Chapter 7: Interval Estimation

7.4 Sample Size

Part I: "How large should the sample size be to estimate a mean, μ?"

To estimate μ within a predetermined margin of error ϵ, take a random sample of size

$$\boxed{n = \frac{z_{\alpha/2}^2 \sigma^2}{\epsilon^2}}$$

where it is assumed that σ^2 is known.

Proof. $100(1-\alpha)\%$ CI for μ is

Note 1. – We sometimes call $\epsilon = z_{\alpha/2}(\sigma/\sqrt{n})$ the **maximum error of the estimate**.

– If the experimenter has no idea about the value of σ^2, it may be necessary to first take a preliminary sample to estimate σ^2.

– Since n must be an integer we need to ____________ the value of n to an integer.

Example 1. Let X equal the excess weight of soap in a "1000- gram" bottle. Assume that the distribution of X is $N(\mu, 169)$. What sample size is required so that we have 95% confidence that the maximum error of the estimate of μ is 1.5?

Example 2. Measurements of the length in centimeters of $n = 29$ fish yielded an average length of $\bar{x} = 16.82$ and $s^2 = 34.9$. Determine the size of a new sample so that $\bar{x} \pm 0.5$ is an approximate 95% confidence interval for μ.

Part II: "How large should the sample size be to estimate a proportion, p?"

To estimate p within a predetermined margin of error ϵ, take a random sample of size

$$\boxed{n = \frac{z^2_{\alpha/2} p^*(1-p^*)}{\epsilon^2}}$$

where p^* is a good guess for p.

Proof. $100(1-\alpha)\%$ CI for p is

What if we do not have a good guess (p^*) for p? Often, we do not have a strong prior idea about p. It is interesting to observe that no matter what value p takes between ______ and ______, it is always true that $p^*(1-p^*) \leq$ ______. (Note that the equality hods when $p^* = 0.5$.) Hence,

Thus, if we want the $100(1-\alpha)\%$ confidence interval for p to be no longer than $y/n \pm \epsilon$, a solution for n that provides this protection is

Note 2. We set $p^* = 0.5$ when we know nothing about p

Example 3. A possible gubernatorial candidate wants to assess initial support among the voters before making an announcement about her candidacy. If the fraction p of voters who are favorable, without any advance publicity, is around 0.15, the candidate will enter the race. From a poll of n voters selected at random, the candidate would like the estimate y/n to be within 0.03 of p. The decision is based on a 95% CI.

1. Find the sample size n need to achieve the desired reliability and accuracy.

2. Suppose that $n =$ ______. voters around the state were selected at random and interviewed and $y = 214$ express support for the candidate. Find an approximate 95% confidence interval for p

3. On the basis of this sample, would the candidate decide to run for office? Explain.

How to determine the sample size when the population is not so large relative to the desired sample size. Suppose that you want to estimate the proportion p of a student body that favors a new policy. How large should the sample be? If you want to be 95% confident that the maximum error of the estimate is $\epsilon = 0.02$.

Such a sample size makes sense at a large university. However, if you are a student at a small college, the entire enrollment could be less than 2401.

Suppose now that we are interested in determining the sample size n from a population of size N that is required to have $100(1 - \alpha)\%$ confidence that the maximum error of the estimate of p is ϵ. We let

Example 4. If the size of the student body is $N = 4000$ and $1 - \alpha = 0.95$, $\epsilon = 0.02$. Find the required sample size.

Example 5. Suppose that a college of N = 3000 students is interested in assessing student support for a new form for teacher evaluation. To estimate the proportion p in favor of the new form, how large a sample is required so that the maximum error of the estimate of p is $\epsilon = 0.03$ with 95% confidence?

Example 6. Let p equal the proportion of college students who favor a new policy for alcohol consumption on campus. How large a sample is required to estimate p so that the maximum error of the estimate of p is 0.04 with 95% confidence when the size of the student body is

1. N = 1500?

2. N = 15,000??

3. N = 25,000?