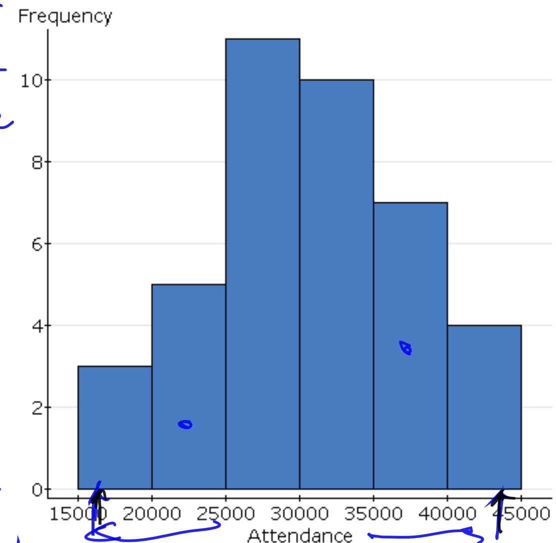
Packet 10: One-Sample Confidence Interval for the Population Mean Textbook pages: 589 – 596

After completing this material, you should be able to:

- construct a confidence interval for the ~~proportion~~ population mean using the appropriate format.
- state when it is valid to use this procedure.
- explain what "confidence" means.

Bryan Price, the manager of the Cincinnati Reds, is interested in estimating the average attendance at home games. The attendance numbers reported for a sample of 40 home games in the 2016 season are summarized in the histogram and summary statistics below.

Referring to the summary statistics when appropriate, completely describe the distribution of attendance at home games using complete sentences. The distribution of home game attendance is approximately normal. The average attendance at Reds 40 home games was about 30,405; 50% of the games were attended by 30479 people or more. The minimum attendance was 15616, while the maximum was 43633. The middle 50% of games had attendance between 26203 & 35921.



$\bar{y}$ estimating μ

Summary statistics:

Column	n	Mean	Std. dev.	Min	Q1	Median	Q3	Max
Attendance	40	30404.575	7063.3843	15616	26202.5	30479	35921.5	43633

s estimating σ

Find the interval that is two standard deviations from the mean. Interpret this interval (in context) using either the Empirical Rule or Chebyshev's rule, whichever is most appropriate for this distribution.

approximately normal.

y should be within 2σ of the mean 95% of the time

$(30{,}405 \pm 2(7063))$

$y \in [16{,}278, 44{,}530]$ with 95% confidence

Recall: What is a confidence interval?

A confidence interval (CI) is an estimate for a population parameter as a box (rather than a point - just a number) - a box within which the parameter will fall say 95% of the time

Steps in a Confidence Interval

1. Choose the correct formula
2. Collect data & calculate the interval
3. Interpret the interval (level of confidence, context)

we choose

We want to develop a confidence interval which will be used to estimate the population mean. What will the formula for this interval be? Confidence Interval CI for the population mean is:

parameters estimate statistic spread based on n

μ lies in $\bar{y} \pm t \cdot \frac{s}{\sqrt{n}}$

df = n-1 n sample size

$\boxed{\mu} \in \left[\bar{y} - t_{crit}\frac{s}{\sqrt{n}},\ \bar{y} + t_{crit}\frac{s}{\sqrt{n}}\right]$

Back to the example: Estimate the mean home game attendance for the Cincinnati Reds with 95% confidence. Assume the appropriate critical point is 2.023.

$t_{crit} = 2.023$ t with df = 39 = 40-1

95% confidence

By the way, z_{crit} for in the past, 95% confidence is 1.96

Notice how close 1.96 is to 2.023

formula: $\bar{y} \pm t_{crit} \cdot \frac{s}{\sqrt{n}}$

$\mu \in 30404.575 \pm \underbrace{2.023\left(\frac{7063.3843}{\sqrt{40}}\right)}_{MOE}$

CI: $\mu \in [28145.25,\ 32663.90]$; with 95% confidence, we estimate the mean μ attendance at Red's home games to be somewhere between [28145, 32664].

Explain why the two intervals we have found differ.

The interval [16278, 44530] was 95% CI for y. The interval [28145, 32664] is a 95% CI for μ, based off the sample mean $\bar{y}$. The std dev. of $\bar{y}$ is much smaller than the std. dev. of y (by a factor of $s/\sqrt{n}$); similarly, the CI based on $\bar{y}$ is much smaller than a CI for y.

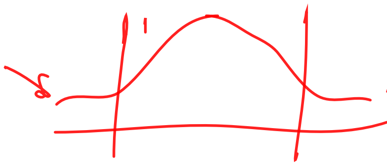
Example: Hoping to lure more shoppers downtown, a city builds a new parking garage in the central business district. The city plans to pay for the structure using parking fees. During a two-month period (41 weekdays), daily fees collected averaged \$126.50 with a standard deviation of \$15.278.

10% in the two tails $n=41$ $df = n-1 = 41-1 = 40$

— Estimate the mean amount of money collected from daily parking fees with 90% confidence.

with a CI - rather than a simple point estimate

Two-tail probability	0.20	0.10	0.05	0.02	0.01	
One-tail probability	0.10	0.05	0.025	0.01	0.005	
Table T df						df
30	1.310	1.697	2.042	2.457	2.750	30
32	1.309	1.694	2.037	2.449	2.738	32
35	1.306	1.690	2.030	2.438	2.725	35
40	1.303	1.684	2.021	2.423	2.704	40
45	1.301	1.679	2.014	2.412	2.690	45



missing something right & left both tails   two tails

formula: $\bar{y} \pm t_{crit}\frac{s}{\sqrt{n}}$

$t_{crit} = 1.684$ | 90% confidence, $df = 40$

In this case: $126.50 \pm \underbrace{1.684\frac{15.278}{\sqrt{41}}}_{4.018 \text{ MOE}}$

C.I. = [122.482, 130.518]

Interpret: With 90% confidence we estimate the parking receipts will be between \$122.48 & \$130.52. \$130 is in there but!

— The consultant who advised the city on this project predicted that parking revenues would average \$130 per day. Based on your interval, do you think the consultant was correct? Explain.

We can't reject a value \$130 per day, because it falls within our interval. However we're a little nervous, because we just barely caught it, on the R.H.S. (right hand side) of the interval.

— Suppose that for budget planning purposes, the city needs a better estimate of the mean daily income from parking fees. How can they get a better estimate?

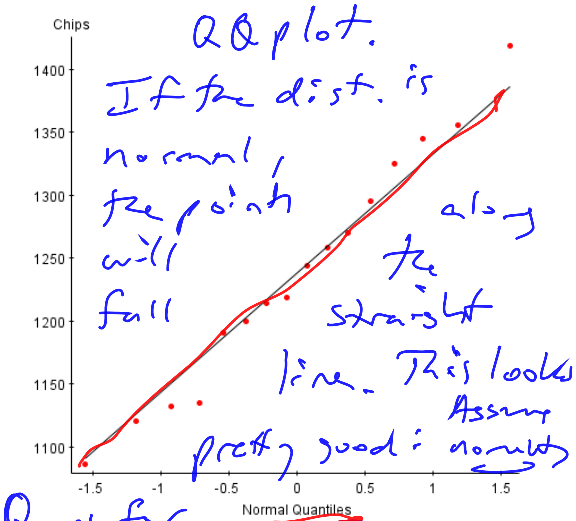
The one thing we have control over is sample size, & since larger sample size shrinks the C.I., we could take a larger sample. We really need a random sample, however; not 41 days in a row.

Example: Prior to an advertising campaign, quality control experts working for Chips Ahoy needed to estimate the average number of chocolate chips in 18oz packages of cookies. To do this, a random sample of 16 bags of cookies were taken from the production line, and the quality control engineer counted the number of chocolate chips in the cookies in each bag.

$n = 16$

— Determine if the required assumptions are satisfied in order to estimate the mean number of chocolate chips in bags of Chips Ahoy cookies.

n=16 is way too low; I'd like 200 bags of cookies (to be safe); but actually 30 would be good enough.
The distribution of chips in a bag is not known.
The QQ plot is going to save us! Q is for



— The sample of 16 bags had an average of 1238.1875 chocolate chips with a standard deviation of 94.282 chips. Estimate the mean number of chocolate chips in 18oz packages of Chips Ahoy cookies with 95% confidence. If the estimate should not be calculated, explain why.
Note: One of the intervals given below is a 90% confidence interval, one corresponds to 95% confidence, and the final corresponds to 99% confidence.

$\bar{y} = 1238.1875$ chips
$s = 94.282$ chips

Possible Intervals	
(1187.9481, 1288.4269)	95%
(1196.8672, 1279.5078)	90%
(1168.7320, 1307.6430)	99%

We proceed, justified by the QQ plot.

C.I.: $\bar{y} \pm t_{crit} \frac{s}{\sqrt{n}}$ $df = 16 - 1 = 15$, 95% confidence

So with 95% confidence, an 18oz bag of Chips Ahoy cookies contains between 1187 & 1289 chips.

— In the 90s, Chips Ahoy aired a commercial claiming "1000 chips in every bag." Comment on the validity of this commercial claim.

Our claim is about $\mu \in [1187, 1289]$; however the standard deviation of the 16 bags is high! About 94 chips. If the true mean were 1187, 1000 is only 2 standard deviations away.
5% of the time we'd ...