EC 320: Introduction to Econometrics
Lecture Notes

David M. Hall
University of Oregon

Fall 2025

1 Session Overview

- Date: 10/2/2025
- Topic: Statistical Review
  - Chapters (Dougherty): R Subsections 1,2,4-7,9,11,12
  - Supplementary: 13-15
- Goals/Objectives:
  1. Review statistics necessary for success in this course.
  2. Introduce assignment number 1 (flashcards).

1.1 Before Class:

1. Dougherty reading (review)
2. After-class reflection finished.
3. R should be downloaded

2 During Class:

2.1 Welcome

- Today we will be covering stats review for two purposes:
  1. Cover statistical material that will be necessary for you to pass this course.
  2. See where your current understanding of the material is.

	Loss	Push	Win	Total
Value	-\$25	\$0	\$25	
Probability	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\sum_{i=1}^{3} p_i = 1\checkmark$

Table 1: Probability of Blackjack Scenarios for the example.

2.1.1 Warm Up:

Example 2.1. 1. Write out, in your own words, your own definition of a "random variable."

2. Suppose you are playing blackjack (21) and have been given two cards: a 3 and an 8 making your current total 11. The dealer is showing a King, and you have x-ray vision so you know that they also have a Jack underneath making their total 20.

   To win, you need to pull a ten. Suppose you decide to hit (no double downs here), and there are three cards left in deck: you happen to know they are an ace, a 9 and a Queen. If you draw an ace, your total is 12 and you lose. If you draw a 9, your total is 20 and you push. If you draw a queen, your total is 21 and you win.

   (a) Create a frequency and probability table containing the (a) value of each of the scenarios assumingg you bet \$25 and (b) the probability of each scenario.

   (b) Calculate the expected value of a hit.

Note that the table should look something like this:
Recall that the definition of expected value is:

Definition 2.1 (Expected Value). $E(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n = \sum_{i=1}^{n} x_ip_i$ where X is a vector of values with each element of vector X denoted by i.

Thus we know that the expected value is 0 since the calculation is simply:

$$-25 * \frac{1}{3} + 0 * \frac{1}{3} + 25 * \frac{1}{3} = 0$$

2.2 Content Section #1:

Time:

Give examples and definition of a random variable.

A random variable is a variable, or we can think of it as an observable attribute of some sort of study. You're used to thinking about them as either "discrete" or "continuous".

Note that we often call the expected value of a random variable its **population mean** which is usually denoted by μ or μ_X.

Some properties of expectations:

- For a function $g(X)$ of the random variable X, $E(g(X)) = g(x_1)p_1 + g(x_2)p_2 + \cdots + g(x_n)p_n = \sum_{i=1}^{n} g(x_i)p_i$.
- $E(X + Y + Z) = E(X) + E(Y) + E(Z)$

- For constant b, $E(bX) = bE(X)$
- $E(b) = b$

Recall that we often find it interesting to look at the population variance of a known random vairable defined as:

Definition 2.2 (Population Variance)**.** $var(X) = \sigma_X^2 = E(X - \mu_X)^2$
$= \sum_i (x_i - \mu_X)^2 p_i$
$= E(X^2) - \mu_X^2 = E(X^2) - [E(X)]^2$

PAUSE FOR QUESTIONS

We might also consider the way two random variables X and Y can interact through covariance and correlation. Recall these definitions:

Definition 2.3 (Population Covariance)**.** $cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$

Recall that if these two variables are *independent* then:

- $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$
- $E(XY) = E(X)E(Y)$
- Population covariance of 0: $E[(X - \mu_X)(Y - \mu_Y)] = 0$
  - This is because $\mu_X = E(X)$ and $\mu_Y = E(Y)$

For your flashcards, write out and understand the proofs related to these facts. We will simply go over them.

- If $Y = V + W$ then $cov(X, Y) = cov(X, V) + cov(X, W)$
- If $Y = bZ$ for constant b then $cov(X, Y) = cov(X, bZ) = b * cov(X, Z)$
- If $Y = b$ for constant b then $cov(X, Y) = 0$ since Y does not have any variance.
- Note further than $cov(X, X) = var(X)$

Pause here: so far we have discussed population parameters. Let's reflect a bit on the difference between population and sample parameters/estimates.

2.3 Activity #1:

Just ask questions, see if we can get to the fact that:

- We usually do not observe a full population, so those parameters (population mean, variance, covariance) will not be known.
- We *will* observe a sample.
  - e.g. We do not observe all people in the U.S. for the ACS or the Census, we observe a *sample* and then infer population statistics from that.
- Since we do not know population parameters, we use *estimators* to attempt to obtain those population parameters.

The sample mean estimator is defined as:

Definition 2.4 (Sample Mean). $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_i^n X_i$

The sample mean has a different variance than the mean of the population itself. The sample mean has variance σ_X^2/n which indicates that, as n gets closer to ∞ the variance decreases and we can be more sure that $\bar{X} = \mu_X$

2.4 Content Section #2:

Time:

If we do not know the population parameter and we want to estimate them, how do we know that the sample estimate is going to be close (and asymptotically so, we hope) to the population parameter?

Specifically, we want two things in an estimator:

1. We want the estimator to be *unbiased*.
2. We want the estimator to be *efficient*.

Sometimes these things are at odds, and so you want to find some sort of balance. See what Dougherty talks about in R.6 about the "loss function."

Definition 2.5 (Unbiased). For an estimator $\bar{X}$ to be an unbiased estimator of the parameter μ_X, it must be the case that $E(\bar{X}) = \mu_X$

For the case of efficiency, we simply want it so that the estimator distribution gets as close to the true value as possible.

Popular estimators of variance and covariance we will use are:

Definition 2.6 (Variance Estimator). $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$

Definition 2.7 (Covariance Estimator). $\hat{\sigma}_{XY} = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$

As we move forward and look at regression estimators, what you'll find is that we find what a population parameter might be as a function of population mean/variance/covariance and estimate that population parameter by inputting the estimators defined above.

2.5 Activity #2 (Optional):

Time:

For the rest of class I want to have an open ended discussion about hypothesis testing, what we remember, etc. They should be going along the following lines:

- In hypothesis testing we have a null and an alternative hypothesis.
- To test the hypothesis, I generate the parameter being tested using a sample and then see if it is relatively close to the null/alternative being presented.
- I can reject or fail to reject the null based on some criterion:
  - If I reject the null, it must be because the statistic used for the test is above the critical value for the test.
  - You have previously used a z statistic I assume? Maybe a t?
  - You can also just see if the confidence interval contains the value you are hypothesis testing for.

Let's draw it out. Draw a normal distribution, have them explain to me the significance level, rejection regions,

2.6 Summary Activity (5 Q's of Reflection):

1.

3 Be Able To

- Write out a be able to - separate but related to the goals of this lecture.

4 Key Words

- Fill in Here