GV300 - Quantitative Political Analysis

University of Essex - Department of Government

Lorenzo Crippa

Week 8 – 18 November, 2019

Communication

This week: Office hour on Thursday, from 11 to 13

Office 5B.153

l.crippa@essex.ac.uk

R and Stata session

Using your preferred statistical software, do the following exercise of descriptive statistics:

1. Download data on sales of Monet paintings from Moodle and import them
2. What kinds of variables do you have? What's the best way to represent them?
3. Use plots and descriptive statistics to synthesize your data.
   - You can use histograms, densities, boxplots as appropriate
   - You should show where the mean and median are
4. Save plots you generated in a specific subfolder

R and Stata session

Now imagine you own a beautiful painting from Monet and want to sell it because you are in desperate need to raise money to finance your postgraduate studies. Luckily, you also have Greene's data and statistical skills.

1. You expect to get 4 million dollars from the sale of your painting. Can you support this hypothesis with your data?
2. You are considering whether to sell your painting in house number one or number two, but the rumour is that on average the two houses sell paintings at the same price. Is this a safe theory based on your evidence?
3. Your painting is signed by no less than Mr. Monet in person. Can you support the theory that signed paintings, on average, are worth more than the average price for a non-signed one?

Using your preferred statistical software, do the following:

1. Draw 30 observations from each of 50 variables normally distributed with mean 0 and standard deviation 1 and obtain the 50 averages (one per each variable)
2. Plot the distribution of the averages. Choose the appropriate way to plot them
3. Program a more general function that draws n random observations from each of X variables normally distributed with mean μ and standard deviation σ and returns a vector with the X means computed

Non-parametric tests

- There are alternatives to the t-test
- Choose your test based on your problem, and not the other way around
- Do not stick to the t-test if your problem requires you to do otherwise

Fisher's exact test – Intuition

When do we use it?

1. When we have two independent samples arranged in a contingency table
2. When in each sample we measure the frequency of success *vs* failure (Bernoulli)
3. When expected frequency of any success or failure is below 5. Otherwise use χ^2 test

Fisher's exact test – Three cases

1. Row totals are fixed, column totals are random (or the other way around)
    - The hypothesis tested is $p_1 = p_2$
2. Both row totals and column totals are random
    - The hypothesis tested is bivariate independence
3. Both row totals and column totals are fixed
    - The hypothesis tested is independence

1. indicate as O_{1S} the *S*uccessful outcome you observe from sample 1 (similarly you have O_{1F}, O_{2S}, O_{2F}).
2. Indicate as $n_{1\cdot}$ ($n_{2\cdot}$) the independent repeated Bernoulli trials from sample 1 (2), each one with success probability p_1 (p_2)

We will have:

	Successes	Failures	Totals
Sample 1	O_{1S}	O_{1F}	$n_{1\cdot}$
Sample 2	O_{2S}	O_{2F}	$n_{2\cdot}$
Totals	$n_{\cdot S}$	$n_{\cdot F}$	n

$$H_0 : p_1 = p_2 = p$$
$$H_1 : p_1 \neq p_2$$

Fisher's exact test – Procedure

- In effect the test asks "what is the probability of having a table as extreme as the one we observe, if the null hypothesis is true"?
- Hypergeometric distribution: $Prob(O_{1S} = x | n_{1.}, n_{2.}, n_{.S}, n_{.F}) = \frac{n_{1.}! n_{2.}! n_{.S}! n_{.F}!}{n! x! O_{1F}! O_{2S}! O_{2F}!}$
- Fisher's exact test rejects $H_0 : p_1 = p_2$ if your observed $O_{1S} \geq q_\alpha$
- Where q_α is chosen from the conditional distribution described above so that $Prob(O_{1S} = x | n_{1.}, n_{2.}, n_{.S}, n_{.F}) = \alpha$ where α is our desired level of significance

Fisher's exact test – Examples

Suppose you have:

	Successes	Failures	Totals
Sample 1	5	4	9
Sample 2	4	2	6
Totals	9	6	15

- Assume totals are fixed
- $H_0 : p_1 > p_2$
- What are the probabilities of all tables that would give us a value as large as or larger than the observed value of $O_{1S} = 5$?

Fisher's exact test – Example

Sample 1	3	6	4	5	5	4	6	3	7	2	8	1	9	0
Sample 2	6	0	5	1	4	2	3	3	2	4	1	5	0	6
p	.017		.151		.378		.336		.108		.011		.000	

- $H_0 : p_1 > p_2$
- What are the probabilities of all tables that would give us a value as large as or larger than the observed value of $O_{1S} = 5$?
- It will be $.378 + .336 + .108 + .011 + .000 = 0.832$
- $H_0 : p_1 < p_2$? $p = .017 + .151 + .378 = .545$

Fisher's exact test – Example

If what we observe were:

	Successes	Failures	Totals
Sample 1	8	1	9
Sample 2	1	5	6
Totals	9	6	15

The probability of having a table as extreme (or more) as the one you observe would be: $p = 0.0107 + 0.0002 = 0.0109$!

Conclusion

All clear? Questions?

Thanks and see you next week!