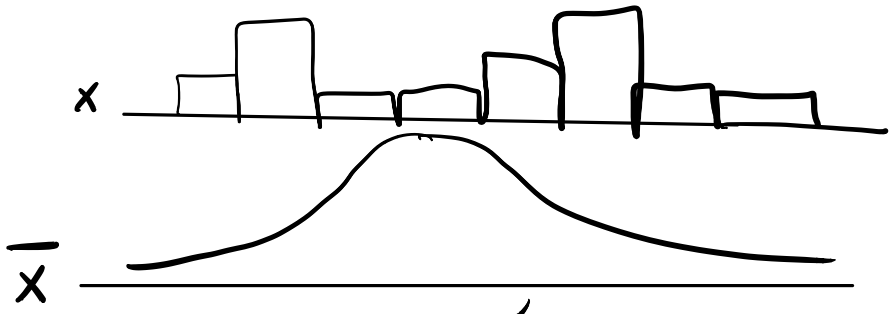
FINAL EXAM: MONDAY 12/14 3:30-5:45 PM

(BLACKBOARD EXAM, LIKE EXAM 1 & 2)

PROJECT (POSTED)

6

7a



CENTRAL LIMIT THM.

SUMMARY OF CONFIDENCE INTERVALS:

INTERVAL IN WHICH THE POPULATION PARAMETER $(\mu, p, \mu_1 - \mu_2, p_1 - p_2)$ IS LIKELY TO LIE.

$$\left[\text{SAMPLE STATISTIC} - z_{\alpha/2}\, \text{S.E},\ \text{SAMPLE STATISTIC} + z_{\alpha/2}\, \text{S.E}\right]$$

LOWER BOUND FOR CONF. INT.

UPPER BOUND FOR CONF. INT.

POPULATION PARAMETER	SAMPLE STATISTIC	STANDARD ERROR FOR SAMPLE STATISTIC
MEAN μ	$\bar{x}$	$\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$
PROPORTION p	$\hat{p} = \frac{x}{n}$	$\sqrt{\frac{pq}{n}} \approx \sqrt{\frac{\hat{p}\hat{q}}{n}}$
DIFF. BETWEEN 2 POP. MEANS $\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
DIFF. BETWEEN 2 POP. PROP. $p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$

PROPERTIES OF THE SAMPLING DISTRIBUTION OF $(\bar{x}_1 - \bar{x}_2)$, THE DIFFERENCE BETWEEN TWO SAMPLE MEANS

When independent random samples of n_1 and n_2 observations have been selected from populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, the sampling distribution of the difference $(\bar{x}_1 - \bar{x}_2)$ has the following properties:

1. The mean of $(\bar{x}_1 - \bar{x}_2)$ is

   $$\mu_1 - \mu_2$$

   and the standard error is

   $$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

   which can be estimated as

   $$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$ when the sample sizes are large.

2. **If the sampled populations are normally distributed,** then the sampling distribution of $(\bar{x}_1 - \bar{x}_2)$ is **exactly** normally distributed, regardless of the sample size.
3. **If the sampled populations are not normally distributed,** then the sampling distribution of $(\bar{x}_1 - \bar{x}_2)$ is **approximately** normally distributed when n_1 and n_2 are both 30 or more, due to the Central Limit Theorem.

§8.6 ESTIMATING THE DIFFERENCE BETWEEN 2 POPULATION MEANS

8.41 Independent random samples were selected from populations 1 and 2. The sample sizes, means, and variances are as follows:

	Population 1	Population 2
Sample Size	35	49
Sample Mean	12.7	7.4
Sample Variance	1.38 (s_1^2)	4.14 (s_2^2)

a. Find a 95% confidence interval for estimating the difference in the population means $(\mu_1 - \mu_2)$.

b. Based on the confidence interval in part a, can you conclude that there is a difference in the means for the two populations? Explain.

$$\left[\text{SAMPLE STATISTIC} - z_{\alpha/2}\, S.E,\ \text{SAMPLE STATISTIC} + z_{\alpha/2}\, S.E\right]$$

$\mu_1 - \mu_2$?

SAMPLE STATISTIC

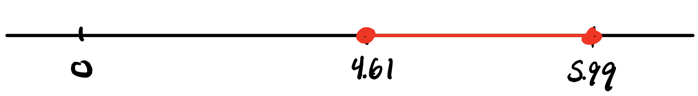
$$(\bar{x}_1 - \bar{x}_2) = 12.7 - 7.4 = 5.3$$

95% CONF INT $\Rightarrow z_{\alpha/2} = 1.96$

$$S.E. = \sqrt{\frac{1.38}{35} + \frac{4.14}{49}}$$

$$\left[5.3 - 1.96\sqrt{\frac{1.38}{35} + \frac{4.14}{49}},\ 5.3 + 1.96\sqrt{\frac{1.38}{35} + \frac{4.14}{49}}\right]$$

$$= \underbrace{[\,4.6100,\ 5.9900\,]}_{(a)}$$



(b) YES, THE CONFIDENCE INTERVAL LIES ENTIRELY ABOVE 0.

THUS, WE ARE 95% CONFIDENT THAT $\mu_1 - \mu_2 > 0$

$(\mu_1 > \mu_2)$

i.e. NOT THE SAME

THERE <u>IS</u> A DIFFERENCE.

8.46 9-1-1 A study was conducted to compare the mean numbers of police emergency calls per 8-hour shift in two districts of a large city. Samples of 100 8-hour shifts were randomly selected from the police records for each of the two regions, and the number of emergency calls was recorded for each shift. The sample statistics are listed here:

	Region 1	Region 2
Sample Size	100	100
Sample Mean	2.4	3.1
Sample Variance	1.44	2.64

Find a 90% confidence interval for the difference in the mean numbers of police emergency calls per shift between the two districts of the city. Interpret the interval.

$$S.E. \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

CONFIDENCE LEVEL (PROB THAT μ IS IN THE INTERVAL ESTIMATE)	α = PROB THAT μ IS NOT IN THE INTERVAL ESTIMATE	$\alpha/2$	$z_{\alpha/2}$
.95	.05	.025	1.96
.90	.10	.05	1.645
.98	.02	.01	2.33
.99	.01	.005	2.58

CONFIDENCE INTERVAL FOR $\mu_1 - \mu_2$: $\left[(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\, S.E.\ ,\ (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\, S.E. \right]$

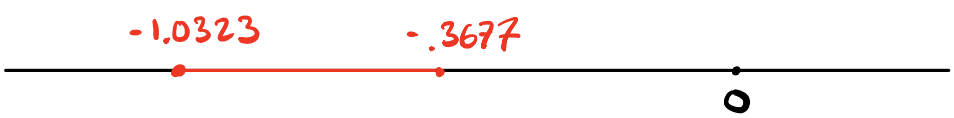
$2.4 - 3.1 = -.7$

$\sqrt{\frac{1.44}{100} + \frac{2.64}{100}}$

1.645

$$\left[-.7 - 1.645\sqrt{\frac{1.44}{100} + \frac{2.64}{100}}\ ,\ -.7 + 1.645\sqrt{\frac{1.44}{100} + \frac{2.64}{100}} \right]$$

$$\left[-1.0323\ ,\ -0.3677 \right]$$



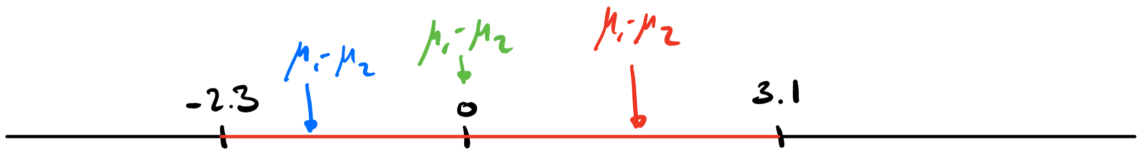
ENTIRE INTERVAL IS NEG

$\Rightarrow \mu_1 - \mu_2 < 0$ (95% CONFIDENCE)

$\mu_1 < \mu_2$ THERE IS A DIFFERENCE.

WHAT IF CONF. INTERVAL FOR $\mu_1 - \mu_2$ CONTAINS 0?

ex. $[-2.3, 3.1]$



THEN IT MAY BE THAT $\mu_1 - \mu_2 < 0$

$\mu_1 - \mu_2 > 0$

$\mu_1 - \mu_2 = 0$

NO DIFFERENCE: $\mu_1 = \mu_2$

§8.7 ESTIMATING THE DIFFERENCE BETWEEN BINOMIAL PROPORTIONS

EXAMPLES:

- The proportion of defective items manufactured in two production lines
- The proportion of male and female voters who favor an equal rights amendment
- The germination rates of untreated seeds and seeds treated with a fungicide

PROPERTIES OF THE SAMPLING DISTRIBUTION OF THE DIFFERENCE $(\hat{p}_1 - \hat{p}_2)$ BETWEEN TWO SAMPLE PROPORTIONS

Assume that independent random samples of n_1 and n_2 observations have been selected from binomial populations with parameters p_1 and p_2, respectively. The sampling distribution of the difference between sample proportions

$$(\hat{p}_1 - \hat{p}_2) = \left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right)$$

has these properties:

1. The mean of $(\hat{p}_1 - \hat{p}_2)$ is

   $$p_1 - p_2$$

   and the standard error is

   $$SE = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

   which is estimated as

   $$SE = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

2. The sampling distribution of $(\hat{p}_1 - \hat{p}_2)$ can be approximated by a normal distribution when n_1 and n_2 are large, due to the Central Limit Theorem.

S.E. FOR $\hat{p}_1$ IS $\sqrt{\frac{p_1q_1}{n_1}} \approx \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}}$

S.E. FOR $\hat{p}_2$ IS $\sqrt{\frac{p_2q_2}{n_2}} \approx \sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}}$

$= \sqrt{(\text{S.E. FOR } \hat{p}_1)^2 + (\text{S.E. FOR } \hat{p}_2)^2}$

Assumption: n_1 and n_2 must be sufficiently large so that the sampling distribution of $(\hat{p}_1 - \hat{p}_2)$ can be approximated by a normal distribution—namely, if $n_1\hat{p}_1$, $n_1\hat{q}_1$, $n_2\hat{p}_2$, and $n_2\hat{q}_2$ are all greater than 5.

$$\left[\text{SAMPLE STATISTIC} - z_{\alpha/2}\,\text{S.E}\,,\ \text{SAMPLE STATISTIC} + z_{\alpha/2}\,\text{S.E}\right]$$

CONF. INT. FOR DIFF. BETWEEN 2 POPULATION PROPORTIONS:

$$\left[(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\ ,\ (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right]$$

8.59 Baseball Fans The first day of baseball comes in late March, ending in October with the World Series. Does fan support grow as the season goes on? Two CNN/*USA Today*/Gallup polls, one conducted in March and one in November, both involved random samples of 1001 adults aged 18 and older. In the March sample, 45% of the adults claimed to be fans of professional baseball, while 51% of the adults in the November sample claimed to be fans.[14]

a. Construct a 99% confidence interval for the difference in the proportion of adults who claim to be fans in March versus November.

b. Does the data indicate that the proportion of adults who claim to be fans increases in November, around the time of the World Series? Explain.

MARCH Pop. 1	NOV. Pop. 2
$\hat{p}_1 = .45$	$\hat{p}_2 = .51$
$\hat{q}_1 = .55$	$\hat{q}_2 = .49$
$n_1 = 1001$	$n_2 = 1001$

CONF. INT. FOR DIFF. BETWEEN 2 POPULATION PROPORTIONS:

$$\left[(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \,,\; (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\underbrace{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} \right]$$

$.45 - .51 = -.06$

$z_{\alpha/2} = 2.58$

$$\sqrt{\frac{(.45)(.55)}{1001} + \frac{(.51)(.49)}{1001}} \approx .0223$$

$$\left[-.06 - (2.58)(.0223) \,,\; -.06 + (2.58)(.0223) \right]$$

$$= \left[-.1175 \,,\; -.0025 \right]$$

ENTIRE INTERVAL LIES BELOW 0

$\therefore$ WE ARE CONFIDENT

$p_1 - p_2 < 0$

$+p_2 \qquad +p_2$

$p_1 < p_2$

$p_2 > p_1$

$$-.1175 \le p_1 - p_2 \le -.0025$$

$$p_2 - .1175 \le p_1 \le p_2 - .0025$$

$$p_1 + .0025 \le p_2 \le p_1 + .1175$$

§8.9 Choosing the Sample Size

Suppose you want to estimate the average lifetime of lightbulbs produced by a particular company. To do this you are going to collect a random sample of lightbulbs produced by this company and record how long each lightbulb lasts to compute the sample mean. You report your findings by stating that the population mean for the lifetime of ALL lightbulbs is approximately the sample mean +/- an "error term".

"Error term" = $z_{\alpha/2} \times$ S.E. for sample statistic.

"Margin of error" = $1.96 \times$ S.E. (default)

Conf. Int. $[\text{Sample statistic} - z_{\alpha/2}\,\text{S.E},\ \text{Sample statistic} + z_{\alpha/2}\,\text{S.E}]$

Error term

$z_{\alpha/2}$ S.E

$z_{\alpha/2}$ S.E

Two ways to make error term smaller:

(1) Lower confidence level

(2) Increase sample size

Sample statistic

$$\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

$$\sqrt{\frac{pq}{n}} \approx \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

S.E. gets smaller as the sample size(s) n (n_1 & n_2) gets larger.

CHOOSING THE SAMPLE SIZE BASED ON THE MAXIMUM DESIRED ERROR TERM B FOR A CONFIDENCE INTERVAL WITH A GIVEN CONFIDENCE LEVEL.

$$\text{ERROR} = z_{\alpha/2} \times S.E. < B \quad (\text{BOUND})$$

EXAMPLE 8.13 Producers of polyvinyl plastic pipe want to have a supply of pipes sufficient to meet marketing needs. They wish to survey wholesalers who buy polyvinyl pipe in order to estimate the proportion who plan to increase their purchases next year. What sample size is required if they want their estimate to be within .04 of the actual proportion with probability equal to .90?

$$z_{\alpha/2} \times S.E. < B$$

$$1.645\sqrt{\frac{pq}{n}} < .04$$

NOTE: IF p & q ARE UNKNOWN WE SET $p = q = .5$ (WORST CASE SCENARIO)

SOLVE FOR n: $$\frac{\cancel{1.645}\sqrt{\frac{(.5)(.5)}{n}}}{\cancel{1.645}} < \frac{.04}{1.645}$$

$$\left(\sqrt{\frac{.25}{n}}\right)^2 < \left(\frac{.04}{1.645}\right)^2$$

$$\cancel{n}\left(\frac{.25}{\cancel{n}}\right) < \left(\frac{.04}{1.645}\right)^2 n$$

$$\frac{.25}{\left(\frac{.04}{1.645}\right)^2} < \frac{\cancel{\left(\frac{.04}{1.645}\right)^2} n}{\cancel{\left(\frac{.04}{1.645}\right)^2}}$$

$$\frac{.25}{\left(\frac{.04}{1.645}\right)^2} < n \quad \Rightarrow \quad 422.8 < n$$

Round up (always!)

$$n \geq 423$$