Topic 7A: The Normal Distribution and Other Continuous Distribution

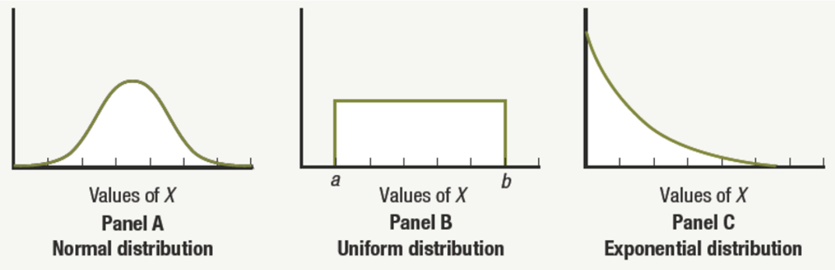
Dr. Lei Pan

School of Accounting, Economics and Finance
Curtin University

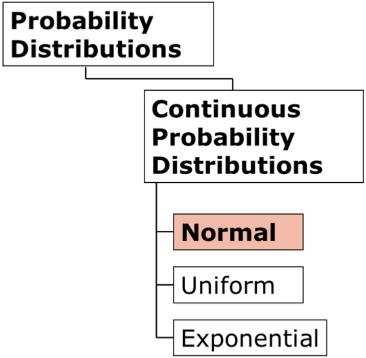
Continuous Probability Distributions (1 of 2)

- 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)

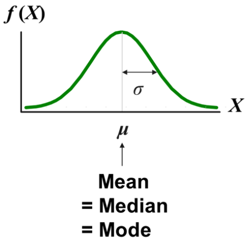


The Normal Distribution (1 of 2)



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:

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

where

e: mathematical constant approximated by 2.71828

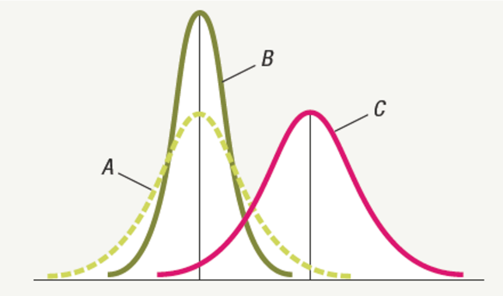
π: mathematical constant π approximated by 3.1415927

μ: population mean

σ: population standard deviation

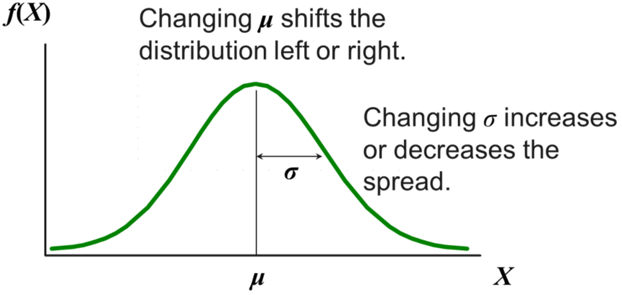
X: any value of the continuous variable

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:

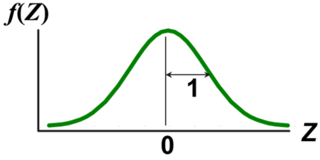
$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-(\frac{1}{2})Z^2}$$

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.



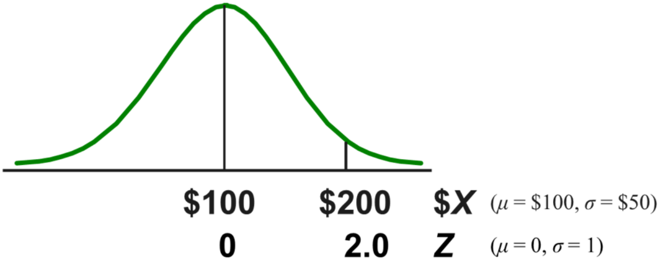
Example

- 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.

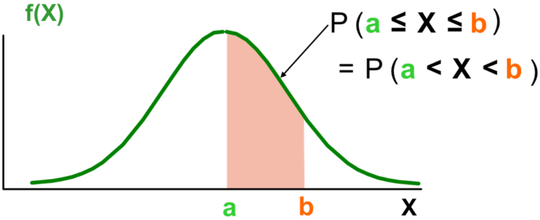
Comparing X and Z units



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

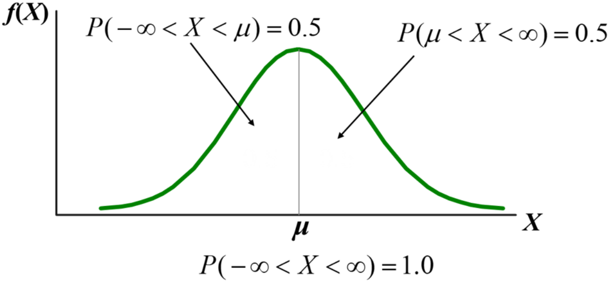
- Probability is measured by the area under the curve.



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

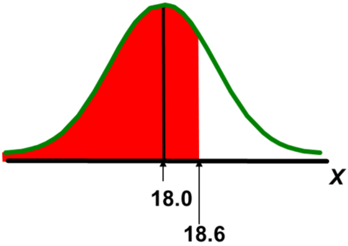
Probability as Area Under the Curve

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



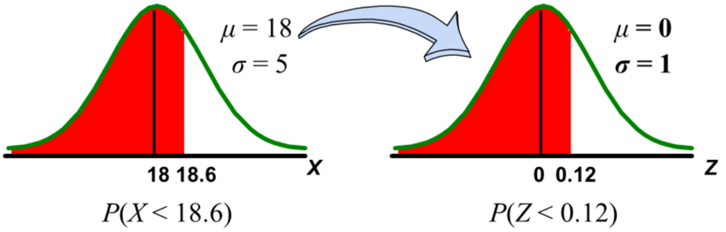
Finding Normal Probabilities

- 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 $P(X < 18.6)$

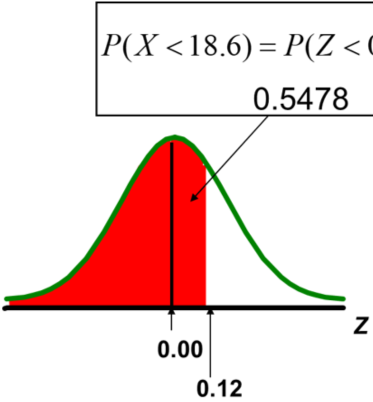


Solution (1 of 2)

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

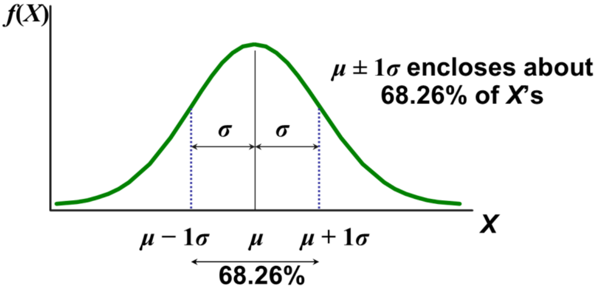


Solution (2 of 2)



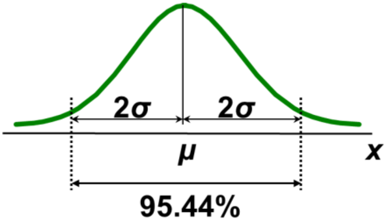
Empirical Rule (1 of 2)

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)

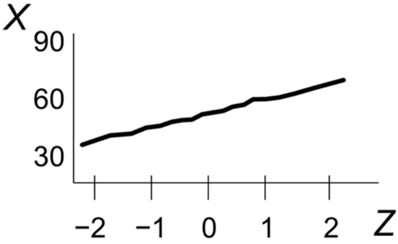
- Observe the distribution of the data set
  - Do approximately 2/3 of the observations lie within mean ± 1 standard deviation?
  - Do approximately 80% of the observations lie within mean ± 1.28 standard deviation?
  - Do approximately 95% of the observations lie within mean ± 2 standard deviation?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

Constructing a Normal Probability Plot

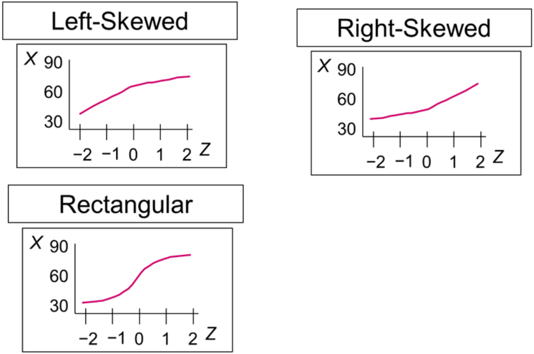
- A **normal probability plot** is a graphical approach used to evaluate if data are normal.
- A common approach is the **quantile-quantile plot**.
  - Arrange data into ordered array.
  - Find corresponding standardized normal quantile values (Z).
  - Plot the pairs of points with observed data values (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:



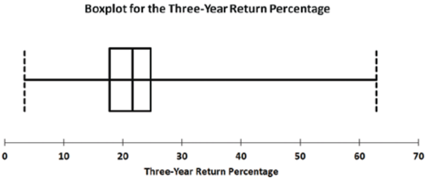
Normal Probability Plots



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)

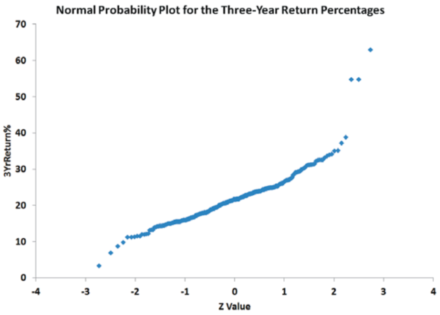
Descriptive Statistics

	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.)

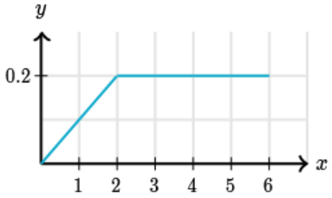
Evaluating Normality an Example: Mutual Fund Returns (5 of 5)

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:

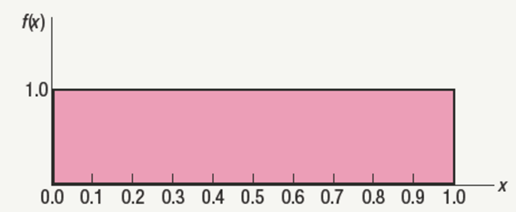
$$\mu = \frac{a+b}{2}$$

- 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

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

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 $P(3.5 < X < 4.5)$.

Curtin University

50 YEARS OF INNOVATION

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:

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

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