Simulation 1

-Quantifying the World-

Lecturer: Darren Homrighausen, PhD

Simulation

Simulation is a crucial part of data science

It can be used to...

- Estimate sampling distributions for intractable estimators (e.g. the bootstrap)
- Compute p-values when the reference distribution for a hypothesis test is unknown
- Find the power/size of a give test
- Develop intuition about or compare methods

Simulation

Suppose we use `R` to find 20 random standard normal draws

Simulation

Suppose we use R to find 20 random standard normal draws

```
x = rnorm(20)
```

How can I change x to have mean 20 and variance 3?

Simulation

Suppose we use R to find 20 random standard normal draws

```
x = rnorm(20)
```

How can I change x to have mean 20 and variance 3?

```
x = x *sqrt(3) + 20
```

Size, Power, Type I and II error

Suppose we have a hypothesis test about a parameter θ

$$H_0 : \theta = \theta_0$$
$$H_A : \theta \neq \theta_0$$

(think of θ_0 as some number like $\theta_0 = 0$)

If we have a hypothesis test based around a test statistic T, we can compute a p-value as

$$P(|T| > t|\text{the null hypothesis } H_0 \text{ is true})$$

where t is some threshold

Size, Power, Type I and II error

A classic example is the T-test

$$T = \frac{\overline{Y}}{\mathrm{SE}(\overline{Y})}$$

If α is the significance level, then the standard threshold is

Size, Power, Type I and II error

A classic example is the T-test

$$T = \frac{\overline{Y}}{\mathrm{SE}(\overline{Y})}$$

If α is the significance level, then the standard threshold is

$$t = t_{\alpha/2,n-1}$$

(The $\alpha/2$ quantile of the t-distribution with $n-1$ degrees of freedom)

Size, Power, Type I and II error

- Size: This is the probability we do not reject the null hypothesis given that it is true

$$\text{size} = P(\text{do not reject } H_0 | H_0 \text{ is true})$$

- Power: This is the probability we reject the null hypothesis given that it is false

$$\text{power} = P(\text{reject } H_0 | H_0 \text{ is false})$$

- Type I Error:

$$1 - \text{size}$$

- Type II Error:

$$1 - \text{power}$$

Size, Power, Type I and II error

For any hypothesis test, we can define these (theoretical) quantities of interest

They can in some cases be exactly computed using probability theory

(Or approximately exactly computed using asymptotic theory)

However, in many cases it is impossible/difficult

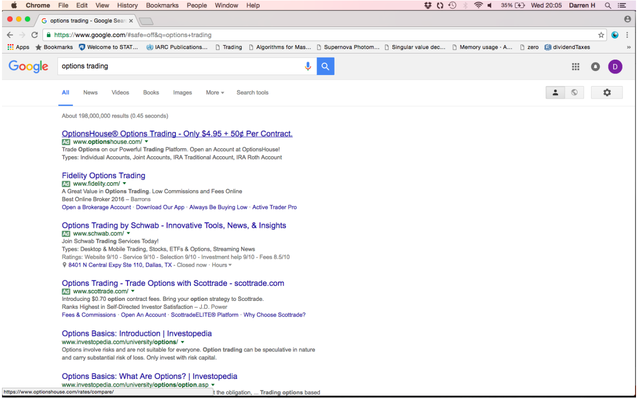
Hence, we tend to use simulation to compute them

Internet advertising

Terminology

When discussing advertising on the internet, there are three very important concepts

- Impressions: A person seeing the page with an ad
- Clicks: A person clicking on that ad
- Conversion: A person doing a particular thing on the site


Chrome File Edit View History Bookmarks People Window Help
35% Wed 20:05 Darren H
options trading - Google Sear
https://www.google.com/#safe=off&q=options+trading
Apps Bookmarks Welcome to STAT... IARC Publications... Trading Algorithms for Mas... Supernova Photom... Singular value dec... Memory usage · A... zero dividendTaxes
Google
D
All News Videos Books Images More Search tools

options trading

About 198,000,000 results (0.45 seconds)

OptionsHouse® Options Trading - Only $4.95 + 50¢ Per Contract.
Ad www.optionshouse.com/
Trade **Options** on our Powerful **Trading** Platform. Open an Account at OptionsHouse!
Types: Individual Accounts, Joint Accounts, IRA Traditional Account, IRA Roth Account

Fidelity Options Trading
Ad www.fidelity.com/
A Great Value in **Options Trading**. Low Commissions and Fees Online
Best Online Broker 2016 – Barrons
Open a Brokerage Account · Download Our App · Always Be Buying Low · Active Trader Pro

Options Trading by Schwab - Innovative Tools, News, & Insights
Ad www.schwab.com/
Join Schwab **Trading** Services Today!
Types: Desktop & Mobile Trading, Stocks, ETFs & Options, Streaming News
Ratings: Website 9/10 - Service 9/10 - Selection 9/10 - Investment help 9/10 - Fees 8.5/10
8401 N Central Expy Ste 110, Dallas, TX - Closed now · Hours

Options Trading - Trade Options with Scottrade - scottrade.com
Ad www.scottrade.com/
Introducing $0.70 **option** contract fees. Bring your **option** strategy to Scottrade.
Ranks Highest in Self-Directed Investor Satisfaction – J.D. Power
Fees & Commissions · Open An Account · ScottradeELITE® Platform · Why Choose Scottrade?

Options Basics: Introduction | Investopedia
www.investopedia.com/university/**options**/
Options involve risks and are not suitable for everyone. **Option trading** can be speculative in nature and carry substantial risk of loss. Only invest with risk capital.

Options Basics: What Are Options? | Investopedia
www.investopedia.com/university/**options**/**option**.asp
https://www.optionshouse.com/rates/compare/
t the obligation, ... **Trading options** based

Advertising considerations

In the particular case of `Adwords` companies can pay for ads in three ways

- By the click
- By the impression
- By the conversion

Companies can change how the ad it written, the keywords, the amount they are willing to pay, ...

Experimental design

Companies are continually trying to tweak their ads so that more people are converted from impressions to clicks

(The rate at which this happens is called the click-thru rate)

This is commonly referred to as A/B testing, but we would know it as hypothesis testing

Let's consider a consulting problem that is really quite common

Consulting problem

A new company wants to advertise on Google and has decided to pay by the impression

This company has come up with two different keyword sets and wants to see which one leads to the highest click-thru rate

The company wants to go with the better ad as quickly as possible

It hired our company to answer the following basic question:

> How many impressions do we need to pay for before we can conclude which ad is better?

Contingency Tables

Analyzing discrete data

A contingency table is a cross-tabulation of discrete variables

In this particular problem, we our table would be

Ad ‘A’	clicks	no clicks	impressions
Ad ‘B’	clicks	no clicks	impressions

Example: We pay for 1000 impressions for each ad and get the following

	clicks	no clicks	impressions
Ad ‘A’	8	992	1000
Ad ‘B’	5	995	1000

Can we answer which ad is better?

Contingency Tables

There are two main ways to answer this question

Both approaches compare the observed counts to the counts expected if there is no difference in the two ads

They differ in the approach to computing a p-value

- Chi-squared test: Uses an asymptotic normal approximation
- Fisher's exact test: Uses an exact hypergeometric distribution

```
contingencyTable = matrix(c(8,1000-8,5,1000-5),nrow=2,byrow=T)
> contingencyTable
 8  992
 5  995
> chisq.test(contingencyTable)$p.value
[1] 0.577861
> fisher.test(contingencyTable)$p.value
[1] 0.5797925
```