Public Policy 529
Summary and Review

Jonathan Hanson

Gerald R. Ford School of Public Policy
University of Michigan

December 6, 2023

Outline

1. Final Exam

2. Reflecting on the Origins of Many Statistical Methods

3. Looking Back and Forward

Outline

Exam

- The final exam will draw from the full course, with a greater focus on post-midterm material. Note what is new since Quiz 2.
- You can expect a set of questions involving interpretation of a bivariate regression table (see practice material).
- Significance tests involving means/proportions. Confidence intervals.
- General knowledge: use of the various tables, understanding of random sampling theory, ability to identify what is the appropriate test given a scenario.

Miscellaneous Tips

- Be ready to interpret the results of significance tests from ANOVA or chi-squared tests if given a table with output.
- Review lecture slides/notes to prepare yourself for questions related to general knowledge.
- Generally, however, questions will come in the format that you have seen on problem sets and exams.
- Pay attention to the measurement level of the variables and sample sizes. These tell you what kind of tests are applicable.

Outline

Statistics and Eugenics

- Many of the statistical methods we have learned in this class were developed by people involved in the eugenics movement.
- Eugenics is the idea that the genetic quality of the human population can be improved by selective breeding.
- Policies to prevent reproduction of "undesirable" traits included prohibitions on marriage, forced sterilizations, and institutionalization.
- Eugenics was central to the ideology of Hitler and the Nazi Party in Germany.

Francis Galton

- A key figure in the modern eugenics movement. Coined the term in 1883. Believed in a scientific basis for white racial superiority.
- He was also instrumental in developing linear regression.
- He graphed father's height (x-axis) and son's height (y-axis) and noted that there was a "co-relation."
- Son's height was a weighted average of father's height and mean height of population.
- Published article in 1886 called "Regression Toward Mediocrity in Hereditary Stature."

Karl Pearson

- Was deeply racist and saw conflict between races as inevitable and necessary to weed out "inferior" races.
- Started the journal *Biometrika* and the journal *Annals of Eugenics*. Was the Galton Chair in National Eugenics at UCL.
- A key figure in the development of mathematical statistics. Foundational thinking for significance testing.
- Created the χ^2 test, the Pearson's r correlation coefficient, and the theory behind p-values.

Ronald Fisher

- Succeeded Pearson as the Galton Chair of National Eugenics at UCL. Advocated for legalizing sterilization for eugenic purposes.
- Believed there were racial differences in the capacity for intellectual development.
- Promoted the significance testing framework and developed the Fisher's Exact test, the F-test, and ANOVA.
- Used these methods to claim that his theories were based on data and scientific objectivity.

Questions

- What lessons can we take from this history?
- To what extent do the origins of these methods call into question the use of these tools?
- What makes research objective? To what extent can statistical methods increase the objectivity of research findings?

Outline

Think Back to Where We Started

- Our goal is to learn things about the world. How does it work?
- We learn about the world through observation, data collection, and analysis of the data.
- The kind of data that we collect determines what kinds of tools we can use for analysis, whether qualitative or quantitative.
- Thus, measurement is at the foundation of everything. What concepts do we want to measure, and how do we measure them?
- Our results are only as good as the quality of our measures!

Representing Data as Variables

- When we measure concepts, we record the data using some kind of metric. The result is a variable.
- The measurement level of our variables – nominal, ordinal, or interval – dictates what kind of analyses are possible.
- We also need to consider whether the data are discrete or continuous, and the number of values that are possible, when making graphs and tables.

  e.g. dichotomous variables are a special class.
- Descriptive statistics for our variables help us understand the data better and to be aware of potential problems.

Linking Measurement Levels to Analysis Type

Dependent	Independent	Type of Analysis
Categorical	Categorical	Cross-tabs with χ^2 or Fisher's Exact; Difference of proportions with z-test;
Interval	Categorical	Mean comparison with t-test ANOVA with F-test
Interval	Interval/Dichot.	Linear regression; Correlation analysis
Dichotomous	Interval/Dichot.	Probit or Logit analysis
Categorical	Interval/Dichot.	Multinomial Logit/Ordered Probit

Probability Theory

- The laws of probability give us the tools we need to be systematic about understanding our level of precision when making statistical inferences.
- Randomness has well-understood properties that tell us how likely it is that the apparent patterns we see in the data were the result of random chance.
- We use this knowledge to develop tests of statistical significance: z-tests, t-tests, χ^2-tests, F-tests, etc.
- All of these tests measure the probability that our sample would produce our statistics (means, proportions, coefficients, etc.) under the scenario of the null hypothesis.

What's Coming Next

- One goal of this course is to provide the foundations for more advanced methods.
- Coming up is a much more thorough treatment of regression and how to use it in more sophisticated ways.
- Additionally, you will learn research design strategies to facilitate causal inference.
- Finally, you will learn techniques to work with data that regression cannot handle (e.g. dichotomous dependent variables).

Remember the Limitations!

- This course is about quantitative analysis, but it's important to remember that many questions cannot be addressed quantitatively.
- Some kinds of questions, particularly those with complex causal mechanisms or highly-contextual processes, are better studied qualitatively.
- Additionally, we cannot always rely on the law of large numbers to discern relationships in the data.
- A combination of qualitative and quantitative approaches is often the most fruitful.

Statistics is Just a Tool

- I encourage you to see statistics as a tool. How you use it is up to you!
- If you find a statistical analysis that seems wrong, then fix the problem and do it the right way.
- At the same time, just because you have a hammer does not mean every problem is a nail.

  The questions you pursue should not be determined by the methods you know but by the methods that work best for that question.

Good Luck!