6.5

Distribution of Sample Means

Central Limit Theorem

6.5

Distribution of Sample Means

A sampling distribution of sample means is a distribution obtained by using the means computed from random samples of a specific size taken from a population.

ex I select 50 samples of a certain size from a population. I compute the means of each sample, $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_{50}$. These means are a sampling distribution of sample means.

Sampling error - the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

Properties of the Distribution of Sample Means

1. The mean of the sample means is the same as the population mean; $\mu_{\bar{X}} = \mu$ (must use all possible sam
2. The standard deviation of the sample means $\sigma_{\bar{X}}$
   $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, note $\sigma_{\bar{X}} < \sigma$.

ex: data 2, 3, 4, 6, 7 ← quiz score, age, # shoes ← entire population

$\mu = 4.4$ $\quad\sigma = 1.8547$

If all samples of size 2 are taken w/ replaceme

Sample	mean	Sample	mean	S	mean	S	m	S	m
2,2	2 ✓	3,2	2.5 ✓	4,2	3 ✓	6,2	4 ✓	7,2	4.5
2,3	2.5 ✓	3,3	3 ✓	4,3	3.5 ✓	6,3	4.5 ✓	7,3	5 ✓
2,4	3 ✓	3,4	3.5 ✓	4,4	4 ✓	6,4	5 ✓	7,4	5.5
2,6	4 ✓	3,6	4.5 ✓	4,6	5 ✓	6,6	6 ✓	7,6	6.5
2,7	4.5 ✓	3,7	5 ✓	4,7	5.5 ✓	6,7	6.5 ✓	7,7	7

frequency distribution

$\overline{X}$	freq.
2	1
2.5	2
3	3
3.5	2
4	3
4.5	4
5	4
5.5	2
6	1
6.5	2
7	1

$$\mu_{\overline{x}} = \frac{2(1) + 2.5(2) + 3(3) + \cdots + 7(1)}{25} = 4.4$$

$$\sigma_{\overline{x}} = \sqrt{\frac{(2-4.4)^2 + (2.5-4.4)^2 + \cdots + (7-4.4)^2}{25}}$$

$$\cong 1.3115$$

$$\sigma / \sqrt{2} = 1.8547 / \sqrt{2} \cong 1.3115$$

We call $\sigma_{\overline{x}}$ the <u>standard error of the mean</u>

The Central Limit Theorem

As the sample size n increases the shape of the distribution of the sample means taken with replacement from a population with mean μ and st. dev. σ will approach a normal distribution.

By the Central Limit Theorem we can find info about sample means like we used the normal distribution.

One difference:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

How big does n need to be? "As the sample size n increas

Two rules:

1. If the original r.v. is normally distributed any size, n, works.
2. If the original r.v. might not be normal $n \geq 30$ (larger n better estimate)

ex) I give an exam, the scores are normally distributed with $\mu = 82$, $\sigma = 9$. I randomly select 10 scores, find the probability that the mean of these 10 scores is above 90.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{90 - 82}{9/\sqrt{10}} \approx 2.81 \text{ from table}$$

$.5 - .4975 = .0025$

$P(\bar{X} > 90) = .0025$

$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ or $z = \frac{X - \mu}{\sigma}$ when to use what?

$\uparrow$ used with sample mean

$\uparrow$ used with individual value

ex Same exam as before, $\mu = 82$, $\sigma = 9$.

a. I randomley select a student. find $P(X > 90)$.

$z = \frac{X - \mu}{\sigma} = \frac{90 - 82}{9}$

b. I randomley select 5 students. find $P(\bar{X} > 90)$.

$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{90 - 82}{9/\sqrt{5}}$