AP Statistics
Unit 06 – Day 01 Notes
Confidence Intervals: The Basics

Name Key
Period______

Statistical Inference: provides methods for drawing conclusions about a population from sample data.

Parameter: a # that describe a population.
* fixed #. * sometimes estimated when we do not know the true population value.

Statistic: a # that describes a sample.
* changes from sample to sample.
* used sometimes to estimate an unknown parameter.

Statistics come from **S**amples, and **P**arameters come from **P**opulations

SAMPLE	POPULATION
statistic	parameter
$\bar{x}$ = mean of a sample	μ = mean of a population
s_x = standard deviation of a sample	σ = standard deviation of a population
$\hat{p} = \frac{x}{n} = \frac{\#\text{successes}}{\#\text{in sample}}$ = sample proportion	p = population proportion

Example 1: Identify the population and the sample, and explain what p and $\hat{p}$ represent.
Oprah put out a survey to find out what percent of people thought it was time for her retirement. Out of the 500 surveys that were sent out only 340 were sent back with "retire" as the response. We want to estimate the thoughts of the general public.

Population: general public

p: % of people who thought she should retire

Sample: 500 surveys that were sent out

$\hat{p} = \frac{340}{500} = 0.68$

Point Estimator: ($\bar{x}, s_x, \hat{p}$) statistic that provides an estimate of a population parameter (μ, σ, p)

Point Estimate: (#) specific value of a point estimator from a sample ($\bar{x}, s_x, \hat{p}$)

Example 2: In each of the following settings, determine the point estimator you would use and calculate the value of the point estimate.

a) Quality control inspectors want to estimate the mean lifetime μ of the AA batteries produced in an hour at a factory. They select a random sample of 50 batteries during each hour of production and then drain them under conditions that mimic normal use. Here are the lifetimes (in hours) of the batteries from one such sample:

16.73	15.60	16.31	17.57	16.14	17.28	16.67	17.28	17.27	17.50	15.46	16.50	16.19
15.59	17.54	16.46	15.63	16.82	17.16	16.62	16.71	16.69	17.98	16.36	17.80	16.61
15.99	15.64	17.20	17.24	16.68	16.55	17.48	15.58	17.61	15.98	16.99	16.93	16.01
17.54	17.41	16.91	16.60	16.78	15.75	17.31	16.50	16.72	17.55	16.46		

point estimator: ($\bar{x}$) sample mean to estimate (μ) population mean lifetime of AA batteries.

point estimate: $\bar{x} = 16.718$ hours

b) What proportion p of U.S. high school student's smoke? The 2011 Youth Risk Behavioral Survey questioned a random sample of 15,425 students in grades 9 to 12. Of these, 2792 said they had smoked cigarettes at least one day in the past month.

point estimator: sample proportion ($\hat{p}$) to estimate the population proportion (p) of US high schoolers who smoke.

point estimate: $\hat{p} = \frac{2792}{15425} = 0.181$

18.1% of US highschoolers smoke

c) The quality control inspectors in part (a) want to investigate the variability in battery lifetimes by estimating the population variance σ^2.

point estimator: sample variance s_x^2 to estimate population variance σ^2 of battery lifetimes.

point estimate: $s_x^2 = 0.441$ hours

C% Confidence Interval: gives an interval of plausible values for a parameter. The interval is calculated from the data and has the form point estimate ± margin of error

- Can be written three different ways: $\bar{x} \pm \mathrm{SD}$, lower # to a higher #, or (low #, high #)
- Gives an interval of plausible values for the parameter

* **Interpreting Confidence Intervals**

We are C% that the interval from _______ to _______ captures the [parameter in context].

Margin of Error: the difference between the point estimate and the true parameter value will be less than the margin of error in C% of all samples, where C is the confidence level.

- Other things being equal, the margin of error of a confidence interval gets smaller as the confidence level C decreases and the sample size n increases

Confidence level C: gives the overall success rate of the method for calculating the confidence interval. That is, for C% of all possible samples, the method would yield an interval that captures the true parameter.

- Use 90% confidence or higher
- 95% confidence is the most commonly used!!! *

* **Interpreting Confidence Levels**

If we take many samples of the same size from this populations, about C % of them will result in an interval that captures the true [parameter in context].

Example 3: Two weeks before a presidential election, a polling organization asked a random sample of registered voters the following question: "If the presidential election were held today, would you vote for candidate A or candidate B?" Based on this poll, the 95% confidence interval for the population proportion who favor candidate A is (0.48, 0.54).

a) Interpret the confidence interval.

We are 95% confident that the interval from 0.48 to 0.54 captures the true proportion of all registered voters who favor candidate A in the election.

b) What is the point estimate that was used to create the interval? What is the margin of error?

point estimate = midpoint of the interval: $\frac{0.48+0.54}{2} = 0.51 = \hat{p}$

margin of error = $0.54 - 0.51 = 0.03$

c) Based on this poll, a political reporter claims that the majority of registered voters favor candidate A. Use the confidence interval to evaluate this claim.

Any value from 0.48 - 0.54 is a plausible value for the population proportion p that favors candidate A. Because there are plausible values of p less than 0.50, the confidence interval does not give convincing evidence to support the reporters' claim that the majority (more than 50%) of registered voters favor candidate A.

ACTIVITY

Interpreting Confidence Levels… more info!

If we take many samples of the same size from this population, about C% of them will result in an interval that captures the actual parameter value.

- WHATS THE PROBABILITY THAT OUR 95% CONFIDENCE INTERVAL CAPTURES THE PARAMETER? It's not 95%! Once we have chosen a sample and have a mean it is either 100% in the confidence interval or it is 0% in the confidence interval.

Critical Value: multiplier that makes the interval wide enough to have the stated capture rate. The critical value depends on both the confidence level C and the sampling distribution of the statistic.

Calculating a Confidence Interval

The confidence interval for estimating a population parameter has the form

$$\text{Statistic} \pm \underbrace{(\text{critical value}) \times (\text{standard deviation of statistic})}_{\text{margin of error}}$$