Homework 3 for Stat Inference

Brian Caffo

About these slides

- These are some practice problems for Statistical Inference Quiz 3
- They were created using slidify interactive which you will learn in Creating Data Products
- Please help improve this with pull requests here (https://github.com/bcaffo/courses)

Load the data set `mtcars` in the `datasets` R package. Calculate a 95% confidence interval to the nearest MPG for the variable `mpg`.

1. What is the lower endpoint of the interval?
2. What is the upper endpoint of the interval?

*** .hint Do `library(datasets)` and then `data(mtcars)` to get the data. Consider `t.test` for calculations. You may have to install the datasets package.

*** .explanation

```
library(datasets); data(mtcars)
round(t.test(mtcars$mpg)$conf.int)
```

```
## [1] 18 22
## attr(,"conf.level")
## [1] 0.95
```

18 22

Suppose that standard deviation of 9 paired differences is 1. What value would the average difference have to be so that the lower endpoint of a 95% students t confidence interval touches zero?

1. Give the number here to two decimal places

*** .hint The t interval is $\bar{x} \pm t_{.975,8} * s/\sqrt{n}$

*** .explanation 0.77

We want $\bar{x} = t_{.975,8} * s/\sqrt{n}$

```
round(qt(.975, df = 8) * 1 / 3, 2)
```

```
## [1] 0.77
```

An independent group Student's T interval is used instead of a paired T interval when:

1. The observations are paired between the groups.
2. *The observations between the groups are naturally assumed to be statistically independent*
3. As long as you do it correctly, either is fine.
4. More details are needed to answer this question

*** .hint A paired interval is for paired observations.

*** .explanation We can't pair them if the groups are independent of each other as well as independent within themselves.

Consider the `mtcars` dataset. Construct a 95% T interval for MPG comparing 4 to 6 cylinder cars (subtracting in the order of 4 - 6) assume a constant variance.

1. What is the lower endpoint of the interval to 1 decimal place?
2. What is the upper endpoint of the interval to 1 decimal place?

*** .hint Use `t.test` with `var.equal=TRUE`

*** .explanation

```
m4 <- mtcars$mpg[mtcars$cyl == 4]
m6 <- mtcars$mpg[mtcars$cyl == 6]
#this does 4 - 6
confint <- as.vector(t.test(m4, m6, var.equal = TRUE)$conf.int)
```

3.2 10.7

If someone put a gun to your head and said "Your confidence interval must contain what it's estimating or I'll pull the trigger", what would be the smart thing to do?

1. *Make your interval as wide as possible*
2. Make your interval as small as possible
3. Call the authorities

*** .hint C'mon. You don't need a hint

*** .explanation This is just an example of what happens to confidence intervals as you increase the confidence level. You want to be quite sure in your interval (i.e. have a large confidence level) and so you would increase the interval's width

Refer back to comparing MPG for 4 versus 6 cylinders. What do you conclude?

1. The interval is above zero, suggesting 6 is better than 4 in the terms of MPG
2. *The interval is above zero, suggesting 4 is better than 6 in the terms of MPG*
3. The interval does not tell you anything about the hypothesis test; you have to do the test.
4. The interval contains 0 suggesting no difference.

*** .hint Refer back to the problem, consider the implications of the interval being larger than 0, double check the order in which things were subtracted and make sure the results make sense in the context of the problem.

*** .explanation The interval was conducted subtracting 4 - 6 and was entirely above zero.

Suppose that 18 obese subjects were randomized, 9 each, to a new diet pill and a placebo. Subjects' body mass indices (BMIs) were measured at a baseline and again after having received the treatment or placebo for four weeks. The average difference from follow-up to the baseline (followup - baseline) was 3 kg/m2 for the treated group and 1 kg/m2 for the placebo group. The corresponding standard deviations of the differences was 1.5 kg/m2 for the treatment group and 1.8 kg/m2 for the placebo group. The study aims to answer whether the change in BMI over the four week period appear to differ between the treated and placebo groups.

1. What is the pooled variance estimate? (to 2 decimal places)

*** .hint The sample sizes are equal, so the pooled variance is the average of the individual variances

*** .explanation

```
n1 <- n2 <- 9
x1 <- -3  ##treated
x2 <- 1  ##placebo
s1 <- 1.5  ##treated
s2 <- 1.8  ##placebo
spsq <- ( (n1 - 1) * s1^2 + (n2 - 1) * s2^2) / (n1 + n2 - 2)
```

2.75

For Binomial data the maximum likelihood estimate for the probability of a success is

1. *The proportion of successes*
2. The proportion of failures
3. A shrunken version of the proportion of successes
4. A shrunken version of the proportion of failures

*** .hint Look back at the notes about likelihood.

*** .explanation The MLE for binomial data is always the proportion of successes.

Bayesian inference requires

1. A type I error rate
2. Setting your confidence level
3. *Assigning a prior probability distribution*
4. Evaluating frequency error rates

*** .explanation All of the other answers discuss frequentist concepts. All Bayesian analyses requiring setting a prior.