9.10. Estimating one proportion.

9.11 2-sample: estimating diff of 2-proportions.

population $\qquad$ sample (survey)

$N >> n$.

n, Y ✓

p: population proportition $\longleftarrow$ $\hat{p} = \frac{Y}{n}$ sample proporti

$$\frac{400,000}{999,999} \qquad \frac{399,999}{999,999}$$

$$\frac{400,000}{999,001}, \quad \frac{399,001}{999,001},$$

When $N >> n$, we assume $Y \sim B(n,p)$

$$Y \approx N(np, npq)$$

$$\hat{p} = \frac{Y}{n} \approx N\left(p, \frac{pq}{n}\right)$$

$$\Rightarrow \frac{\hat{p} - p}{\sqrt{pq/n}} \approx Z$$

$100(1-\alpha)\%$ CI for p: $\quad \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{pq}{n}}$

$$\left(\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}},\ \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

Ex: In a big statewide election, let p be the support rate for candidate A.

Based on a random survey of 1000 people, 520 said they support A.

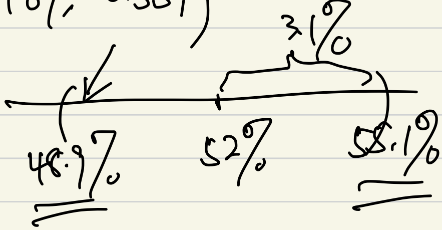
① what's the point est of p? $\hat{p} = 0.52$

② what's the margin of error of your est. at 95% conf level?

$$z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96 \sqrt{\frac{0.52 \cdot 0.48}{1000}} = 0.031$$

③ 95% CI for p: $(0.489, 0.551)$

$z_{0.025}$



④ How large does n need to be, so you can be 95% conf. that A is wining based on $\hat{p} = 0.52$?

$$z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \le \varepsilon$$

$$n \geq \frac{z_{\frac{\alpha}{2}}^2 \hat{p}\hat{q}}{\varepsilon^2} = \frac{1.96^2 \cdot 0.52 \cdot 0.48}{0.02^2}$$

$$= 2398$$

Ex. Go to a region.. want to est p. How large n need to be such that your est. has a m.o.e < 0.03 at 90% conf. level?

$$z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \le \varepsilon$$

$$n \geq z_{\frac{\alpha}{2}}^2 \frac{\hat{p}\hat{q}}{\varepsilon^2} = \frac{1.645^2 \cdot \frac{1}{2} \cdot \frac{1}{2}}{0.03^2}$$

$$= \frac{z_{\frac{\alpha}{2}}^2}{4\varepsilon^2} = 752$$

§9.11. difference of 2 population proportion.

popu 1 $\quad p_1$, $\quad n_1 \quad Y_1$ ✓ $\quad \hat{p}_1 = \frac{Y_1}{n_1}$

popu 2 $\quad p_2$, $\quad n_2 \quad Y_2$ ✓ $\quad \hat{p}_2 = \frac{Y_2}{n_2}$

Goal: To est $p_1 - p_2$.

point estimator: $\quad \hat{p}_1 - \hat{p}_2$

sampling dist of $\hat{p}_1 - \hat{p}_2$: $\approx N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$

$$N\left(p_1, \frac{p_1 q_1}{n_1}\right) \qquad N\left(p_2, \frac{p_2 q_2}{n_2}\right)$$

$$\Rightarrow \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \sim Z.$$

$100(1-\alpha)\%$ CI for $p_1 - p_2$.

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

		parts	def	
Ex.	existing process	1500	75	$\hat{p}_1 = 0.05$
	new process	2000	80	$\hat{p}_2 = 0.04$

Find a 90% CI for the true diff of the proportions of the defectives for the existing and new processes.

$$\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$= (0.05 - 0.04) \pm 1.645 \sqrt{\frac{0.05 \cdot 0.95}{1500} + \frac{0.04 \cdot 0.96}{2000}}$$

$$= 0.01 \pm 0.0117 = (-0.0017, 0.0217)$$

$\underbrace{\mu, \ \mu_1 - \mu_2,}_{z \text{ or } t} \quad \underbrace{p, \ p_1 - p_2}_{z}$

all symm.

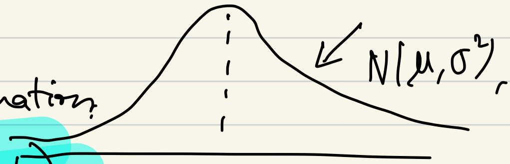
all C.I's: $(\text{point est.}) \pm K \cdot \underline{s.e(\text{point est.})}$

9.6. Prediction Interval.

Ex: $X_1, X_2, \cdots X_n \sim N(\mu, \sigma^2)$ σ^2 known

Goal: To predict a future observation X_0

point est: $\bar{X}$



variance of your estimation

$$\frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1+\frac{1}{n}\right)$$

variance of $\bar{X}$ as an est of μ.

variance of an individual observation from the popu.

$100(1-\alpha)\%$ prediction interval for an indiv. response x_0:

$$\left(\bar{X} - z_{\frac{\alpha}{2}}\,\sigma\sqrt{1+\frac{1}{n}},\ \bar{X} + z_{\frac{\alpha}{2}}\cdot\sigma\cdot\sqrt{1+\frac{1}{n}}\right)$$

$$\frac{X_0 - \bar{X}}{\sigma\sqrt{1+\frac{1}{n}}} \sim Z.$$

Ex: $n=50$, $\bar{X}=260$, assume $\sigma = 25$.

for the next mortgage applicant, ~~predict~~ Find a 95% prediction interval for his loan amount?

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \sigma \sqrt{1+\frac{1}{n}}$$

$$= 260 \pm 1.96 \cdot 25 \sqrt{1+\frac{1}{50}}$$

$$= 260 \pm 49.5$$

$$= (210.5, 309.5)$$