Deisgn of Experiments & Hypothesis Testing

Felipe José Bravo Márquez

April 16, 2021

Design of Experiments

- sfsdf

Hypothesis Testing

- When we want to test whether some assumed **property** about a population is contrasted with a statistical sample we use a **hypothesis test**.
- The test consists of the following hypotheses:
  - **Null Hypothesis** H_0: Symbolizes the current situation. What has been considered real up to the present.
  - **Alternative Hypothesis** H_a: it is the alternative model that we want to consider.
- The idea is to find enough **statistical evidence** to reject H_0 and be able to conclude H_a.
- If we do not get enough statistical evidence **we fail to reject** H_0

Hypothesis Testing (2)

Methodology to Perform a Hypothesis Test

- Choose a null hypothesis H_0 and alternative H_a.
- Set a test significance level α.
- Calculate a statistic T from the data.
- The statistic T is usually a standardized value that we can check in a distribution table.
- Define a rejection criterion for the null hypothesis. It is usually a critical value c.

Hypothesis Testing (3)

- Example: It is known that the average number of hours of monthly Internet use in Chile is 30 hours.
- Suppose we want to show that the average is different from that value.
- We would have that $H_0 : \mu = 30$ and $H_a : \mu \neq 30$
- Let's set $\alpha = 0.05$ and collect 100 observations.
- Suppose we get $\overline{X_n} = 28$ and $s = 10$
- One way to test is to construct a confidence interval for μ and see if H_0 is in the interval.

```
> 28-qt(p=0.975,99)*10/sqrt(100)
[1] 26.01578
> 28+qt(p=0.975,99)*10/sqrt(100)
[1] 29.98422
```

- The interval would be the acceptance zone of H_0 and anything outside of this would be my rejection region.
- Since 30 is in the rejection region, I reject my null hypothesis with 5% confidence.

Hypothesis Testing (4)

- Another way to perform the test is to compute the statistic $T = \frac{\overline{X_n} - \mu_o}{\frac{s}{\sqrt{n}}}$
- In this case it would be

$$T = \frac{28 - 30}{\frac{10}{\sqrt{100}}} = -2$$

- Since $H_a : \mu \neq 30$, we have a two-sided test, where the acceptance region is.

$$t_{n-1,1-\alpha/2} < T < t_{n-1,\alpha/2}$$

```
 > qt(0.025,99)
[1] -1.984217
> qt(0.975,99)
[1] 1.984217
```

- Since T is in the rejection region, we reject the null hypothesis.

Hypothesis Testing (5)

- Generally, in addition to knowing whether we reject or fail to reject a null hypothesis we want to quantify the evidence we have against it.
- A **p-value** is defined as the probability of obtaining an outcome at least as extreme as that observed in the data given that the null hypothesis is true.
- “Extreme” means far from the null hypothesis and favorable for the alternative hypothesis.
- If the **p-value** s less than the significance level α, we reject H_0
- Example:

```
> data(iris)
> mu<-3 # null hypothesis
> alpha<-0.05
> n<-length(iris$Petal.Length)
> xbar<-mean(iris$Petal.Length)
> s<-sd(iris$Petal.Length)
> se<-s/sqrt(n)
> t<-(xbar-mu)/(s/sqrt(n))
> pvalue<-2*pt(-abs(t),df=n-1)
> pvalue
[1] 4.94568e-07 # is less than 0.05 then we reject H0
```

Hypothesis Testing (6)

- The elegant way to do it in R:

```
> t.test(x=iris$Petal.Length,mu=3)

One Sample t-test

data:  iris$Petal.Length
t = 5.2589, df = 149, p-value = 4.946e-07
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 3.473185 4.042815
sample estimates:
mean of x
    3.758
```

Hypothesis Testing (7)

- We have two types of errors when we perform a hypothesis test
- Type I error: it is when we reject the null hypothesis when it is true.
- This error is equivalent to the significance level α.
- Type II error: is when the null hypothesis is false but we do not have statistical evidence to reject it.
- To mitigate type I errors we generally use smaller values of α.
- To mitigate type II errors we generally work with larger samples.
- There is a trade-off between type I and type II errors.

	Retain H_0	Reject H_0
H_0 true	✓	type I
H_1 true	type II error	✓

Statistical Power

Critics to Hypothesis Testing

FOUR CARDINAL RULES OF STATISTICS by Daniela Witten

- ONE: CORRELATION DOES NOT IMPLY CAUSATION. Yes, I know you know this, but it's so easy to forget! Yeah, YOU OVER THERE, you with the p-value of 0.0000001 — yes, YOU!! That's not causation.
- No matter how small the p-value for a regression of IQ onto shoe size is, that doesn't mean that big feet cause smarts!! It just means that grown-ups tend to have bigger feet and higher IQs than kids.
- So, unless you can design your study to uncover causation (very hard to do in most practical settings — the field of causal inference is devoted to understanding the settings in which it is possible), the best you can do is to discover correlations. Sad but true.
- TWO: A P-VALUE IS JUST A TEST OF SAMPLE SIZE. Read that again — I mean what I said! If your null hypothesis doesn't hold (and null hypotheses never hold IRL) then the larger your sample size, the smaller your p-value will tend to be.
- If you're testing whether mean=0 and actually the truth is that mean=0.000000001, and if you have a large enough sample size, then YOU WILL GET A TINY P-VALUE.
- Why does this matter? In many contemporary settings (think: the internet), sample sizes are so huge that we can get TINY p-values even when the deviation from the null hypothesis is negligible. In other words, we can have STATISTICAL significance w/o PRACTICAL significance.

FOUR CARDINAL RULES OF STATISTICS by Daniela Witten

- Often, people focus on that tiny p-value, and the fact that the effect is of **literally no practical relevance** is totally lost.
- This also means that with a large enough sample size we can reject basically ANY null hypothesis (since the null hypothesis never exactly holds IRL, but it might be “close enough” that the violation of the null hypothesis is not important).
- Want to write a paper saying Lucky Charms consumption is correlated w/blood type? W/a large enough sample size, you can get a small p-value. (Provided there’s some super convoluted mechanism with some teeny effect size. . . which there probably is, b/c IRL null never holds)
- THREE: SEEK AND YOU SHALL FIND. If you look at your data for long enough, you will find something interesting, even if only by chance! In principle, we know that we need to perform a correction for multiple testing if we conduct a bunch of tests.
- But in practice, what if we decide what test(s) to conduct AFTER we look at data? Our p-value will be misleadingly small because we peeked at the data. Pre-specifying our analysis plan in advance keeps us honest. . . but in reality, it’s hard to do!!!
- Everyone is asking me about the mysterious and much-anticipated fourth rule of statistics. The answer is simple: we haven’t figured it out yet.... that’s the reason we need to do research in statistics

References I