Chp 19: Inference About a Population Proportion

☆ necessary when the data we collect has only "success" or "failure". (for example, ① yes/No questionnaire or ② coin toss)

sample proportion: $\hat{p} = \dfrac{\#\text{ successes in sample}}{\text{total }\#\text{ individuals in sample}}$

Sampling Distribution of Sample Proportion

SRS, from large population, of size n that has proportion p of successes. $\hat{p} = \dfrac{\#\text{successes in sample}}{n}$

- Then mean of sampling distribution is p.
- s.d. $= \sqrt{\dfrac{p(1-p)}{n}}$
- As n increases, $\hat{p} \sim N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$ approximately.

☆ $\hat{p}$ is unbiased estimator for p (like $\bar{x}$ is unbiased estimator for μ in previous chapters)

Ex

Do college students pray? A study of religious practices among college students interviewed a sample of 127 students; 107 of the students said that they prayed at least once in a while.

(a) Describe the population and explain in words what the parameter p is.

(b) Give the numerical value of the statistic $\hat{p}$ that estimates p.

Chp 19 (cont)

Ex 2

Watching online video. About 75% of young adult Internet users (ages 18 to 29) watch online video. Suppose that a sample survey contacts an SRS of 1000 young adult Internet users and calculates the proportion $\hat{p}$ in this sample who watch online video.

(a) What is the approximate distribution of $\hat{p}$?

(b) If the sample size were 4000 rather than 1000, what would be the approximate distribution of $\hat{p}$?

Large-Sample CI for p

SRS of size n from large population w/ unknown proportion of successes p. ~CI for p is

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

* use CI only when # successes + # failures in sample are both at least 15, and for large n.

why z^*, not t^*? because s.d. depends only on $\hat{p}$, we can approximate distribution better as normal curve

Chp 19 (cont)

Ex 3

How common is SAT coaching? A random sample of students who took the SAT college entrance examination twice found that 427 of the respondents had paid for coaching courses and that the remaining 2733 had not.[6] Give a 99% confidence interval for the proportion of coaching among students who retake the SAT.

Plus Four CI for p

same as last CI except $\hat{p} = \dfrac{\#\text{ successes in sample} + 2}{n+4}$

and new sample size is $n+4$

☆ use this CI when want at least 90% confidence and $n \geq 10$, w/ any # successes & failures.

☆ This is necessary to make CI more accurate when sample success rate is too far away from 0.5

☆ It turns out to get true 90% CI, we'd need $n = 646$ but only $n = 11$ for plus-four CI. (based on numerical work done w/ computer).

Chp 19 (cont)

Ex 4

Teens' MySpace profiles. Over half of all American teens (ages 12 to 17 years) have an online profile, mainly on MySpace. A random sample of 487 teens with profiles found that 385 included photos of themselves.[11]

(a) Give the 95% large-sample confidence interval for the proportion p of all teens with profiles who include photos of themselves.

(b) Give the plus four 95% confidence interval for p. If you express the two intervals in percents, rounded to the nearest tenth of a percent, how do they differ? (The plus four interval always pulls the results away from 0% or 100%, whichever is closer. Even though the condition for using the large-sample interval is met, the plus four interval is more trustworthy.)

margin of error

$$m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

How do we choose $\hat{p}$ to calculate n for specified m?

① use guessed p^* based on $\hat{p}$ from pilot study or past experience.

② use $p^* = 0.5$ (this gives most conservative m)

⇒ to get CI level C for population w/ proportion p,

use $$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$$

Chp 19 (cont)

Ex 5

Canadians and doctor-assisted suicide. A Gallup Poll asked a sample of Canadian adults if they thought the law should allow doctors to end the life of a patient who is in great pain and near death if the patient makes a request in writing. The poll included 270 people in Québec, 221 of whom agreed that doctor-assisted suicide should be allowed.[12]

(a) What is the margin of error of the large-sample 95% confidence interval for the proportion of all Québec adults who would allow doctor-assisted suicide?

(b) How large a sample is needed to get the common ±3 percentage point margin of error? Use the previous sample as a pilot study to get p^*.

Significance Test for proportion p

SRS size n from large population w/ p unknown.

$H_0: p = p_0$ $\quad z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

1. $H_a: p > p_0$ $\quad$ pvalue $= P(Z \geq z)$
2. $H_a: p < p_0$ $\quad$ pvalue $= P(Z \leq z)$
3. $H_a: p \neq p_0$ $\quad$ pvalue $= 2P(Z \geq |z|)$

☆ Use when n is large enough so that np_0 & $n(1-p_0)$ are both 10 or more.

Chp 19 (cont)

Ex 6

Spinning pennies. Spinning a coin, unlike tossing it, may not give heads and tails equal probabilities. I spun a penny 200 times and got 83 heads. How significant is this evidence against equal probabilities?

Ex 7

No test. Explain why we can't use the z test for a proportion in these situations:

(a) You toss a coin 10 times in order to test the hypothesis H_0: $p = 0.5$ that the coin is balanced.

(b) A college president says, "99% of the alumni support my firing of Coach Boggs." You contact an SRS of 200 of the college's 15,000 living alumni to test the hypothesis H_0: $p = 0.99$.