Statistical Methods for Insurance: Randomisation

Di Cook & Souhaib Ben Taieb, Econometrics and Business Statistics, Monash University

W1.C1

Overview of the class

- Topics
- Assessment
- Resources
- Instructors, tutors

Topics

- Topic 1: Simulation of games for decision strategies (2 weeks))
- Topic 2: Statistical distributions for decision theory (1.5 weeks)
- Topic 3: Linear models for credibility theory (1.5 weeks)
- Topic 4: Compiling data to problem solve (2 weeks)
- Topic 5: Bayesian statistical thinking (1.5 weeks)
- Topic 6: Temporal data and time series models
- Topic 7: Modeling risk and loss, with data and using randomization to assess uncertainty (2 weeks)

Assessment

- Final exam: 70%
- Tutorials/labs: 25%, Weekly reports due Monday noon after the lab
- Class exercises: 5%
- ETC5242 students: Labs 15%, Project report and presentation 10%

Resources

- Web site: `http://dicook.github.io/Statistical_Thinking/`
- Moodle
- Statistics online textbook
- Accuarial online curriculum/exam material
- Software: R, RStudio Desktop

Instructors

- Instructors:
  - Professor Di Cook
  - Dr Souhaib Ben Taieb
- Tutors:
  - Earo Wang (working with Di on PhD)
  - Nathaniel Tomasetti (worked with Di for Honors, working with Dr Catherine Forbes on PhD)

What is randomness?

- Coin flip
- Die roll
- Your sporting team wins
- Gender of a baby
- Rain tomorrow
- Stock price in an hour from now
- Lightning strike
- Pipe burst

Your turn

We are going to play a game of "Stump the Professor".

1. Flip a coin. If it shows up tails do A first, if it shows up heads to B first.

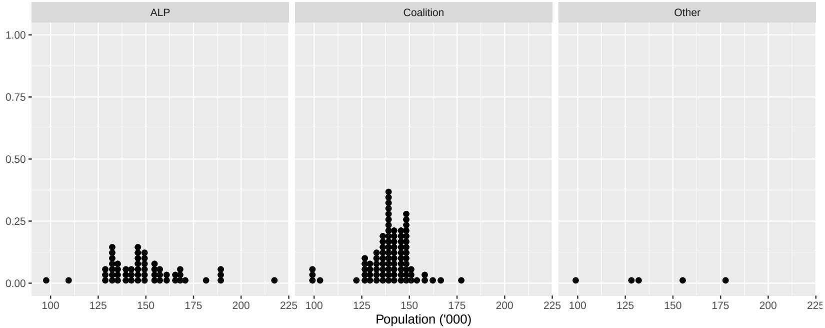
```
A. Write down a sequence of heads and tails that you
might expect to come from TEN flips of a coin
B. Now flip a coin TEN times, and write down the
outcomes
```

2. Enter these in the online sheet at `http://bit.ly/ETC2420-1` (Remember whether you entered the coin flip sequence first or the made up sequence.)
3. Now I am going to look at what you entered, and guess if sequence was made up, or actual outcomes from coin flips.
4. You record how many times I get it right.

Example: a look at the Australian electoral distribution

- Results of 2013 election from Australian Electoral Commission web site
- 2011 Census data from the Australian Bureau of Statistics
- Combined demographics of electorate with political representation
- Interactive application, in R package `eechidna`

How to use randomization to understand probability



Your turn

- What is the difference (roughly) in population between the biggest and smallest electorates?
- What is the relative worth of a voter in the electorate with the largest population, compared to a voter in the electorate with the smallest population?

Politics

- Ideally all electorates have exactly the same number of people.
- Geography can interfere with this, e.g an electorate cannot be part in Tasmania and part in Victoria.
- The Australian Electoral Commission will adjust geographic boundaries before each election to adjust for population changes as measured in the most recent Census.

Compute averages

```
#> # A tibble: 2 x 3
#>     PartyGp      m     s
#>       <chr>  <dbl> <dbl>
#> 1       ALP 149245 20167
#> 2 Coalition 139534 12476
```

Statistical thinking

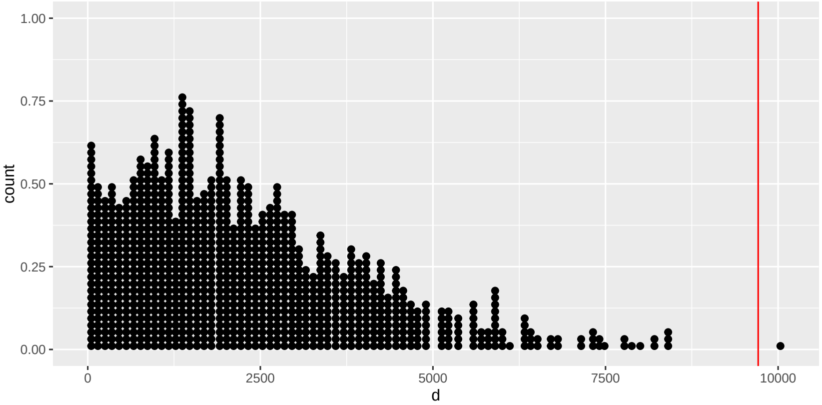
- The means are different
- How big is this difference?
- How likely is this difference to have arisen by chance?

We could use a two-sample t-test to answer these, but here is how to do the equivalent by randomization.

Procedure

1. Compute the statistic for the data (e.g. absolute value of mean difference)
2. Shuffle the group labels (e.g. put the MP party names into a hat, mix them around, draw them and assign to new electorate)
3. Compute the statistic for this shuffled data
4. Repeat steps 2, 3 many times
5. Examine how often the value of the data statistic, or a larger value occurs

Let's do it



Let's also count the number of times that we see a bigger difference by chance. It is 1.

What does this mean?

If we oberve a difference this large 1 out of 1000 random shuffles, is it likely to see this electorate distribution by chance?

Caveats

Let's wait until the next Census results are in (after August this year) and the latest election results, to compare populations of electorates again.

Next steps

- Next few classes we will
  - review statistical hypothesis testing
  - go through a few more examples of using permutation to make statistical significance tests
  - discuss random number generators, and simulating events
  - compute probabilities of events, and
- Lab this week will be getting started with using R - the top data science software available today.

Share and share alike