3 Expectation, variance, covariance and correlation

3.1 Expected value of a Random Variable – $E(X)$ or μ_X

- An expectation is a measure of central tendency.
- The expected value of a random variable X, denoted $E(X)$ or μ_X, is the average value of the random variable over many repeated draws.

  - Discrete case: $E(X) = \Sigma_{i=1}^{N} x_i P(X = x_i)$

  - Continuous case: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$

  Example: What grade would student A get from EE425?

x_i	$P(X = x_i)$	$x_i P(X = x_i)$
$0(F)$	0.05	0
$1(D)$	0.05	0.05
$2(C)$	0.25	0.50
$3(B)$	0.40	1.20
$4(A)$	0.25	1.00
Total $(\Sigma_{i=1}^{N})$	1.00	$2.75(\approx B-)$

3.2 Properties of Expected Values

1. For any constant c, $E(c) = c$
2. For any constant a and $b, E(aX + b) = aE(X) + b$
3. If $\{a_1, a_2, ..., a_n\}$ are constants and $\{X_1, X_2, ..., X_n\}$ are random variables, then

$$E(a_1X_1 + a_2X_2 + ... + a_nX_n).$$

   Or using summation notation

$$E\sum_{i=1}^{n} a_iX_i = \sum_{i=1}^{n} a_iE(X_i).$$

   ** It is important to be reminded that $E(X^2) \neq E(X)E(X)$!

3.3 Conditional and Marginal Expectations

- Example: What grade would student A get from EE425 given the number of hours/week she spent on studying EE425.

$P(X = x_i, H = h_i)$					
$x_i \backslash h_i$	0	3	6	9	12
$0(F)$	0.0	0.0	0.0	0.0	0.0
$1(D)$	0.1	0.1	0.0	0.0	0.0
$2(C)$	0.0	0.1	0.1	0.0	0.0
$3(B)$	0.0	0.05	0.15	0.1	0.05
$4(A)$	0.0	0.0	0.05	0.1	0.1

where x_i indicates grade, h_i indicates number of hours/week spent on studying EE425.

- What are the expectations of grade (x_i) and hours (h_i)?
- Are (x_i) and (h_i) independent? why or why not?
- What is the conditional expectation of x_i given $h_i = 9$?

3.4 Variance and Standard Deviation of a Random Variable

- The variance and standard deviation measure the "variability", the "dispersion" or the "spread" of a probability distribution

$$\begin{aligned} Var(X) \text{ or } \sigma^2 &= E[(X-\mu_x)^2] \\ &= \\ &= \\ &= \\ &= \\ &= E(X^2)-\mu^2. \end{aligned}$$

- The standard deviation, denoted $sd(X)$, is the positive square root of the variance:

$$sd(X) \text{ or } \sigma = +\sqrt{Var(X)}.$$

Exercise:

- What would happen if variance = 0?

-Depict probability distributions with different values of variance.

3.5 Properties of Variances

1. $Var(X) = 0$ if, and only if, there is a constant c, such that $P(X = c) = 1$, in which case, $E(X) = c$.

2. For any constand a and b, $Var(aX + b) = a^2 Var(X)$

3.6 Properties of Standard Deviations

1. For any constant c, $sd(c) = 0$

2. For any constants a and b,

$$sd(aX + b) = |a| sd(X).$$

   In particular, if $a > 0$, then $sd(aX) = a \times sd(X)$.

Exercise: What is $Var(aX + b) =?$, $Var(aX + bY) =?$

- Well, how do we compare and assess one random variable against another? One way to do this is through standardization.
- You may have heard of the Z-score

$$Z = \frac{X - \mu_X}{\sigma}.$$

- If X is normally distributed, then Z would be normally distributed. We can use the well-known Z statistic table for hypothesis-testing.
- In many cases, where the sample size gets very large, $Z = \frac{X-\mu_X}{\sigma}$ is approximately normally distributed regardless of the orginal distribution of X.

  Exercise:

  - What is $E(Z)$ and $Var(Z)$?

3.7 Covariance and Correlation of Two Variables

- We talked about joint distribution and independence earlier. How do these two concepts relate to covariance and correlation?
- When the movement of one variable can give some information about the movement of another variable, these two variables are dependent.
- We can also say that they are "correlated" or their "covariance" is not zero.
- Covariance and correlation measure the amount of *linear* association between variables.
- **It is worth noting that zero correlation does not imply independence.

$$\begin{aligned}
\text{covariance or } Cov(X,Y) \text{ or } \sigma_{X,Y} &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= \\
&= \\
&= \\
&= E(XY) - \mu_X\mu_Y.
\end{aligned}$$

Exercise: What's the unit of $Cov(X,Y)$?

- Correlation – makes the unit of dependence more standardized.

$$Corr(X,Y) \text{ or } \rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}.$$

- $\rho_{X,Y} = 1$; perfect positive linear relationship
- $\rho_{X,Y} = -1$; perfect negative linear relationship
- $\rho_{X,Y} = 0$; no linear relationship.

---

4 Estimators and desirable properties of estimators

- In statistics and econometrics, we hardly have the complete information of the population.
- Most data that we deal with are from a subset of the population or a "sample".
- We would like to learn about the "population" as we can using the "sample" that we have.

  - It is important to identify the population of interest

  - Once the population is identified, we can specify the model for the population relationship of interest.

- Examples of Estimators...

Population Parameter	Estimator(s)
population mean (μ)	$\bar{X} = \frac{1}{N}\Sigma_{i=1}^{N} X_i$
population variance (σ^2)	$S^2 = \frac{1}{N-1}\Sigma_{i=1}^{N}(X_i - \bar{X})^2$
$\alpha, \beta_1, \beta_2, \beta_N$	$\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_N$

  - Miguel and Kremer(2004) "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities."

$$\text{school attendance rate} = \alpha + \beta_1 \times \text{deworming} + \beta_2 x_2 + ... + \beta_N x_N$$

  How do we know the true value of $\alpha, \beta_1, \beta_2, \beta_N$? Without having the population and the correct model, it may be impossible to know the true value of $\alpha, \beta_1, \beta_2, \beta_N$. But the econometricians try their best to come up with the estimators, often denoted $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_N$. The better the estimator, the more it satisfies the desirable properties.

4.1 Desirable properties of estimators

- From now on, let's denote the population parameter of interest "θ" and its estimator "W". The desirable properties of estimators are: unbiasedness, efficiency and consistency.

1. Unbiasedness – the expected value of the estimator is equal to the value of the parameter it tries to estimate.
- $E(W) = \theta$
- $Bias(W) \equiv E(W) - \theta$
- Exercise: is $\bar{X}$ a biased estimator? What about X_1 ?

---

2. Efficiency – an estimator with a lower variance is said to be "more efficient" than another estimator with a higher variance

- If $Var(W_1) \leq Var(W_2)$, then W_1 is a more efficient estimator of θ than W_2.
- Exercise: Which estimator is more efficient, $\bar{X}$ or X_1?

3. Consistency – when the sample size gets large, the estimater W can do better and better in estimating θ.

- Large sample properties can also be called "asymptotic" properties.
- For W_n, which is an estimator of θ based on a sample $X_1, X_2, ..., X_n$ of sample size n. Then, W_n is a consistent estimator of θ, if for every $\varepsilon > 0$,

$$P(|W_n - \theta| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

if not, then we can say that W_n is inconsistent.

4.2 The concept of Mean Squared Error (MSE)

- is used to compare different biased estimators.

$$MSE(W) = E[(W - \theta)^2]$$

Exercise: show that $MSE(W) = Var(W) + [Bias(W)]^2$.