Quiz 1 Materials

The quiz will be 45mins pencil-and-paper, no computers. I will do the questions in a way that you don't really need one. The topics will be a) linear algebra, b) statistics, c) linear regression, d) causality, e) doing all this on computer

The more detailed topics are:

a) linear algebra. Topics: matrix multiplication, basic understanding of inverses, and singularity. Examples:

- multiply matrices $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$
- is the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ singular? Tell this based on either determinants/eigenvalues/whatever other singularity conditions you can come up with.

b) statistics: Bayes theorem, distributions, probability mass function, probability density function, cumulative distribution function. Distributions: Bernoulli, Poisson, Binomial, Normal, Uniform. Expectation, variance.

Examples:

- There were 200 1st class passengers of 2000 passengers on Titanic. 140 of these survived, out of 700 survivors in total. Given a person was saved, what is the probability that she traveled in the 1st class? (Numbers just made up.)
- We have a discrete distribution:

X	Probability
0	0.5
1	0.25
2	0.25

Sketch the pmf and cdf of this distribution. Compute it's expectation, variance.

- You are analyzing server failures in a large server park. What kind of distribution you may choose to describe the number of servers failing each day?

  Answer: Poisson. We did not have time to talk about it in class. Check out OS 2.5.2, p 156.

c) linear regression: dummy variables, interactions

Examples:

- We estimate a linear regression model explaining personal income by gender and location (rural, town, large city):

$$y_i + \alpha_0 + \alpha_1 \cdot \text{sex}_i + \alpha_2 \cdot \text{location}_i + \alpha_3 \cdot \text{location}_i \cdot \text{sex}_i + \epsilon_i$$

The results are (just made up)

parameter	value	t-value	signif.
Intercept	4.7	11.33	***
sex = male	0.2	4.14	***
location = town	0.1	1.3	
location = city	0.3	2.8	**
male · town	0.2	1.0	
male · city	0.1	1.8	*
$R^2 = 0.22$			

What is the (predicted) income for women in towns? For men in big cities? Can we say that men always earn more than women? Is the income in towns significantly (i.e. significant in the statistical sense) larger than in the rural areas?

d) causality: differences-in-differences, 3 different ways to explain correlation as causality.

Examples:

- Consider data

Treatment	Time	Outcome
0	0	10
0	1	20
1	0	33
1	1	33

Find dif-in-dif estimator of the treatment.

Answer: $(33 - 33) - (20 - 10) = -10$.

- We analyze schooling outcome of refugees and run following linear regression:

$$y_i = \beta_0 + \beta_1 R_i + \boldsymbol{\beta}_3' \boldsymbol{x}_i + \epsilon_i$$

where y is schooling outcome (test score), R is refugee status (0 or 1), and $\boldsymbol{x}$ are other covariates. We find $\beta_1 = -11.3$ and it's standard error $\text{sd}_{\beta 1} = 2.2$.

(a) is this estimate statistically significant?
Hint: compute t value

(b) can we conclude that fleeing conflict lowers your test scores?
No: the regression does not establish causality.

(c) can we conclude that refugees are performing worse?
Yes: this is what regression does: compares refugees with non-refugees.

(d) can you provide the 3 different explanations how causality may go here? Try to tell convincing stories. a) (refugee → lower test score): fleeing lowers the score as you can't attend school for a while, maybe you also suffer from trauma; b) (lower test score → refugee): less-skilled people flee, more skilled stay put (perhaps they have more resources); c) (something else causes both): conflicts tend to occur in regions with lagging-behind education, hence refugee children struggle at school.

e) Do this on computer. Use R or python (or something else if you wish and can). Create matrices, manipulate matrices. This involves writing a tiny amount of code on paper. Small mistakes are not a problem but you should get the basics right! In particular, ensure that you do matrix multiplication, not elementwise-multiplication if that's what's required.

Examples:

- create matrix $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$

  In R:

  ```
  matrix(1:4, 2, 2, byrow=TRUE)
  ```

  python:

  ```
  np.array([[1, 2], [3, 4]])
  ```

- multiply this matrix with $\begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix}$
- Assume `A` is a matrix. Transpose it.
- Compute eigenvalues of matrix `B`