MIT 14.32 **J. Angrist (angrist@mit.edu)**
Spring 2023

Problem Set 1

Due: Thursday, February 23

A. Prob/Stats Problems

1. Consider three random variables, denoted X, Y, and Z.

   (a) Suppose Z is Bernoulli. Prove that $E[Y|X] = E[E(Y|X,Z)|X]$ by writing the right-hand-side as a weighted average of two conditional expectations. What general property does this illustrate?

   (b) Suppose that $E[X] = 0$ and the conditional expectation function (CEF) for Y given X is $E[Y|X] = X^2$. Show that the mean (expectation) of Y is the variance of X.

   (c) Continue assuming $E[X] = 0$ and that $E[Y|X] = X^2$, so that Y and X are dependent. Show that it may nevertheless be possible that Y and X are uncorrelated. Give an example where this is surely true. (Hint: this requires only a little math.)

2. Complete the proofs omitted from LN1

   (a) Consider the distribution of earnings in the population of American workers, denoted by random variable Y, with mean $E[Y] = \mu$. Suppose you'd like a single number c that gives the best prediction of Y in the sense of minimizing mean-squared prediction error (MSE), defined as $MSE_Y(c) = E[Y - c]^2$. Show that

   $$MSE_Y(c) = \sigma_Y^2 + [\mu - c]^2,$$

   where σ_Y^2 is the variance of Y_i. (As noted in LN1, in view of this, we say that "MSE equals variance plus bias-squared.")

   (b) Consider two random variables, Y and X. Use the law of iterated expectations to show that $E(Y|X)$ is the minimum mean-squared error (MSE) predictor of Y given X.

   (c) Show that for any two random variables, X and Y, the variance of Y can be decomposed as:

   $$\sigma_Y^2 = E[\sigma_{Y|X}^2] + V[E(Y|X)], \qquad (1)$$

   where $\sigma_{Y|X}^2 = E\{(Y - E[Y|X])^2|X\}$ is the *conditional* variance of Y given X and the second term is the variance of the CEF. Equation (1) is called the analysis of variance (ANOVA) formula. Interpret this formula in words.

   (d) Prove Chebyshev's inequality, which says that for any random variable, X, for which the mean and variance exist, and for any positive constant, c:

   $$P(|X - \mu_X| \geq c\sigma_X) \leq {}^1/_{c^2}$$

   In other words, the probability that <u>any</u> random variable is more than c standard deviations away from its mean is less than ${}^1/_{c^2}$.

3. Let $\bar{Y}_n$ denote the sample mean of Y_i in a sample of size n from a population with mean μ and variance σ^2. Consider two estimators of μ: $\hat{\alpha}_1 = \left(\frac{n-1}{n}\right)\bar{Y}_n$ and $\hat{\alpha}_2 = \bar{Y}_n/2$.

   (a) Estimator $\hat{\theta}_n$ is said to be a*n unbiased* estimator of parameter θ when $E[\hat{\theta}_n] = \theta$. Show that $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are both *biased* estimators of μ. How does the bias of each vary as function of sample size?

   (b) Derive formulas for the standard errors of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ as a function of μ, σ^2, and n. For $n > 2$, which estimator is more <u>precise</u>? Which is more <u>biased</u>?

(c) For what values of μ and n does the biased estimator, $\hat{\alpha}_2$, have lower MSE than the (unbiased) sample mean? Give brief intuition for this. Why might this fact be useful?

(d) (More challenging) Consider estimators of μ of the form $\hat{\alpha}_k = k\bar{Y}_n; k > 0$. Define the optimal k as that which minimizes MSE. How does the optimal k vary as a function of μ and n? Give brief intuition for this.

4. Let $\bar{Y}_n$ denote the sample mean of Bernoulli Y_i, computed in a sample of size n. Let $\mu = E[Y_i]$.

   (a) Show that $\bar{Y}_n$ is an unbiased estimator of the probability that $Y_i = 1$.

   (b) Show that the standard error of the sample mean in this case can be estimated using only the sample mean and the sample size. Is this true in general?

   (c) (More challenging) Estimator $\hat{\theta}_n$ is said to be a *consistent* estimator of parameter θ when $plim\, \hat{\theta}_n = \theta$. Use Chebyshev's inequality to prove that the sample mean is a consistent estimator of the population mean for any Y_i (not just Bernoulli). This is a law of large numbers.

B. Empirical Warm-up

1. When workers lose their jobs, US states provide them with time-limited weekly unemployment insurance (UI) payments to make up for lost earnings. Economists have long been debated whether more generous UI (as was offered during the pandemic recession of 2020) reduces the incentive to work. Table 3 in Woodbury and Spiegelman (1987, posted in Canvas Module A) reports the results of two social experiments meant to encourage Unemployment Insurance (UI) recipients to return to work. In the Employer Experiment, any UI recipient finding employment for at least 4 months received a voucher worth $500 to his or her employer. In the Claimant Experiment, any UI recipient finding employment for at least 4 months received $500 directly.

   (a) For each experiment, test the hypothesis that bonuses decreased the proportion of UI claimants who exhausted their benefits. Compute the standard error needed for the denominator of the test statistic under two scenarios: (i) the experiment has no effect and (ii) the experiment has an effect.

   (b) For each experiment, pick a significance level and test the hypothesis that the experiment reduced weeks of insured unemployment in the first unemployment spell using a one-tailed and two-tailed test. Which test seems seems more sensible in this case?

2. This problem asks you to conduct a series of *sampling experiments* using Stata

   (a) Draw 500 random samples each with a sample size of 8 from a Normal distribution. Next, increase the sample size to 32. Finally, increase the sample size to 128. Plot histograms of the sampling distributions of the sample mean for each of these three sample sizes. Now repeat your experiments (and plots) for three samples drawn from a Bernoulli distribution with parameter p (your choice of value).

   (b) Your experiments produce a "sample of sample means." Compute the mean and variance of the 500 sample means generated by each experiment and compare them to the sampling distribution mean and variance predicted by statistical theory. Does the variance of the sample means (i.e., the sampling variance) decrease with sample size at the rate predicted by the theory? Does Normality of the underlying data seem to matter for this?

C. Working with NHIS Data

Does health insurance help keep you healthy? Table 1.1 in *Mastering 'Metrics* compares the health and demographic characteristics of married couples with and without health insurance. Health is measured on a scale of 1-5, with 5 being the best.

1. Using the information in the published table, calculate the t-statistic for the null hypothesis that there is no difference between the health of husbands who have health insurance and the health of those who don't. Is this difference significantly different from zero?

2. Panel B of Table 1.1 shows that husbands with and without health insurance differ in many ways. The difference in health between the "Some HI" and "No HI" groups may shrink when computed within more homogeneous groups. To investigate this, download the Stata data set and program posted on Canvas; the posted program applies the selection criteria used to produce the sample that generates Table 1.1. Note also that the MM table was computed using survey weights; you should should use weights too).

   With data and code in hand, explore and interpret the following conditional comparisons:

   (a) Replicate the first 3 columns of MM Table 1.1 (which report statistics for husbands)

   (b) Restrict your sample to employed college graduates (those with 16+ years of schooling). How does this restriction change the health gap by insurance status?

   (c) Show that employed college graduates are generally healthier than the rest of your husband sample and that the insured are more likely than the uninsured to be employed college graduates. How do these facts help interpret the change in insurance gaps as you move from (a) to (b)?