

Descriptive Inference Examples for Single Variables Using Confidence Intervals

Brady T. West

Example 1: Confidence Intervals for Proportions

Research Question:

What **proportion** of non-Hispanic African-Americans age 18+ in the U.S. in 2015-2016 had systolic blood pressure greater than 130 mmHg?

Inference Approach:

Provide a **95% confidence interval estimate** (*lower and upper limits*) for this population proportion

Step 1: Estimate the Population Proportion

- **Number** of black respondents with **non-missing data** on first systolic blood pressure measurement: 1135 *sample size*
- **Best Points Estimate:** *Assuming simple random sample of black adults, sample proportion is $\frac{465}{1135} = 0.4097$*
- **Interpretation:** Estimate 40.97% of all such black adults in 2015-16 had systolic blood pressure greater than 130 mmHg

Step 2: Compute Estimated Standard Error

Standard error = $\sqrt{\text{sampling variance of sample proportion}}$
= standard deviation of the sampling distribution of all possible sample proportions if repeated samples of 1135 were obtained

$$\text{standard error} = \sqrt{\frac{p(1-p)}{n}}$$

$$\textbf{Estimated standard error} = \sqrt{\frac{0.4097(1-0.4097)}{1135}} = 0.0146$$

Step 3: Form the Confidence Interval

Best Estimate \pm Margin of Error

Best Estimate \pm “a few” (estimated) standard errors

Large sample size \rightarrow 95% multiplier = 1.96

Lower Limit: $0.4097 - 1.96 \times 0.0146 = 0.3811$

Upper Limit: $0.4097 + 1.96 \times 0.0146 = 0.4383$

Inference about the Population Proportion

95% confidence interval for the population proportion of non-Hispanic African-Americans age 18+ in U.S. in 2015-2016 with systolic blood pressure greater than 130 mmHg is:
(0.3811, 0.4383)

- “95% confidence” → 95% of intervals formed this way expected to cover the true population proportion!
- **Inference:** if hypothesized proportion was 0.35, 95% confidence interval suggests 0.35 **not a plausible value**

Example 2: Confidence Intervals for

Research Question:

What was the **mean** systolic blood pressure for non-Hispanic African-Americans age 18+ in U.S. in 2015-2016?

Inference Approach:

Provide a **95% confidence interval estimate** (*lower and upper limits*) for this population mean

Step 1: Estimate the Population Mean

- **Number** of black respondents with **non-missing data** on first systolic blood pressure measurement: 1135 *sample size*
- **Best Point Estimate:** *Assuming simple random sample of black adults, sample mean is 128.252 mmHg*
- **Interpretation:** Our estimate of the mean systolic blood pressure for all such black adults in 2015-2016 is 128.252 mmHg

Step 2: Compute Estimated Standard Error

Standard error = $\sqrt{\text{sampling variance}}$ of sample mean
= standard deviation of the sampling distribution of all possible sample means if repeated samples of 1135 were obtained

Sample standard deviation of the 1135 blood pressure measurements is 19.958 mmHg

$$\text{Estimated standard error} = \frac{19.958}{\sqrt{1135}} = 0.592 \text{ mmHg}$$

$$\text{Standard error} = \frac{s}{\sqrt{n}}$$

Step 3: Form the Confidence Interval

Best Estimate \pm Margin of Error

Best Estimate \pm “a few” (estimated) standard errors

Large sample size \rightarrow 95% multiplier = 1.96

Lower Limit: $128.252 - 1.96 \times 0.592 = 127.091$ mmHg

Upper Limit: $128.252 + 1.96 \times 0.592 = 129.413$ mmHg

Inference about the Population Mean

95% confidence interval for the population mean systolic blood pressure of non-Hispanic African-Americans age 18+ in U.S. in 2015-2016 is:

(127.091 mmHG, 129.413 mmHg)

- “95% confidence” → 95% of intervals formed this way expected to cover the true population proportion!
- **Inference:** if hypothesized mean was 128 mmHg, 95% confidence interval suggests 128 is a **plausible value**

Alternative Inferential Approaches

Hypothesis testing approach

→ reject or fail to reject null hypotheses about specific values of population mean or proportion of interest, at a certain significance level (e.g. 5%)

Confidence intervals

→ advantage of providing range of plausible values for population mean or proportion of interest, with a certain confidence level (e.g., 95%)