8.6 Large-Sample Confidence Intervals

General Setup:

- We now consider any estimator $\hat{\theta}$ of some target parameter θ, where $\hat{\theta}$ has an approximately normal sampling distribution with mean θ and standard deviation $\sigma_{\hat{\theta}}$.

Then

- Hence Z is (approximately) a pivotal quantity.
- Let's derive a $100(1-\alpha)\%$ two-sided CI for θ:

Note: Often $\sigma_{\hat{\theta}}$ will depend on unknown parameters.

- We will replace it by an estimate $\hat{\sigma}_{\hat{\theta}}$.
- If the estimate is "good" (in the sense that $\hat{\sigma}_{\hat{\theta}} \approx \sigma_{\hat{\theta}}$ with high probability for large n) then:

Four Examples of Large-Sample CIs

- We earlier derived four unbiased estimators for various sampling situations:

1. $\overline{Y}$ for μ (one random sample)
2. $\hat{p}$ for p (one Bernoulli random samples)
3. $\overline{Y}_1 - \overline{Y}_2$ for $\mu_1 - \mu_2$ (two independent samples)
4. $\hat{p}_1 - \hat{p}_2$ for $p_1 - p_2$ (two indep. Bernoulli samples)

- By the CLT, each of these estimators has an ____________ __________ sampling distribution when __________________.

Recall: For large n,

For large n_1, n_2:

- So approximate $100(1-\alpha)\%$ large-sample CIs for these parameters are:

1

2

(3)

(4)

Example 1: In a 1983 poll of 1250 randomly selected Americans, 238 reported using seat belts. In a 1992 poll of 1251 random Americans, 876 reported using seat belts. Find a 95% CI for the difference in the true proportions wearing seat belts in 1983 and in 1992.

Meaning of "95 % confidence": If we took many random samples from these populations & constructed a 95 % CI for $p_1 - p_2$ each time, then about 95 % of the CIs we obtained would contain the true value of $p_1 - p_2$.

Exercise: A sample of 60 summer days measuring the number of ships passing near a power-plant location showed a mean of 7.2 ships per day with a sample variance of 8.8. A sample of 90 winter days yielded a mean of 4.7 and a variance of 4.9. Show that a 90 % CI for the difference between the true summer mean and true winter mean is (1.762, 3.238).

$1-\alpha =$ $\frac{\alpha}{2} =$

Conclusion:

8.7 Sample-size Determination

- For CIs of the form $\hat{\theta} \pm Z_{\alpha/2}\,\hat{\sigma}_{\hat{\theta}}$, the term $Z_{\alpha/2}\,\hat{\sigma}_{\hat{\theta}}$ is called the <u>bound</u> B (or the margin of error) of the CI.

- In some studies, collecting observations is expensive or time-consuming.

- Often the researcher wishes to know <u>ahead of time</u> what size sample she must take to achieve some desired margin of error.

<u>Note:</u> The estimated standard error $\hat{\sigma}_{\hat{\theta}}$ cannot be calculated before we collect the sample data.

- For determining sample size, we plug in a "guess" for σ or p, based on previous knowledge or expert opinion.

<u>Example 1:</u> We seek a 95% CI for the mean number of hunting days for licensed hunters in SC, with a margin of error of 2 days. How many hunters must we sample? (We believe σ is around 10.)

Example 2: An experiment wishes to estimate (with 90% confidence) the proportion of people who respond to a stimulus, with a margin of error of at most 0.04. Previous belief is that this proportion is around 0.6.

- For CIs about a proportion, if there is no prior knowledge about p, a guess of 0.5 should be used. This will yield the most conservative (largest) value for n.