Introduction to Econometrics

Lecture 2: Review of Probability

Outline

Random Variables and Probability Distributions

Expected Values, Mean and variance

Two Random Variables

The Normal Chi-Squared, Student t and F Distributions

Random Sampling and the Distribution of the Sample Average

Large-Sample Approximations to Sampling Distributions

Outline

Random Variables and Probability Distributions

Expected Values, Mean and variance

Two Random Variables

The Normal Chi-Squared, Student t and F Distributions

Random Sampling and the Distribution of the Sample Average

Large-Sample Approximations to Sampling Distributions

Probabilities, the Sample Space, and Random Variables

- Outcomes: The mutually exclusive potential results of a random process
- Probability of an outcome: The proportion of the time that the outcome occurs in the long run
- Sample space: The set of all possible outcomes
- Event: A subset of the sample space
- Random variable: A numerical summary of a random outcome
- Examples: gender of the next person you meet after class, number of times your computer crashes while writing a term paper

Probability Distribution of a Discrete Random Variable

- Discrete random variables take on only a discrete set of values
- Probability distribution: List of all possible values of the variable and the probabilities that each value will occur

$$P[Y = y_k], \qquad k = 1, \ldots, K$$

- The probability of an event is given by the sum of the probabilities of the outcomes associated with it

$$P[Y = y_2 \text{ or } Y = y_3] = P[Y = y_2] + P[Y = y_3]$$

- Cumulative probability distribution: Probability that the random variable is less than or equal to a particular value

$$P[Y \leq y_k], \qquad k = 1, \ldots, K$$

TABLE 2.1 Probability of Your Computer Crashing *M* Times

	Outcome (number of crashes)				
	0	**1**	**2**	**3**	**4**
Probability distribution	0.80	0.10	0.06	0.03	0.01
Cumulative probability distribution	0.80	0.90	0.96	0.99	1.00

Probability Distribution of a Continuous Random Variable

- Continuous random variables take on a continuum of possible values
- Cumulative distribution function (CDF): Probability that the random variable is less than or equal to a particular value

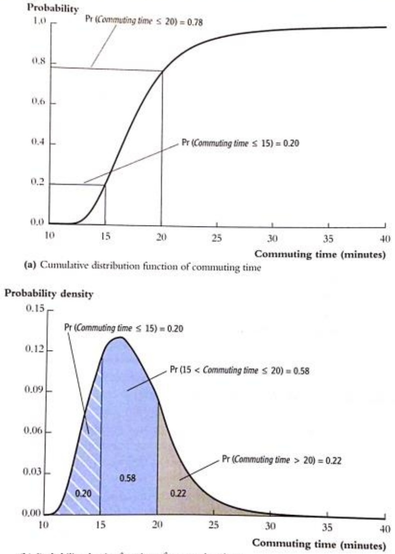
$$F(y) = P[Y \leq y]$$

- Probability density function (PDF): Area under the PDF between any two points is the probability that the random variable falls between those two points

$$f(y) = \frac{d}{dy} F(y)$$

$$\int_{y_1}^{y_2} f(y)\, dy = P[y_1 < Y < y_2]$$

FIGURE 2.2 Cumulative Distribution and Probability Density Functions of Commuting Time



(a) Cumulative distribution function of commuting time

(b) Probability density function of commuting time

Figure 2.2a shows the cumulative probability distribution (or c.d.f.) of commuting times. The probability that a commuting time is less than 15 minutes is 0.20 (or 20%), and the probability that it is less than 20 minutes is 0.78 (78%). Figure 2.2b shows the probability density function (or p.d.f.) of commuting times. Probabilities are given by areas under the p.d.f. The probability that a commuting time is between 15 and 20 minutes is 0.58 (58%) and is given by the area under the curve between 15 and 20 minutes.

Outline

The Expected Value of a Random Variable

- The long-run average value of the random variable over many repeated trials or occurrences
- For a discrete random variable computed as a probability-weighted average of the possible outcomes:

$$\mu_Y = E[Y] = \sum_{k=1}^{K} y_k P[Y = y_k]$$

- For a continuous random variable computed as a PDF-weighted average of the possible outcomes:

$$\mu_Y = E[Y] = \int y\, f(y) dy$$

TABLE 2.1 Probability of Your Computer Crashing M Times

	Outcome (number of crashes)				
	0	1	2	3	4
Probability distribution	0.80	0.10	0.06	0.03	0.01
Cumulative probability distribution	0.80	0.90	0.96	0.99	1.00

$$E[M] = 0.80 \times 0 + 0.10 \times 1 + 0.06 \times 2 + 0.03 \times 3 + 0.01 \times 4 = 0.35$$

The Standard Deviation and Variance

- Measures of the dispersion of a probability distribution
- Variance:

$$\sigma_Y^2 = \mathrm{Var}[Y] = E\left[\left(Y - \mu_y\right)^2\right] = E[Y^2] - (E[Y])^2$$

- Standard deviation:

$$\sigma_Y = \mathrm{std.\,Dev}[Y] = \sqrt{\mathrm{Var}[Y]}$$

- Standard deviation is in natural units, making it more convenient to report

TABLE 2.1 Probability of Your Computer Crashing *M* Times

	Outcome (number of crashes)				
	0	**1**	**2**	**3**	**4**
Probability distribution	0.80	0.10	0.06	0.03	0.01
Cumulative probability distribution	0.80	0.90	0.96	0.99	1.00

$$Var\,[M] = 0.80 \times (0 - 0.35)^2 + 0.10 \times (1 - 0.35)^2 + 0.06 \times (2 - 0.35)^2 + 0.03 \times (3 - 0.35)^2 + 0.01 \times (4 - 0.035)^2 = 0.6475$$

$$\text{std. Dev}[Y] = \sqrt{0.6475} = 0.805$$

Mean and Variance of a Linear Function of a Random Variable

- Suppose Y is a linear function of X :

$$Y = a + bX$$

- Then the mean and variance of Y are given by

$$\mu_Y = a + b\mu_X$$

And

$$\sigma_Y^2 = b^2\sigma_X^2$$

- Example: after-tax and pre-tax earnings (a = 2000, b = 0.8)

Other Measures of the Shape of a Distribution

- Skewness: Measure of the lack of symmetry of a distribution

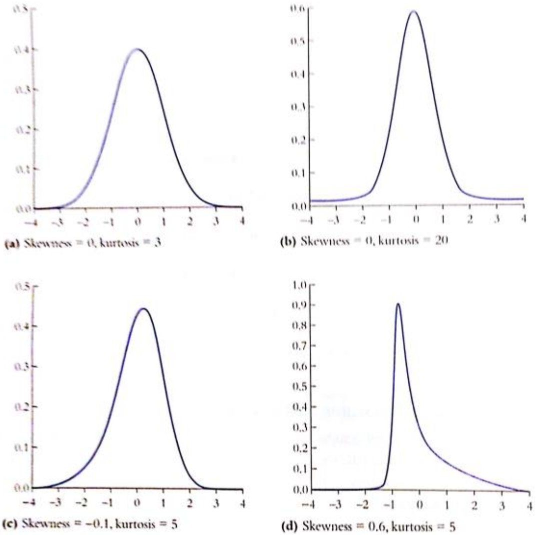
$$Skewness = \frac{E[(Y-\mu_Y)^3]}{\sigma_Y^3}$$

- Kurtosis: Measure of the thickness/heaviness of the tails of a distribution

$$Kurtosis = \frac{E[(Y-\mu_Y)^4]}{\sigma_Y^4}$$

- Moments: r^{th} moment of Y is $E[Y^r]$

FIGURE 2.3 Four Distributions with Different Skewness and Kurtosis



All of these distributions have a mean of 0 and a variance of 1. The distributions with skewness of 0 (a and b) are symmetric; the distributions with nonzero skewness (c and d) are not symmetric. The distributions with kurtosis exceeding 3 (b–d) have heavy tails.

Outline

Joint, Marginal, and Conditional Distributions

- Joint probability distribution: Probability that the random variables simultaneously take on certain values

$$P[X = x, \qquad Y = y]$$

- Marginal probability distribution: Another name for the probability distribution of a single random variable

$$P[Y = y] = \sum_{k=1}^{K} P[X = x_k, \qquad Y = y]$$

- Conditional distribution: Distribution of Y conditional on X taking on a specific value

$$P[Y = y \mid X = x] = \frac{P[X = x,\ Y = y]}{P[X = x]}$$

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

$$P[Y = 0 \mid X = 0] = \frac{0.15}{0.30} = 0.5$$
$$P[Y = 0 \mid X = 1] = \frac{0.07}{0.70} = 0.1$$

$$P[Y = 1 \mid X = 0] = \frac{0.15}{0.30} = 0.5$$
$$P[Y = 1 \mid X = 1] = \frac{0.63}{0.70} = 0.9$$

TABLE 2.3 Joint and Conditional Distributions of Computer Crashes (M) and Computer Age (A)

A. Joint Distribution

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
Old computer ($A = 0$)	0.35	0.065	0.05	0.025	0.01	0.50
New computer ($A = 1$)	0.45	0.035	0.01	0.005	0.00	0.50
Total	0.80	0.10	0.06	0.03	0.01	1.00

B. Conditional Distributions of *M* given *A*

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
$\Pr(M \mid A = 0)$	0.70	0.13	0.10	0.05	0.02	1.00
$\Pr(M \mid A = 1)$	0.90	0.07	0.02	0.01	0.00	1.00

Conditional Expectation and Variance

- Conditional expectation:

$$E[Y \mid X = x] = \sum_{k=1}^{K} y_k P[Y = y_k \mid X = x]$$

- Conditional variance:

$$Var[Y \mid X = x] = \sum_{k=1}^{K} y_k - E[Y \mid X = x]^2 \; P[Y = y_k \mid X = x]$$

- Law of iterated expectations:

$$E[Y] = E\big[E[Y \mid X]\big]$$

- Law of total variance:

$$Var[Y] = E\big[Var[Y \mid X]\big] + Var\big[E[Y \mid X]\big]$$

TABLE 2.3 Joint and Conditional Distributions of Computer Crashes (*M*) and Computer Age (*A*)

A. Joint Distribution

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
Old computer ($A = 0$)	0.35	0.065	0.05	0.025	0.01	0.50
New computer ($A = 1$)	0.45	0.035	0.01	0.005	0.00	0.50
Total	0.80	0.10	0.06	0.03	0.01	1.00

B. Conditional Distributions of *M* given *A*

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
$\Pr(M \mid A = 0)$	0.70	0.13	0.10	0.05	0.02	1.00
$\Pr(M \mid A = 1)$	0.90	0.07	0.02	0.01	0.00	1.00

$E[M \mid A = 0] = 0.56$ $\qquad Var[M \mid A = 0] = 0.99$

$E[M \mid A = 1] = 0.14$ $\qquad Var[M \mid A = 1] = 0.22$

$E[M] = 0.35$ $\qquad Var[M] = 0.6475$

Independence

- Random variables X and Y are independent if knowing the value of one provides no information about the other:

$$P[Y = y \mid X = x] = P[Y = y]$$

for all values of x and y

- If X and Y are independent, then

$$P[X = x \mid Y = y] = P[X = x]P[Y = y]$$

for all values of x and y

Covariance and Correlation

- Measures of the extent to which two random variables move together (linearly)
- Covariance:

$$\sigma_{XY} = Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$
$$= E[XY] - E[X]E[Y]$$

- Correlation:

$$\rho = Corr[X, Y] = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- Independence implies uncorrelatedness but not the other way around

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

$\mu_X = 0.70, \ \sigma_X = 0.448$
$\mu_Y = 0.78, \ \sigma_Y = 0.414$

$\sigma_{XY} = 0.15 \times (0 - 0.70)(0 - 0.78) + 0.15 \times (0 - 0.70)(1 - 0.78) +$
$0.07 \times (1 - 0.70)(0 - 0.78) + 0.63 \times (1 - 0.70)(1 - 0.78) = 0.084$

$\rho = \frac{0.084}{0.448 \times 0.414} = 0.442$

The Mean and Variance of Sums of Random Variables

- Suppose *W* is a linear function of *X* and *Y*

$$W = a + bX + cY$$

- Then the mean and variance of *W* are given by

$$\mu_W = a + b\mu_X + c\mu_Y$$

And

$$\sigma_W^2 = b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\sigma_{XY}$$

- Example: course grade as a function of midterm and final exam scores $(a = 25, b = 0.25, c = 0.50)$

Outline

The Normal Distribution

- Normal distribution with mean μ and variance σ^2: $N(\mu, \sigma^2)$
- Standard normal distribution: $N(0, 1)$
- Standardization: If $Y \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

FIGURE 2.5 The Normal Probability Density

The normal probability density function with mean μ and variance σ^2 is a bell-shaped curve, centered at μ. The area under the normal p.d.f. between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95. The normal distribution is denoted $N(\mu, \sigma^2)$.

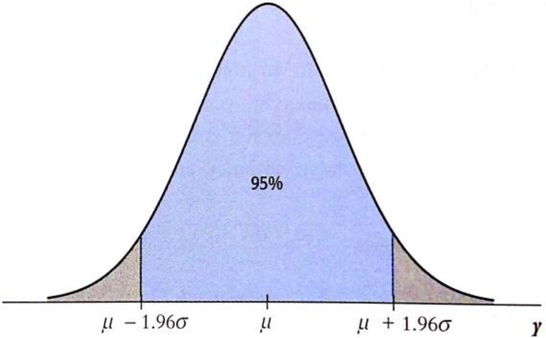
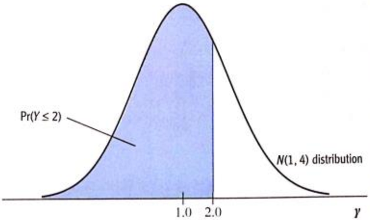
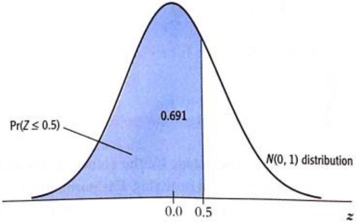


FIGURE 2.6 Calculating the Probability That $Y \leq 2$ When Y Is Distributed $N(1, 4)$

To calculate $\Pr(Y \leq 2)$, standardize Y, then use the standard normal distribution table. Y is standardized by subtracting its mean ($\mu = 1$) and dividing by its standard deviation ($\sigma = 2$). The probability that $Y \leq 2$ is shown in Figure 2.6a, and the corresponding probability after standardizing Y is shown in Figure 2.6b. Because the standardized random variable, $(Y - 1)/2$, is a standard normal (Z) random variable, $\Pr(Y \leq 2) = \Pr\left(\frac{Y-1}{2} \leq \frac{2-1}{2}\right) = \Pr(Z \leq 0.5)$. From Appendix Table 1, $\Pr(Z \leq 0.5) = \Phi(0.5) = 0.691$.



(a) $N(1, 4)$



(b) $N(0, 1)$

The Chi-Squared Distribution

- $Z_1, \dots, Z_m$ are independent and $Z_i \sim N(0, 1)$
- X_m^2 distribution is the distribution of

$$\sum_{i=1}^{m} Z_i^2$$

- m is called the degrees of freedom

The Student t Distribution

- $Z \sim N(0,1)$
- $W \sim X_m^2$
- Z and W are independent
- t_m distribution is the distribution of

$$\frac{Z}{\sqrt{\frac{W}{m}}}$$

- m is called the degrees of freedom

The F Distribution

- $W \sim X_m^2$
- $V \sim X_n^2$
- W and V areindependent
- $F_{m.n}$ distribution is the distribution of

$$\frac{\sqrt{\frac{W}{m}}}{\sqrt{\frac{V}{n}}}$$

- m and n are called the degrees of freedom

,

Outline

Random Sampling

- Simple random sampling: n objects are selected at random from a population

$$Y_1, \dots, Y_n$$

- Identically distributed: $Y_1, \dots, Y_n$ have the same marginal distribution
- Independently and identically distributed (IID): $Y_1, \dots, Y_n$ are drawn from the same distribution and are independent from each other
- Example: heights of five randomly chosen students

The Sampling Distribution of the Sample Average

- Sample average:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- $Y_1, \ldots, Y_n$ are random variables $\Longrightarrow$ $\bar{Y}$ is a random variable
- Mean and variance of $\bar{Y}$

$$E[\bar{Y}] = \mu_Y$$

$$Var[\bar{Y}] = \frac{\sigma_Y^2}{n}$$

$$Std, Dev[\bar{Y}] = \frac{\sigma_Y}{\sqrt{n}}$$

- Example: average height of the five randomly chosen students

Outline

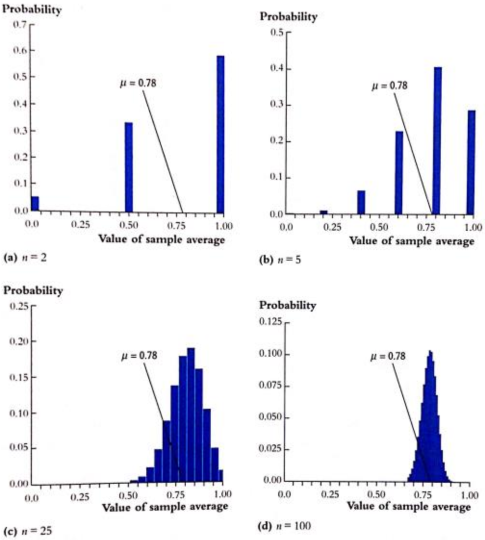
Exact Versus Approximate Distribution

- Sampling distributions play a central role in the development of statistical and econometric procedures
- Two approaches to characterizing sampling distributions: exact and approximate
- Exact approach: Derive a formula that holds exactly for any n (finite-sample distribution)
- Approximate approach: Derive a large-sample approximation (asymptotic distribution)
- Exact distributions are often difficult to derive (or require too many assumptions), and we typically have a “large enough” n

The Law of Large Numbers and Consistency

- Law of large numbers: Under general conditions, $\bar{Y}$ will be near μ_Y with very high probability when *n* is large
- One set of conditions for this result is that $Y_1, \dots, Y_n$ are IID with $E[Y_i] = \mu_Y$ and $Var[Y_i] < \infty$
- The property that $\bar{Y}$ is near μ_Y with increasing probability as *n* increases is called convergence in probability or consistency

FIGURE 2.8 Sampling Distribution of the Sample Average of n Bernoulli Random Variables

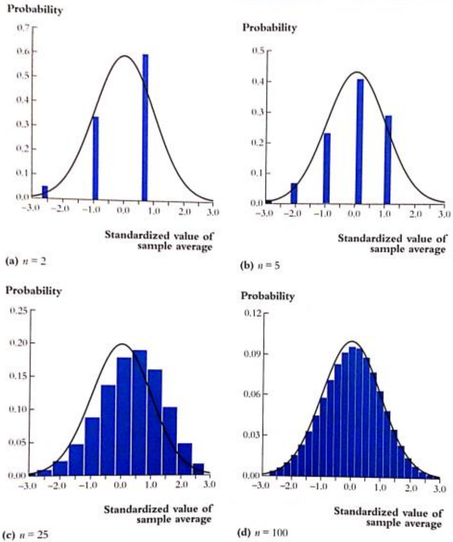


The distributions are the sampling distributions of $\overline{Y}$, the sample average of n independent Bernoulli random variables with $p = \Pr(Y_i = 1) = 0.78$ (the probability of a short commute is 78%). The variance of the sampling distribution of $\overline{Y}$ decreases as n gets larger, so the sampling distribution becomes more tightly concentrated around its mean $\mu = 0.78$ as the sample size n increases.

The Central Limit Theorem

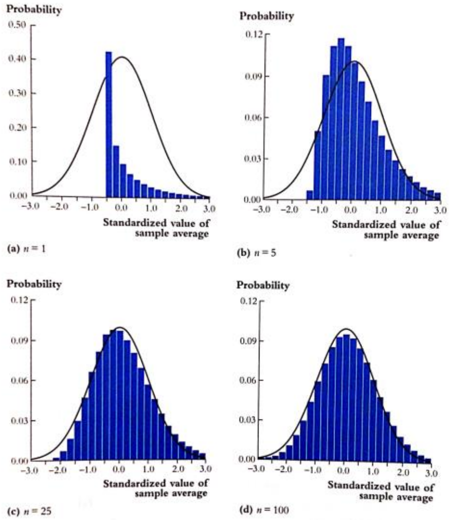
- Central limit theorem: Under general conditions, the distribution of $\bar{Y}$ is well approximated by the normal distribution $N\left(\mu_{\bar{Y}}, \sigma_{\bar{Y}}^2\right)$ when n is large
- Similarly, the distribution of $\frac{\bar{Y}-\mu_{\bar{Y}}}{\sigma_{\bar{Y}}}$ is well approximated by the standard normal distribution $N(0,1)$ when n is large
- When is n large enough? It depends on the underlying distribution of Y_i

FIGURE 2.9 Distribution of the Standardized Sample Average of n Bernoulli Random Variables with $p = 0.78$



The sampling distribution of $\bar{Y}$ in Figure 2.8 is plotted here after standardizing $\bar{Y}$. This plot centers the distributions in Figure 2.8 and magnifies the scale on the horizontal axis by a factor of $\sqrt{n}$. When the sample size is large, the sampling distributions are increasingly well approximated by the normal distribution (the solid line), as predicted by the central limit theorem. The normal distribution is scaled so that the height of the distributions is approximately the same in all figures.

FIGURE 2.10 Distribution of the Standardized Sample Average of n Draws from a Skewed Distribution



The figures show the sampling distribution of the standardized sample average of n draws from the skewed (asymmetric) population distribution shown in Figure 2.10a. When n is small ($n = 5$), the sampling distribution, like the population distribution, is skewed. But when n is large ($n = 100$), the sampling distribution is well approximated by a standard normal distribution (solid line), as predicted by the central limit theorem. The normal distribution is scaled so that the height of the distributions is approximately the same in all figures.

$$
\begin{aligned}
E[Y] &= E[a + bX] \\
&= \sum_{k=1}^{K} (a + bx_k) P[X = x_k] \\
&= a + b \sum_{k=1}^{K} x_k P[X = x_k] \\
&= a + bE[X]
\end{aligned}
$$

$$
\begin{aligned}
E[(Y - E[Y])^2] &= E[b^2 (X - E[X])^2] \\
&= b^2 E[(X - E[X])^2]
\end{aligned}
$$

$$
\begin{aligned}
E[Y] &= \sum_{k=1}^{K} y_k P[Y = y_k] \\
&= \sum_{k=1}^{K} y_k \sum_{m=1}^{M} P[X = x_m, Y = y_k] \\
&= \sum_{k=1}^{K} y_k \sum_{m=1}^{M} \frac{P[X = x_m, Y = y_k]}{P[X = x_m]} P[X = x_m] \\
&= \sum_{k=1}^{K} y_k \sum_{m=1}^{M} P[Y = y_k \mid X = x_m] P[X = x_m] \\
&= \sum_{k=1}^{K} y_k P[Y = y_k \mid X = x_m] \sum_{m=1}^{M} P[X = x_m] \\
&= \sum_{m=1}^{M} P[X = x_m] \, E[Y \mid X = x_m] \\
&= E\big[E[Y \mid X]\big]
\end{aligned}
$$

$$Var[Y] = E[(Y - E[Y])^2]$$

$$= E\big[E[(Y - E[Y])^2 \mid X]\big]$$

$$= E\left[E\left[\big((Y - E[Y \mid X]) + (E[Y \mid X] - E[Y])\big)^2 \mid X\right]\right]$$

$$= E\big[E[(Y - E[Y \mid X])^2 \mid X]\big] + E\big[E[(E[Y \mid X] - E[Y]^2) \mid X]\big]$$

$$+2E\big[E[(Y - E[Y \mid X])(E[Y \mid X] - E[Y]) \mid X]\big]$$

$$= E\big[E[(Y - E[Y \mid X])^2]\big] + E[(E[Y \mid X] - E[Y]^2)]$$

$$= E\big[Var[Y \mid X]\big] + Var\big[E[Y \mid X]\big]$$

$$E[W] = E[a + bX + cY]$$

$$= a + bE[X] + cE[Y]$$

$$E[(W - E[W])^2] = E\left[\big(b(X - E[X]) + c(Y - E[Y])\big)^2\right]$$

$$= b^2E[(X - E[X])^2] + c^2E[(Y - E[Y])^2]$$

$$+bcE[(X - E[X])(Y - E[Y])]$$

$$E[\bar{Y}] = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[Y_i]$$

$$= \frac{n\mu_Y}{n}$$

$$= \mu_Y$$

$$Var[\bar{Y}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var[Y_i]$$

$$= \frac{n\sigma_Y^2}{n^2}$$

$$= \frac{\sigma_Y^2}{n}$$