

Confidence Intervals, Hypothesis Tests, and p-values

Homework Problems

Rob McCulloch

1. The Audit
2. The Distribution of the Sample Mean
 - 2.1. Problem: Cereal Audit
3. Estimating a Normal Mean
 - 3.1. Problem: Normal Mean Confidence Interval
 - 3.2. Problem: USA Mean Return
 - 3.3. Problem: Tokyo Level
4. The Confidence Interval for a Bernoulli p
 - 4.1. Problem: CI-for-proportion-of-up-ticks
 - 4.2. Problem: Sample Size for Acceptable Error
5. The Improved Cereal Process
6. Testing a Normal Mean
7. p-values
8. p-values and testing
 - 8.1. Problem: Testing the USA Mean
 - 8.2. Problem: Testing Equality of Proportions
 - 8.3. Problem: AB Testing

2.1. Problem: Cereal Audit

Suppose you are in charge of the “cereal box filling process”.

You have lots of data on your process and are very confident that the amount of cereal put in each box can be described as iid draws from the normal distribution with mean $\mu = 345$ and standard deviation $\sigma = 15$.

You are about to be audited by an inspector who will take a sample of 5 boxes and compute the average amount put in the 5 boxes.

If that average is greater than 370 or less than 330 you will be in trouble!!

(a)

Let \bar{Y} denote the average of 5 weights the auditors *will* get.

What is the distribution of \bar{Y} ?

(b)

What is the probability that you pass the audit?

3.1. Problem: Normal Mean Confidence Interval

In the notes we computed the 95% confidence for μ using the 500 observations from the cereal box data.

We got a \pm of 1.37.

(a)

Get the cereal.csv data from the webpage check that the sample standard deviation of the weights (s_y) is 15.33 and the sample mean \bar{y} is 344.22.

Check that the 95% confidence interval has a \pm of $1.37 = 2 * \frac{15.33}{\sqrt{500}}$.

(b)

Now let's get these number from the software.

In Excel try:

Click in an empty cell and then:

/Formulas/More Functions/Statistical/confidence.norm.

alpha: .05 (for a 95% interval)

Standard_dev: 15.33 (sample sd)

Size: 500 (sample size)

Or, if you are using R, try:

```
> cd = read.csv("http://www.rob-mcculloch.org/data/cereal.csv")  
> t.test(cd$weights)
```

The width of the R confidence interval divided by 2 should be about 1.37.

Due to rounding "error" it will not be exactly the same.

3.2. Problem: USA Mean Return

Assuming the returns on usa are iid normal, get the 95% confidence for the true mean return.

(use the data conret.csv)

3.3. Problem: Tokyo Level

Get the data `tokyo_sub.csv` from the webpage.

This data is a time-series of daily levels of a Japanese stock index.

(a)

Do the time-series plot of the levels.

Do the time-series plot of the difference of the levels.

(for example the first two values are 10743 and 10760, so the first difference is $10760 - 10743 = 17$)

Do the histogram of the difference of the levels.

Which one could be iid normal?

(b)

For the rest of this question, assume we are modeling the differences of the levels as iid normal.

Let's call this variable D .

$$D_t \sim N(\mu_D, \sigma_D^2).$$

So, for example, D_2 turned out to be 17.
(note that there is no D_1 !!).

Give the 95% confidence interval for μ_D .

(c)

Using the sample mean of the differences as your estimate of μ_D and the sample standard deviation of the differences as your estimate of σ_D , *plug the estimates in* and get a 95% interval for the next difference, that is, what is D_{101} ?

(d)

Using the sample mean of the differences as your estimate of μ_D and the sample standard deviation of the differences as your estimate of σ_D , *plug the estimates in* and get a 95% interval for the next value of the daily level.

Note:!! $L_{101} = L_{100} + D_{101}$.

(e)

What does the confidence interval in part (b) tell us about our plug in procedures in parts (c) and (d)?

4.1. Problem: CI-for-proportion-of-up-ticks

Suppose we are modeling whether or not a stock price goes up each month as iid Bernoulli(p).

We observed that out of the last 99 months, the price went up 62 time so that our estimate of p is $62/99 = 0.63$.

What is the 95% confidence interval for $p = \text{Prob}(\text{price goes up})$?

Is it big?

4.2. Problem: Sample Size for Acceptable Error

Suppose you think an election is close.

You think that if you take a poll, you are likely to get a \hat{p} (sample proportion) for Candidate A close to .5.

Since the election is close, you are thinking the usual $\pm .03$ for sample sizes of about 1,000 will be too big.

What sample size do you need to have a \pm of .01?

8.1. Problem: Testing the USA Mean

Previously we computed the 95% confidence interval for the true mean return of the USA portfolio using the data from the file `conret.csv`.

(a)

Test the hypothesis $H_0 : \mu = 0$.

What is the associated p-value?

(b)

Test the hypothesis $H_0 : \mu = .1/12$.

What is the associated p-value?

8.2. Problem: Testing Equality of Proportions

(a)

We have a random sample of 1000 women voters.

550 of them say they will vote Democrat.

What is your 95% confidence interval for the true proportion of women voters who would vote Democrat?

(b)

We have a random sample of 1000 male voters.

450 of them say they will vote Democrat.

What is your 95% confidence interval for the true proportion of male voters who would vote Democrat?

(c)

Do you think the true proportion of women voters that would vote Democrat is equal to the true proportion for men?

Let p_{men} be the true population proportion for men and p_{women} be the true population proportion for women.

Here is the R output for testing $H_0 : p_{men} = p_{women}$.

```
> prop.test(c(450,550),n=c(1000,1000))  
  
2-sample test for equality of proportions with continuity correction  
  
data:  c(450, 550) out of c(1000, 1000)  
X-squared = 19.602, df = 1, p-value = 9.537e-06  
alternative hypothesis: two.sided  
95 percent confidence interval:  
 -0.14460645 -0.05539355  
sample estimates:  
prop 1 prop 2  
 0.45   0.55
```


8.3. Problem: AB Testing

In Data Science estimating the difference in means or proportions in two different situations is often called “A/B Testing”.

What Is A/B Testing?

I cut this from a webpage

(<https://www.optimizely.com/optimization-glossary/ab-testing/>):
A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better. AB testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

Suppose for example you are considering two different versions of a webpage and you want to see how the “click-through” depends on the version.

In one version an add is placed near the top of the page and in the other version it is placed near the bottom.

A click-through means someone who visits the page clicks on the add.

p_T :

Probability of a “click-through” if you place your ad at the top of the web-page.

p_B :

Probability of a “click-through” if you place your ad at the bottom of the web-page.

Things like Google analytics allow you to randomly move the ad around the page so that you can estimate the difference $p_T - p_B$!

In the internet age, the opportunity to experiment has exploded.

Suppose you randomly assigned the page position of your ad and 138 out of 5009 times the ad was on top you got a click-through and 97 out of 5116 times the ad was on the bottom you got a click through.

- ▶ click-through rate on top: $138/5009 = 0.02755041$.
- ▶ click-through rate on bottom: $97/5116 = 0.01896013$.

The difference is $0.02755041 - 0.01896013 = 0.00859028$

So, it looks like you are getting a much better click-through rate with the ad on top, but, what is the uncertainty?

Here is the R output for prop.test applied to the data:

```
> x
[1] 138  97
> n
[1] 5009 5116
> prop.test(x,n)

      2-sample test for equality of proportions with continuity correction

data:  x out of n
X-squared = 7.8636, df = 1, p-value = 0.005044
alternative hypothesis: two.sided
95 percent confidence interval:
 0.002517929 0.014662640
sample estimates:
      prop 1      prop 2 
0.02755041 0.01896013
```

What does the output of prop.test tell us about the difference in click-through rates?

Are they “significantly” different?