bar{x}$ of five trout taken at random is between 8 and 12 inches?

$\mu = 10.2 \quad \sigma = 1.4 \quad n = 5$

(Hint: Use $P(8 < \bar{x} < 12)$ to find the z scores to find the probability using the new formula.)

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

$$P(8 < \bar{x} < 12) = \frac{8 - 10.2}{\frac{1.4}{\sqrt{5}}} < z < \frac{12 - 10.2}{\frac{1.4}{\sqrt{5}}}$$

```
(8-10.2)/(1.4/√5
     -3.513821107
(12-10.2)/(1.4/
      2.874944542
```

$P(-3.51 < z < 2.87)$

0.9979
-0.0001

$\boxed{0.9978}$

Jan 3-11:39 AM

c) Why are the probabilities from (a) and (b) so different, a = 0.8433 and b = 0.9977?

Think about the formulas....

$a = \sigma \qquad b = \frac{\sigma}{\sqrt{n}}$

Jan 3-11:42 AM

The *standard error* is the standard deviation of a sampling distribution. For the x sampling distribution,

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

Jan 3-11:44 AM

Theorem: The Central Limit Theorem for ANY Probability Distribution

x = random variable for **ANY** distribution

μ = population mean

σ = standard deviation parameter

$\bar{x}$ = sample mean corresponding to random samples of size n from the x distribution as n increases without limit.

So... x can have any distribution whatsoever, but that as the sample size gets larger and larger the distribution of $\bar{x}$ will approach a normal distribution. Back to the Law of Large Numbers.

$\boxed{n \geq 30}$

Jan 3-12:33 PM

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

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

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

Using the Central Limit Theorem to convert the $\bar{x}$ distribution to the standard normal distribution.

where n is the sample size ($n \geq 30$),

μ is the mean of the x distribution,

σ is the standard deviation of the x distribution.

Jan 3-11:27 AM

Example 2: Bacteria in Raw Milk

A certain strain of bacteria occurs in all raw milk. Let x be the bacteria count per milliliter of milk. The health department has found that if the milk is not contaminated, then x has a distribution that is more or less mound-shaped and symmetrical. The mean of the x distribution is μ = 2500, and the standard deviation is σ = 300. In a large commercial dairy, the health inspector takes 42 random samples of the milk produced each day. At the end of the day, the bacteria count in each of the 42 samples is averaged to obtain the sample mean bacteria count $\bar{x}$.

SRS
42
$n \geq 30$

a) Assuming the milk is not contaminated, what is the distribution of $\bar{x}$?

n = 42

$\mu_{\bar{x}} = 2500$

$\sigma_{\bar{x}} = \frac{300}{\sqrt{42}}$

Is that bigger than 30? Does the central limit theorem apply here?

yes yes

Jan 3-12:39 PM

b) Assuming the milk is not contaminated, what is the probability that the average bacteria count $\bar{x}$ for one day is between 2350 and 2650 bacteria per milliliter?

(HINT: Find the z score using $P(2350 \le \bar{x} \le 2650)$)

$$P(2350 < \bar{x} < 2650) = \frac{2350-2500}{\frac{300}{\sqrt{42}}} < z < \frac{2650-2500}{\frac{300}{\sqrt{42}}}$$

```
2350-2500
               -150
300/√42
        46.29100499
-150/46.29
         -3.2404407
2650-2500
               150
300/√42
        46.29100499
150/46.29
          3.2404407
```

$$= \frac{-150}{46.29} < z < \frac{150}{46.29}$$

$$P(-3.24 < z < 3.24) = 0.9994 - 0.0006 = \boxed{0.9988}$$

Jan 3-12:48 PM

HOW TO FIND PROBABILITIES REGARDING $\bar{x}$

n = sample size

μ = mean of x distribution

σ = standard deviation of x distribution

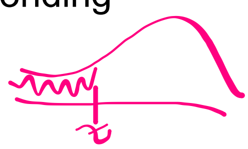
---

1. If the x distribution is normal, then the $\bar{x}$ distribution is normal.

2. Even if the x distribution is not normal, if the sample size $\boxed{n \ge 30}$, then, by the central limit theorem, the $\bar{x}$ distribution is approximately normal.

3. Use the z score formula with sigma dividied by root n to convert x-bar to z score. $\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

4. Use the standard normal distribution to find the corresponding probabilities of events regarding $\bar{x}$.



Jan 3-12:50 PM

Bias and Variability

A sample statistic is unbiased if the mean of its sampling distribution equals the value of the parameter being estimated.

sample $\bar{x}$

The spread of the sampling distribution indicates the variability of the statistic. The spread is affected by the sampling method and the sample size. Statistics from larger random samples have spreads that are smaller!!!

pop.

Random

$n \geq 30$

Jan 3-12:54 PM

HW: pg. 339: 1, 5, 7, 9, 11, 13, 17

1. standard deviation

5. a) normal, $\mu_{\bar{x}} = 8$; $\sigma_{\bar{x}} = 2$. b) 0.50 c) 0.3085 d) No, about 30% of all such samples have means exceeding 9.

7. a) 30 or more. b) no

9. The second. The standard error of the first is σ/10, while that of the second is σ/15, where σ is the standard deviation of the original x distribution.

11. a) $\mu_{\bar{x}} = 15$; $\sigma_{\bar{x}} = 2.0$; 0.3413 b) $\mu_{\bar{x}} = 15$; $\sigma_{\bar{x}} = 1.75$; 0.3729

    c) The standard deviation of part (b) is smaller, resulting in a narrower distribution.

13. a) 0.2643 b) 0.0026 c) No; yes

17. a) 0.1020 b) 224 c) 0.0014 d) 0.8849. Unlikely.

Jan 3-12:57 PM