Notes for Tutors: Probability, Data, Variables & Moments

In-Class Quiz

In the Canvas folder you will find an accompanying set of questions for the in-class quiz for this Class. These include the required statistical tables at the end. The instructions for the students are at the top of page 1. You should distribute the questions in advance of the Class and give the students enough time to prepare. In the class itself you should administer the quiz closed-book. The questions are not hard and don't require much in the way of derivations or calculations and the students should be able to get through them quickly.

For reference, the answers to the handful of numerical questions are

19. (a) about 0.68, (b) about 0.95, (c) almost 1 (0.997)
20. (a) 0, (b) 0.1587, (c) 0.7699, (d) 0, (e) -0.8416, (f) 1.6445 (g) 1.34
21. (a) 0, (b) 0.6306, (c) 0.2518, (d) 2, (e) -4.5249, (f) -6.9346 and 2.9346 (g) -6.0223 and 2.0223
23. (a) 0.0461, (b) 2.0449, (c) 1
25. (a) $\frac{3}{24}, \frac{8}{24}, \frac{8}{24}, \frac{5}{24}$, (b) $\frac{6}{24}\frac{10}{24}\frac{8}{24}$, (c) $\frac{2}{8}, \frac{4}{8}, \frac{2}{8}$ (d) $\frac{1}{8}, \frac{4}{8}, \frac{6}{8}, 1$ (e) No

You may want to only ask a subset of questions - use your judgement.

Much of the rest of the tutorial should be spend addressing any errors uncovered in the quiz and responding to any questions which arise.

Discussion Points

The following are some potential discussion points you may go over with the students.

Probability

- The frequentist approach to probability.
- The meaning and uses of conditional probability.
- The meaning and uses of independence.
- Bayes Rule and the idea of learning from observation.
- The meaning of terms like "probability" and "likelihood".

Data, Variables & Moments

- Variability and the idea that data have a distribution.
- The intuition behind the idea of an expectation. After all, the expected value might be impossible (take a value which doesn't lie in the support of the variable).
- Quantiles and their relation to CDF's
- Some mathematical puzzles:
  - if a variable has infinite support why is the area under the PDF equal to 1?
  - how can some variables not have an expected value?
- Means and variances for binary/indicator variables.
- The idea of correlation and association.

Notes for Tutors: Statistics

Preamble

The statistical inference we will look at in Prelims only concerns means and always assumes iid random sampling and large enough samples that the Central Limit Theorem can be applied. Students do not need to know about small sample tests and will only need to use Standard Normal tables. Hypothesis testing will only be of the two-tailed variety. The treatment of hypothesis tests focus on p-values rather than critical values - the Examination will reflect this so it is important that students can calculate and interpret p-values correctly.

In-Class Quiz

In the Canvas folder you will find an accompanying set of questions for the in-class quiz for this Class. These include the required statistical tables at the end. The instructions for the students are at the top of page 1. You should distribute the questions in advance of the Class and give the students enough time to prepare. In the class itself you should administer the quiz closed-book. The questions are not hard and don't require much in the way of derivations or calculations and the students should be able to get through them quickly. Much of the rest of the tutorial should be spend addressing any errors uncovered in the quiz and responding to any questions which arise.

You may want to only ask a subset of questions - use your judgement. Note that the question on confidence intervals and hypothesis testing are close parallels. This is deliberate. If you are happy that the students are on top of the material and see the relationship between them then you might decide to focus more on one than the other.

For reference, there is only one numerical questions, Q6 under "Confidence Intervals" and the answer is 1.28.

Discussion Points

The following are some potential discussion points you may go over with the students.

Statistical Theory

- What it means to say that a statistic has a sampling distribution.
- The relationship between
  - the population distribution of a variable,
  - the distribution of that variable in a sample, and
  - the sampling distribution of the mean of that variable.
- The difference between a "standard deviation" and a "standard error"

Statistical Inference

- The interpretation of a confidence interval (focus on the sampling variation of the interval, not the population parameter).
- The concept of testing "under the null".
- The meaning of a p-value.
- Why we focus on controlling the Type 1 error when carrying out a statistical test. (How could we formulate the test statistic under the alternative?).
- Why the different treatments of the calculation of the standard error in the case of binary/indicator variables.

Notes for Tutors: Time Series & Causal Inference

Preamble

Two topics this week: time series and causal inference. Neither of these deal with estimation.

The aims of the Time Series material are modest: understand the importance of serial correlation; the main empirical features of time-series; an introduction to the notion of stationary; some basic TS models (all based on an AR(1) so students should focus on the AR(1)); some empirical mistakes to avoid: "spurious" and "nonsense" correlation (on JD's advice a distinction is drawn) and mean reversion.

The causal inference is entirely "population-first" (last, and always). The main ideas are: potential outcomes; selection bias; RCTs and quasi/natural experiments; LATE. The latter (ahem) is examinable so in the lectures and the demo and in the Collection there are worked examples to demystify it through practice. NB even though these is no estimation/statistical inference discussed here, in the exam a question on causal inference can be mixed in with some statistical inference on the difference between means/proportions - students should not be encouraged to compartmentalise too much.

In-Class Quiz

In the Canvas folder you will find an accompanying set of questions for the in-class quiz for this Class. The instructions for the students are at the top of page 1. You should distribute the questions in advance of the Class and give the students enough time to prepare. In the class itself you should administer the quiz closed-book. The questions for this week don't require any calculations and the students should be able to get through them quickly. Much of the rest of the tutorial should be spend addressing any errors uncovered in the quiz and responding to any questions which arise.

Discussion Points

The following are some potential discussion points you may go over with the students.

Time Series

- The idea of non-independence in the data (autocorrelation)
- Stationarity and why it matters
- The TS models we cover as special cases of the AR(1)
- Pitfalls of high observed correlation between TS variables and mean reversion.

Causal Inference

- Understanding potential outcomes.
- The idea that treatment needs to be independent of potential outcomes (not observed outcomes)
- How RCT's work (you can revise the material covered in "Moments" on independence/mean independence.
- LATE and compliance types.