Topic 5: Probability Distributions

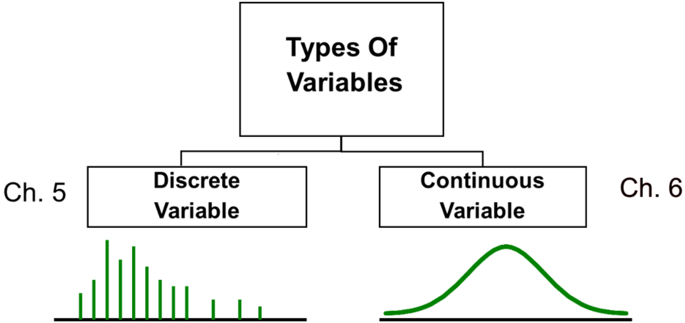
Dr. Lei Pan

School of Accounting, Economics and Finance
Curtin University

Definitions

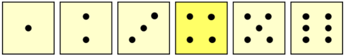
- **Discrete** variables produce outcomes that come from a counting process (e.g. number of classes you are taking).
- **Continuous** variables produce outcomes that come from a measurement (e.g. your height, or your weight).

Type of Variables



Discrete Variables

- Can only assume a countable number of values
  - Example 1: Roll a dice twice. Let X be the number of times 4 occurs(then X could be 0, 1, or 2 times).



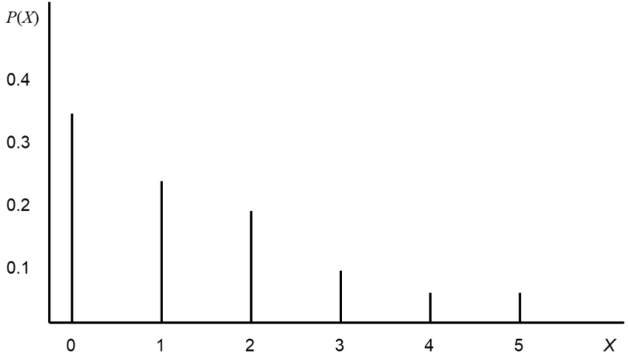
  - Example 2: Toss a coin 5 times. Let X be the number of heads (then $X = 0, 1, 2, 3, 4$ or 5).

Probability Distribution for a Discrete Variable

- A **probability distribution for a discrete variable** is a mutually exclusive listing of all possible numerical outcomes for that variable and a probability of occurrence associated with each outcome.

Interruptions Per Day In Computer Network	Probability
0	0.35
1	0.25
2	0.20
3	0.10
4	0.05
5	0.05

Probability Distribution Are Often Represented Graphically



Discrete Variables Expected Value

- The **expected value** (or mean) **of a discrete random variable** (weighted average) is
  - a measure of central tendency
  - the mean of a discrete random variable

$$\mu = E(X) = \sum_{i=1}^{N} x_i P(X = x_i)$$

Example

Interruptions Per Day In Computer Network (x_i)	Probability $P(X=x_i)$	$x_iP(X=x_i)$
0	0.35	(0)(0.35) = 0.00
1	0.25	(1)(0.25) = 0.25
2	0.20	(2)(0.20) = 0.40
3	0.10	(3)(0.10) = 0.30
4	0.05	(4)(0.05) = 0.20
5	0.05	(5)(0.05) = 0.25
	1.00	$\mu = E(X) = 1.40$

Variance of A Discrete Random Variable

- **Variance** of a discrete variable – *definition* formula

$$\sigma^2 = \sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i)$$

- **Variance** of a discrete variable – *calculation* formula

$$\sigma^2 = \sum_{i=1}^{N} X_i^2 P(X_i) - E(X)^2$$

where $E(X)$: expected value of the discrete random variable X

X_i: the ith outcome of the discrete random variable X

$P(X_i)$: probability of the ith occurrence of X

Example

Toss 2 coins, X = No. of heads, calculate variance.

Answer:

$$\begin{aligned} E(X) &= (0 \times 0.25) + (1 \times 0.50) + (2 \times 0.25) \\ &= 1.0 \\ \sigma^2 &= \sum_{i=1}^{N} X_i^2 P(X_i) - E(X)^2 \\ &= (0^2 \times 0.25 + 1^2 \times 0.5 + 2^2 \times 0.25) - (1^2) \\ &= 0.5 \end{aligned}$$

Standard Deviation of A Discrete Random Variable

- **Standard deviation** of a discrete variable

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N} [X_i - E(X)]^2 P(X_i)}$$

where $E(X)$: expected value of the discrete random variable X
X_i: the ith outcome of the discrete random variable X
$P(X_i)$: probability of the ith occurrence of X

Covariance

- The **covariance** measures the direction of a linear relationship between two variables.
  - Covariance – definition formula

$$\sigma_{XY} = \sum_{i=1}^{N} [X_i - E(X)][Y_i - E(Y)]P(X_i Y_i)$$

  - Covariance – calculation formula

$$\sigma_{XY} = \sum_{i=1}^{N} X_i Y_i P(X_i Y_i) - E(X)E(Y)$$

where

$X_i Y_i$: the ith outcome of the discrete random variables X and Y respectively

$P(X_i Y_i)$: probability of the ith occurrence of X and Y

Example: Compute the Mean for Investment Returns

Return per $1,000 for two types of investments.

		Investment	
$P(X_iY_i)$	Economic condition	Passive Fund X	Aggressive Fund Y
0.2	Recession	-$25	-$200
0.5	Stable Economy	+50	+60
0.3	Expanding Economy	+100	+350

- $E(X) = \mu_X = (-25) \times (0.2) + (50) \times (0.5) + (100) \times (0.3) = 50$
- $E(Y) = \mu_Y = (-200) \times (0.2) + (60) \times (0.5) + (350) \times (0.3) = 95$

Example: Compute the Standard Deviation for Investment Returns

		Investment	
$P(X_iY_i)$	Economic condition	Passive Fund X	Aggressive Fund Y
0.2	Recession	-$25	-$200
0.5	Stable Economy	+50	+60
0.3	Expanding Economy	+100	+350

$$\sigma_X = \sqrt{(-25)^2 \times (0.2) + (50)^2(0.5) + (100)^2(0.3)} - (50)^2$$
$$= 43.30$$
$$\sigma_Y = \sqrt{(-200)^2 \times (0.2) + (60)^2(0.5) + (350)^2(0.3)} - (95)^2$$
$$= 193.71$$

Curtin University

Example: Compute the Covariance for Investment Returns

		Investment	
$P(X_iY_i)$	Economic condition	Passive Fund X	Aggressive Fund Y
0.2	Recession	-\$25	-\$200
0.5	Stable Economy	+50	+60
0.3	Expanding Economy	+100	+350

$$\sigma_{XY} = \sum_{i=1}^{N} X_i Y_i P(X_i Y_i) - E(X)E(Y)$$

$$= [((-25) \times (-200) \times 0.2) + (50 \times 60 \times 0.5) + (100 \times 350 \times 0.3)] - [(50) \times (95)]$$

$$= 8205$$

Interpreting the Results

- The aggressive fund has a higher expected return (mean), but much more risk.

$$\mu_Y = 95 > \mu_X = 50$$
$$\sigma_Y = 193.71 > \sigma_X = 43.30$$
$$\sigma_{XY} = 8250$$

- The covariance of 8,250 indicates that the two investments are positively related and will vary in the same direction.
- Correlation – A standardised version of covariance.

A Little More About the Expected Value

- **Expected value of the sum of two random variables**
  - measures of central tendency
  - mean of the sum of two random variables

$$
\begin{aligned}
E(X+Y) &= E(X)+E(Y) \\
E(a) &= a \\
E(aX+bY) &= aE(X)+bE(Y)
\end{aligned}
$$

where a and b are constants

A Little More About Variance and Standard Deviation

- **Variance of the sum of two random variables**
  - measures of variation
  - directly related to the standard deviation

$$\mathrm{Var}(X + Y) = \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY}$$

- **Standard deviation of the sum of two random variables**
  - measures of variation
  - directly related to the variance

$$\sigma_{XY} = \sqrt{\sigma^2_{X+Y}}$$

Curtin University

YEARS 50 OF INNOVATION

Portfolio Expected Return and Portfolio Risk

- **Portfolio**: a combined investment in two or more assets
- **Portfolio expected return**
  - measure of central tendency
  - mean return on investment

$$E(P) = wE(X) + (1 - w)E(Y)$$

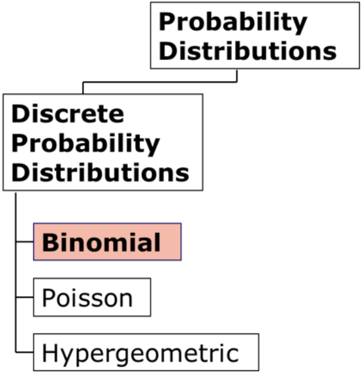
- **Portfolio risk**: measure of the variation of investment returns

$$\sigma_P = \sqrt{w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 + 2w(1-w)\sigma_{XY}}$$

where

w: portion of portfolio value in asset X
$(1 - w)$: portion of portfolio value in asset Y

Probability Distributions



Binomial Distribution (1 of 2)

Four essential properties of the binomial distribution:

1. A fixed number of observations, or *trials*, n.
   - e.g. 15 tosses of a coin; ten light bulbs taken from a warehouse
2. Each observation is categorized as to whether or not the “event of interest” occurred.
   - e.g. head or tail in each toss of a coin; defective or not defective light bulb
   - Since these two categories are mutually exclusive and collectively exhaustive
     - When the probability of the event of interest is represented as p, then the probability of the event of interest not occurring is $1 - p$

Binomial Distribution (2 of 2)

3. Constant probability for the event of interest occurring (p) for each observation
   - Probability of getting a tail is the same each time we toss the coin
4. Observations are independent
   - The outcome of one observation does not affect the outcome of the other
   - Two sampling methods deliver independence
     - Infinite population without replacement
     - Finite population with replacement

Possible Applications for the Binomial Distribution

- A manufacturing plant labels items as either defective or acceptable.
- A firm bidding for contracts will either get a contract or not.
- A marketing research firm receives survey responses of "yes I will buy" or "no I will not".
- New job applicants either accept the offer or reject it.

The Binomial Distribution Counting Techniques

- Suppose the event of interest is obtaining heads on the toss of a fair coin. You are to toss the coin three times. In how many ways can you get two heads?
- Possible ways: HHT, HTH, THH, so there are three ways you can getting two heads.
- This situation is fairly simple. We need to be able to count the number of ways for more complicated situations.

Rule of Combinations

The number of combinations of selecting X objects out of n objects is:

$$\binom{n}{X} = \frac{n!}{X!(n-X)!}$$

where

$$n! = (n)(n-1)(n-2)...(2)(1)$$
$$X! = (X)(X-1)(X-2)...(2)(1)$$
$$0! = 1 \text{ (by definition)}$$

Example

- How many possible 3 scoop combinations could you create at an ice cream parlor if you have 31 flavors to select from and no flavor can be used more than once in the 3 scoops?
- The total choices is $n = 31$, and we select $X = 3$.

$$\binom{31}{3} = \frac{31!}{3!(31-3)!} = \frac{31 \cdot 30 \cdot 29 \cdot 28!}{3 \cdot 2 \cdot 1 \cdot 28!} = 31 \cdot 5 \cdot 29 = 4,495$$

Binomial Distribution Formula

$$P(X) = \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X}$$

where

$P(X)$: probability of X successes in n trials, with probability of success p on each trial

X: number of 'successes' in sample ($X = 0, 1, 2, ..., n$)

n: sample size (number of trials or observations)

p: probability of 'success'

$(1-p)$: probability of failure

Example

Flip a coin four times, let $X = \#$ heads

$$\begin{aligned} n &= 4 \\ p &= 0.5 \\ 1 - p &= (1 - 0.5) = 0.5 \\ X &= 0, 1, 2, 3, 4 \end{aligned}$$

Calculating Example 1

What is the probability of one success in five observations if the probability of an event of interest is 0.1?

$$X = 1, n = 5, \text{ and } p = 0.1$$

$$\begin{aligned} P(X) &= \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X} \\ &= \frac{5!}{1!(5-1)!} (0.1)^1 (1-0.1)^{5-1} \\ &= (5)(0.1)(0.9)^4 \\ &= 0.3281 \end{aligned}$$

Calculating Example 2

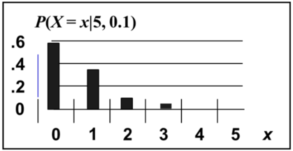
Suppose the probability of purchasing a defective computer is 0.02. What is the probability of purchasing 2 defective computers in a group of 10?

$$X = 2, n = 10, \text{ and } p = 0.02$$

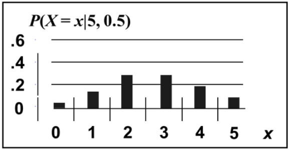
$$\begin{aligned} P(X) &= \frac{n!}{X!(n-X)!} p^X (1-p)^{n-X} \\ &= \frac{10!}{2!(10-2)!} (0.02)^2 (1-0.02)^{10-2} \\ &= (45)(0.0004)(0.8508) \\ &= 0.0153 \end{aligned}$$

The Binomial Distribution Shape

- The shape of the binomial distribution depends on the values of p and n.
  - Case 1: $n = 5$ and $p = 0.1$



  - Case 2: $n = 5$ and $p = 0.5$



Binomial Distribution Characteristics

- Mean

$$\mu = E(X) = np$$

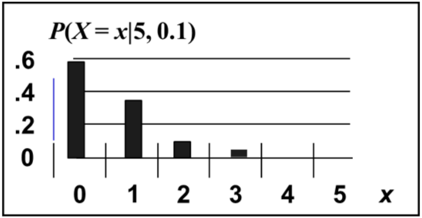
- Variance and standard deviation

$$\sigma^2 = np(1-p)$$
$$\sigma = \sqrt{np(1-p)}$$

where

n: sample size
p: probability of success
$(1-p)$: probability of failure

Example



$$\mu = np = (5)(0.1) = 0.5$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{(5)(0.1)(1-0.1)}$$
$$= 0.6708$$

Demonstration – Binomial Distribution

- What the distribution look like for different values of the parameters?
- Calculating probabilities from the distribution. (Using Excel)
- Mean and variance of the distribution.

You Do It

A student decides to guess on the multiple choice section of his ECOM5005 final exam. The section contains 50 questions and each question has 5 possible answers.

(a) Find the expected number of correct responses.

(b) Find the standard deviation of the number of correct responses (keep 4 decimal points).

Poisson Distribution

- You use the **Poisson distribution** when you are interested in the number of times an event occurs in a given **area of opportunity**.
- An **area of opportunity** is a continuous unit or interval of time, volume, or such area in which more than one occurrence of an event can occur.
    - The number of scratches in a car's paint.
    - The number of mosquito bites on a person.
    - The number of computer crashes in a day.

Poisson Distribution Formula

- The distribution has one parameter λ, which is the mean or expected number of events per interval.

$$P(X) = \frac{e^{-\lambda}\lambda^{x}}{X!}$$

where

$P(X)$: the probability of X events in a given interval
λ: expected number of events in the given interval
e: base of the natural logarithm system (2.71828...)

Poisson Distribution Characteristics

- **Mean**

$$\mu = \lambda$$

- **Variance and standard deviation**

$$\sigma^2 = \lambda$$

$$\sigma = \sqrt{\lambda}$$

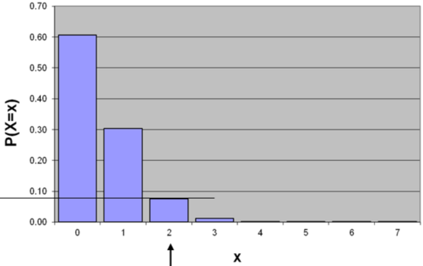
where λ: expected number of events

Graph of Poisson Probabilities

Graphically:

$\lambda = 0.50$

X	$\lambda = 0.50$
0	0.6065
1	0.3033
2	0.0758
3	0.0126
4	0.0016
5	0.0002
6	0.0000
7	0.0000

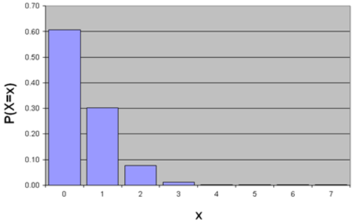


$P(X = 2 \mid \lambda = 0.50) = 0.0758$

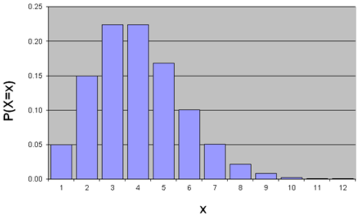
Poisson Distribution Shape

- The shape of the Poisson Distribution depends on the parameter λ.

$\lambda = 0.50$



$\lambda = 3.00$



Curtin University

Demonstration – Poisson Distribution

- What the distribution look like for different values of the parameters?
- Calculating probabilities from the distribution. (Using Excel)
- Mean and variance of the distribution.

Hypergeometric Distribution

- It is a discrete probability distribution.
- The random variable is the number of successes in a sample of n observations from a finite population without replacement.
- Outcomes of trials are dependent.
- Concerned with finding the probability of 'X' successes in the sample where there are 'A' successes in the population.

Hypergeometric Distribution Formula

$$P(X) = \frac{\binom{A}{X}\binom{N-A}{n-X}}{\binom{N}{n}}$$

where

$P(X)$: probability of X successes, given n, N and A

N: population size

A: number of successes in the population

$N - A$: number of failures in the population

n: sample size

X: number of successes in the sample

$n - X$: number of failures in the sample

Hypergeometric Distribution Characteristics

- **Mean**

$$\mu = E(x) = \frac{nA}{N}$$

- **Variance and standard deviation**

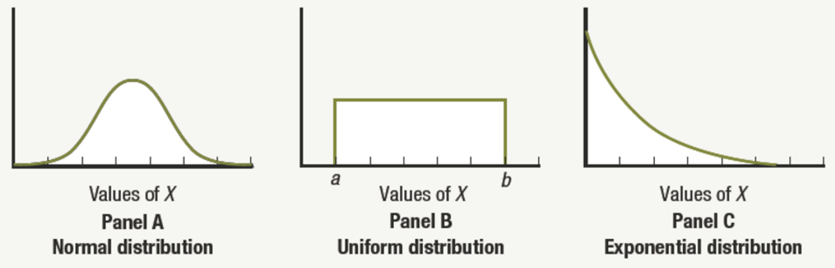
$$\sigma = \sqrt{\frac{nA(N-A)}{N^2}} \cdot \sqrt{\frac{N-n}{N-1}}$$

where $\sqrt{\frac{N-n}{N-1}}$ is called the **finite population correction factor** from sampling without replacement from a finite population

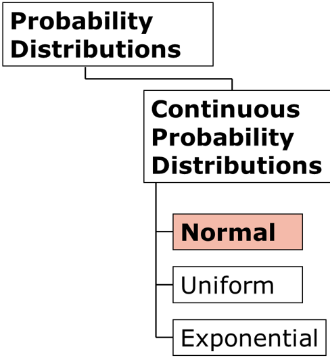
Continuous Probability Distributions (1 of 2)

- A **continuous random variable** is a variable that can assume any value on a continuum (can assume an uncountable number of values).
  - thickness of an item
  - time required to complete a task
  - height, in cm
- These can potentially take on any value depending only on the ability to precisely and accurately measure.

Continuous Probability Distributions (2 of 2)

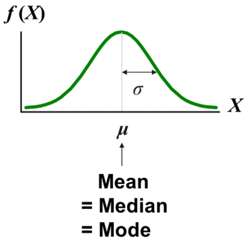


The Normal Distribution (1 of 2)



The Normal Distribution (2 of 2)

- Bell shaped
- Symmetrical
  - Mean, median and mode are equal.
  - Central location is determined by the mean, μ.
  - Spread is determined by the standard deviation, σ.
  - The random variable has an infinite theoretical range: $+\infty$ to $-\infty$.



The Normal Distribution Probability Density Function

The formula for the normal probability function is:

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{X-\mu}{\sigma}\right]^2}$$

where

e: mathematical constant approximated by 2.71828

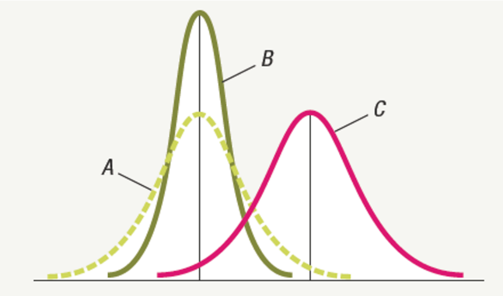
π: mathematical constant π approximated by 3.1415927

μ: population mean

σ: population standard deviation

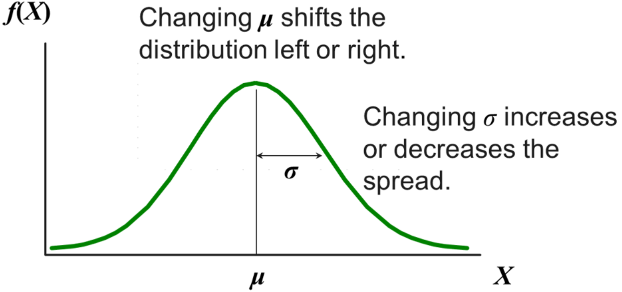
X: any value of the continuous variable

Different Normal Distributions by Varying the Parameters μ and σ



- A and B have the same mean but different standard deviations.
- A and C have different means but same standard deviation.

The Normal Distribution Shape



Demonstration - Normal Distribution

- What the distribution look like for different values of the parameters?
- Calculating probabilities from the distribution. (Using Excel)
- Mean and variance of the distribution.

Translation to the Standardized Normal Distribution

- Any normal distribution (with any mean and standard deviation combination) can be transformed into the standardized normal distribution (Z).
- To compute normal probabilities need to transform X units into Z units.
- Translate from X to the standardized normal (the "Z" distribution) by subtracting the mean of X and dividing by its standard deviation:

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

The Standardised Normal Probability Density Function

The formula for the standardised normal probability density function is:

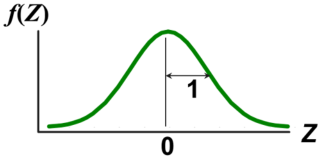
$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{2})Z^2}$$

where

e: mathematical constant approximated by 2.71828
π: mathematical constant approximated by 3.1415927
Z: any value of the standardised normal distribution

The Standardised Normal Distribution

- Also known as the "Z distribution".
- Mean is 0.
- Standard deviation is 1.
- Values above the mean have positive Z-values.
- Values below the mean have negative Z-values.



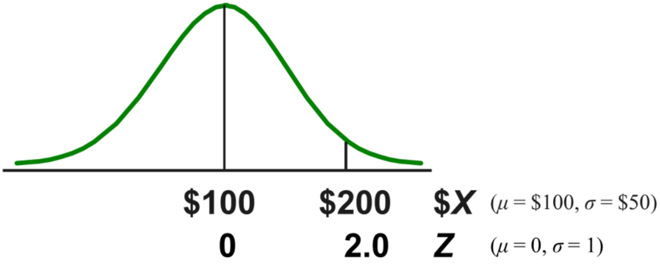
Example

- If X is distributed normally with mean of \$100 and standard deviation of \$50, the Z value for $X = \$200$ is

$$Z = \frac{X - \mu}{\sigma} = \frac{\$200 - \$100}{\$50} = 2.0$$

- This says that $X = \$200$ is two standard deviations (2 increments of \$50 units) above the mean of \$100.

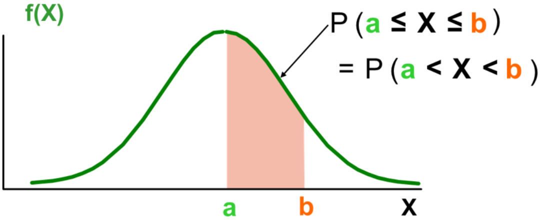
Comparing X and Z units



Note that the shape of the distribution is the same, only the scale has changed. We can express the problem in the original units (X in dollars) or in standardised units (Z).

Finding Normal Probabilities

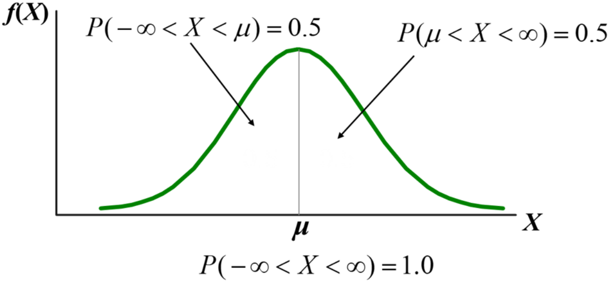
- Probability is measured by the area under the curve.



- Note that the probability of any individual value is zero since the X axis has an infinite theoretical range: $+\infty$ to $-\infty$.

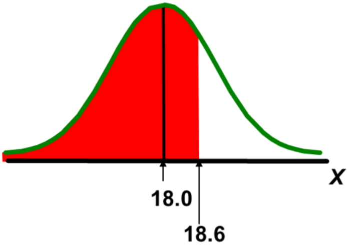
Probability as Area Under the Curve

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below.



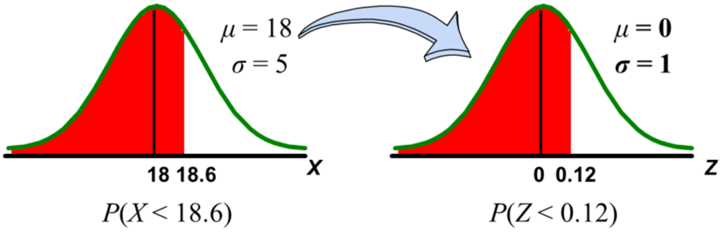
Finding Normal Probabilities

- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with a mean of 18.0 seconds and a standard deviation of 5.0 seconds. Find $P(X < 18.6)$

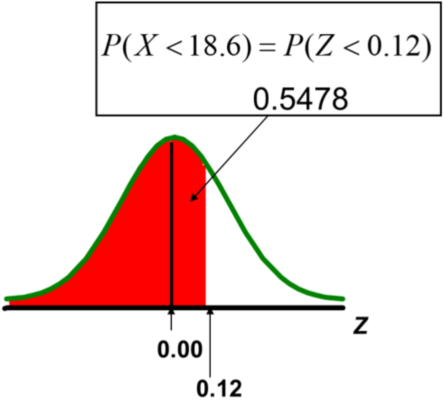


Solution (1 of 2)

$$Z = \frac{X - \mu}{\sigma} = \frac{18.6 - 18.0}{5.0} = 0.12$$

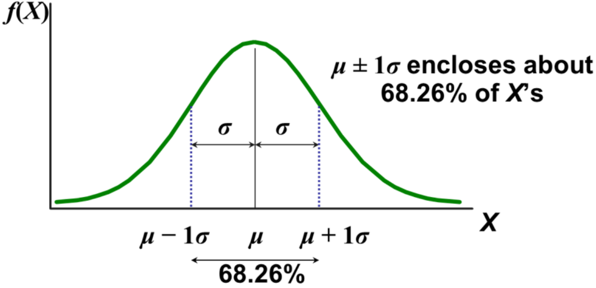


Solution (2 of 2)



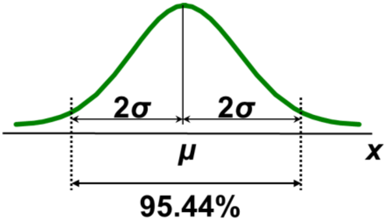
Empirical Rule (1 of 2)

What can we say about the distribution of values around the mean? For any normal distribution:



Empirical Rule (2 of 2)

- $\mu \pm 2\sigma$ covers about 95.44% of X's.
- $\mu \pm 3\sigma$ covers about 99.73% of X's.



Evaluating Normality (1 of 2)

- Not all continuous distributions are normal.
- It is important to evaluate how well the data set is approximated by a normal distribution.
- Construct charts and observe their appearance:
  - For small- or moderate-sized data sets, construct a stem-and-leaf display or a boxplot to check for symmetry.
  - For large data sets, does the histogram or polygon appear bell-shaped?
- Calculate descriptive summary measures
  - Do the mean, median and mode have similar values? (Remember there may be no unique mode or there may be multiple modes.)
  - Is the interquartile range approximately 1.33σ?
  - Is the range approximately 6σ?

Evaluating Normality (2 of 2)

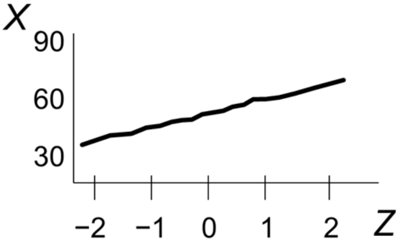
- Observe the distribution of the data set
  - Do approximately 2/3 of the observations lie within mean ± 1 standard deviation?
  - Do approximately 80% of the observations lie within mean ± 1.28 standard deviation?
  - Do approximately 95% of the observations lie within mean ± 2 standard deviation?
- Evaluate normal probability plot
  - Is the normal probability plot approximately linear (i.e. a straight line) with positive slope?

Constructing a Normal Probability Plot

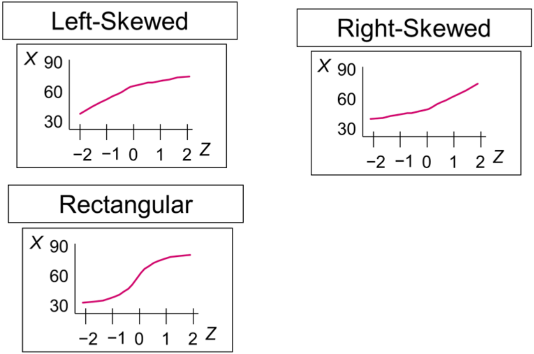
- A **normal probability plot** is a graphical approach used to evaluate if data are normal.
- A common approach is the **quantile-quantile plot**.
  - Arrange data into ordered array.
  - Find corresponding standardized normal quantile values (Z).
  - Plot the pairs of points with observed data values (X) on the vertical axis and the standardized normal quantile values (Z) on the horizontal axis.
  - Evaluate the plot for evidence of linearity.

The Normal Probability Plot Interpretation

A normal probability plot for data from a normal distribution will be approximately linear:



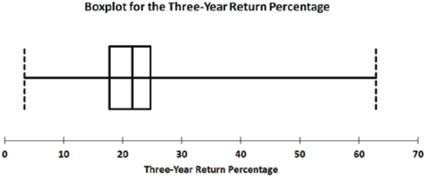
Normal Probability Plots



Nonlinear plots indicate a deviation from normality.

Evaluating Normality an Example: Mutual Fund Returns (1 of 5)

Five-Number Summary	
Minimum	3.39
First quartile	17.76
Median	21.65
Third quartile	24.74
Maximum	62.91



The boxplot is skewed to the right. (The normal distribution is symmetric.)

Evaluating Normality an Example: Mutual Fund Returns (2 of 5)

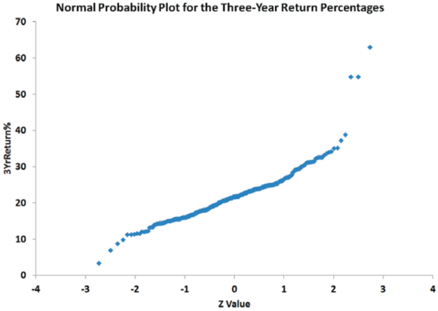
Descriptive Statistics

	3YrReturn%
Mean	21.84
Median	21.65
Mode	21.74
Minimum	3.39
Maximum	62.91
Range	59.52
Variance	41.2968
Standard Deviation	6.4263
Coeff. of Variation	29.43%
Skewness	1.6976
Kurtosis	8.4670
Count	318
Standard Error	0.3604

Evaluating Normality an Example: Mutual Fund Returns (3 of 5)

- The mean (21.84) is approximately the same as the median (21.65). (In a normal distribution the mean and median are equal.)
- The interquartile range of 6.98 is approximately 1.09 standard deviations. (In a normal distribution the interquartile range is 1.33 standard deviations.)
- The range of 59.52 is equal to 9.26 standard deviations. (In a normal distribution the range is 6 standard deviations.)
- 77.04% of the observations are within 1 standard deviation of the mean. (In a normal distribution this percentage is 68.26%.)
- 86.79% of the observations are within 1.28 standard deviations of the mean. (In a normal distribution this percentage is 80%.)
- 96.86% of the observations are within 2 standard deviations of the mean. (In a normal distribution this percentage is 95.44%.)
- The skewness statistic is 1.698 and the kurtosis statistic is 8.467. (In a normal distribution, each of these statistics equals zero.)

Evaluating Normality an Example: Mutual Fund Returns (4 of 5)



Plot is not a straight line and shows the distribution is skewed to the right. (The normal distribution appears as a straight line.)

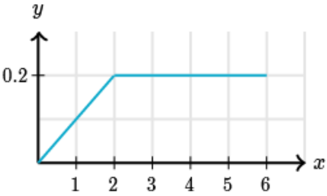
Evaluating Normality an Example: Mutual Fund Returns (5 of 5)

Conclusions:

- The returns are right-skewed.
- The returns have more values concentrated around the mean than expected.
- The range is larger than expected.
- Normal probability plot is not a straight line.
- Overall, this data set greatly differs from the theoretical properties of the normal distribution.

You Do It

Consider the density curve below.



Find the probability that x is less than 2.

The Uniform Distribution (1 of 2)

- The **uniform distribution** is a probability distribution that has equal probabilities for all possible outcomes of the random variable.
- It is also called a **rectangular distribution**.

The Uniform Distribution (2 of 2)

The continuous uniform distribution

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq X \leq b \\ 0, & \text{otherwise} \end{cases}$$

where

$f(x) =$ value of the density function at any X value
$a =$ minimum value of X
$b =$ maximum value of X

Characteristics of the Uniform Distribution

- The mean of the uniform distribution is:

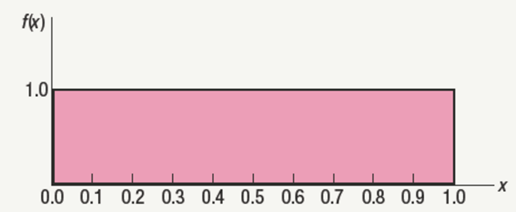
$$\mu = \frac{a+b}{2}$$

- The standard deviation is:

$$\sigma = \sqrt{\frac{(b-a)^2}{12}}$$

Uniform Distribution Probability Density Function

Probability density function for a uniform distribution with $a = 0$ and $b = 1$.



The Exponential Distribution (1 of 2)

- Exponential distribution is a continuous distribution which is right skewed and ranges from 0 to $+\infty$.
- Often used to model the **length of time between two occurrences** of random or independent events (the time between events).
  - time between trucks arriving at an unloading dock
  - time between transactions at an ATM machine
  - time between phone calls to the main operator

The Exponential Distribution (2 of 2)

Defined by a single parameter, its mean λ, the expected number of events per interval

$$P(X < A) = 1 - e^{-\lambda A}$$

where

$e =$ mathematical constant approximated by 2.71828
$\lambda =$ the expected number of events in interval
$X =$ an exponential random variable where $0 \leq X \leq \infty$

Example

Customers arrive at the service counter at the rate of 15 per hour.What is the probability that the arrival time between consecutive customers is less than three minutes?

Solution:

- The mean number of arrivals per hour is 15, so $\lambda = 15$.
- Three minutes is 0.05 hours, so $A = 0.05$.
- $P(X < 0.05) = 1 - e^{-\lambda A} = 1 - e^{-(15)(0.05)} = 0.5276$.
- So there is a 52.76% probability that the arrival time between consecutive customers is less than three minutes.

Normal Approximation to the Binomial Distribution (1 of 3)

- The binomial distribution is a discrete distribution whereas the normal distribution is continuous.
- To use the normal to approximate the binomial, accuracy is improved if you use a continuity correction.
- Example: X is discrete in a binomial distribution, so $P(X = 4)$ can be approximated with a continuous normal distribution by finding $P(3.5 < X < 4.5)$.

Normal Approximation to the Binomial Distribution (2 of 3)

- **General rule**: The normal distribution can be used to approximate the binomial distribution if:

$$np \geq 5 \text{ and } n(1-p) \geq 5$$

- The closer p is to 0.5, the better the normal approximation to the binomial.
- The larger the sample size n, the better the normal approximation to the binomial.

Normal Approximation to the Binomial Distribution (3 of 3)

- The mean and standard deviation of the binomial distribution are:

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

- Transform binomial to normal using the formula:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{np(1-p)}}$$

Curtin University

Example (1 of 2)

If $n = 1000$ and $p = 0.2$, what is $P(X \leq 180)$?

Solution: Approximate $P(X \leq 180)$ using a continuity correction adjustment: $P(X \leq 180.5)$

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{180.5 - (1000)(0.2)}{\sqrt{(1000)(0.2)(1-0.2)}} = -1.54$$

Example (2 of 2)

So, $P(Z \leq -1.54) = 0.0618$