大數據時代必備的商業決策利器
活用統計做管理

Source: 經理人, 2014. Sep. no. 118

活用統計觀念

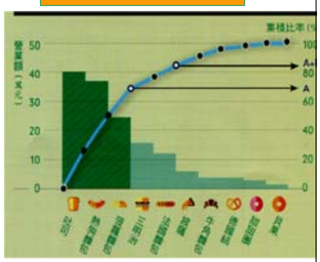
■ 問題: 麵包店

➢ 烤箱產能有限，該選擇生產哪些種類的麵包?

各商品營業額的比率

麵包種類	營業額	相對次數	累積相對比率
吐司	40	0.27 (40/150=0.27)	27%
熱狗麵包	38	0.25	52% (0.27+0.25=0.52)
菠蘿麵包	25	0.17	69%
三明治	15	0.10	79%
法國麵包	12	0.08	87%
披薩	6	0.04	91%
牛角麵包	5	0.03	94%
德國結	5	0.03	97%
甜甜圈	3	0.02	99%
貝果	1	0.01	100%
合計	150 (萬元/月)		

各商品營業額的柏拉圖



Source: 經理人, 2014. Sep. no. 118

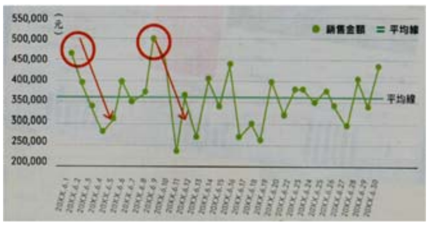
活用統計觀念

■ 問題: 營業額波動幅度大

六月份營業額

	銷售金額 (元)
20XX.6.1	462,014
20XX.6.2	394,031
20XX.6.3	338,740
20XX.6.4	289,039
20XX.6.5	319,714
⋮	⋮
20XX.6.30	427,403
平均	361,367.67

六月份營業額折線圖



Source: 經理人, 2014. Sep. no. 118

活用統計觀念

推出廣告的期望值 =「發生機率×預期結果」的總和
=50%×3500+50%×(−1000)
= 1750 + (−500)
=1250

不推出廣告的期望值=0

決定推出新廣告

通過期望值的計算，你預測推出廣告可能得到1250萬的報酬，但不推出廣告就不會有報酬，所以應該選期望值較高的方案。

Source: 經理人, 2014. Sep. no. 118

Chapter 6

The Normal Distribution
(Part 1)

常見的連續型機率分配

- 連續型隨機變數
  - 意義
  - 密度函數
- 均勻分配
  - 均勻分配之意義
  - 期望值與變異數
- 常態分配
  - 常態分配之意義
  - 期望值與變異數
  - 標準化隨機變數
- 中央極限定理
  - Sampling
  - Sampling Error
  - Distribution of Sample Means
- The Normal Approximation to the Binomial Distribution

Chapter 6 Overview

Introduction

6-1 Normal Distributions

6-2 Applications of the Normal Distribution

6-3 The Central Limit Theorem

6-4 The Normal Approximation to the Binomial Distribution

Chapter 6 Objectives

1. Identify distributions as symmetric or skewed.
2. Identify the properties of a normal distribution.
3. Find the area under the standard normal distribution, given various *z* values.
4. Find probabilities for a normally distributed variable by transforming it into a standard normal variable.

Chapter 6 Objectives

5. Find specific data values for given percentages, using the standard normal distribution.
6. Use the central limit theorem to solve problems involving sample means for large samples.
7. Use the normal approximation to compute probabilities for a binomial variable.

連續隨機變數的機率密度函數

- 連續隨機變數的機率密度函數

設X為連續隨機變數，其值為$a \le X \le b$，若$f(x)$滿足下列二條件：

① $f(x) \ge 0$

② $\int_a^b f(x)dx = 1$

則$f(x)$為X的機率密度函數(probability density function)，簡稱*pdf*。

Supplement

Uniform Distribution

(均勻分配)

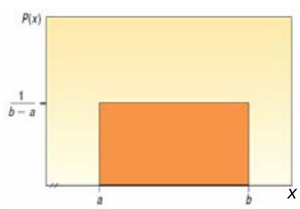
均勻分配

- 令隨機變數X的可能值之範圍為區間(a,b)，且呈均勻分配，則其機率密度函數為

$$f(x)=\frac{1}{b-a}, a<x<b$$

$$E(X)=\frac{a+b}{2}$$

$$Var(X)=\frac{(b-a)^2}{12}$$



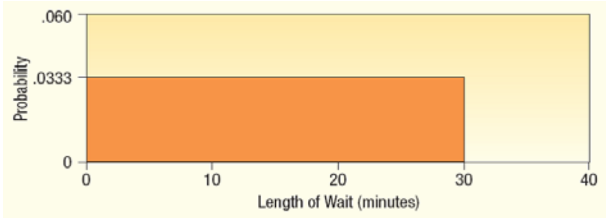
The Uniform Distribution - Example

Southwest Arizona State University provides bus service to students while they are on campus. A bus arrives at the North Main Street and College Drive stop every 30 minutes between 6 A.M. and 11 P.M. during weekdays. Students arrive at the bus stop at random times. The time that a student waits is uniformly distributed from 0 to 30 minutes.

1. Draw a graph of this distribution.
2. How long will a student “typically” have to wait for a bus? In other words what is the mean waiting time? What is the standard deviation of the waiting times?
3. What is the probability a student will wait more than 25 minutes?
4. What is the probability a student will wait between 10 and 20 minutes?

The Uniform Distribution - Example

1. Draw a graph of this distribution.



The Uniform Distribution - Example

2. Show that the area of this distribution is 1.00

The times students must wait for the bus is uniform over the interval from 0 minutes to 30 minutes, so in this case *a* is 0 and *b* is 30.

$$\text{Area} = (\text{height})(\text{base}) = \frac{1}{(30-0)}(30-0) = 1.00$$

The Uniform Distribution- Example

3. How long will a student “typically” have to wait for a bus? In other words what is the mean waiting time? What is the standard deviation of the waiting times?

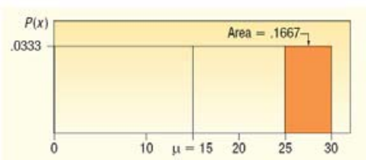
$$\mu = \frac{a+b}{2} = \frac{0+30}{2} = 15$$

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(30-0)^2}{12}} = 8.66$$

The Uniform Distribution- Example

4. What is the probability a student will wait more than 25 minutes?

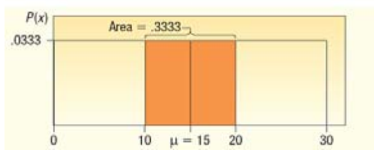
$$P(25 < \text{wait time} < 30) = (\text{height})(\text{base}) = \frac{1}{(30-0)}(5) = 0.1667$$



The Uniform Distribution - Example

5. What is the probability a student will wait between 10 and 20 minutes?

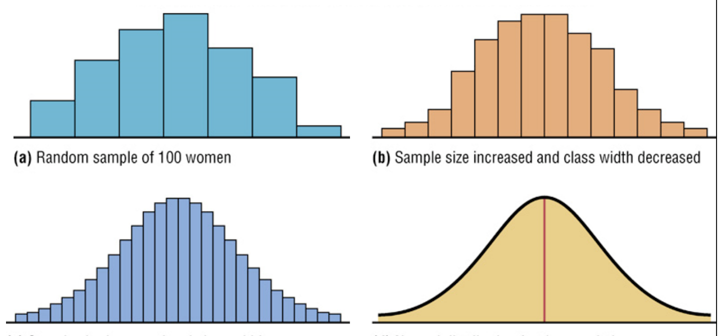
$$P(10 < \text{wait time} < 20) = (\text{height})(\text{base}) = \frac{1}{(30-0)}(10) = 0.3333$$



Section 6-1

Normal Distribution

Figure 6-1
Histograms and Normal Model for the Distribution of Heights of Adult Women



(a) Random sample of 100 women

(b) Sample size increased and class width decreased

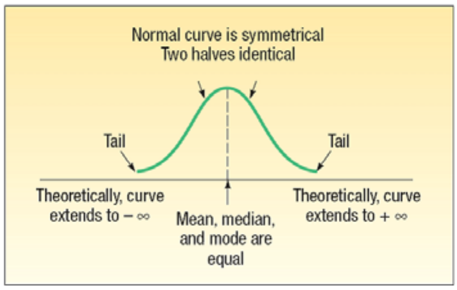
(c) Sample size increased and class width decreased further

(d) Normal distribution for the population

6.1 Normal Distributions

- Many continuous variables have distributions that are bell-shaped and are called **approximately normally distributed variables**.
- The theoretical curve, called the **bell curve** or the **Gaussian distribution**, can be used to study many variables that are not normally distributed but are approximately normal.

The Normal Distribution - Graphically



Normal Distributions

The mathematical equation for the normal distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

where

$e \approx 2.718$

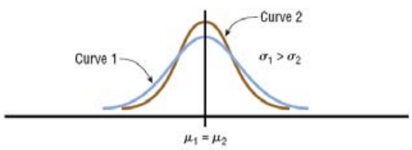
$\pi \approx 3.14$ → $N.D.(\mu, \sigma^2)$ or $N.(\mu, \sigma^2)$

μ = *population mean*

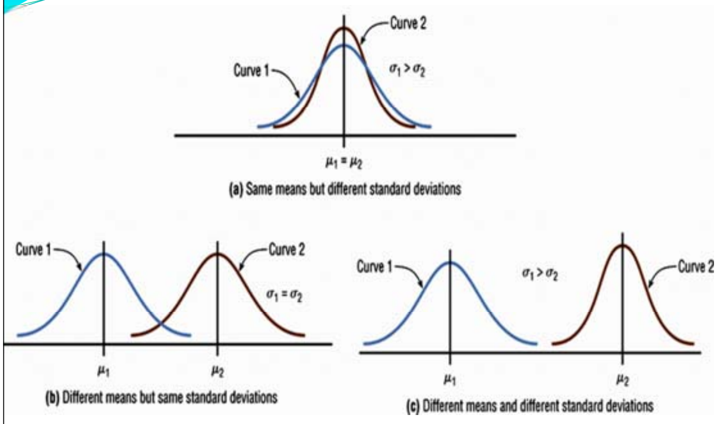
σ = *population standard deviation*

Normal Distributions

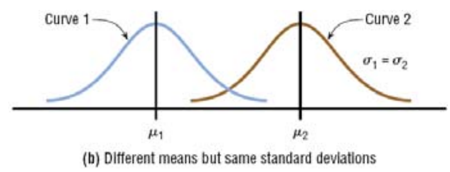
- The shape and position of the normal distribution curve depend on two parameters, the **mean** and the **standard deviation**.
- Each normally distributed variable has its own normal distribution curve, which depends on the values of the variable's mean and standard deviation.

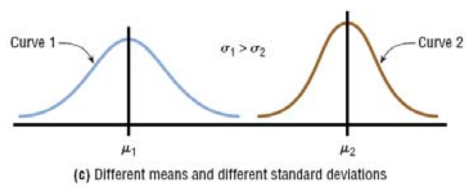


(a) Same means but different standard deviations



Normal Distributions





Normal Distribution Properties

- The normal distribution curve is **bell-shaped**.
- The curve is **symmetrical about the mean**, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.
- The mean, median, and mode are equal and located at the center of the distribution.
- The normal distribution curve is **unimodal** (i.e., it has only one mode).

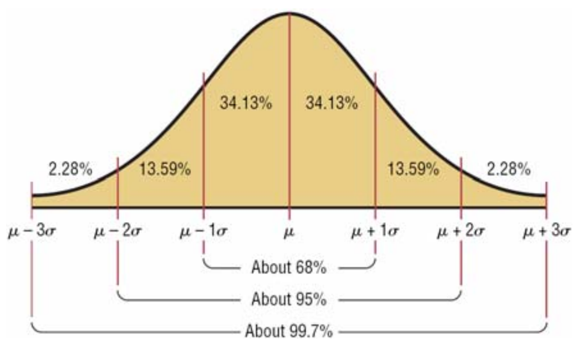
Normal Distribution Properties

- The curve is continuous—i.e., there are no gaps or holes. For each value of *X*, here is a corresponding value of f(*x*) or *Y*.
- The curve **never touches the *x* axis**. Theoretically, no matter how far in either direction the curve extends, it never meets the *x* axis—but it gets increasingly closer.

Normal Distribution Properties

- The total area under the normal distribution curve is equal to 1.00 or 100%.
- The area under the normal curve that lies within
  - 1 stdev. of the mean is approximately 68%.
  - 2 stdev. of the mean is approximately 95%.
  - 3 stdev of the mean is approximately 99.7%.

Normal Distribution Properties



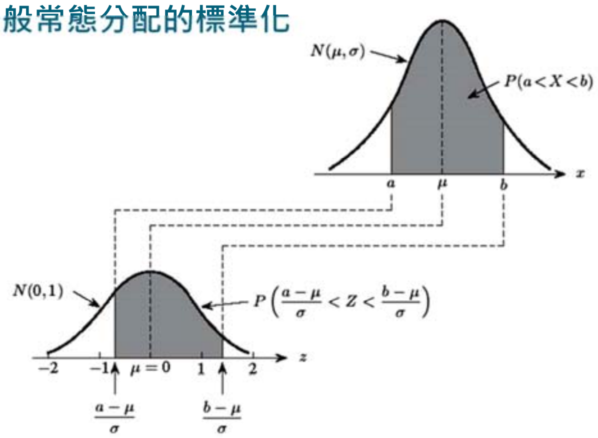
Standard Normal Distribution

- Since each normally distributed variable has its own mean and standard deviation, the shape and location of these curves will vary.
- Find P(a < x < b)

$$P(a < x < b) = \int_a^b f(x)\,dx$$

- The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1.

一般常態分配的標準化



z value (Standard Value)

The *z* value is the number of standard deviations that a particular *X* value is away from the mean. The formula for finding the z value is:

$$z = \frac{\text{value - mean}}{\text{standard deviation}}$$

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

The Normal Distribution – Example

- The weekly incomes of shift foremen in the glass industry follow the normal probability distribution with a mean of \$1,000 and a standard deviation of \$100.
  - What is the z value for the income, let's call it *X*, of a foreman who earns \$1,100 per week? For a foreman who earns \$900 per week?

For X = \$1,100:

$$z = \frac{X-\mu}{\sigma} = \frac{\$1{,}100 - \$1{,}000}{\$100} = 1.00$$

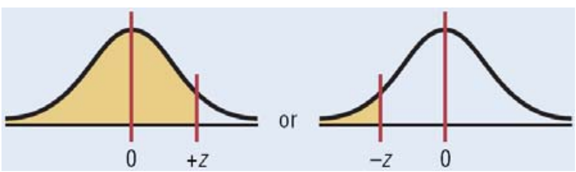
For X = \$900:

$$z = \frac{X-\mu}{\sigma} = \frac{\$900 - \$1{,}000}{\$100} = -1.00$$

Area under the Standard Normal Distribution Curve

1. To the left of any *z* value:

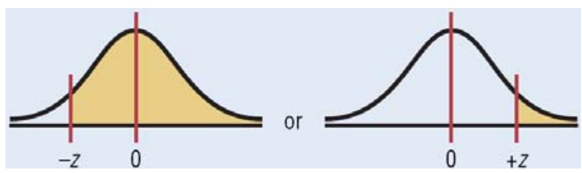
Look up the *z* value in the table and use the area given.



Area under the Standard Normal Distribution Curve

2. To the right of any *z* value:

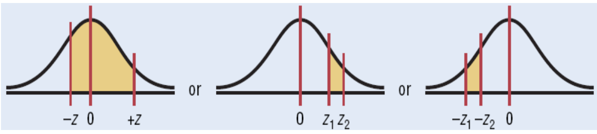
Look up the *z* value and subtract the area from 1.



Area under the Standard Normal Distribution Curve

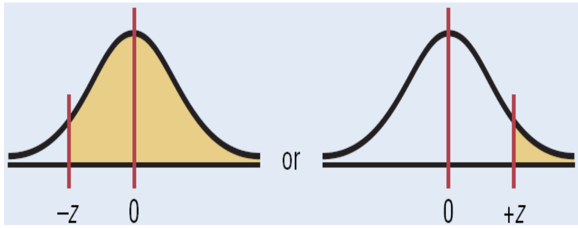
3. Between two *z* values:

Look up both *z* values and subtract the corresponding areas.



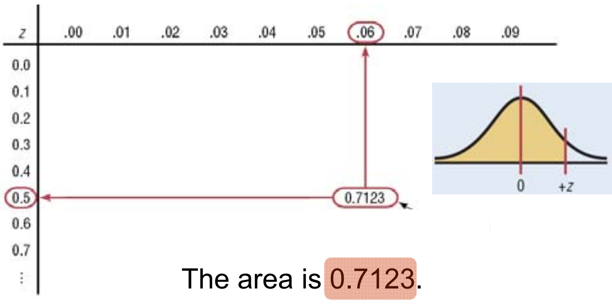
標準常態的特性

- $P(Z \le 0) = 0.5$
- $P(Z < -z) = 1 - P(Z \le z) = P(Z \ge z)$



Example 1: Area under the Curve

Find the area to the left of $z = 0.56$.

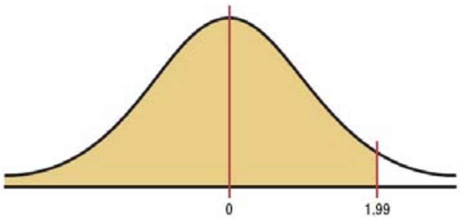


The area is 0.7123.

i.e. $P(Z<0.56)=0.7123$

Example 2: Area under the Curve

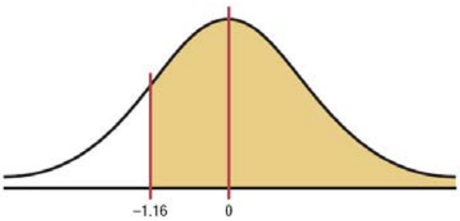
Find the area to the left of $z = 1.99$.



The value in the 1.9 row and the .09 column of Table E is .9767. The area is .9767.

Example 3: Area under the Curve

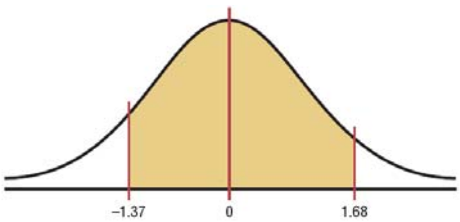
Find the area to right of $z = -1.16$.



The value in the -1.1 row and the .06 column of Table E is .1230. The area is 1 - .1230 = .8770.

Example 4: Area under the Curve

Find the area between $z = 1.68$ and $z = -1.37$.



The values for $z = 1.68$ is .9535 and for $z = -1.37$ is .0853. The area is .9535 - .0853 = .8682.

Example 5: Probability

- Find the probability:

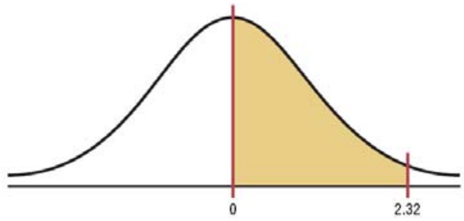
  (a) $P(0 < z < 2.32)$ = 0.4898

  (b) $P(z < 1.73)$ = 0.8827

  (c) $P(z > 1.98)$ = 0.0239

Example 5: Probability

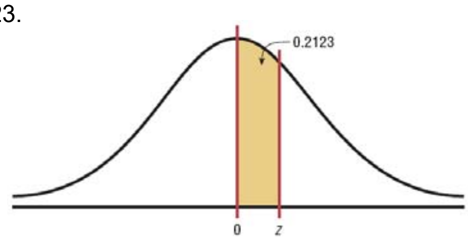
a. Find the probability: $P(0 < z < 2.32)$



The values for z = 2.32 is .9898 and for z = 0 is .5000. The probability is .9898 - .5000 = .4898.

Example 5: Probability

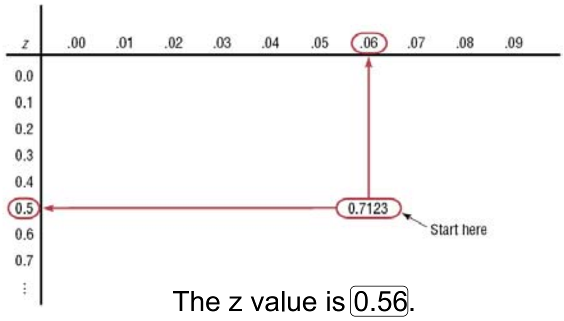
Find the z value such that the area under the standard normal distribution curve between 0 and the z value is 0.2123.



Add .5000 to .2123 to get the cumulative area of .7123. Then look for that value inside Table E.

Example 5: Probability

Add .5000 to .2123 to get the cumulative area of .7123. Then look for that value inside Table E.



The z value is $\boxed{0.56}$.

Eercises

- P(-1.2 < Z < 2.85) = 0.8827
- P(0.23 < Z < 2.03) = 0.3878
- P(-2.33 < Z < -0.50) = 0.2986
- P(|Z|< 0.25) = 0.1974
- P(|Z|> 1.50) = 0.1336
- P(a < Z < 1.58) = 0.8028
  - a = －1.08
- P(1.25 < Z < b) = 0.1001
  - b = 2.54
- P(-2.58 < Z < c) = 0.5822
  - c = 0.22

標準化 (z-score)

設已知$X \sim N(\mu=10, \sigma^2=4)$，試求

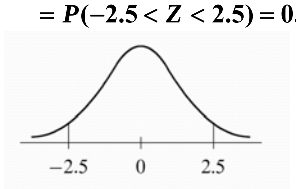
(1) $P(5 < X < 15)$

(2) $P(X > 16)$

解：

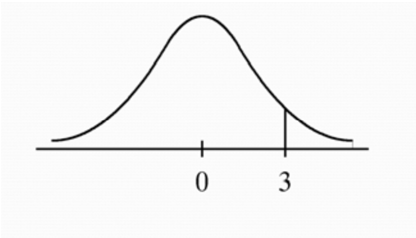
$$(1)\ P(5 < X < 15) = P\left(\frac{5-10}{2} < Z < \frac{15-10}{2}\right)$$

$$= P(-2.5 < Z < 2.5) = 0.9876$$



標準化

$$(2)\ P(X>16)=P\left(Z>\frac{16-10}{2}\right)=P(Z>3)=0.0013$$



Exercises 6-1

- Section 6-1 (p.322)
  - **Homework: 19-21, 25, 35, 48, 50, 53-56**
  - 9-12, 14, 15, 19, 20, 25, 30, 35, 37, 42, 47, 49, 50, 51-56, 60

Section 6-2

Applications of the Normal Distributions

Applications of the Normal Distributions

- The standard normal distribution curve can be used to solve a wide variety of practical problems. The only requirement is that the variable be normally or approximately normally distributed.
- For all the problems presented in this chapter, you can assume that the variable is normally or approximately normally distributed.

Applications of the Normal Distributions

- To solve problems by using the standard normal distribution, transform the original variable to a standard normal distribution variable by using the *z* value formula.
- This formula transforms the values of the variable into standard units or *z* values. Once the variable is transformed, then the Procedure Table and Table E in Appendix C can be used to solve problems.

常態分配應用

例 4.25

- 假設台灣地區國中生的智商(IQ)為一常態分配，且已知平均數為100，標準差為16。今隨機自該地區抽出一位國中生，試問
  - 該生智商為120的 z score.
  - (1) 該生智商超過120的機率？
  - (2) 該生智商介於100~120間的機率？

常態分配應用

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

解：令 X 表智商，則 $X \sim N(100, 256)$

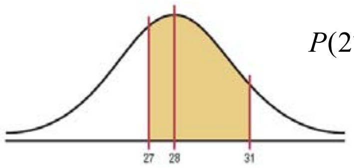
$$(1)\ P(X>120)=P\left(Z>\frac{120-100}{16}\right)=P(Z>1.25)=0.1056$$

$$(2)\ P(100<X<120)=P\left(\frac{100-100}{16}<Z<\frac{120-100}{16}\right)$$
$$=P(0<Z<1.25)$$
$$=0.3944$$

Example 6-7: Newspaper Recycling

Each month, an American household generates an average of 28 pounds of newspaper for garbage or recycling. Assume the standard deviation is 2 pounds. If a household is selected at random, find the probability of its generating between 27 and 31 pounds per month. Assume the variable is approximately normally distributed.

Step 1: Draw the normal distribution curve.

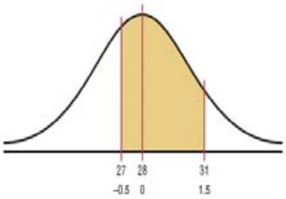


$$P(27<X<31)=?$$

Example 6-7: Newspaper Recycling

Step 2: Find *z* values corresponding to 27 and 31, and the area between the z values.

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

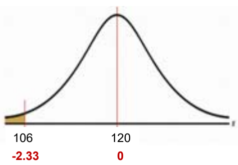


$$P(27<X<31)$$
$$=P\left(\frac{27-28}{2}<\frac{x-\mu}{\sigma}<\frac{31-28}{2}\right)$$
$$=P(-0.5<z<1.5)$$
$$=0.9332-0.3085$$
$$=0.6247$$

Ans: The probability is 62%.

Example 6-8: Amount of Electricity Used by a PC

A desktop PC used 120 watts of electricity per hour based on 4 hours of use per day the variable is approximately normally distributed and the standard deviation is 6. If 500 PCs are selected, approximately how many will use less than 106 watts of power

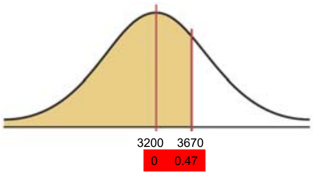


$$P(X<106)$$
$$=P\left(\frac{X-\mu}{\sigma}<\frac{106-120}{6}\right)$$
$$=P(z<-2.33)=0.0099$$
$$500\times 0.0099\cong 5$$

Ans: Approximately 5 PCs use less than 106 watts.

Example: Holiday Spending

A national survey found that women spend on average \$3200 for the Christmas holidays. Assume the standard deviation is \$1000. Find the percentage of women who spend less than \$3670. Assume the variable is normally distributed.

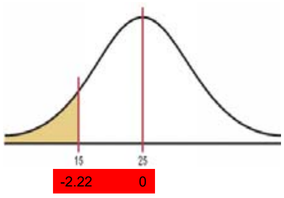


$$P(X<3670)$$
$$=P\left(\frac{X-\mu}{\sigma}<\frac{3670-3200}{1000}\right)$$
$$=P(z<0.47)$$
$$=0.6808$$

Ans: 68% of women spend less than \$3670

Example: Emergency Response

The American Automobile Association reports that the average time it takes to respond to an emergency call is 25 minutes. Assume the variable is approximately normally distributed and the standard deviation is 4.5 minutes. If 80 calls are randomly selected, approximately how many will be responded to in less than 15 minutes?



$$P(X<15)$$
$$=P\left(\frac{X-\mu}{\sigma}<\frac{15-25}{4.5}\right)$$
$$=P(z<-2.22)$$
$$=0.0132$$
$$80\times 0.0132=1.056$$

Ans: Approximately 1 call will be responded to in under 15 minutes.

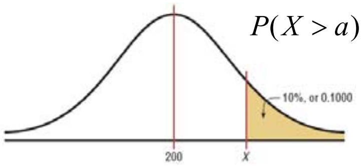
Finding data values given specific probability

Example 6-9: Police Academy

To qualify for a police academy, candidates must score in the top 10% on a general abilities test. The test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify. Assume the test scores are normally distributed.

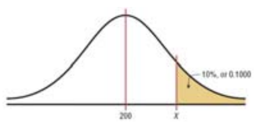
Step 1: Draw the normal distribution curve.

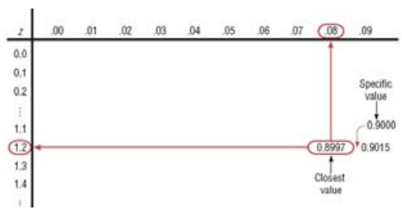
$$P(X > a) = 0.1, \text{ find } a = ?$$



Example 6-9: Police Academy

Step 2: Subtract 1 - 0.1000 to find area to the left, 0.9000.





Step 3: Find *X*.

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

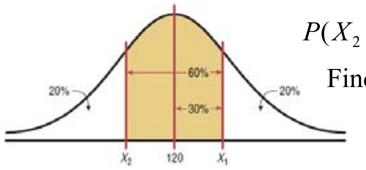
$$X = \mu + z\sigma$$

The cutoff, the lowest possible score to qualify, is 226.

Example 6-10: Systolic Blood Pressure

For a medical study, a researcher wishes to select people in the middle 60% of the population based on blood pressure. If the mean systolic blood pressure is 120 and the standard deviation is 8, find the upper and lower readings that would qualify people to participate in the study.

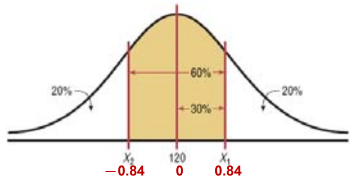
Step 1: Draw the normal distribution curve.



$$P(X_2 < X < X_1) = 0.6$$

Find X_1 and X_2.

Example 6-10: Systolic Blood Pressure



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

Area to the left of the positive z_1: $0.5000 + 0.3000 = 0.8000$.
Using Table E, $z_1 \approx 0.84$. $X_1 = 120 + 0.84(8) = 126.72$

Area to the left of the negative z_2: $0.5000 – 0.3000 = 0.2000$.
Using Table E, $z_2 \approx -0.84$. $X_2 = 120 - 0.84(8) = 113.28$

The middle 60% of readings are between 113 and 127.

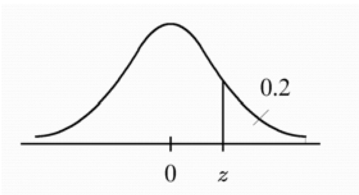
常態分配應用

例 4.26

- 假設某高中數學競試成績為一常態分配，已知平均分數為60分，標準差10分。若成績採四等第計分，最高分數前 20%以 A 計等第，其次20%以B計等第，再其次40%以C計等第，最後20%以D計等第。試問
  - 多少分以上才能得到A等第？
  - 多少分以下得到D等第？
  - 若其中某生得到B等第，其分數介於多少分之間？
  - 第25個百分位數為多少分？

解：令X 表分數，則$X \sim N(60, 100)$

(1) $P(X > x) = 0.2$



$P(Z > z) = 0.2$，則採內插法如下

$$\frac{z - 0.84}{0.85 - 0.84} = \frac{0.8 - 0.7995}{0.8023 - 0.7995}$$

$$z = 0.84 + \frac{5}{28}(0.85 - 0.84) = 0.842$$

即 $\frac{x - 60}{10} = 0.842$

$x = 60 + 0.842\ (10) = 68.42$

故 68.42分以上才能得到A等第。

(2) $P(X < x) = 0.2$

$p\left(Z < \frac{x-60}{10}\right) = 0.2$，查附錄表二採內插法得

$z = -0.842$

即 $\frac{x-60}{10} = -0.842$，

$x = 60 + (–0.842) \times 10$

$= 51.58$

故 51.85 分以下得 D 等第。

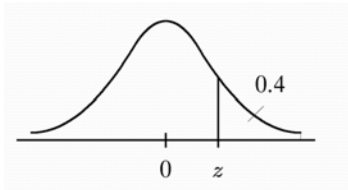
(3) 設分數介於 x_L, x_U之間得 B 等第，則分數 x_U以上得 A 等第，且$P(X > x_L)$ = 0.4 (因A, B等第合計佔40%)。

故$P\left(Z > \dfrac{x_L - 60}{10}\right) = 0.4$，查附錄表二採內插法得

z = 0.253

即 $\dfrac{x_L - 60}{10} = 0.253$， x_L = 60 + 0.253(10) = 62.53

由(1)已知68.42分以上得A等第，即 x_U = 68.42所以該生分數介於62.53與68.42之間。



(4) 設第25個百分位數為 x，則表示

$P(X< x)$ = 0.25 而

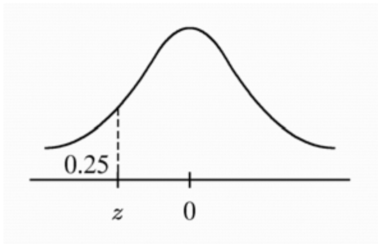
$p\left(Z < \dfrac{X - 60}{10}\right) = 0.25$ 採內插法如下

$$\frac{z-(-0.68)}{-0.67-(-0.68)} = \frac{0.25-0.2483}{0.2514-0.2483}$$

$$z = (-0.68) + \frac{17}{31}(0.01) = -0.675$$

故 x = 60 + (–0.675) (10) = 53.25

所以第25個百分位數為53.15分



Normal Distributions

- A normally shaped or bell-shaped distribution is only one of many shapes that a distribution can assume; however, it is very important since many statistical methods require that the distribution of values (shown in subsequent chapters) be normally or approximately normally shaped.
- There are a number of ways statisticians check for normality. We will focus on three of them.

Checking for Normality

- Histogram
- Outliers
- Pearson's coefficient (PC) of Skewness or Pearson's Index (PI) of Skewness

$$PC = \frac{3(\overline{X} - MD)}{s}$$

- Other Tests
  - Normal Quantile Plot
  - Chi-Square Goodness-of-Fit Test
  - Kolmogorov-Smikirov Test
  - Lilliefors Test

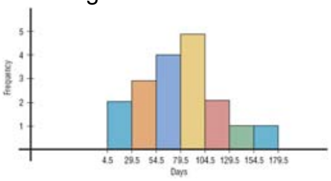
Example 6-11: Technology Inventories

A survey of 18 high-technology firms showed the number of days' inventory they had on hand. Determine if the data are approximately normally distributed.

5 29 34 44 45 63 68 74 74
81 88 91 97 98 113 118 151 158

Method 1: Construct a Histogram.

Class	Frequency
5–29	2
30–54	3
55–79	4
80–104	5
105–129	2
130–154	1
155–179	1



The histogram is approximately bell-shaped.

Example 6-11: Technology Inventories

Method 2: Check for Skewness.

5 29 34 44 45 63 68 74 74
81 88 91 97 98 113 118 151 158

$$\bar{X} = 79.5, MD = 77.5, s = 40.5$$

$$\text{PI} = \frac{3(\bar{X} - MD)}{s} = \frac{3(79.5 - 77.5)}{40.5} = 0.148$$

The PI is not greater than 1 or less than 1, so it can be concluded that the distribution is not significantly skewed.

Example 6-11: Technology Inventories

Method 3: Check for Outliers.

5 29 34 44 45 63 68 74 74
81 88 91 97 98 113 118 151 158

Five-Number Summary: 5 - 45 - 77.5 - 98 - 158

Q1 – 1.5(IQR) = 45 – 1.5(53) = -34.5

Q3 – 1.5(IQR) = 98 + 1.5(53) = 177.5

No data below -34.5 or above 177.5, so no outliers.

Example 6-11: Technology Inventories

A survey of 18 high-technology firms showed the number of days' inventory they had on hand. Determine if the data are approximately normally distributed.

5 29 34 44 45 63 68 74 74
81 88 91 97 98 113 118 151 158

Conclusion:

- The histogram is approximately bell-shaped.
- The data are not significantly skewed.
- There are no outliers.

Thus, it can be concluded that the distribution is approximately normally distributed.

Exercises

- Section 6-1
  - **Homework: 19-21, 25, 35, 48, 50, 53-56**
  - 9-12, 14, 15, 19, 20, 25, 30, 35, 37, 42, 47, 49, 50, 51-56, 60
- Section 6-2
  - **Homework: 21, 36-38, 42**
  - 1, 5, 10, 13, 14, 17, 21, 27, 36-38, 42