Stat 145, Mon 1-Nov-2021 -- Mon 1-Nov-2021

Monday, November 01st 2021

Due:: WW ch06Part3 due at 11 pm

Monday, November 1st 2021

Wk 10, Mo

Topic:: Inference scenarios

Read:: Lock5 Chapter 6

HW((WW ch06Part4 due Wed.

Example data:

1. Case: summary data is all we know

   Means are for number of beetle larvae per stem in oat crop

Group	n	x-bar	s
Control	13	3.47	1.21
Malathion	14	1.36	0.52

   summary info

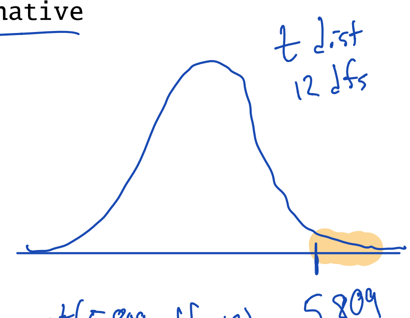
   Friday: constructed a CI for $\mu_C - \mu_M$

   Test hypothesis that mu_C-mu_M = 0 vs. one-sided alternative

   $H_0: \mu_C - \mu_M = 0$ vs. $H_a: \mu_C - \mu_M > 0$

   test stat: $\bar{x}_C - \bar{x}_M = 3.47 - 1.36 = 2.11$

   standardized: $t = \dfrac{2.11 - 0}{\sqrt{\frac{1.21^2}{13} + \frac{0.52^2}{14}}} \approx 5.809$



   P value: 1 - pt(5.809, df=12)

2. CaffeineTaps data

On the following pages, I have included a problem (taken mostly literally) from the Lock 5 text, Chapter 6. There are five, in all, and no clues besides the problem statement and the actual data are available for deciding on a statistical procedure for addressing the question.

1. A recent study compared 298 children with Autism Spectrum Disorder to 1507 randomly selected control children without the disorder. Of the children with autism, 20 of the mothers had used antidepressant drugs during the year before pregnancy or the first trimester of pregnancy. Of the control children, 50 mothers had used antidepressant drugs. Is there a significant association between prenatal expsure to antidepressant medicine and the risk of autism?

$H_0: p_C - p_A = 0$ $\qquad H_a: p_C - p_A \neq 0$

2-proportion, hyp. test

Because we are doing hypothesis testing, we not only need

$$\hat{p}_1 = \frac{20}{298} \quad \text{and} \quad \hat{p}_2 = \frac{50}{1507},$$

but the pooled proportion as well:

$$\hat{p} = \frac{20+50}{298+1507} = \frac{70}{1805}$$

The standardized test statistic is

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}_1 - \hat{p}_2}} = \frac{20/298 - 50/1507}{\sqrt{\left(\frac{70}{1805}\right)\left(\frac{1735}{1805}\right)\left(\frac{1}{298} + \frac{1}{1507}\right)}} \doteq 2.7724$$

We get our P-value from

$$\left(1 - \text{pnorm}(2.7724)\right) \cdot 2 = 0.00556$$

At each of the significance levels $\alpha = 0.1, 0.05, 0.01$, we reject the null hypothesis and conclude the evidence of an association between prenatal exposure to antidepressants and autism is statistically significant.

2. A story spoiler gives away the ending early. Does having a story spoiler diminish the suspense, harming enjoyment by the reader? A study investigated this question in the following way. For twelve different short stories, the researchers created created a second version containing a spoiler paragraph at the beginning that discussed the story and revealed the outcome. Both versions were read and rated on a scale from 1 to 10 (10 being the highest enjoyment rating) by at least 30 people; the overall ratings can be found in the data frame **StorySpoilers**. Is there a difference in mean overall enjoyment rating based on whether or not there is a spoiler?

Story	1	2	3	4	5	6	7	8	9	10	11	12
with spoiler	4.7	5.1	7.9	7.0	7.1	7.2	7.1	7.2	4.8	5.2	4.6	6.7
original	3.8	4.9	7.4	7.1	6.2	6.1	6.7	7.0	4.3	5.0	4.1	6.1

diffs.: 0.9 0.2 0.5 -0.1 0.9

gives $\bar{x}_{diff}$ s_{diff}

Matched-pairs hyp. test

$H_0: \mu_{Diff} = 0$ $H_a: \mu_{Diff} \neq 0$

The twelve differences: 0.9, 0.2, 0.5, -0.1, 0.9, 1.1, 0.4, 0.2, 0.5, 0.2, 0.5, 0.6

have mean $\bar{x}_{diff} = 0.4917$ and std. deviation $s_{diff} = 0.3476$.

The standardized test statistic is

$$t = \frac{\bar{x}_{diff} - 0}{SE} \doteq \frac{0.4917}{0.3476/\sqrt{12}} \doteq 4.900$$

Our P-value is

$$2 \cdot (1 - pt(4.9, df = 11)) \doteq 0.000472.$$

3. Do males or females eat more fiber? Use the **Fiber** variable in the data frame **NutritionStudy** to estimate, at the 92% level, the difference.

2-sample t, CI

(not on Test 2)

From

```
> favstats(Fiber ~ Sex, data = NutritionStudy),
```

we learn

	$\bar{x}$	s	n
Female	12.692	5.403	273
Male	13.414	4.843	42

Our point estimate of $\mu_F - \mu_M$ is $\bar{x}_F - \bar{x}_M = -0.722$

We calculate approximate SE:

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{5.403^2}{273} + \frac{4.843^2}{42}} \doteq 0.8157$$

Our critical value comes from a df = 41 t-distribution:

$$t^* = qt(0.96, df = 41) \doteq 1.795$$

Thus our 92% CI for $\mu_F - \mu_M$ is

$$(\bar{x}_F - \bar{x}_M) \pm (t^*)(SE) = -0.722 \pm (1.795)(0.8157),$$

or

$$(-2.186, 0.742).$$

4. The U.S. Food and Drug Administration has a limit for mercury content in fish of 1.0 ppm. For each lake in the **FloridaLakes** data the `AvgMercury` variable gives the average mercury level for a sample of large-mouth bass from that lake. Does it seems there is evidence that lakes in Florida, on average, have permissible mercury levels in the bass population?

1-sample t, hypothesis test

$H_0: \mu = 1.0 \qquad H_a: \mu > 1.0$

```
> favstats(~ AvgMercury, data = FloridaLakes)
```

tells us $\bar{x} = 0.527$, $s = 0.341$, $n = 53$

Our standardized test statistic is

$$t = \frac{\bar{x} - 1.0}{0.341/\sqrt{53}} = -10.098$$

Since the alternative hypothesis is right-tailed, this number (-10.098) falls on the wrong side of 0 to be statistically significant. Our P-value is > 0.5, so we fail to reject the null hypothesis.

5. The dataset **ICUAdmissions**, includes information on 200 patients admitted to an Intensive Care Unit. One of the variables, `Status`, indicates whether each patient lived or died (1 means the patient lived). Give a 88% confidence interval for the survivor rate.

1-prop. CI

There are 40 survivors in the 200 patients, so $\hat{p} = \frac{40}{200} = 0.2$.

For the purpose of a CI, we approximate the std. error

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{200}} = \sqrt{\frac{(0.2)(0.8)}{200}} \doteq 0.0283$$

And our critical value

$$z^* = \text{qnorm}(0.94) \doteq 1.5548$$

So, our 88% CI is

$$\hat{p} \pm (z^*)(SE) = 0.2 \pm (1.5548)(0.0283)$$

or

$$(0.156, 0.244).$$