A quick review of calculus

- Use calculus to find Min or Max of functions
- Multi-value functions

$$z = -5x^2 + 10xy - 20x - 7y^2 + 240y - 5300$$

- Find the x & y that gives the Max z

$\frac{\partial z}{\partial x} = 0$ and $\frac{\partial z}{\partial y} = 0$ are the first order conditions (FOCs)

- Solve simultaneouslv for x & y

$$\frac{\partial z}{\partial x} = -10x + 10y - 20 = 0$$

$$\frac{\partial z}{\partial y} = 10x - 14y + 240 = 0$$

- Solve simultaneously: $y^* = 55 \quad x^* = 53$
- Plug in x* and y* into z to get the value $z^* = 770$
- Is it a maximum or minimum? Need 2nd order conditions (don't worry here). Here it is the maximum.

Summations

- We'll use summation notation a lot

  If $x = \{8, -3, 4, 2, -1\}$, $n = 5$

  then $\sum_{i=1}^{n} x_i = 8 + (-3) + 4 + 2 + (-1) = 10$

  If $y = \{0, 0, 10, -1, -1\}$, then $\sum y = 8$

  If a is constant, then $\sum_{i=1}^{n} a = \underbrace{a + a + a + \ldots\ldots + a}_{n \text{ times}} = na$

  e.g. $\sum_{i=1}^{99} 10 = \underbrace{10 + 10 + \ldots\ldots + 10}_{99 \text{ times}} = 99 \cdot 10 = 990$

- Note: $\sum x_i \cdot \sum y_i$

  $= 10 \cdot 8$

  $\neq \sum x_i \cdot y_i$

  $= (8 \cdot 0) + (-3 \cdot 0) + (4 \cdot 10) + (2 \cdot -1) + (-1 \cdot -1) = 39$

Summations

- A few important properties

$$\sum ax_i = a\sum x_i$$

$$\sum a = n \bullet a$$

$$\sum x_i y_i = \sum y_i x_i$$

$$\sum (ax_i + by_i) = a\sum x_i + b\sum y_i$$

$$\sum \sum x_i = n\sum x_i$$

Review of Probability Theory

1) Random variable and distribution

- Outcomes- Potential results
- Sample Space- All possible outcomes
- Probability- Proportion of the time that an outcome occurs
- Event- Set of one or more outcomes (subset of sample space)
- e.g.

Grade	Probability
A 4.0	.15
B 3.0	.40
C 2.0	.30
D 1.0	.10
F 0.0	.05

- A,B,C,D,F are each outcomes, together they are sample space
- $\sum prob = 1$ and each prob $\in [0,1]$

1) Random variable and distribution

- Let V=the event that a random student gets an A or B
  In this example "Grade" is a discrete random variable

$$P(V) = P(A) + P(B) = .15 + .40 = .55$$

- Two types of random variables (RV):
  - Discrete random variable: takes only a discrete set of values, like 0,1,2…,
  - Continuous random variable: takes on a continuum of possible values.

1) Random variable and distribution

- Probability Distribution

  For a discrete random variable, the probability distribution lists all the probabilities (e.g. our grade example)

Outcomes	x_1	x_2	…	x_k
Probability	p_1	p_2	…	p_k

1) Random variable and distribution

- For continuous random variable, we have an equation called probability density function that does a similar thing. It provides the probability that the RV takes on a value within an interval.

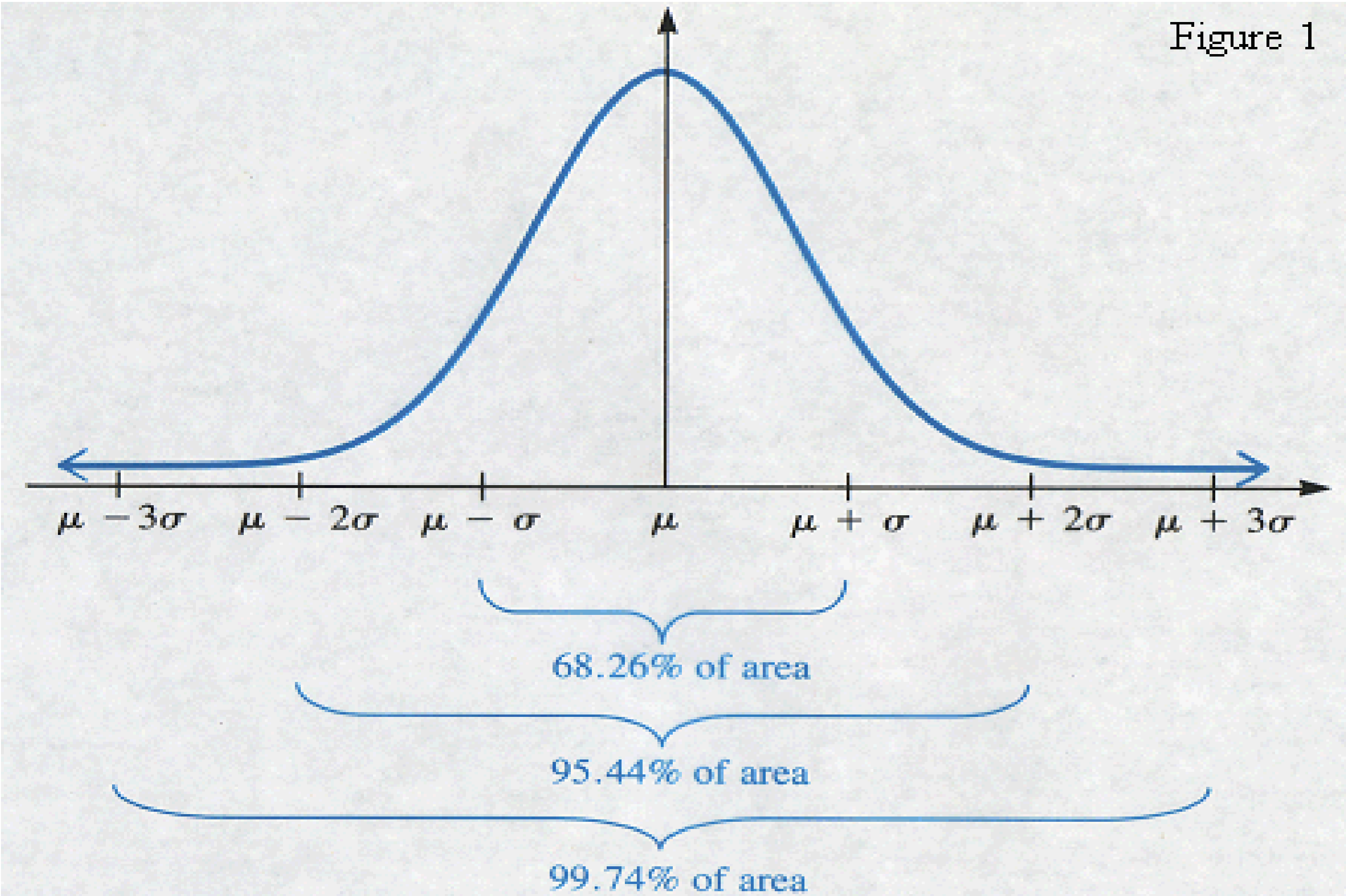
- E.g. Normal Dist:

$$\phi(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

  - When $\mu = 0, \sigma = 1$, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

1) Random variable and distribution

- Normal probability density



1) Random variable and distribution

- CDF (cumulative density function)
- If X is a random variable, then its CDF is defined for any real number x by

$$F(x) = \Pr(X \leq x)$$

- So, $F(-\infty) = 0;\ F(\infty) = 1$, and F takes values between 0 and 1
- If a<b, then F(a) $\leq$ F(b), (that is, F is non-decreasing in its argument)
- If a<b, then Pr(a<X<b)=F(b)-F(a)

1) Random variable and distribution

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

FIGURE 2.1 **Probability Distribution of the Number of Computer Crashes**

The height of each bar is the probability that the computer crashes the indicated number of times. The height of the first bar is 0.8, so the probability of 0 computer crashes is 80%. The height of the second bar is 0.1, so the probability of 1 computer crash is 10%, and so forth for the other bars.

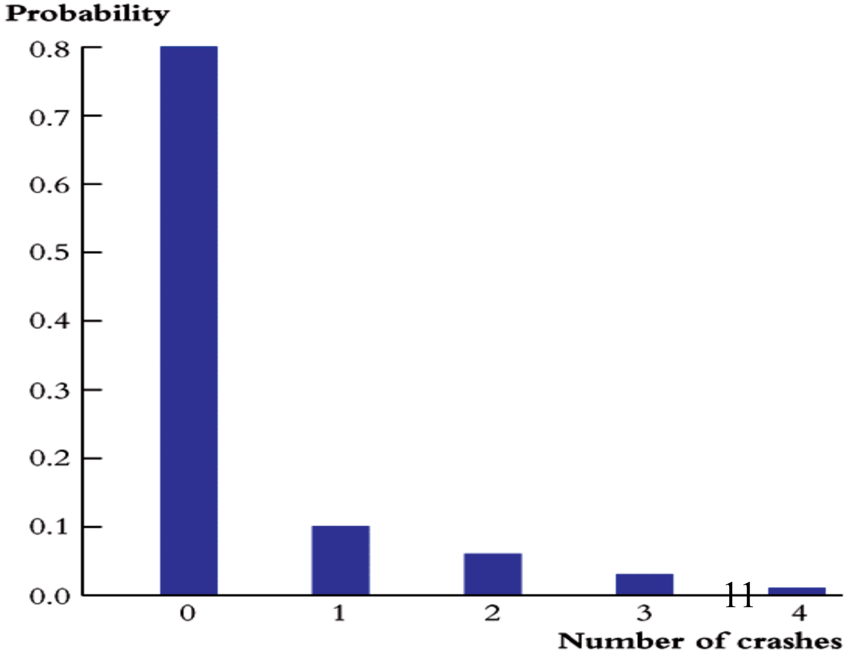
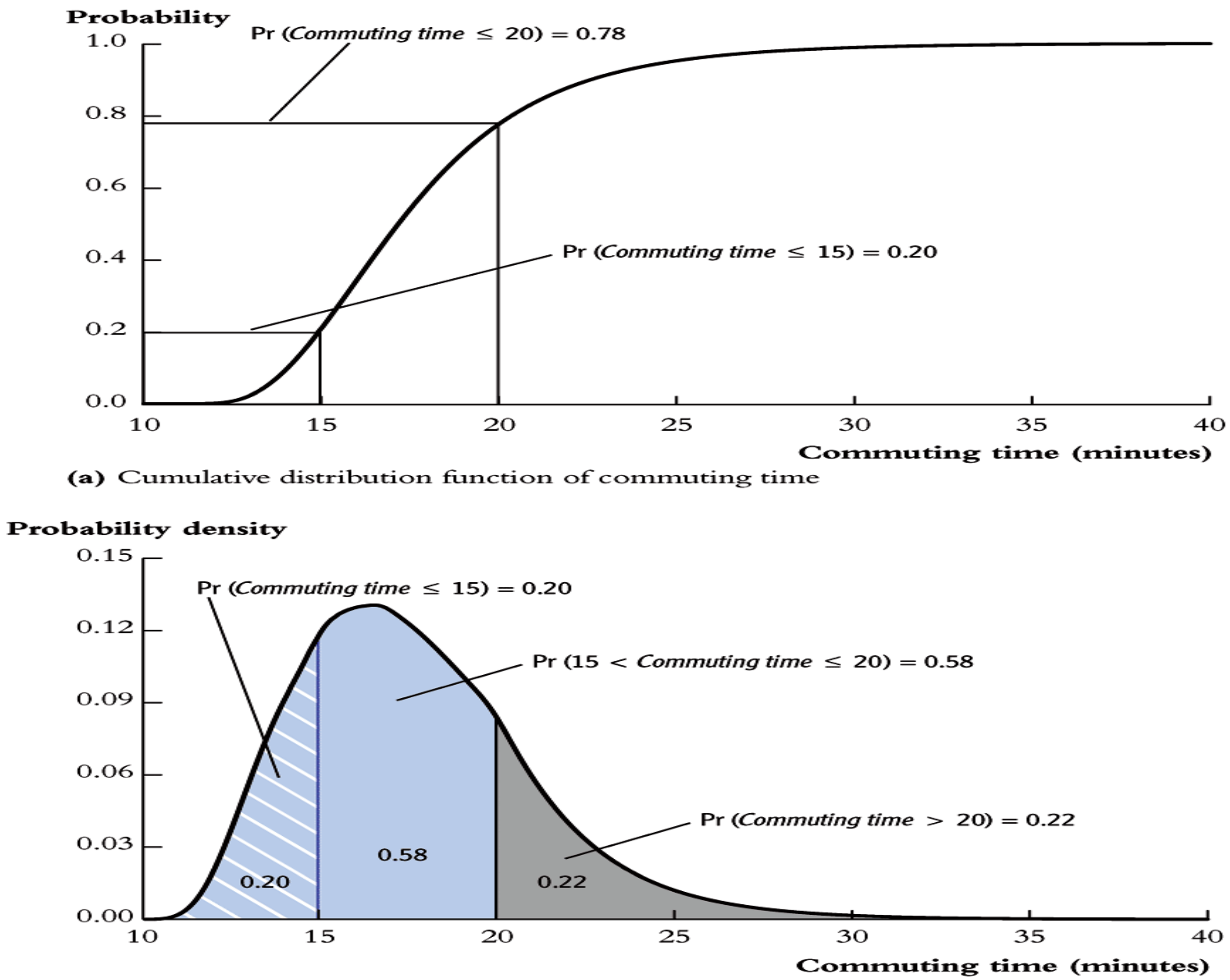


FIGURE 2.2 **Cumulative Distribution and Probability Density Functions of Commuting Time**



(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.

2) Expectation, Mean, and Variance

- Expected value of RV X is the average value of the RV over many repeated trials.

$$E(X) = \mu_X = mean\ of\ X$$

- For discrete random variables:

$$E(X) = \mu_X = \sum_i x_i \bullet p(x_i)$$

-

Grade		Probability
A	4.0	.15
B	3.0	.40
C	2.0	.30
D	1.0	.10
F	0.0	.05

$$E(X) = \sum x_i * P(x_i)$$

$$= 4.0 * .15 + 3.0 * .40 + \ldots\ldots + 0 * .05$$

$$= 2.5$$

2) Expectation, Mean, and Variance

- For a continuous variable

$$E(X) = \mu_X = \int x \cdot f(x)dx,$$

where $f(x)$ is the probability density function (PDF).

2) Expectation, Mean, and Variance

- Variance & Standard deviation are measures of "spread"

$$var(X) = \sigma_X^2 = E\left[(X - \mu_X)^2\right] = E(X^2) - \mu_X^2$$

- The standard deviation is simply

$$\sigma_X = \sqrt{\sigma_X^2}$$

- For a discrete RV,

$$var(X) = \sigma_X^2 = \sum_i (x_i - \mu_X)^2 \cdot p(x_i)$$

- For a continuous RV,

$$var(X) = \int (x - \mu_X)^2 \cdot f(x)dx$$

2) Expectation, Mean, and Variance

- Example:

Grade		Probability
A	4.0	.15
B	3.0	.40
C	2.0	.30
D	1.0	.10
F	0.0	.05

$$\sigma_X^2 = (4-2.5)^2 \bullet 0.15 + (3-2.5)^2 \bullet 0.40 + \ldots$$

$$= 1.05$$

$$\sigma_X = \sqrt{1.05} = 1.025$$

2) Expectation, Mean, and Variance

- Suppose

$$Y = a + bX$$
$$E(Y) = E(a + bX) = a + bE(X)$$
$$Var(Y) = Var(a + bX) = 0 + b^2 Var(X)$$

- E.g. Consumption & Income

$$C = 1000 + 0.9I \qquad E(I) = 4000 \qquad Var(I) = 800$$

- Then

$$E(C) = 1000 + 0.9E(I) = 1000 + 0.9 \cdot 4000 = 4600$$

$$var(C) = 0.9^2 var(I) = 0.81 \cdot 800 = 648$$

2) Expectation, Mean, and Variance

- Measures of Shape of Dist

  1)Mean (Central tendency)

  2)Variance & Standard deviation (Dispersion)

  3) skewness = measure of asymmetry of a distribution

$$skewness = \frac{E\left[(X-\mu_X)^3\right]}{\sigma_X^3}$$

- *Skewness* = 0: distribution is symmetric
- *Skewness* > (<) 0: distribution has long right (left) tail

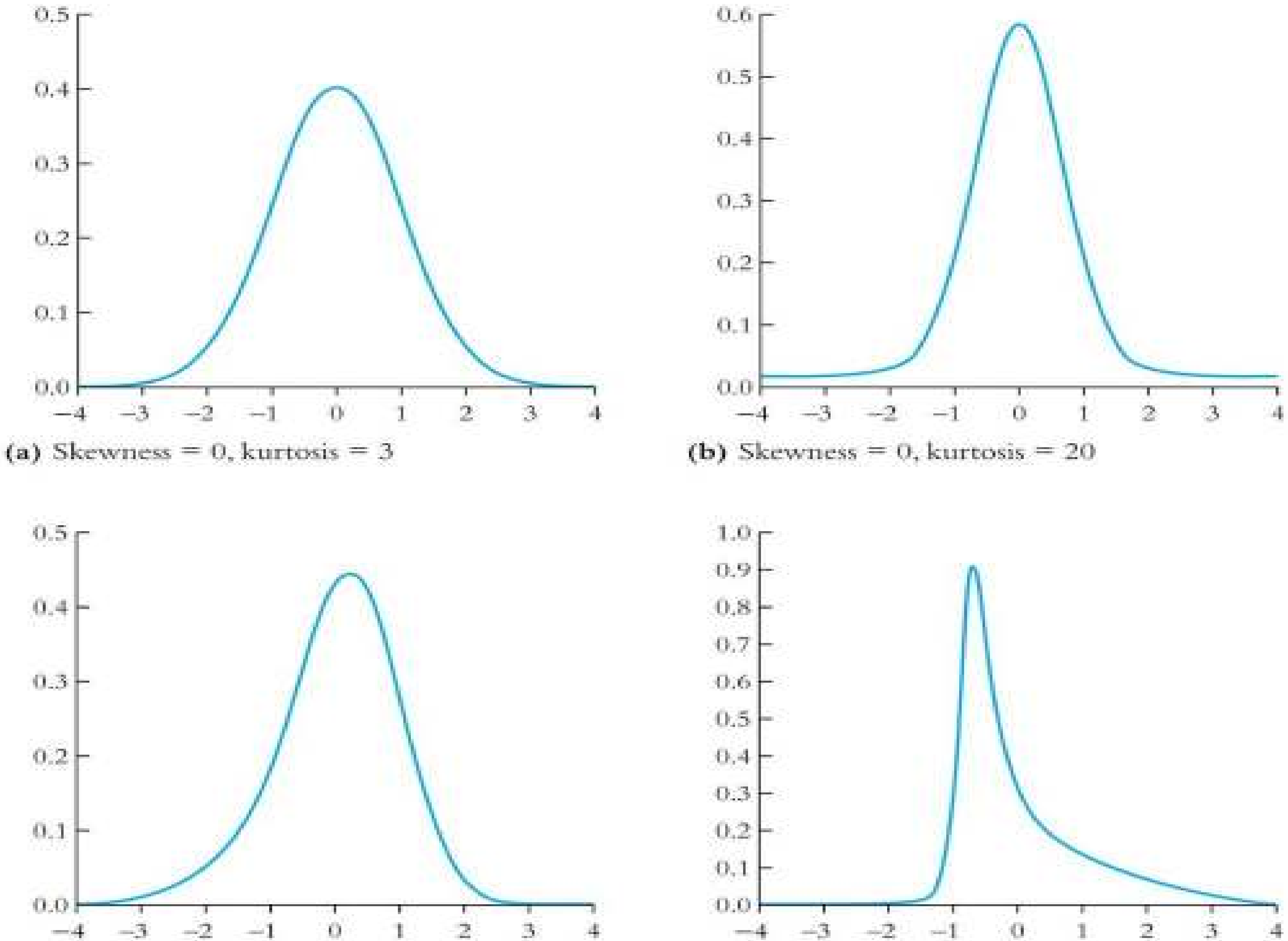
- 4) kurtosis = measure of mass in tails
  = measure of probability of large values

$$kurtosis = \frac{E\left[(X-\mu_X)^4\right]}{\sigma_X^4}$$

- kurtosis = 3: normal distribution
- kurtosis > 3: heavy tails ("*leptokurtotic*")

2) Expectation, Mean, and Variance

➢



(a) Skewness = 0, kurtosis = 3

(b) Skewness = 0, kurtosis = 20

(c) Skewness = −0.1, kurtosis = 5

(d) Skewness = 0.6, kurtosis = 5

3) Two Random Variables

- Consider 2 RVs X & Y

  e.g. X is Income, Y is Gender

  e.g. X is Parent height , Y is Child's height

  e.g. X is Price, Y is Qty demand

- Joint Prob Dist

- Marginal Dist

- Conditional Dist

3) Two Random Variables

- Joint distribution : Random variables X and Y have a ***joint distribution***, i.e., the probability that the random variables **simultaneously** take on certain values, say x and y.
- The time to commute to school, and the weather

Time\weather	0 (raining)	1 (not raining)
40 min	0.08	0.15
50 min	0.20	0.25
60 min	0.28	0.04

- Joint "and"

$$P(X = x, Y = y)$$
$$e.g \quad P(X = 60, Y = 0) = .28$$

3) Two Random Variables

- ***Marginal probability distribution*** of a random variable Y is just another name for its probability distribution.
- This term is used to distinguish the distribution of Y alone and another random variable, say X.
- $$P(Y = y) = \sum_i P(X = x_i, Y = y)$$
  $$P(Y = 0) = \sum_i P(X = x_i, Y = 0) = 0.56$$
-

Time\weather	0 (raining)	1 (not raining)	Marginal of time
40 min	0.08	0.15	0.23
50 min	0.20	0.25	0.45
60 min	0.28	0.04	0.32
Marginal of weather	0.56	0.44	1

3) Two Random Variables

- Conditional Distribution:

$$P(X = x \mid Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$P(Y = y \mid X = x) = \frac{P(X=x, Y=y)}{P(X=x)}$$

- What is the probability of X=60 given that Y=1?

$$P(X = 60 \mid Y = 1) = \frac{0.04}{0.44} = 0.09$$

- What is the probability of Y=1 given that X=60?

$$P(Y = 1 \mid X = 60) = \frac{0.04}{0.32} = 0.125$$

3) Two Random Variables

Time\weather	0 (raining)	1 (not raining)	Marginal of time
40 min	0.08	0.15	0.23
50 min	0.20	0.25	0.45
60 min	0.28	0.04	0.32
Marginal of weather	0.56	0.44	1

Time	Condition distribution given whether =0
40 min	0.08/0.56=0.143
50 min	0.20/0.56=0.357
60 min	0.28/0.56=0.5
sum	1

Time	Condition distribution given whether =1
40 min	0.15/0.44=0.340
50 min	0.25/0.44=0.568
60 min	0.04/0.44=0.091
sum	1

3) Two Random Variables

- X & Y are independent if

$$P(X = x \mid Y = y) = P(X = x)$$

for all the possible x and y.

- i.e. "given" Y does not change the probability
- Since

$$P(X = x \mid Y = y) = \frac{P(x{=}x, Y{=}y)}{P(Y{=}y)}$$

then X & Y are independent of

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

for all the possible x and y.

3) Two Random Variables

- Let's check our time/weather example

Time\weather	0 (raining)	1 (not raining)	Marginal of time
40 min	0.08	0.15	0.23
50 min	0.20	0.25	0.45
60 min	0.28	0.04	0.32
Marginal of weather	0.56	0.44	1

- $$P(X = 60 \mid Y = 1) = 0.09$$

$$P(X = 60) = 0.32$$

$$0.09 \neq 0.32$$

so clearly X & Y are NOT independent

3) Two Random Variables

- Conditional Expectation is the mean of Y given X=x:

$$E(Y \mid X = x) = \sum y_i \cdot P(Y = y_i \mid X = x)$$

- Conditional Expectation is the mean of X given Y=y:

$$E(X \mid Y = y) = \sum x_i \cdot P(X = x_i \mid Y = y)$$

- If X and Y are independent, then

$E(Y \mid X) = E(Y)$ and $E(X \mid Y) = E(X)$

3) Two Random Variables

Time/weather example

Time	Condition distribution given whether =0
40 min	0.08/0.56=0.143
50 min	0.20/0.56=0.357
60 min	0.28/0.56=0.5
sum	1

Time	Condition distribution given whether =1
40 min	0.15/0.44=0.340
50 min	0.25/0.44=0.568
60 min	0.04/0.44=0.091
sum	1

$$E(X \mid Y = 0) = 40 \cdot \frac{0.08}{0.56} + 50 \cdot \frac{0.20}{0.56} + 60 \cdot \frac{0.28}{0.56} = 53.57$$

$$E(X \mid Y = 1) = 40 \cdot \frac{0.15}{0.44} + 50 \cdot \frac{0.25}{0.44} + 60 \cdot \frac{0.04}{0.44} = 47.5$$

3) Two Random Variables

- Suppose that X is a random variable:

$$X = \begin{cases} x_1 \text{ with probability } p_1 \\ x_2 \text{ with probability } p_2 \\ \quad\cdots \\ x_k \text{ with probability } p_k \end{cases}$$

- Suppose that Y is a random variable:

$$Y = \begin{cases} y_1 \text{ with probability } q_1 \\ y_2 \text{ with probability } q_2 \\ \quad\cdots \\ y_m \text{ with probability } q_m \end{cases}$$

3) Two Random Variables

- **Define** a random variable $E(Y \mid X)$ as:

$$E(Y \mid X) = \begin{cases} E(Y \mid X = x_1) \text{ with probability } p_1 \\ E(Y \mid X = x_2) \text{ with probability } p_2 \\ \cdots \\ E(Y \mid X = x_k) \text{ with probability } p_k \end{cases}$$

This is just how we define the random variable $E(Y \mid X)$

- **Define** a random variable $E(X \mid Y)$ as

$$E(X \mid Y) = \begin{cases} E(X \mid Y = y_1) \text{ with probability } q_1 \\ E(X \mid Y = y_2) \text{ with probability } q_2 \\ \cdots \\ E(X \mid Y = y_m) \text{ with probability } q_m \end{cases}$$

This is just how we define the random variable $E(X \mid Y)$

3) Two Random Variables

- Law of Iterated Expectations (LIE):

$$E(Y) = E[E(Y \mid X)]$$

$$\underbrace{E\Bigg[\underbrace{E(Y \mid X)}_{\text{take the expectation with respect to } Y \text{ given } X}\Bigg]}_{\text{take the expectation with respect to } X}$$

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

- In the discrete case

$$E(X) = \sum[E(X \mid Y = y_i) \cdot p(Y = y_i)]$$

$$E(Y) = \sum[E(Y \mid X = x_i) \cdot p(X = x_i)]$$

3) Two Random Variables

- Let's check our time/weather example

Time\weather	0 (raining)	1 (not raining)	Marginal of time
40 min	0.08	0.15	0.23
50 min	0.20	0.25	0.45
60 min	0.28	0.04	0.32
Marginal of weather	0.56	0.44	1

- $$E(X) = \sum x_i \cdot P(X = x_i) = 40 \cdot 0.23 + 50 \cdot 0.45 + 60 \cdot 0.32 = 50.9$$

- We show: $$E(X \mid Y = 0) = 40 \cdot \frac{0.08}{0.56} + 50 \cdot \frac{0.20}{0.56} + 60 \cdot \frac{0.28}{0.56} = 53.57$$

$$E(X \mid Y = 1) = 40 \cdot \frac{0.15}{0.44} + 50 \cdot \frac{0.25}{0.44} + 60 \cdot \frac{0.04}{0.44} = 47.5$$

- By LIE, $$\begin{aligned} E(X) &= \sum E(X \mid Y = y_i) \cdot P(Y = y_i) \\ &= E(X \mid Y = 0) \cdot p(Y = 0) + E(X \mid Y = 1) \cdot p(Y = 1) \\ &= 53.57 \cdot 0.56 + 47.5 \cdot 0.44 \\ &= 50.9 \end{aligned}$$

3) Two Random Variables

- Properties of conditional expectation
- $E[c|X] = c$, where c is a constant
- $E[aY + b|X] = aE(Y|X) + b$
- $E[f(X)|X] = f(X)$
- $E[f_1(X) \cdot f_2(Y)|X] = f_1(X) \cdot E[f_2(Y)|X]$
- $E[Y|X] = E(Y)$ if X and Y are independent
- $E(Y) = E[E[Y|X]]$ (law of iterated expectation)
- Example $E(U|X) = 0$ implies that $E(U) = 0$ and $E(UX) = 0$

Proof: $E(U) = E[E(U|X)] = E(0) = 0$ (law of iterated expectation)

$E(UX) = E[E(UX|X)]$ (law of iterated expectation)

$= E[X \cdot E(U|X)] = E[X \cdot 0] = 0$ ■

3) Two Random Variables

- Conditional Variance:

$$Var(Y \mid X = x) = E\left[[Y - E(Y \mid X = x)]^2 \middle| X = x \right]$$

$$Var(X \mid Y = y) = E\left[[X - E(X \mid Y = y)]^2 \middle| Y = y \right]$$

- Discrete case:

$$var(Y \mid X = x) = \sum \left\{ [y_i - E(Y \mid X = x)]^2 \bullet P(Y = y_i \mid X = x) \right\}$$

$$var(X \mid Y = y) = \sum \left\{ [x_i - E(X \mid Y = y)]^2 \bullet P(X = x_i \mid Y = y) \right\}$$

- If X and Y are independent

$$var(Y \mid X = x) = var(Y)$$

$$var(X \mid Y = y) = var(X)$$

3) Two Random Variables

➢ Conditional Variance:

Time	Condition distribution given whether =0
40 min	0.08/0.56=0.143
50 min	0.20/0.56=0.357
60 min	0.28/0.56=0.5
sum	1

$$E(X \mid Y=0) = 40 \cdot \frac{0.08}{0.56} + 50 \cdot \frac{0.20}{0.56} + 60 \cdot \frac{0.28}{0.56} = 53.57$$

$$var(X \mid Y=0)$$

$$= (40-53.57)^2 \cdot \frac{0.08}{0.56} + (50-53.57)^2 \frac{0.20}{0.56} + (60-53.57)^2 \frac{0.28}{0.56}$$

$$= 51.53$$

3) Two Random Variables

Econ 3334

- Covariance: (linear) statistical relationship between variables

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

- Correlation:

$$corr(X,Y) = \rho_{XY} = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

- $$-1 \leq corr \leq 1$$

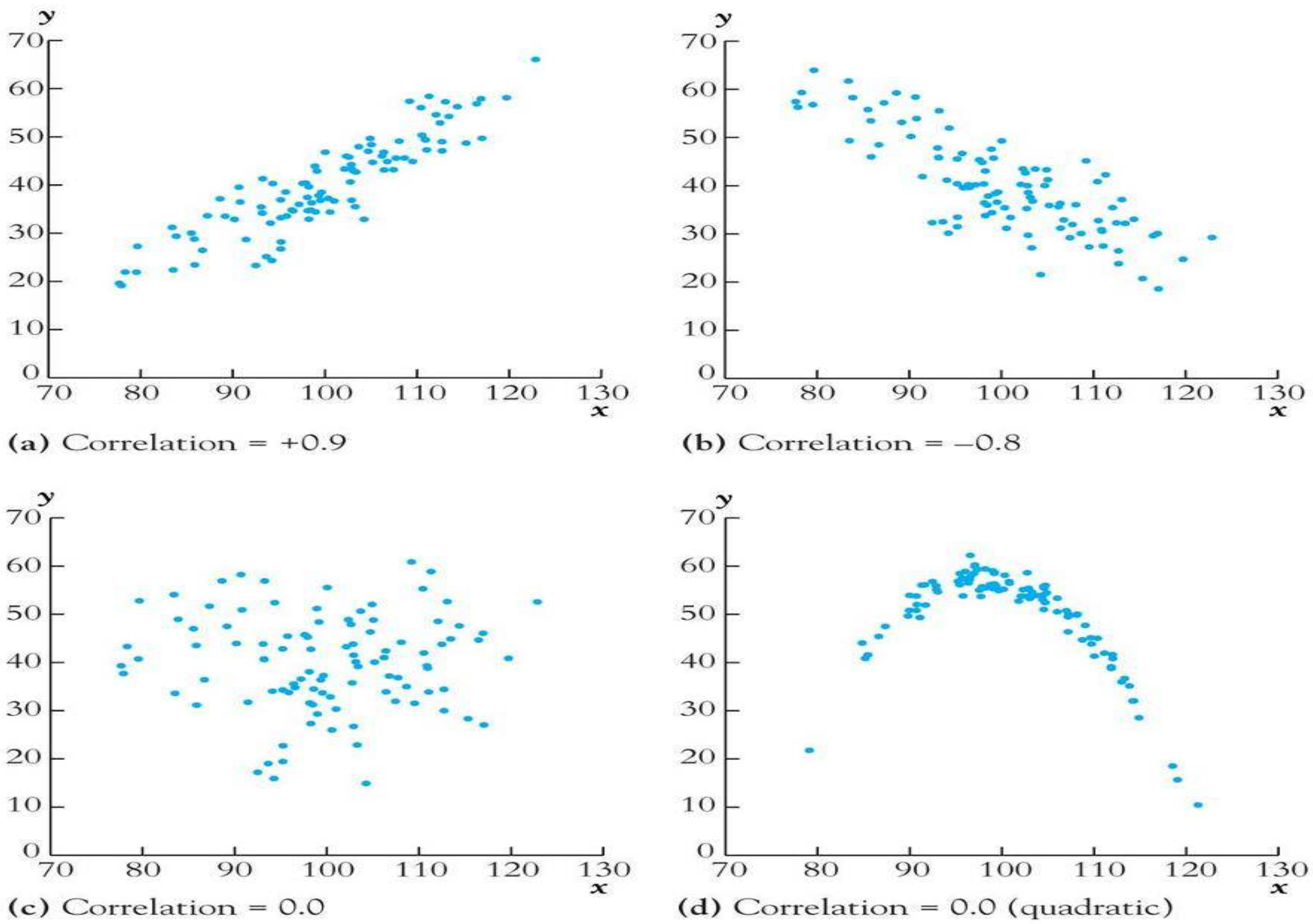
- If X and Y are independent, then corr(X,Y)=0

3) Two Random Variables

- corr(X,Y) = 1 mean perfect positive linear association
- corr(X,Y) = –1 means perfect negative linear association
- corr(X,Y) = 0 means no linear association

3) Two Random Variables

The correlation coefficient measures linear association



3) Two Random Variables

- Mean and variance of sums of RVs

$$E(aX + bY) = aE(X) + bE(Y)$$

$$var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab \cdot cov(X, Y)$$

- If X and Y are independent

$$var(aX + bY) = a^2 var(X) + b^2 var(Y)$$