MATH 3307
Lesson 16

Sampling Distributions

Let's look at the "sampling distribution of the means" for a set of data – this is the same as the "distribution of the sample means".

Consider the population consisting of the values 3,5,9,11 and 14.

μ = 8.4　　　　σ = 4.4497

Consider the population consisting of the values 3,5,9,11 and 14.

Let's take samples of size 2 from this population.

The set above is the **sampling distribution of size 2** for this population.

List all the possible pairs from 3,5,9,11 and 14 and find their means.

pairs	$\bar{x}$
3,5	4
3,9	6
3,11	7
3,14	8.5
5,9	7
5,11	8
5,14	9.5
9,11	10
9,14	11.5
11,14	12.5

Mean of the Sample Means:

$$\mu_{\bar{x}} = \underline{8.4},$$

$$\sigma_{\bar{x}} = \underline{2.569}$$

Standard Deviation of the Sample Means

Consider the population consisting of the values 3,5,9,11 and 14.

What about for the sampling distribution of size 3?

sets	$\bar{x}$
3,5,9	5.667
3,5,11	6.333
3,5,14	7.333
3,9,11	7.667
3,9,14	8.667
5,9,11	8.333
5,9,14	9.333
9,11,14	11.333
3,11,14	9.333
5,11,14	10

$\mu_{\bar{x}} = \underline{8.4}$,

$\sigma_{\bar{x}} = \underline{1.713}$

Comparison of Mean and Standard Deviation.

Compare $\mu_{\bar{x}}$(the mean of the sample means) to μ .

What do you notice about $\sigma_{\bar{x}}$?

The mean of the sample means is equal to the mean of the population.

The standard deviation of the sample means varies with sample size.

Suppose that $\bar{x}$ is the mean of a simple random sample of size *n* drawn from a large population. If the population mean is μ and the population standard deviation is σ , then the mean of the sampling distribution of $\bar{x}$ is $\mu_{\bar{x}}$ = μ and the standard deviation of the sampling distribution is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

If our original population has a normal distribution, the sample mean's distribution is $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Calculating $\mathrm{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ will give the probability that a randomly selected sample (with given criteria) would have the described mean.

An **unbiased statistic** is a statistic used to estimate a parameter in such a way that mean of its sampling distribution is equal to the true value of the parameter being estimated.

We consider the above values to be unbiased estimates of our distribution.

The Central Limit Theorem

The **Central Limit Theorem** states that if we draw a simple random sample of size *n* from any population with mean μ and standard deviation σ , when *n* is large the sampling distribution of the sample mean *x* is close to the normal distribution $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Mean: μ
stnd dev: $\sigma / \sqrt{n}$

In simple terms, we can use a normal distribution probability with above criteria, when the sample size (n) is sufficiently large.

Large Numbers

Determining whether *n* is large enough for the central limit theorem to apply depends on the original population distribution. The more the population distribution's shape is from being normal, the larger the needed sample size will be.

A rule of thumb is that $n > 30$ will be large enough.

Examples: $\mu = 490 \quad \sigma = 80$

The mean TOEFL score of international students at a certain university is normally distributed with a mean of 490 and a standard deviation of 80. Suppose groups of 30 students are studied. Find the mean and the standard deviation for the distribution of sample means.

Mean: $\mu_{\bar{x}} = \mu = 490$

Stnd dev: $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 80/\sqrt{30} = 14.605$

Determine the probability that a randomly selected sample, described above, would have an average score of less than 500.

$P(X < 500)$

$x, \mu, \sigma/\sqrt{n}$

```
pnorm(500,490,80/sqrt(30))
1] 0.7532186
```

Examples: Popper 11

$\mu = 20$ $\sigma = 4$

A waiter estimates that his average tip per table is $20 with a standard deviation of $4. If we take samples of 9 tables at a time, calculate the following probabilities when the tip per table is normally distributed.

1. What is the mean of the sample means? $\mu_{\bar{x}} = \mu = 20$

   a. 10 b. 20 c. 24 d. 29

2. What is the standard deviation of the sample means? $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{9}}$

   a. 1.333 b. 0.444 c. 4 d. 16

3. What is the probability that the average tip for one sample is less than $21?

   a. 0.5987 b. 0.2262 c. 0.7734 d. 0.4013

4. What is the probability that the average tip for one sample is more than $21?

   a. 0.5987 b. 0.2266 c. 0.7734 d. 0.4013

5. What is the probability that the average tip for one sample is between $19 and $21?

   a. 0.5468 b. 0.0000 c. 0.1974 d. 0.3721

#3
```
> pnorm(21,20,4/sqrt(9))
[1] 0.7733726
> 1-pnorm(21,20,4/sqrt(9))
[1] 0.2266274
> pnorm(21,20,4/sqrt(9))-pnorm(19,20,4/sqrt(9))
[1] 0.5467453
```
#4
#5

Sampling Proportions

$\hat{p}$: sample proportion p: population proportion

When *X* is a binomial random variable (with parameters *n* and *p*) the statistic $\hat{p}$, the sample proportion, is equal to $\frac{X}{n}$.

$\hat{p} = \frac{X}{n} = \frac{\text{success}}{\text{total}}$

The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$, and the standard deviation is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. *

Mean of the sample proportions.
Standard deviation of the sample proportions.

Calculating $\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ will give the probability that a randomly selected sample, as described above, will have the described probability of success (or X successes out of size n).*

* See conditions on the next slide

Conditions for Sampling Proportions

If our population size is at least 10 times the sample size, the standard deviation of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. *(Note: If the population size is not given, we can assume this condition has been met.)*

We can use the normal approximation for the sampling distribution of $\hat{p}$ when $np \geq 10$ and $n(1-p) \geq 10$.

Example:

A laboratory has an experimental drug which has been found to be effective in 93% of cases. The lab decides to test the drug of a sampling distribution of 150 mice. Determine the mean and standard deviation for this sample distribution.

mean: $\mu_{\hat{p}} = p = .93$

stnd dev: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} =$

```
> sqrt(.93*(1-.93)/150)
[1] 0.02083267
```

Determine the probability that, in a randomly selected sample, the drug will be effective on at least 135 of the mice.

$x = 135$

$\hat{p} = \frac{x}{n} = \frac{135}{150}$

$\rightarrow$ 135 or more $\rightarrow$ use 1 −

$\hat{p} = \frac{x}{n}$ $\quad p \quad \sqrt{\frac{p(1-p)}{n}}$

```
1-pnorm(135/150,0.93,sqrt(.93*(1-.93)/150))
1] 0.9250728
```

Example: Popper 11

A large high school has approximately 1,200 seniors. The administration of the school claims that 82% of its graduates are accepted into colleges. If a simple random sample of 100 seniors is taken, what is the mean and the standard deviation of the sampling distribution?

6. Mean of the sample distribution: $\mu_{\hat{p}} = p = 0.82$

a. 0.0683 b. 82 c. 0.0833 d. 0.8200

7. Standard Deviation of the sample distribution: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

a. 0.0384 b. 0.0082 c. 8.2000 d. 0.0111

```
sqrt(.82*(1-.82)/100)
1] 0.03841875
```

Example, continued: Popper 11

```
> 100*.82
[1] 82
> 100*(1-.82)
[1] 18
```

8. Based on the population and sample size, can probabilities be calculated using a normal distribution?

a. Yes   b. No, np < 10   c. No, n(1-p) < 10   d. No, np < 10 & n(1-p) < 10

9. What is the probability that at most 64 of the sample will be accepted into college?

```
pnorm(64/100,.82,sqrt(.82*(1-.82)/100))
1] 1.398346e-06
```

a. ≈ 0   b. ≈ 1   c. 0.4286   d. 0.5714

10. What is the probability that between 75 and 85 will be accepted into college?

a. 0   b. 0.0440   c. 0.7485   d. 0.8433

```
pnorm(85/100,.82,sqrt(.82*(1-.82)/100))-pnorm(75/100,.82,sqrt(.82*(1-.82)/100))
1] 0.7483347
```

Comparison of Sampling Means and Sampling Proportions:

Sampling Means:

You will be given the following information: sample size (n), population mean (μ), and a population standard deviation (σ).

You can apply $N\left(\mu, {}^{\sigma}/_{\sqrt{n}}\right)$ if n > 30, or if you a normal distribution is specified.

Sampling Proportions:

You will be given the following information: sample size (n), population proportion (p).

You can apply $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ if (1) population > 10n, (2) np ≥ 10, and (3) n(1-p) ≥ 10

Example:

$\mu = 7 \quad \sigma = 1.5$

It is estimated that mean sleep per night (in hours) is 7 hours with a standard deviation of 90 minutes, and is normally distributed.

Doctors believe that the ideal amount of sleep a person should have is in between 8 and 10 hours per night. You select a sample of 40 people. What is the probability that this sample has a mean sleep time within the ideal limits?

$n = 40$

$P(8 \le x \le 10)$

```
pnorm(10,7,1.5/sqrt(40))-pnorm(8,7,1.5/sqrt(40))
[1] 1.24133e-05
```

Second Example:

$p = .90$

In a population of 10,000 people, it is determined that 90% of people do not get enough sleep each night. You select a sample of 100 people to test this statistic. What is the probability that, within the sample you selected, between 85 and 95 people do not get enough sleep?

$10 \cdot n = 10 \cdot 100 = 1000 <$ population size.

$n \cdot p \geq 10$

$n \cdot (1-p) \geq 10$

```
> 100*.90
[1] 90
> 100*(1-.90)
[1] 10
> pnorm(95/100,.90,sqrt(.90*(1-.90)/100))-pnorm(85/100,.90,sqrt(.90*(1-.90)/100))
[1] 0.9044193
>
```