Regression and Other Stories: Ch 4.4 - 4.7

When conducting data analysis, we want to avoid too strong of conclusions based on the data. Hypothesis testing and error analysis were developed to quantify this issue.

Statistical Significance

- A common, binary, decision rule that **is not** recommended is based on *statistical significance.*

- Statistical significance is defined as

- Statistical significance decisions for a regression coefficient

- More generally, an estimate is not statistically significant

Hypothesis Tests

The Bozeman school district has identified the proportion of covid tests that are positive as an important metric for reopening or closing schools. While describing the hypothesis test, think about how to construct a test focused on the positive test rate.

- **Estimate:**

- **Null and alternative hypothesis:** choose the null and alternative hypothesis.

- **test statistic:**

- **confidence interval:** T

- **p-value:** describes the deviation of the data from the null, Example: I wasn't able to find reliable data from Gallatin County, but I did get data from Virginia Tech https://ready.vt.edu/dashboard.html.

We will treat the last 7 days of tests as 7 data points

```
tests <- c(181,181,406,326,311,307,260)
positives <- c(32,15,74,58,84,50,46)
positive_rates <- positives / tests
positive_rates
```

```
## [1] 0.17679558 0.08287293 0.18226601 0.17791411 0.27009646 0.16286645 0.17692308
```

- **Estimate:** define parameter of interest, test statistic, standard error, and associated degrees of freedom.

- **Null and alternative hypothesis:** choose the null and alternative hypothesis.

- **confidence interval:** The confidence interval is $\hat{\theta} \pm t_{n-1}^{.975} se(\hat{\theta}) =$

- **test statistic:** the t-score

- **p-value:** describes the deviation of the data from the null, formally the probability of observing something at least as extreme as the observed test statistic.

```
t.test(positive_rates, mu = .1, alternative = 'two.sided')
```

```
##
##  One Sample t-test
##
## data:  positive_rates
## t = 3.6819, df = 6, p-value = 0.01031
## alternative hypothesis: true mean is not equal to 0.1
## 95 percent confidence interval:
##  0.1253835 0.2259693
## sample estimates:
## mean of x
## 0.1756764
```

The interpretation would be that there is evidence to reject the null hypothesis that the true positive rate is .1. Furthermore, a confidence interval for the positive rate is (0.125, 0.226).

Type 1 / Type 2 errors The authors state that they don't like the idea of Type 1 and Type 2 errors.

- **type 1 error:**

- **type 2 error:**

The authors state that the fundamental problem with type 1 and type 2 errors is that in many problems the null hypothesis cannot be true.

Type I and type II errors are based on a deterministic (binary) approach to science that might be appropriate for large effects.

Type M (magnitude) and Type S (sign) errors Both a type M and type S error could occur when making a claim.

A **type S error:**

A **type M error:**

The current publishing incentives, *statistical significance*, can lead to type M errors. In particular the requirement for statistical significance creates a lower bound on the estimated effect size.

The authors don't tend to use NHST in their own work, and neither do I. They state that just about every treatment will have *some effect*, and few regression coefficients will be *exactly zero*.

A major issue with the use of NHST is when researchers seek to confirm a hypothesis (say hypothesis A), by coming up with an alternative hypothesis (hypothesis B). Then if hypothesis B is rejected, this is used as evidence

Problems with Statistical Signficance

First of all, there is a disconnect between significance in the common vernacular and the statistical vernacular.

Moving Beyond Hypothesis Testing

One reason that hypothesis testing is still widely used is that there are not clear, widely accepted alternatives.