

7.4 Choosing a sample size

- confidence intervals are given by the general formula

$$\text{point estimate} \pm \text{margin of error}$$

- margin of error is based on the confidence level and the standard error
- for a sample proportion we have the confidence interval

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

z is the z -score for the appropriate confidence level. we can use the normal distribution based on the CLT

- for the sample mean we have

$$\bar{x} \pm t \left(\frac{s}{\sqrt{n}} \right)$$

here the t score comes from the t distribution and is used when the st deviation is unknown and the sample size is assumed to be small. the value for t is based on the confidence level and the degrees of freedom (which is a function of the sample size). as n gets larger t approaches z .

- question : can we use the formulas for confidence intervals to find the sample size needed?
- remember that for the sample proportion we have the assumption

$$n\hat{p} \geq 15 \quad \text{and} \quad n(1 - \hat{p}) \geq 15$$

- we also have the equations for the margin of error. these can be solved for n

$$n = \hat{p}(1 - \hat{p}) \left(\frac{z}{m} \right)^2 \quad \text{for sample proportion of binomial data} \quad (1)$$

$$n = \left(\frac{ts}{m} \right)^2 \quad \text{for sample mean of quantatative data} \quad (2)$$

- remember that the t score is found based on the sample size. since this is unknown we will use $t = 2$ as an initial estimate.
- in practice the values of \hat{p} and s are unknown before a study is done. we'll look at some examples for some guidelines.
- Example : the U would like to improve the performance of their incoming freshmen. the university will do a study and fund some programs in a specific department if a majority of incoming freshmen think they need help from that department. one of the questions on the survey will be "do you want/need help improving your math skills?". assuming that we want a margin of error equal to 0.04 for the study

- (a) assuming we know that a prior study found that 58% answered yes to a similar question, find n .
 - (b) assuming that we think the study will not recommend the money be spent in the math department.
 - (c) the study actually found that 48.4% answered yes and $n = 1096$. what was the actual confidence interval.
- Ex 10: a social scientist would like to determine education levels in Africa. how large should the study be to determine this given a margin of error of 1 year?
- summary :
 - (a) sample size equations are given by (1) and (2)
 - (b) increasing the margin of error will decrease the sample size
 - (c) increasing the confidence level will increase the sample size
 - (d) as the variability in the data increases, so will the required sample size.
 - (e) an organization may need to consider the financial aspects of running a study, meaning that if a smaller sample size is needed (for whatever reason), they may need to settle for a larger margin of error.