GV300 Quantitative Political Analysis

Week 7
Hypothesis testing

Dominik Duell (University of Essex)

November 13, 2019

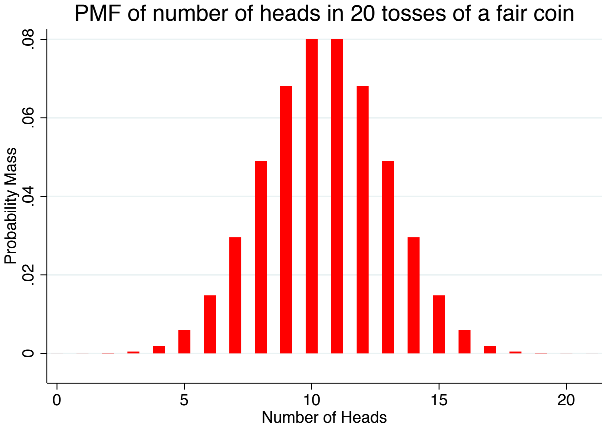
Hypothesis testing and probability

Consider the regression output below:

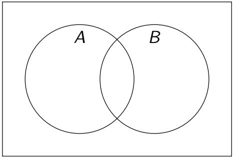
```
. reg voteLabour income;

      Source |       SS       df       MS              Number of obs =   13500
-------------+------------------------------           F(  1, 13498) = 2214.01
       Model |   475.32687     1   475.32687           Prob > F      =  0.0000
    Residual |   2897.8935 13498  .214690584           R-squared     =  0.1409
-------------+------------------------------           Adj R-squared =  0.1408
       Total |  3373.22037 13499  .249886686           Root MSE      =  .46335

------------------------------------------------------------------------------
  voteLabour |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |  -.0092478   .0001965   -47.05   0.000     -.009633   -.0088625
       _cons |   .9457103    .010503    90.04   0.000      .925123    .9662976
------------------------------------------------------------------------------
```

It features the result of two hypothesis tests? Where are these results? Which hypothesis is tested precisely?

- Let's start again with a by now well known example:
  - Say we flip a fair coin 20 times and we are interested in the number of heads– that's random variable H
  - What is $E[H] = 10$
  - What is the PMF?

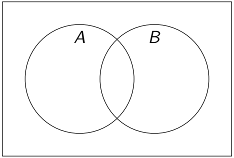

PMF of number of heads in 20 tosses of a fair coin
Probability Mass
0
.02
.04
.06
.08
0
5
10
15
20
Number of Heads

- a p-value is an expression of a conditional probability
- we learn: **assume that this is the distribution of our outcome of interest, what is then the probability of an outcome as extreme as our hypothesized outcome X?**



- In this space, each point is a pair p_H, h_o
  - Event A: true distribution is p_H
  - Event B: we observe h_o

- Hypothesis test considers $p(B|A)$
- Not the same as $P(A|B)$, needs far more information
- But we actually want to know $P(A|B)$, we want to know the probability that true distribution is p_H given that we observe a particular outcome h_0



- Consider:
  - Event B: 15 or more heads
  - Event A: the coin is fair – same PMF as before
- We do not know $P(A)$ and $P(B)$ but we can calculate

$$P(B|A)$$

- If we observe more than 15 heads, the probability of this occurring given a fair coin is about 2.8%
- Usually we fix level of significance, α and ask: what is the largest value of h_o that occurs with a probability less than α

What did we do here?

- Call event A the hypothesis: the coin is fair
- Observing B, we reject this hypothesis – observing 15 heads out of 20 coins is just too extreme of an outcome to could have come from a fair coin.
- **This is not a statement about something being true or false!**

Here is another example:

income

	clinton	trump	other/no answer
under $30,000 **17%**	**53%**	41%	6%
$30k-$49,999 **19%**	**51%**	42%	7%
$50k-$99,999 **31%**	46%	**50%**	4%
$100k-$199,999 **24%**	47%	**48%**	5%
$200k-$249,999 **4%**	48%	**49%**	3%
$250,000 or more **6%**	46%	**48%**	6%

24537 respondents

- Call event A the hypothesis: poor voters are more likely to vote for Clinton than Trump
- The data used here gives you: observing event B (that many poor voters vote for Clinton), can we rejecting the hypothesis? – maybe but whats the distribution of the test statistic?
- **Again, this is not a statement about something being true or false!**

Summary of logic of hypothesis testing:

- Generally, comparison of the *actual* statistic of interest computed from our sample and what we would **expect** the statistic to look like computed from the population
  - A specific example: comparison of the actual relationship between X and Y in our sample with the relationship between X and Y in the underlying population
- We judge the probability by which we think we found a relationship in our sample that is close to what we would expect to find in the population by the **p-value**:
  - probability that the statistic computed from our sample is arising by chance

Test distribution

Know your distribution

- We need the population distribution of our variable or the distribution of our test statistic to be able to assess our hypothesis, how do we get there?
- Combinations, recall coin flipping example above – often complicated
- Easy when population normally distributed – not the case in many applications
- When non-normal population, **Central Limit Theorem** helps many times and we are back to the well understood normal distribution

Law of large numbers

- Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables with mean μ and standard deviation σ
- If we estimate the population mean μ with the sample mean $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ ...
- with a sufficiently large sample ...
  - we can get arbitrarily close to μ or
  - $plim(\overline{X}_n) = \mu$
    which is another way of saying: $\lim_{n \to \infty} P(|\overline{X}_n - \mu|) < \epsilon) = 1$ for any small ϵ
- In other words, the **law of large numbers** states that the **average** of realized values from a large number of experiments (samples) is close to the expected value of the underlying population

Law of large numbers

- We can restate the law of large numbers for any arbitrary statistic, that is any arbitrary function $f(x)$, we may be interested in:

$$plim \frac{1}{n} \sum_{i=1}^{n} f(x_i) = E[f(x)]$$

Central limit theorem

- Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with mean μ and standard deviation σ and $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Then,

$$Z_n = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}$$

  has an **asymptotic** standard normal distribution with mean 0 and standard deviation 1
- where asymptotic means that the distribution of Z_n approximates the standard normal distribution very, very, very closely

Central limit theorem

- We refer to Z_n as the standardized version of $\overline{X_n}$
- In other words, the **central limit theorem** states that the **sampling distribution** of the mean of any iid random variable approximates the normal distribution with increasing sample size
- Regardless of the population distribution of Xs, $Z_n \sim N(0,1)$
- We could have looked at the non-standardize version $\overline{X_n}$, that is not divide by σ or

$$L_n = \sqrt{n}\overline{X}_n - \mu$$

where $L_n \sim N(0,\sigma)$ but we will mostly work with Z_n

Law of large numbers and central limit theorem

- Law of large numbers:
  - the average of many measurements is more accurate than a single measurement
  - As n grows large, the probability that $\overline{X}_n$ is close to μ is 1
- Central limit theorem:
  - As n grows, the distribution of Z_n converges to the normal distribution with $N(0, \sigma^2)$
  - CTL implies approximation that becomes better with growing n

Important test distributions

Normal distribution

- denote normal distribution with mean μ and variance σ^2, $N(\mu, \sigma^2)$
- continuous distribution
- density at point x:

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

What is the normal distribution good for?

- standard/z-score:
  - to convert to normally distributed scores
  - $z = \frac{x - \mu}{\sigma}$
  - where μ is the population mean
  - note, based on assumption about population μ and σ
  - usual rule of thumb, when $n > 30$, sample standard deviation s approximates σ

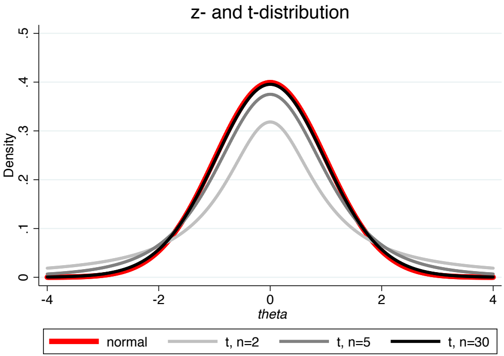
If we know population dispersion!

t-distribution

Should we do not know the population dispersion

- t-score:
  - $t = \sqrt{n}\frac{\overline{x}-\mu}{s} = \frac{\overline{x}-\mu}{s_{\overline{x}}}$
  - where μ is the population mean, $\overline{x}$ the sample mean, s the sample standard deviation, and n the sample size
  - based on assumption about population μ only

Normal and t-distribution



What else is great about the normal distribution?

- You can scale it, or if $X \sim N(\mu, \sigma^2)$ then $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- You can combine random variables, or if X and Y are independent, with $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, then $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
- You can work with squared objects: If $X_1, \ldots, X_n$ are iid random variables which are distributed normally, then $\sum_{i=1}^{n} X_i^2$ is also distributed normal!
- Also holds true for ratios, proportions of random variables

What else is great about the normal distribution?

Also, if $X \sim N(\mu, \sigma^2)$

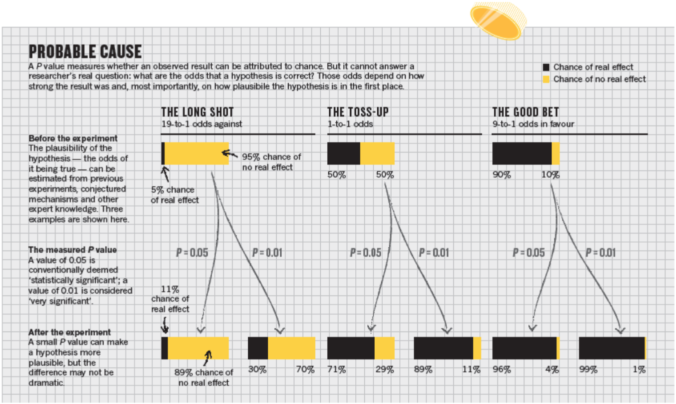
- with 68% probability, X lies between $\mu - \sigma$ and $\mu + \sigma$
- with 95% probability, X lies between $\mu - 2\sigma$ and $\mu + 2\sigma$
- with 99% probability, X lies between $\mu - 3\sigma$ and $\mu + 3\sigma$

Test procedure and terminology

Summary of standard procedure

- Generate a meaningful hypothesis
- Find a valid test statistic
- Derive the distribution of the test-statistic:
  - Based on theory
  - Exact
  - Simulated
- From distribution obtain/make:
  - Critical value
  - p-value
  - Rejection decision

A cautionary note on p-values



Source: Nuzzo (2015), p.2

A cautionary note on p-values

- p-values do not say anything about **size** of effect associated with the test statistic
- Assumption of **random** sample from the population is crucial
- p-value of .001 does not say "an effect occurs with probability .999"
- we need to know the prior odds of an effect
- the more implausible the original hypothesis, the higher the probability of a type I error – independent of p-value
- **When we learn whether something did not happen by chance does not mean we learned anything bout why something happened! No causal effect established!**

p-values and statistical significance

- The statement that a statistic (e.g., describing a relationship between variables X and Y) is **statistical significant** is arbitrary: I rests on ...
  - the researchers statement of the null hypothesis (a theoretical construct)
  - the chosen level of significance that defines the statistic's critical value

More terminology

- **Null hypothesis** vs **alternative hypothesis**
- Type 1 error:
    - Reject null even though it is true
    - **Level of significance** of a test α is probability of a Type 1 error
- Type 2 error:
    - Failing to reject null even though it is false
    - β probability of a Type 2 error
    - **Power of a Test**: $1 - \beta$