Practicing Statistical Inference on One-sample Categorical Data

Ryan Miller

Introduction

A brief recap of what we've been doing:

- Statisticians analyze data in hopes of learning something about a broader population
  - **Estimation** - use sample data to estimate some characteristic of the population
  - **Hypothesis testing** - use sample data to disprove a theory about the population

Introduction

A brief recap of what we've been doing:

- Statisticians analyze data in hopes of learning something about a broader population
  - **Estimation** - use sample data to estimate some characteristic of the population
  - **Hypothesis testing** - use sample data to disprove a theory about the population
- These methods are *complementary*, and both focus on the *variability* present in the sample data
  - Researchers will often use both methods in their analysis of the same application
  - The results of one method will align with those of the other

Example #1

In a survey of 1002 US adults conducted by Pew Research in Dec 2016, 64% said they think "made-up news" is a significant problem.

1) Why would reporting a *confidence interval* be useful in this application?
2) Use data in this survey to find a 95% confidence interval estimate of the population characteristic of interest
3) How should we interpret this interval?
4) What does "95% confidence" mean?
5) Based upon this interval, do you believe a hypothesis test would conclude the majority of US adults think "made-up news" is a problem?

Example #1 (solution)

1) Why would reporting a *confidence interval* be useful in this application?

- While 64% of *this sample* thought made-up news was a problem, we don't expect *exactly* 64% of *all US adults* to believe this. So, we should report an interval estimate to convey the uncertainty that exists in these data.

Example #1 (solution)

2) Use data in this survey to find a 95% confidence interval estimate of the population characteristic of interest

- Confidence intervals take the form:
  Point Estimate $\pm$ Margin of Error

Example #1 (solution)

2) Use data in this survey to find a 95% confidence interval estimate of the population characteristic of interest

- Confidence intervals take the form:
  Point Estimate $\pm$ Margin of Error
  - We can use the Normal model $\hat{p} \sim N(p, \sqrt{p(1-p)/n})$ to come up with the proper MOE

Example #1 (solution)

2) Use data in this survey to find a 95% confidence interval estimate of the population characteristic of interest

- Confidence intervals take the form:
  Point Estimate $\pm$ Margin of Error
  - We can use the Normal model $\hat{p} \sim N(p, \sqrt{p(1-p)/n})$ to come up with the proper MOE
- For a single proportion we used the Normal model suggested by CLT to come up with the formula: $\hat{p} \pm z^* \sqrt{\frac{p(1-p)}{n}}$

$$0.64 \pm 1.96 * \sqrt{\frac{.64(1-.64)}{1002}} = (0.61, 0.67)$$

Example #1 (solution)

3) How should we interpret this interval?

- We are 95% confident that between 61% and 67% of US adults believe "made-up news" is a problem
- Alternatively, we consider anything between 61% and 67% to be a plausible value for the proportion of all adults who believe made "made-up news" is a problem

Example #1 (solution)

4) What does “95% confidence” mean?

- The confidence level describes the *method* used to create the interval. So, if we used this method many different times, we’d expect the resulting intervals to contain the truth 95% of the time (if they are valid 95% confidence intervals)
- In other words, we are confident in the success rate of the approach used to construct this interval

Example #1 (solution)

5) Based upon this interval, do you believe a hypothesis test would conclude the majority of US adults think "made-up news" is a problem?

- Yes, this test would use the null hypothesis $H_0 : p = 0.5$, and the 95% confidence interval of (0.61, 0.67) suggests 50% is not a plausible value for the population. Thus, we expect the p-value of this test will be less than 0.05 (the complement of the confidence level)

Example #2

A local New Jersey newspaper report published in August 2002 raised the issue of racial bias in the issuance of speeding tickets on the New Jersey Turnpike. In one month, 324 speeding tickets were issued with 81 going to black drivers. Only 16% of registered drivers in New Jersey are black.

1) Use a Z-test to assess whether the proportion of tickets that went black drivers is unexpectedly high relative the proportion of NJ drivers who are black.
2) Does this test prove that racial profiling is being used?
3) What other information would you want to know in this situation to accurately interpret these data?

Example #2 (solution)

1) Use a Z-test to assess whether the proportion of tickets that went black drivers is unexpectedly high relative the proportion of NJ drivers who are black.

- $H_0 : p = 0.16$, we observed $\hat{p} = 81/324 = 0.25$
- $Z = \frac{0.25-0.16}{\sqrt{.16(1-.16)/324}} = 4.42$
- Comparing this Z-value to the Standard Normal curve, the two-sided p-value is nearly zero
- These data are extremely incompatible with the null hypothesis that the proportion of black drivers ticketed on the New Jersey Turnpike equals the proportion of registered New Jersey drivers who are black

Example #2 (solution)

2) Does this test prove that racial profiling is being used?

- No, the hypothesis test only rules out random chance as a possibility. We can be confident that black drivers make up a higher proportion of drivers ticketed on the New Jersey Turnpike than their share of registered drivers in New Jersey, but we do not know why.

Example #2 (solution)

3) What other information would you want to know in this situation to accurately interpret these data?

- Following up on the answer to Question #2, we'd want to know about the fraction of drivers who use the New Jersey Turnpike that are black (since there's no reason to expect this matches proportion of registered drivers in New Jersey).

Example #3

Statistical inference is often used by manufacturers for quality control. In the 1960s, Rockford IL was among the largest fastener manufacturing centers in North America, leading the city to proclaim itself "Screw Capital of the World". Consider a large factory in Rockford produces screws in batches. Company policies stipulate that each batch must have a defect rate lower than 2%.

1) To avoid inspecting every screw in these batches, it is more practical to inspect only samples of screws from each batch. How large should these samples for you to be comfortable that the sample proportions will follow a Normal model?
2) Suppose a sample of 600 screws contains 6 defective screws. Are you convinced that the corresponding batch has a defect rate lower than 2%?
3) Considering the hypothesis test used to answer Question #2, what would a Type I and Type II error represent? Which error would be more damaging to the company?

Example #3 (solution)

1) To avoid inspecting every screw in these batches, it is more practical to inspect only samples of each batch. How large should these samples for you to be comfortable that the sample proportions will follow a Normal model?

- The sample size condition to use a Normal model in this scenario is $np \geq 10$ and $n(1-p) \geq 10$; if we focus on $p = 0.02$, we'd need a sample size of at least $n = 500$

Example #3 (solution)

2) Suppose a sample of 600 screws contains 6 defective screws. Are you convinced that the corresponding batch has a defect rate lower than 2%?

- $H_0 : p = 0.02$ vs $H_A : p < 0.02$
- We observed $\hat{p} = 6/600 = 0.01$, under our null hypothesis this corresponds to a Z-value of $Z = \frac{0.01-0.02}{\sqrt{.02(1-.02)/600}} = -1.75$
- The one-sided p-value is 0.04 (note the two-sided p-value is 0.08)
- I'd feel comfortable saying this batch has a defect rate of less than 2%
  - Some people might want to use a stricter evidence threshold

Example #3 (solution)

3) Considering the hypothesis test used to answer Question #2, what would a Type I and Type II error represent? Which error would be more damaging to the company?

- Recall that a Type I error is rejecting H_0 when H_0 is true
    - In this scenario, that means deciding a batch meets the company's requirement when it really has a defect rate of 2% or higher

Example #3 (solution)

3) Considering the hypothesis test used to answer Question #2, what would a Type I and Type II error represent? Which error would be more damaging to the company?

- Recall that a Type I error is rejecting H_0 when H_0 is true
  - In this scenario, that means deciding a batch meets the company's requirement when it really has a defect rate of 2% or higher
- Recall that a Type II error is not rejecting H_0 when H_0 is false
  - In this scenario, that means deciding a batch has a defect rate that could be 2% or higher, when it really meets the requirement

Example #3 (solution)

3) Considering the hypothesis test used to answer Question #2, what would a Type I and Type II error represent? Which error would be more damaging to the company?

- Recall that a Type I error is rejecting H_0 when H_0 is true
  - In this scenario, that means deciding a batch meets the company's requirement when it really has a defect rate of 2% or higher
- Recall that a Type II error is not rejecting H_0 when H_0 is false
  - In this scenario, that means deciding a batch has a defect rate that could be 2% or higher, when it really meets the requirement
- If Type I errors happen often enough, the company might be in legal trouble
  - While Type II errors lead to the company needing to spend more time doing second inspections, etc.