ECONOMICS 3304:
Statistics

Brief Statistics Review!

- Basics
- Sampling
- Estimation
- Hypothesis Testing

1

ECONOMICS 3304:
Basics: Random Variables

- X is a random variable if it represents a random draw from some population.
  - Discrete random variable can take on only selected, countable values (e.g., coin toss, rainy day).
  - Continuous random variable can take on any value in a real interval (e.g., time measured outcomes, height, etc.)
- Associated with each random variable is a probability distribution.

2

Basics: Standardized Variables

- Sometimes random variables represent the culmination of random events.
  - Life Expectancy (genes, nutrition, exercise, etc.)
  - Intelligence (heredity, parenting, schooling, etc.)
  - Stock Price (economy, dividends, forecasts, etc.)
- Standardizing these types of random variables can be important because
  - Common unit of measurement.
  - Similar probability distributions.

3

Standardized Variables: Common Measurement Unit

- How to standardize random variables?

$$Z = \frac{X - \mu}{\sigma}$$

- The new standardized random variable Z has a mean of 0 and standard deviation equal to 1.
- Ex: The IQ of a randomly selected person can be considered a random variable drawn from a population with mean 100 and std dev of 15.

4

Standardized Variables: Central Limit Theorem

- The ***central limit theorem*** states that a random variable X will tend to have a normal distribution, if it reflects the culmination of a large number of independent random variables.
- Culmination typically is interpreted as the summation or average of random variables.

$$\sum X_i = X_1 + X_2 + \ldots + X_N$$

$$\overline{X} = \tfrac{1}{N}(X_1 + X_2 + \ldots + X_N)$$

5

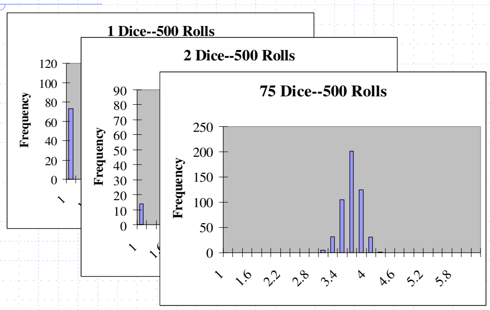
Standardized Variables: The Central Limit Theorem

Let $X_1, X_2, \ldots, X_N$ be independent random variables each with a distribution that has a mean of μ and variance of σ^2. Let $\overline{X}$ be the average of the random variables: $\overline{X} = \frac{1}{N}\sum X_i$.

Then $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{N}\right)$ approximately.

6

Central Limit Theorem: Histograms for Six-Sided Dice



7

Standardized Variables: The Central Limit Theorem

Suppose that $\overline{\mathrm{X}} = \frac{1}{N}\sum X_i$ is standardized with its mean μ and standard deviation $\sigma / \sqrt{\mathrm{N}}$ then the resulting standardized Z statistic will converge to a normal distribution with mean 0 and variance of 1.

Then $\mathrm{Z} = \dfrac{\overline{\mathrm{X}} - \mu}{\sigma / \sqrt{N}} \sim N(0,1)$ approximately.

8

Statistics Review: Sampling

- A **sample** is the part of the data the econometrician actually observes. It can be contrasted with the **population**, which is the complete set of data.
- Why bother taking samples?
  - Expensive to collect data on the entire population.
  - Sample data can be used to make statistical inferences about the population.

9

Statistics Review: Sampling

- **Selection Bias** refers to any sample systematically differing from the population it is supposed to represent.
- **Self-selection Bias** is a similar form of bias that can occur when examining data where individuals have chosen to be in a particular group.
- **Survivor Bias** occurs when choosing a sample from a current population to make inferences about a past population.
- **Nonresponse bias** refers to the systematic refusal of some groups to participate in an experiment or survey.

10

Random Samples and Sampling

- For a random variable X, repeated draws from the same population can be labeled as $X_1, X_2, \ldots, X_N$
- If every combination of N sample points has an equal chance of being selected, this is a random sample.
- A random sample is a set of independent, identically distributed (i.i.d.) random variables.

11

Statistics Review: Estimators and Estimates

- Typically, we can't observe the full population, so we must make inferences base on estimates from a random sample
- An ***estimator*** is just a mathematical formula for estimating a population *parameter* from sample data
- An ***estimate*** is the actual number the formula produces from the sample data

12

Examples of Estimators

- Suppose we want to estimate the parameter for mean of a population. Can calculate the sample average for our sample:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} X_i$$

- Similar estimate can be made for the median parameter.

$$Med\left[X\right] = Med\left[X_1,\ldots,X_N\right]$$

13

Estimators & Estimates: Average Value of Single-Family Homes in Diamond Bar, CA

- A random sample of 22 single family homes in Diamond Bar, California is drawn from a population of 20,000 homes.

$$\overline{X} = \frac{\$425{,}000 + \ldots + \$788{,}888}{22}$$
$$= \$565{,}829$$

- Another random sample will likely yield an estimate with a different value.

14

Estimators: Sampling Distributions

- The **sampling distribution** of a statistic is the probability distribution that describes the population of all possible values of the estimator.
- For sample average, the sampling distribution will tend to be normal due to the central limit theorem.
- For the sample average, also …

$$\text{Mean of } \overline{X} = \mu$$

$$\text{Std.Dev.of } \overline{X} = \frac{\sigma}{\sqrt{N}}$$

15

Estimators: Sampling Distributions

- The Mean of the Sampling Distribution
  - Gauges how close the sample statistic is to the population parameter. An **unbiased estimator** occurs when the mean of the sampling distribution equals the population parameter.
- Standard Deviation of the Sampling Distribution
  - The accuracy of an estimator can be gauged by its standard deviation. The smaller the standard deviation, the higher probability the estimate will be close to the mean.

16

Statistics Review: Hypothesis Tests

- Sample data can be used to support or refute theories about the value of population parameters.
- For example, whether the average price of single family homes in Diamond Bar, CA is $400,000.
- A null hypothesis and alternative hypothesis are proposed.

$$H_0 : \mu = \$400{,}000$$
$$H_A : \mu \neq \$400{,}000$$

17

Hypothesis Tests: The *t*-Distribution

- The central limit theorem shows a standardized variable converges to a normal distribution with mean 0 and variance 1.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{N}} \sim N(0,1)$$

- The t-distribution does not rely on knowing the parameter for standard deviation.

$$t = \frac{\overline{X} - \mu}{s / \sqrt{N}} \sim t_{(N-1)}$$

18

Hypothesis Tests: The *t*-Distribution

- We have a good estimate of μ, would like a good estimate of σ^2
- Can use the sample variance given below – note division by n-1, not n, since mean is estimated too – if know μ can use n

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Hypothesis Tests

- Constructing the Test Statistic
  - An estimator is used to test the null hypothesis, which is called a test statistic.

$$t = \frac{\overline{X} - \mu}{s / \sqrt{N}}$$
$$= \frac{\$565{,}829 - \$400{,}000}{\$116{,}596 / \sqrt{22}} \sim t_{(N-1)}$$
$$= 6.671$$