Quantitative Literacy:
Thinking Between the Lines

Crauder, Noell, Evans, Johnson

Chapter 6:
Statistics

1

Chapter 6: Statistics
Lesson Plan

▸ Data summary and presentation: Boiling down the numbers

▸ The normal distribution: Why the bell curve?

▸ The statistics of polling: Can we believe the polls?

▸ Statistical inference and clinical trials: Effective drugs?

6.3 The statistics of polling: Can we believe the polls?

Learning Objectives:

▸ Understand margins of error and confidence levels in polls.

  ▸ Basic terms: Margin of error, confidence interval, and confidence level

  ▸ Polls: Margin of error, confidence interval, and confidence level

  ▸ How big should the sample be?

6.3 The statistics of polling: Can we believe the polls?

▸ The **margin of error** of a poll expresses how close to the true result (the result for the whole population) the result of the poll can be expected to lie.

▸ To find the **confidence interval,** adjust the result of the poll by adding and subtracting the margin of error.

▸ The **confidence level** of a poll tells the percentage of such polls in which the confidence interval includes the true result.

Polls and Margin of Error

Suppose that, based on random sampling, a poll reports the percentage of the population having a certain property (e.g., planning to vote for a certain candidate) with a margin of error m. Assuming that this margin is based on a 95% confidence level, we can say that if we conducted this poll 100 times, then we expect about 95 of those sample results to be within m percentage points of the true percentage having that property.

6.3 The statistics of polling: Can we believe the polls?

▸ **Example:** Explain the meaning of a poll that says 33% of Americans approve of what Congress is doing, with a margin of error of 4% and confidence level of 90%

▸ **Solution:** In 90% of such polls, the reported approval of Congress will be within four percentage points of the true approval level.

▸ Thus, we can be 90% confident that the true level lies in the confidence interval between 29% and 37%.

Margin of Error

For a 95% level of confidence, we can estimate the margin of error when we poll n people using:

$$\text{Margin of error} \approx \frac{100}{\sqrt{n}} \%$$

Here, the symbol $\approx$ means "is approximately equal to."

▸ **Example:** A recent Oricon fashion survey asked 900 people, "Which Japanese male celebrity looks best in sneakers?" The winner was Kimura Takuya. What is the approximate margin of error for a 95% confidence level?

▸ **Solution:**

With n = 900:

$$\text{Margin of error} \approx \frac{100}{\sqrt{n}}\% = \frac{100}{\sqrt{900}} = 3.3\%$$

We can be 95% confident that our poll result is within 3.3 percentage points of the true value.

6.3 The statistics of polling: Can we believe the polls?

▸ **Example:** The Kaiser Family Foundation polled 1294 residents of Orleans Parish in New Orleans in 2008 and found that 41% of the residents who had lived through Hurricane Katrina in 2005 report that their lives are still disrupted.

1. The poll surveyed 1294 people. What is the approximate margin of error for a 95% confidence interval?

2. The poll of 1294 people found that 41% of respondents still had disrupted lives. Can we conclude with certainty that no more than 45% of residents' lives are still disrupted?

3. Suppose instead that the poll of 1294 people had found that 52% still had disrupted lives. Explain what we could conclude from this result. Could we assert with confidence that a majority of residents' lives are still disrupted by Katrina?

4. Suppose we wish to have a margin of error of two percentage points. Approximately how many people should we interview?

6.3 The statistics of polling: Can we believe the polls?

▸ **Solution:**

1. With $n = 1294$:

$$\text{Margin of error} \approx \frac{100}{\sqrt{n}} = \frac{100}{\sqrt{1294}} = 2.8\%$$

2. Our answer to part 1 tell us that we can be 95% confident that the poll number of 41% is within 2.8 percentage points of the true percentage of all residents whose lives are still disrupted from Katrina. Thus, it is very likely that the true value is:

   between $41 - 2.8 = 38.2\%$ and $41 + 2.8 = 43.8\%$

   Because the whole interval is below 45%, we can be quite confident (at a 95% level) that no more than 45% of residents' lives are still disrupted.

   On the other hand, we cannot make this conclusion with absolute certainty.

▸ **Solution (cont.):**

3. We can be 95% confident that the poll number of 52% is within 2.8 percentage points of the true percentage of all residents whose lives are still disrupted by Katrina. Thus, it is very likely that the true value is:

    between $52 - 2.8 = 49.2\%$ and $52 + 2.8 = 54.8\%$

Most of this interval falls above 50%, so we continue to have good reason to think that a majority of residents' lives are still disrupted.

But, because a portion of the interval falls below 50%, we should be more cautious in drawing conclusions.

▸ **Solution (cont.):**

4. Substitute 2% for the margin of error:

$$\text{Margin of error} = 2 = \frac{100}{\sqrt{n}} \quad \text{or} \quad \sqrt{n} = 50$$

Hence, $n = 2500$.

We should interview about 2500 people.

Note that one Harris Poll with a 95% confidence level and a margin of error of 2% surveyed 2415 people—very close to the 2500 given by the formula.

Sample Size

For a 95% level of confidence, the sample size needed to get a margin of error of *m* percentage points can be approximated using:

$$\text{Sample size} \approx \left(\frac{100}{m}\right)^2$$

6.3 The statistics of polling: Can we believe the polls?

▸ **Example:** What sample size is needed to give a margin of error of 4% with a 95% confidence level?

▸ **Solution:** We use the approximate formula with $m = 4$:

$$\text{Sample size} \approx \left(\frac{100}{m}\right)^2 = \left(\frac{100}{4}\right)^2 = 625$$

Chapter 6 Statistics: **Chapter Summary**

▶ **Data summary and presentation:** Boiling down

  ▶ Four important measures in descriptive statistics:

    mean, median, mode, and standard deviation

▶ **The normal distribution:** Why the bell curve?

  ▶ A plot of normally distributed data: the *bell-shape* curve.

  ▶ The z-score for a data point

  ▶ The Central Limit Theorem

Chapter 6 Statistics: **Chapter Summary**

‣ **The statistics of polling:** Can we believe the polls?

  ‣ Polling involves: a margin of error, a confidence level, and a confidence interval.

‣ **Statistical inference and clinical trials:** Effective drugs?

  ‣ Statistical significance and p-values.

  ‣ Positive correlated, negative correlated, uncorrelated or linearly correlated