UP THRU §8.6

EXAM 2 TOMORROW 5.4 HYPERGEOMETRIC DISTRIBUTION (COUNTING)

6.1-6.3. NORMAL PROB. DISTR.

STAND. NORM. PROB. DISTR. z $\mu = 0, \sigma = 1$

& ALL OTHER NORMAL DISTR. $z = \dfrac{x - \mu}{\sigma}$

TABLE.

6.4 APPROXIMATING BINOMIAL R.V. WITH NORMAL R.V. $\mu = np$, $\sigma = \sqrt{npq}$

$$P(x \le x_0) \approx P(x_{NORM} \le x_0 + .5)$$

INTEGER

$$P(x < x_0) \approx P(x_{NORM} \le x_0 - .5)$$

7.4, 5. CENTRAL LIMIT THEOREM

↳ SAMPLING DISTRIBUTIONS FOR $\bar{x}$ & Σx

SAMPLE MEANS SAMPLE TOTALS

NORMALLY DISTRIBUTED WITH

$\bar{x}$: MEAN μ Σx: MEAN $n\mu$

S.E. $= \dfrac{\sigma}{\sqrt{n}}$ S.E $= \sqrt{n}\,\sigma$

7.6 SAMPLE PROPORTIONS $\hat{p}$ MEAN p

S.E. $\sqrt{\dfrac{pq}{n}}$

§ 8.3-4 POINT ESTIMATION & MARGIN OF ERROR

$z_{\alpha/2}$

$$\mu \approx \bar{X} \pm 1.96 \text{ S.E} = \bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

$$p \approx \hat{p} \pm 1.96 \text{ S.E.} = \bar{x} \pm 1.96 \sqrt{\frac{pq}{n}}$$

§ 8.5 INTERVAL ESTIMATION - CONFIDENCE INTERVALS

95% CONF. INT.

CONF. COEFF.	$z_{\alpha/2}$
99%	2.58
98%	2.33
95%	1.96
90%	1.645

$$\mu \approx \bar{X} \pm z_{\alpha/2} \text{ S.E.} \approx \bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$p \approx \hat{p} \pm z_{\alpha/2} \text{ S.E.} \approx \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

WORST CASE: $\hat{p} = \hat{q} = .5$

§ 8.6 ESTIMATING DIFFERENCE BETWEEN POP. MEANS

$$\mu_1 - \mu_2 \approx \bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2} \text{ S.E.} \quad *$$

* S.E. $= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Z-VALUES: TABLE: 2 DECIMAL VALUES

FORMAT: BLACKBOARD → CONTENT

- EXAM 2 QUESTIONS (PDF)
- EXAM 2 ← INPUT ANSWERS HERE

IGNORE THESE #'s.

1. QUESTION 1a
2. QUESTION 1b
3. QUESTION 2a
4. QUESTION 2b

CONFIDENCE INTERVAL

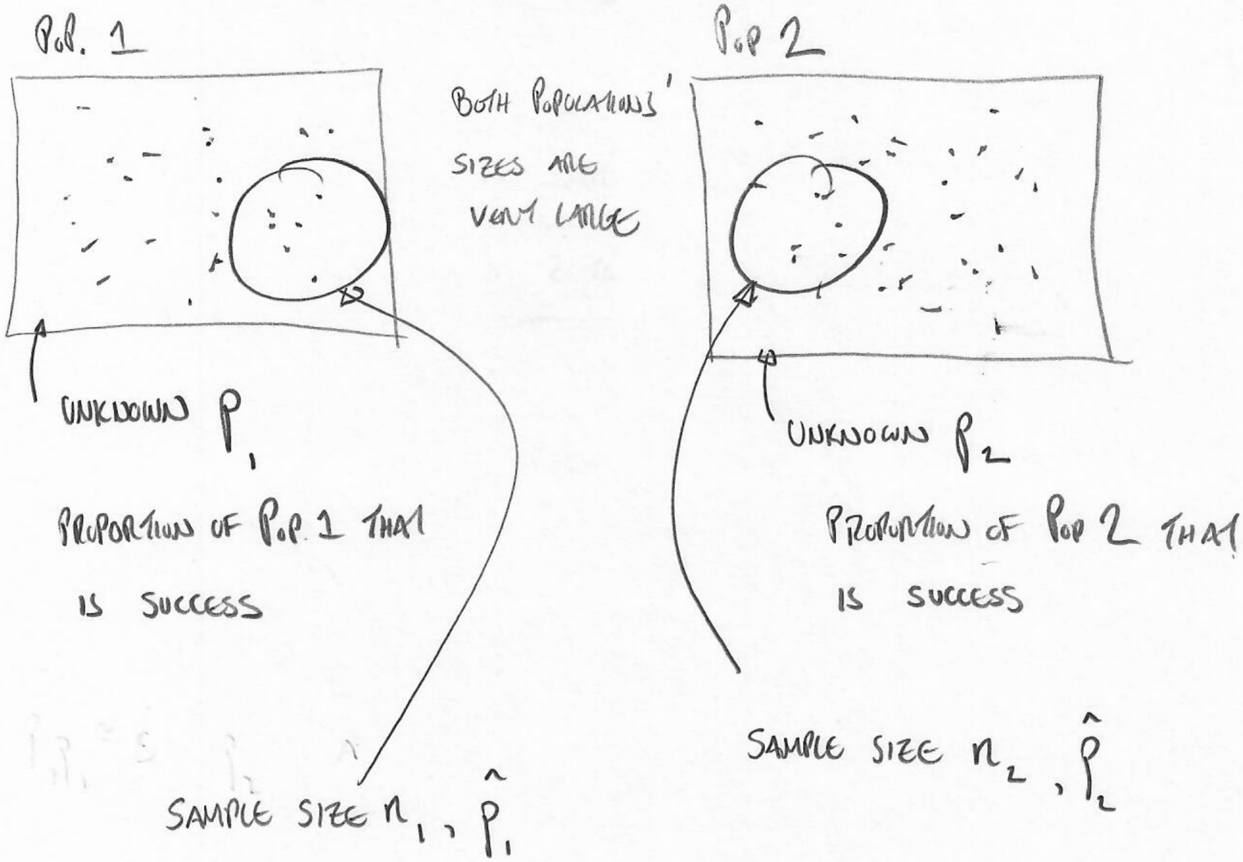
$\bar{x} \pm \#$

SMALLEST VALUE IN INT. $\bar{x} - \#$

LARGEST VALUE IN INT. $\bar{x} + \#$

ANSWERS: DECIMALS, NOT PERCENTS.

§ 8.7 ESTIMATING THE DIFFERENCE BETWEEN POPULATION PROPORTIONS



C.L.T. $\Rightarrow$ $\hat{p}_1$, $\hat{p}_2$ (SAMPLE PROPORTIONS) ARE R.V. VARIABLE

THAT ARE APPROXIMATELY NORMAL IF

$$n_1\hat{p}_1 \geq 5 \qquad n_2\hat{p}_2 \geq 5$$
$$n_1\hat{q}_1 \geq 5 \qquad n_2\hat{q}_2 \geq 5$$

WITH S.E. $\sqrt{\dfrac{p_i q_i}{n_i}} \approx \sqrt{\dfrac{\hat{p}_i \hat{q}_i}{n_i}}$ FOR $i = 1, 2$.

& MEAN p_1 & p_2, RESPECTIVELY.

NOW : $\left(\hat{p}_1 - \hat{p}_2\right)$ IS A RANDOM VARIABLE

(DIFFERENCE OF NORMALLY DISTR. R.V.)

& IT IS NORMALLY DISTRIBUTED WITH

MEAN : $p_1 - p_2$ (UNBIASED)

S.E. : $\sqrt{\dfrac{p_1 q_1}{n} + \dfrac{p_2 q_2}{n_2}}$

$$\approx \sqrt{\dfrac{\hat{p}_1 \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \hat{q}_2}{n_2}}$$

AS LONG AS :

$n_1 \hat{p}_1 \geq 5$ $\quad n_2 \hat{p}_2 \geq 5$

$n_1 \hat{q}_1 \geq 5$ $\quad n_2 \hat{q}_2 \geq 5$.

ex. Suppose 2 samples are taken from 2 different populations.

School District 1

$n_1 = 80$

$\hat{p}_1 = .83$ graduate

School District 2

$n_2 = 70$

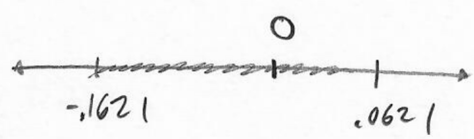
$\hat{p}_2 = .88$ graduate

Estimate $p_1 - p_2$ as a point estimate w/ margin of error.

$$p_1 - p_2 \approx \hat{p}_1 - \hat{p}_2 \pm 1.96\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$\approx .83 - .88 \pm 1.96\sqrt{\frac{(.83)(.17)}{80} + \frac{(.88)(.12)}{70}}$$

$$\approx -.05 \pm .1121$$



Lower bound for interval : $-.1621$

Upper bound for interval : $.0621$

Yes / No : Can you conclude that there is a difference between p_1 & p_2.

No.

8.55

$n_1 = 800$

$\hat{p}_1 = \dfrac{x_1}{n_1} = \dfrac{337}{800}$

$\hat{p}_1 = .42125$

$\hat{q}_1 = 1 - \hat{p}_1 = .57875$

$n_2 = 640$

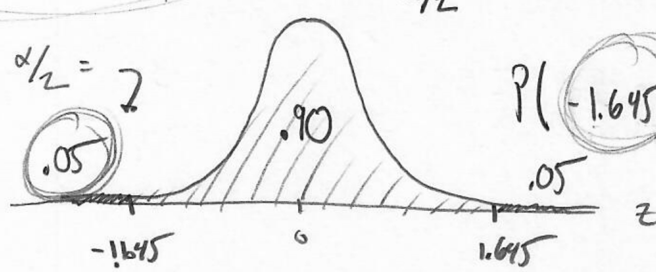
$\hat{p}_2 = \dfrac{x_2}{n_2} = \dfrac{374}{640}$

$\hat{p}_2 = .584375$

$\hat{q}_2 = .415625$

$\alpha = 1 - .9 = .1$

90% CONF. COEFF. $\rightarrow z_{\alpha/2} = 1.645$



$P(-1.645 \le z \le 1.645) = .90$

$$p_1 - p_2 \approx \hat{p}_1 - \hat{p}_2 \pm 1.645 \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n}}$$

$$\approx .42125 - .584375 \pm 1.645 \sqrt{\frac{(.42125)(.57875)}{800} + \frac{(.584375)(.415625)}{640}}$$

$$\approx \boxed{-.163125 \pm .043030 \quad [-.206155, -.120095]}$$

NEG.

$p_1 - p_2 < 0 \implies p_1 < p_2$

ASSUMPTIONS :

$n_1 p_1 \geq 5$

$\downarrow$

$\approx n_1 \hat{p}_1 = (800)(.42125) \geq 5$ ✓

$n_1 q_1 \approx n_1 \hat{q}_1 = (800)(.57875) \geq 5$ ✓

$n_2 p_2 \approx n_2 \hat{p}_2 = (640)(.584375) \geq 5$ ✓

$n_2 q_2 \approx n_2 \hat{q}_2 = (640)(.415625) \geq 5$ ✓

8.65

$n_1 = 200$ $\quad$ $n_2 = 450$

$\hat{p}_1 = .93$ $\quad$ $\hat{p}_2 = .96$

$\hat{q}_1 = .07$ $\quad$ $\hat{q}_2 = .04$

$$p_1 - p_2 \approx \hat{p}_1 - \hat{p}_2 \pm \boxed{z_{\alpha/2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

$$p_1 - p_2 \approx (.93 - .96) \pm \boxed{2.58} \sqrt{\frac{(.93)(.07)}{200} + \frac{(.96)(.04)}{450}}$$

1.645

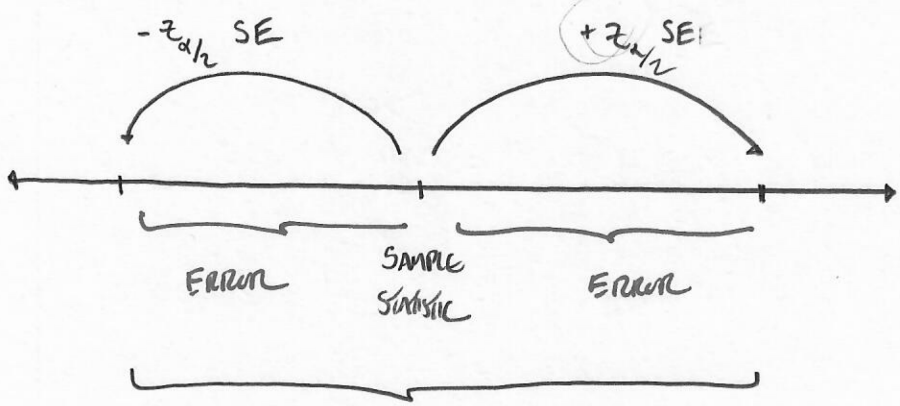
$\approx -.03 \pm .0523$ $\quad$ $\pm .0333$

$$\underbrace{[-.0823 \,,\, .0223]}$$

CONTAINS 0 $\Rightarrow$ CANNOT CONCLUDE THAT THERE IS A DIFFERENCE.

§ 8.9 CHOOSING THE SAMPLE SIZE:

CONFIDENCE INTERVAL:



LENGTH / WIDTH OF CONFIDENCE INTERVAL

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

SHORTER / SMALLER CONF. INTERVAL $\Rightarrow$ MORE ACCURACY

TO GET SMALLER / SHORTER CONF. INTERVAL, MAKE S.E. SMALLER

TO MAKE THE S.E. SMALLER, INCREASE THE SAMPLE SIZE(S).

SUPPOSE WE WANT A BOUND ON OUR ERROR.

i.e. WE WANT $\underbrace{z_{\alpha/2} \text{ S.E.}}_{\text{ERROR}} \leq B$ BOUND

SOLVING THIS INEQUALITY FOR n.

WE NEED TO CHOOSE SAMPLE SIZE(S) LARGE ENOUGH.

ex. SUPPOSE WE WANT TO PERFORM THE EXPERIMENT FROM ex 8.65 AGAIN BUT OBTAIN AN ERROR ($z_{\alpha/2}$ S.E.), WITH CONF. COEFF .99, LESS THAN .01.

BOUND

$$z_{\alpha/2} \; S.E. \leq .01$$

$$2.58 \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \leq .01$$

WE HAVE VALUES FOR $\hat{p}_1$ & $\hat{p}_2$

BECAUSE WE'VE DONE THIS EXPERIMENT BEFORE.

$\hat{p}_1 \approx .93 \qquad \hat{p}_2 \approx .96$

ASSUME $n_1 = n_2$. $(= n)$

SOLVE FOR n: $$2.58 \sqrt{\frac{(.93)(.07)}{n} + \frac{(.96)(.04)}{n}} \leq .01$$

$$\sqrt{\frac{(.93)(.07) + (.96)(.04)}{n}} \leq \frac{.01}{2.58}$$

$$\frac{(.93)(.07) + (.96)(.04)}{n} \leq \left(\frac{.01}{2.58}\right)^2$$

$$(.93)(.07) + (.96)(.04) \leq n \left(\frac{.01}{2.58}\right)^2$$

$$\frac{(.93)(.07) + (.96)(.04)}{\left(\frac{.01}{2.58}\right)^2} \leq n$$

$$6889.374 \leq n$$

SAMPLE SIZES NEED BE AT LEAST

SIZE $\underline{\underline{6,890}}$

ESTIMATE $p_1 - p_2$

TO WITH 1% OF

TRUE VALUE.