# 統計學（一）

## 第六章（續） 抽樣分佈
## (Sampling Distributions )

授課教師：唐麗英教授

國立交通大學 工業工程與管理學系
聯絡電話：(03)5731896
e-mail：litong@cc.nctu.edu.tw

2013

# 本課程內容參考書目

- **教科書**
  - Mendenhall, W., & Sincich, T. (2007). *Statistics for engineering and the sciences*, 5th Edition. Prentice Hall.
- **參考書目**
  - Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2009). *Basic business statistics: Concepts and applications*, 11th Edition. Upper Saddle River, N.J: Pearson Prentice Hall.
  - Larson, H. J. (1982). *Introduction to probability theory and statistical inference*, 3rd Edition. New York: Wiley.
  - Miller, I., Freund, J. E., & Johnson, R. A. (2000). *Miller and Freund's Probability and statistics for engineers*, 6th Edition. Upper Saddle River, NJ: Prentice Hall.
  - Montgomery, D. C., & Runger, G. C. (2011). *Applied statistics and probability for engineers*, 5th Edition. Hoboken, NJ: Wiley.
  - Watson, C. J. (1997). *Statistics for management and economics*, 5th Edition. Englewood Cliffs, N.J: Prentice Hall.
  - 林惠玲、陳正倉（2009），「統計學：方法與應用」，第四版，雙葉書廊有限公司。
  - 唐麗英、王春和（2013），「從範例學MINITAB統計分析與應用」，博碩文化公司。
  - 唐麗英、王春和（2008），「SPSS 統計分析 14.0 中文版」，儒林圖書公司。
  - 唐麗英、王春和（2007），「Excel 2007統計分析」，第二版，儒林圖書公司。
  - 唐麗英、王春和（2005），「STATISTICA6.0與基礎統計分析」，儒林圖書公司。
  - 陳順宇（2004），「統計學」，第四版，華泰書局。
  - 彭昭英、唐麗英（2010），「SAS123」，第七版，儒林圖書公司。

# 抽樣分佈
# (Sampling Distributions)

# Sampling Distributions

- **Recall: Parameter and statistic (參數與統計量)**
  - A quantity computed from the observation in a **population** is called a **parameter**.
  - A quantity computed from the observation in a **sample** is called a **statistic**.

- **Example:**
  1) $\mu$ and $\sigma$ are parameters.
  2) $\bar{X}$ and $S$ are statistics.

國立交通大學開放式課程

# Sampling Distributions

- **Def : Sampling Distribution 抽樣分佈**
  - The **probability distribution** of a **statistic** that results when random sample of size **n** are repeatedly drawn from a given population is called the **sampling distribution** of the **statistic**.
  - **統計量之機率分佈**稱為「**抽樣分佈**」

- **Example：**
  - The distribution of **sample mean** $\bar{X}$ is one type of **sampling distribution.** The distribution of **sample proportion** $\hat{p}$ is another type of **sampling distribution.**

# 利用蒙地卡羅模擬法模擬抽樣分佈 (Approximating a Sampling Distribution by Monte Carlo Simulation)

# Approximating a Sampling Distribution

- **Example 1: See Example 6.18 on pages 238.**
  - Simulate the sampling distribution of the sample mean

$$\overline{Y} = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5} \qquad (n=5)$$

for a sample of n=5 observations drawn from the uniform probability distribution shown in Figure 6.12. Note that the uniform distribution has mean $\mu$=0.5. Repeat the procedure for n=15, 25, 50, and 100.

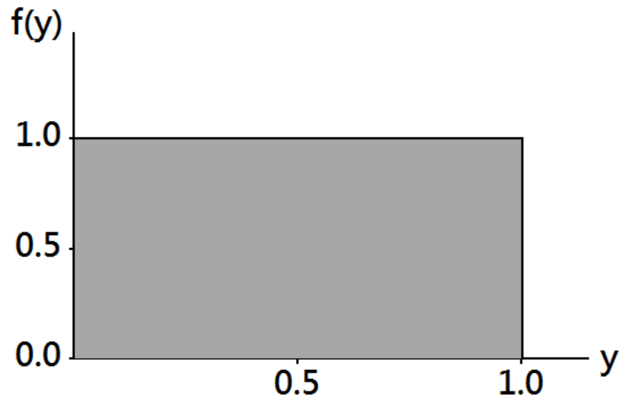
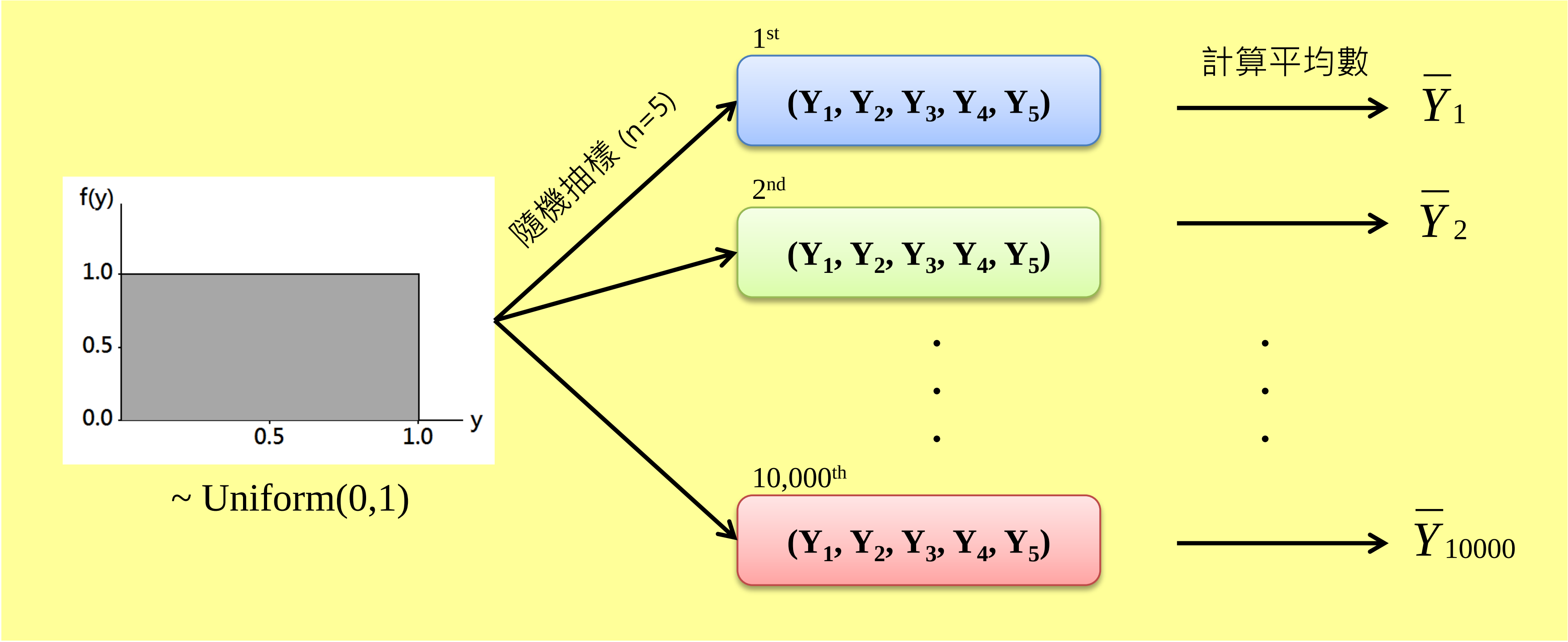



Figure 6.12
Uniform distribution of Example 6.18

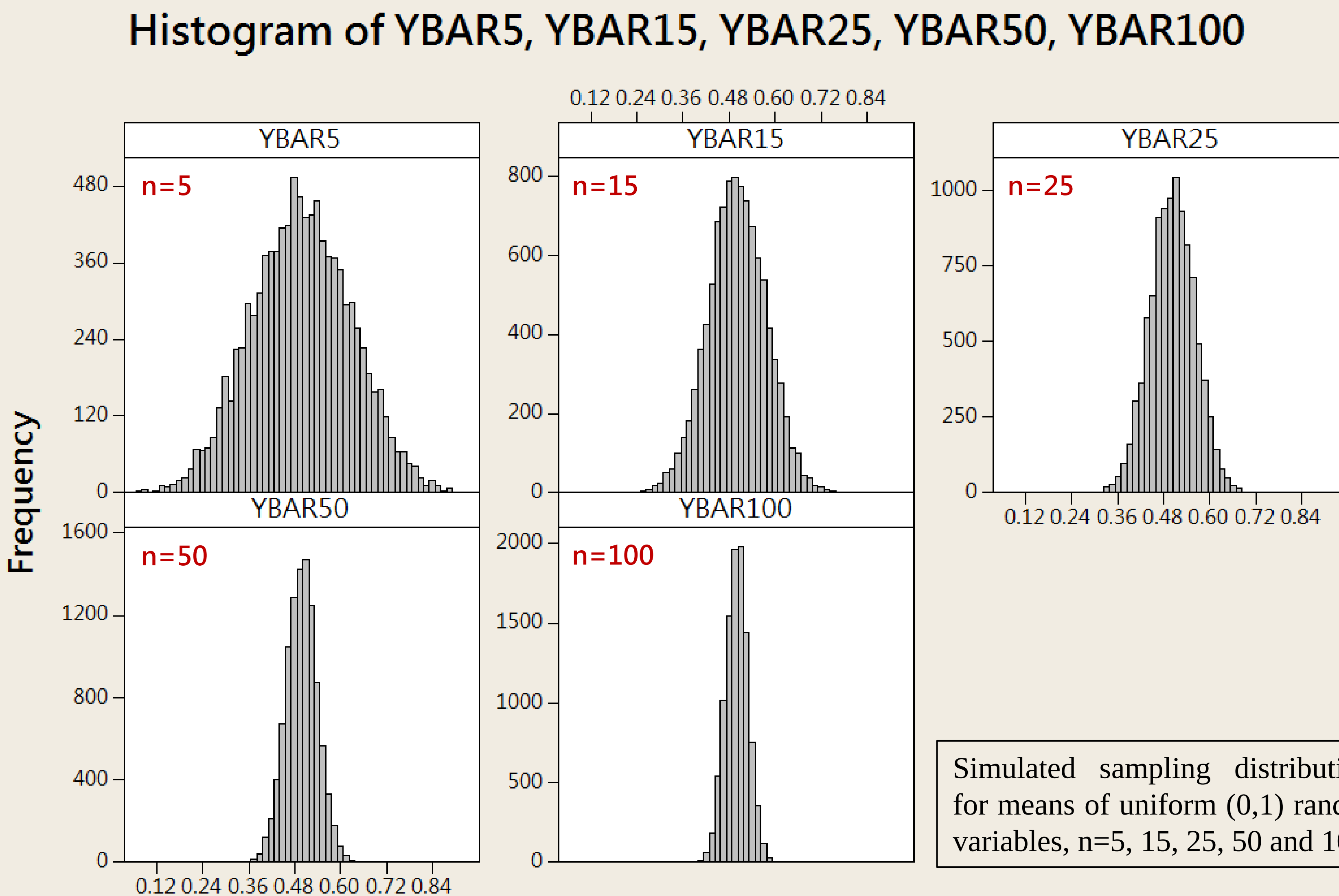
Y ~ Uniform(0,1)

# Approximating a Sampling Distribution

- **Simulation procedure**
  - Use Minitab to obtain 10,000 random samples of size n=5 from the uniform probability distribution, over the interval (0,1), and compute $\bar{\mathbf{Y}}$ for each sample. Then, plot a histogram for the 10,000 values of $\bar{\mathbf{Y}}$ . Repeat the procedure for n=15, 25, 50, and 100.



# Simulated sampling distributions for means of uniform (0,1)




Simulated sampling distributions for means of uniform (0,1) random variables, n=5, 15, 25, 50 and 100

# Approximating a Sampling Distribution

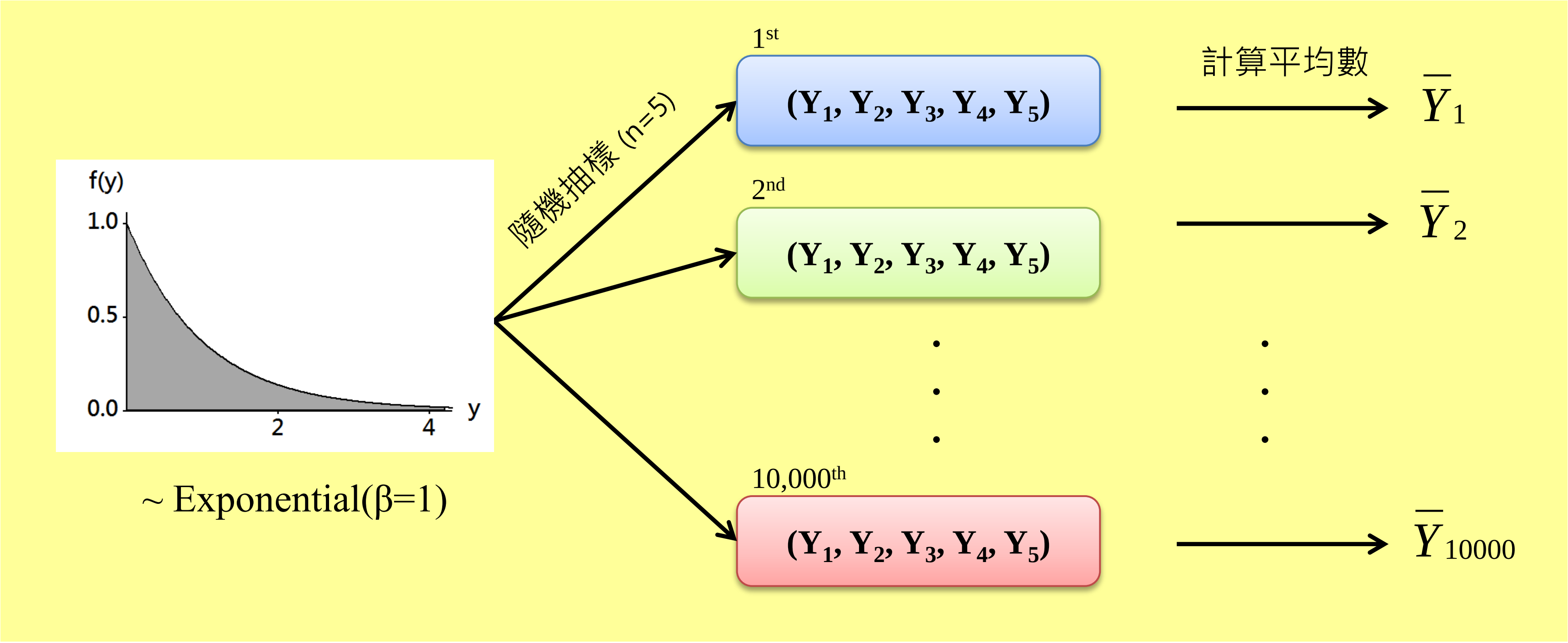
The **histogram** of the simulated **sampling distribution of $\bar{X}$** based on independent random samples from ***Uniform*** distribution is close to **Normal** distribution. Note that:

1) The values of $\bar{X}$ tend to cluster(群聚) about the ***mean*** of the **uniform** distribution.

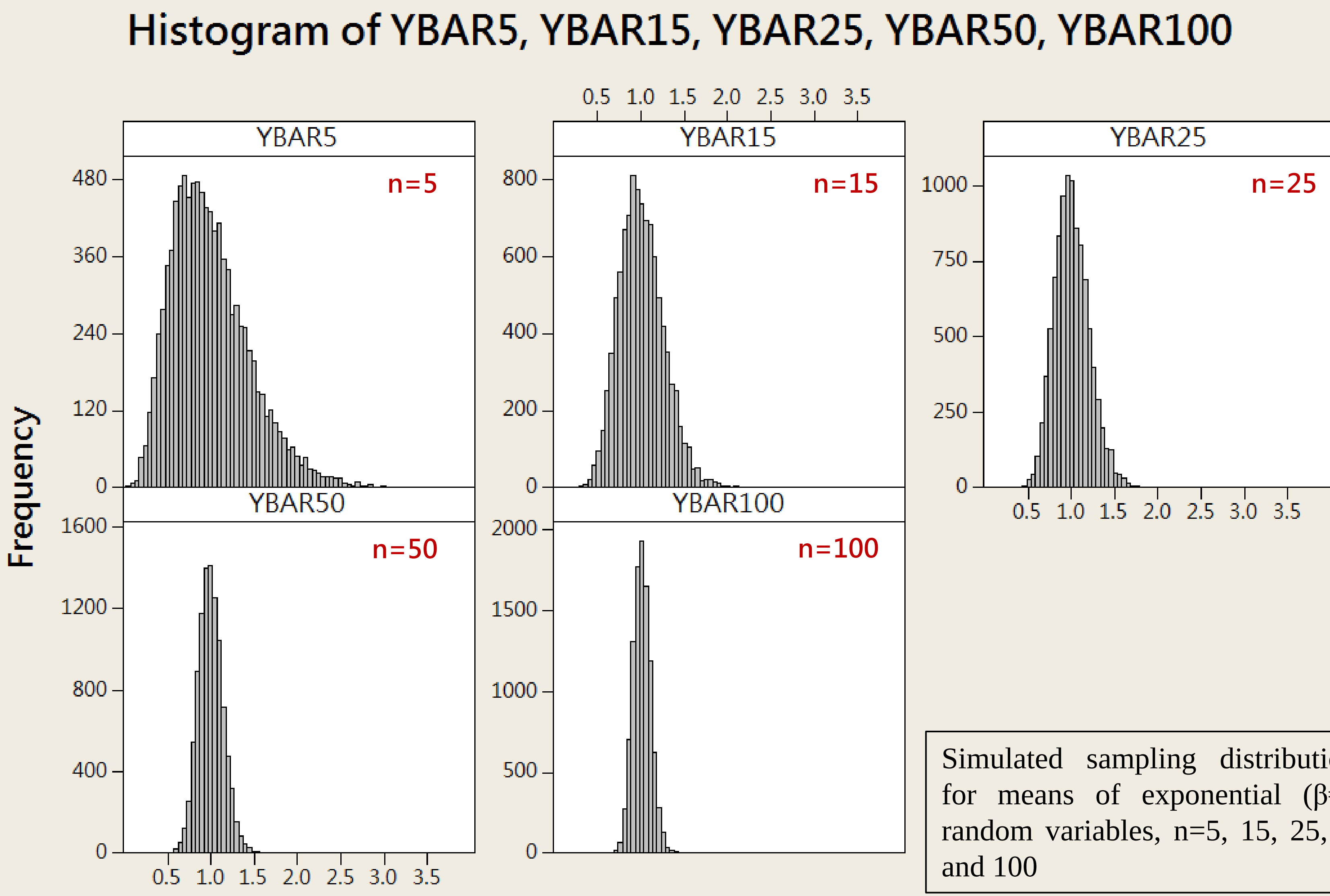
2) As the sample size ***n*** **increases**, there is **less variation** in the **sampling distribution of $\bar{X}$** and the ***shape*** of the **sampling distribution of $\bar{X}$** tends toward the ***shape*** of the ***Normal*** distribution.

# Approximating a Sampling Distribution

- **Example 2: See Example 6.19 on pages 239.**
  - The **histogram** of the simulated **sampling distribution of $\bar{X}$** based on independent random samples from ***Exponential*** distribution is close to **Normal** distribution. **其餘結論與上例相同。**



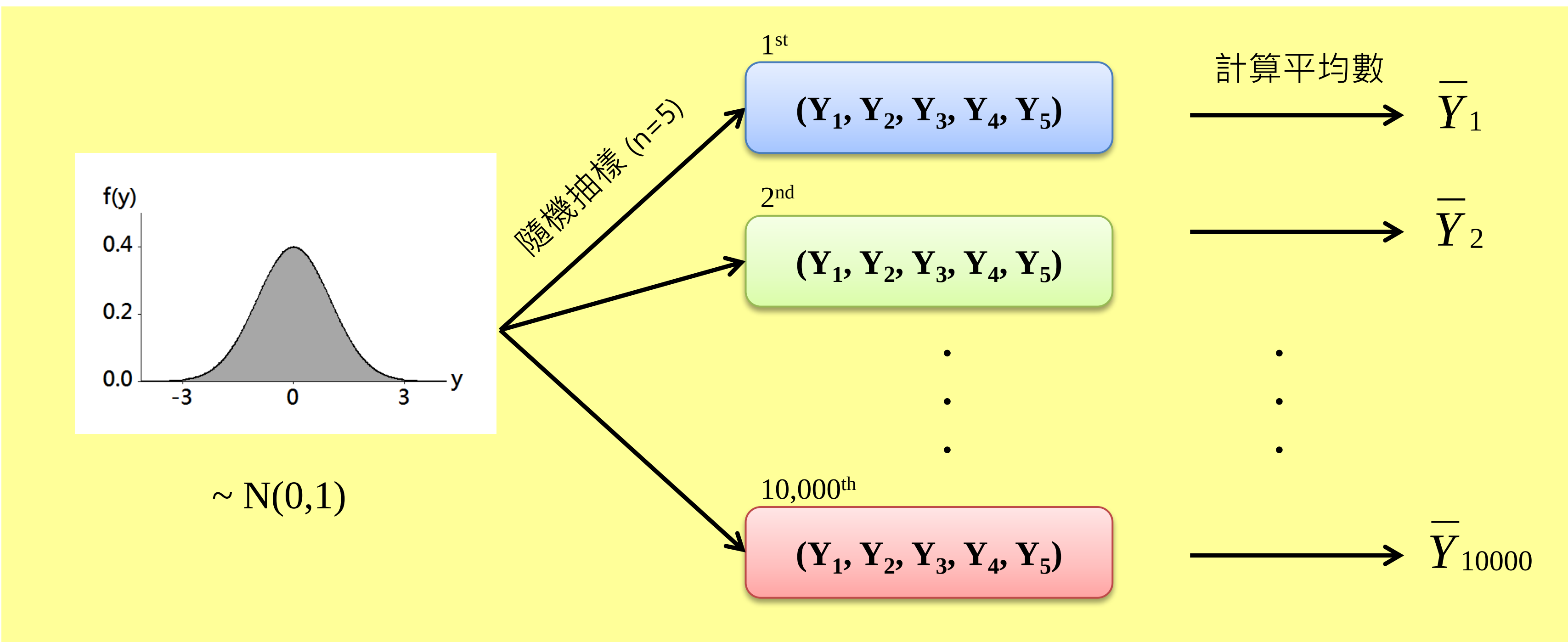
# Simulated sampling distributions for means of exponential (β=1)



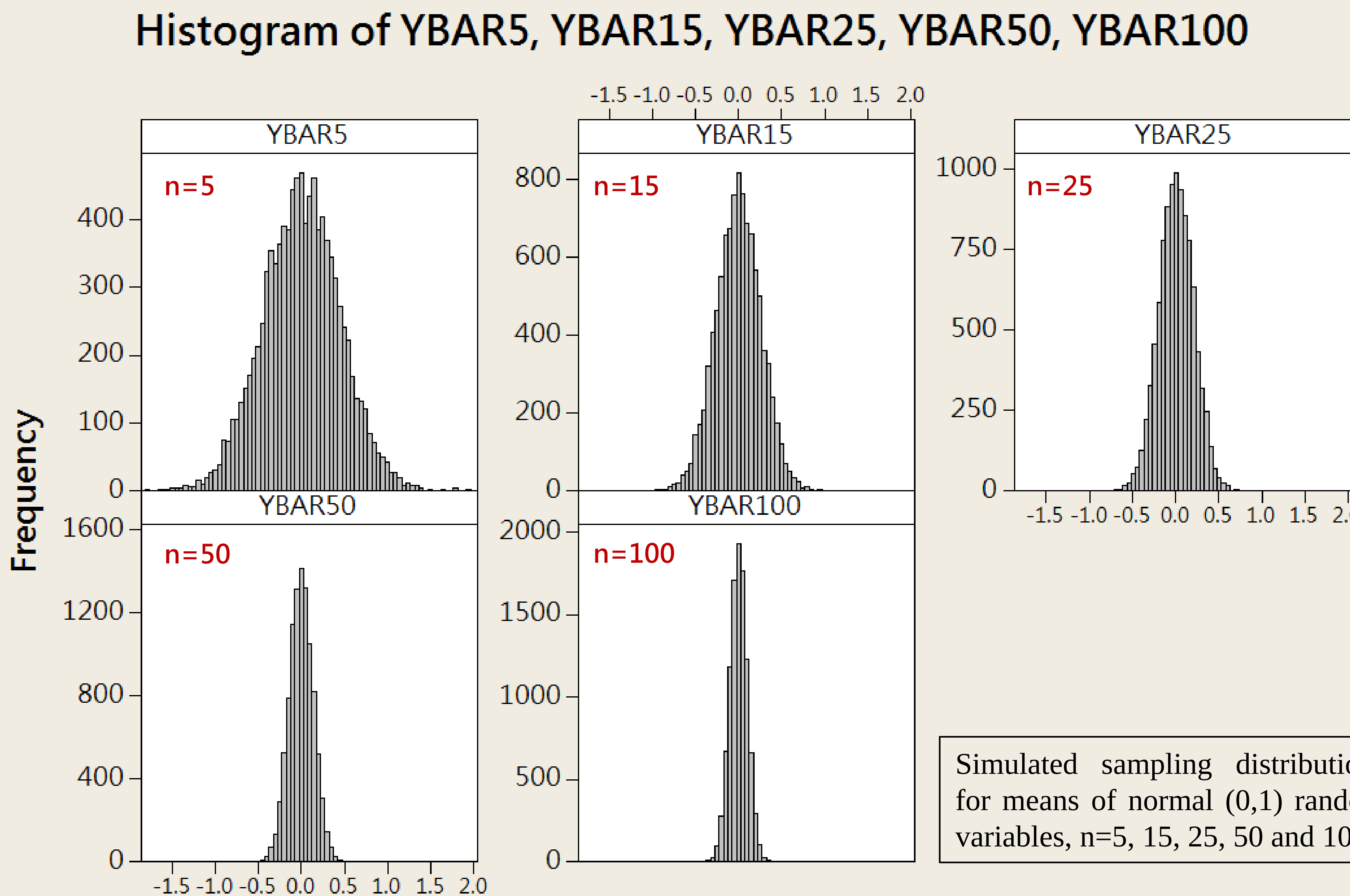

Simulated sampling distributions for means of exponential (β=1) random variables, n=5, 15, 25, 50 and 100

# Approximating a Sampling Distribution

- **Example 3: Normal distribution**
  - The **histogram** of the simulated **sampling distribution of $\bar{X}$** based on independent random samples from ***Normal*** distribution is close to **Normal** distribution. 其餘結論與上例相同。



# Simulated sampling distributions for means of normal (0,1)




Simulated sampling distributions for means of normal (0,1) random variables, n=5, 15, 25, 50 and 100

# 抽樣分布的特性與樣本平均數
# (The Sampling Distributions of Means and Sums)

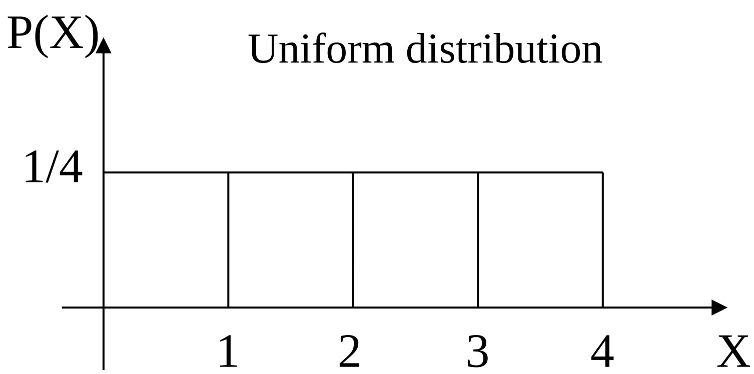
# Sampling Distribution of the Sample Mean

1) **What is the Sampling Distribution of the Sample Mean, $\bar{X}$ (When σ is known) ？**

   – **How the statistic $\bar{X}$ behaves in repeated sampling ?**

# Sampling Distribution of the Sample Mean

- **Example 1 :**
  - Suppose a population consists of four numbers (N=4): 1,2,3,4
  - Since the four values are distinct, the population probability distribution assigns an equal probability of **1/4** to each value of x in the population.




| x | P(x) | x*P(x) | $x^2P(x)$ |
|---|---|---|---|
| 1 | 1/4 | 1/4 | 1/4 |
| 2 | 1/4 | 2/4 | 4/4 |
| 3 | 1/4 | 3/4 | 9/4 |
| 4 | 1/4 | 4/4 | 16/4 |

The **population mean** and **variance** are

$$\mu = \sum_{i=1}^{n} P(x_i) = 2.5 \qquad \sigma^2 = \sum_{i=1}^{N} x^2 P(x_i) - \mu^2 = 7.5 - 2.5^2 = 1.25$$

# Sampling Distribution of the Sample Mean

- **Step 1 :** Now, take a **random sample** of size **2 with replacement** from the population. How many possible samples are there?

4 * 4 = 16  possible samples

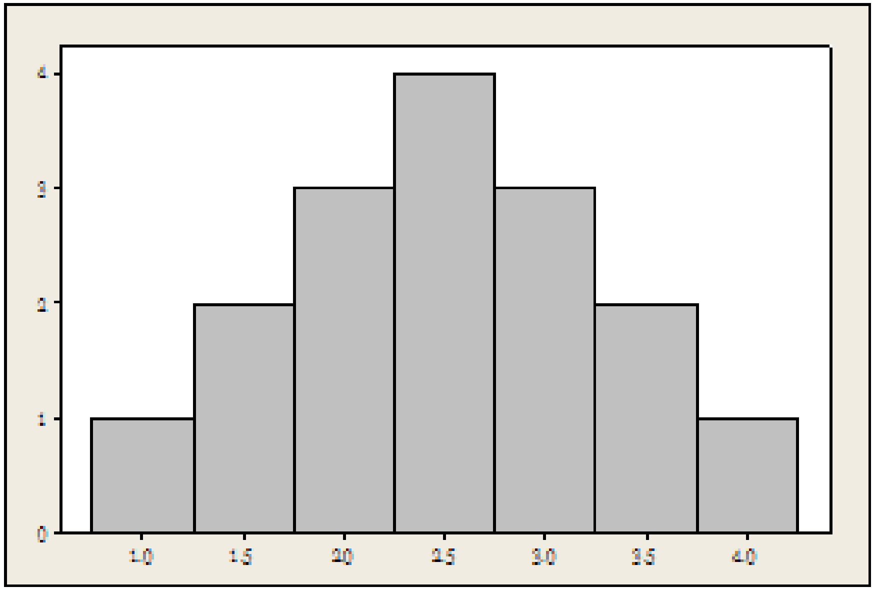
| Sample | Sample mean $\bar{X}$ | Sample | Sample mean $\bar{X}$ |
|---|---|---|---|
| (1 , 1) | 1 | (3 , 1) | 2 |
| (1 , 2) | 1.5 | (3 , 2) | 2.5 |
| (1 , 3) | 2 | (3 , 3) | 3 |
| (1 , 4) | 2.5 | (3 , 4) | 3.5 |
| (2 , 1) | 1.5 | (4 , 1) | 2.5 |
| (2 , 2) | 2 | (4 , 2) | 3 |
| (2 , 3) | 2.5 | (4 , 3) | 3.5 |
| (2 , 4) | 3 | (4 , 4) | 4 |

# Sampling Distribution of the Sample Mean

- **Step 2:** Now construct the **probability distribution** for the sample mean $\bar{X}$.

| $\bar{X}$ | $P(\bar{X})$ | $\bar{X} * P(\bar{X})$ | $\bar{X}^2 * P(\bar{X})$ |
|---|---|---|---|
| 1 | 1/16 | 1/16 | 1/16 |
| 1.5 | 2/16 | 3/16 | 4.5/16 |
| 2 | 3/16 | 6/16 | 12/16 |
| 2.5 | 4/16 | 10/16 | 25/16 |
| 3 | 3/16 | 9/16 | 27/16 |
| 3.5 | 2/16 | 7/16 | 24.5/16 |
| 4 | 1/16 | 4/16 | 16/16 |
| | **1** | **40/16** | **110/16** |

Close to Normal Distribution




For this probability distribution of $\bar{X}$, the **mean** of the **sample mean $\bar{X}$** and the **variance** of the sample mean $\bar{X}$ are :

$\mu_{\bar{X}} =$

$\sigma^2_{\bar{X}} =$

# Sampling Distribution of the Sample Mean

- **Conclusions :**

*1)* $\mu_{\overline{X}} =$

*2)* $\sigma^2_{\overline{X}} =$ _____ , where **n** is the sample size.

3) Whether the distribution of the original population in normal or not, the distribution of the **sample mean** is close to **Normal**.

# Central Limit Theorem 中央極限定理

- **The Central Limit Theorem (C.L.T.) 中央極限定理**

If random sample of n observations are drawn from a population with mean μ and standard deviation σ，then when **n** is **large** ($n \geq 30$), the **sampling distribution** of $\bar{X}$ is approximately normally distributed with $\mu_{\bar{X}} = \mu$, and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{, if } n \geq 30$$

The **approximation** will become more and more **accurate** as **n** becomes large.

**Remark:** If the **population** is **normal**, then the distribution of the sample mean $\bar{X}$ will always be **normal**, regardless of the sample size (n).

# Sampling Distribution of the Sample Mean

- **Example 1 :**
  - Suppose that X follows a distribution with mean $\mu$=10 and variance $\sigma^2$= 4. A sample of size 25 are drawn from this population. What is the distribution of $\overline{X}$ ?

# Application of the Central Limit Theorem

- **Example 2 :**
  - The **average** vitamin B-2 content of a certain brand of vitamins is **30** mg with a **standard deviation** of **2** mg. A quality control inspector selects **36** pills for testing. What is the probability that the **average** vitamin B-2 content of these 36 pills is **less than 28 mg** ?

# Application of the Central Limit Theorem

- **Example 3 :**
  - If a 1-gallon can of a certain kind of paint covers on the average 513.3 square feet with a standard deviation of 31.5 square feet, what is the probability that the mean area covered by a sample of 40 of these 1-gallon cans will be anywhere from 510.0 to 520.0 square feet ?

# Sampling Distribution of the Sample Proportion

- **What is the Sampling Distribution of the Sample Proportion, $\hat{p}$**
  - p：Population Proportion
  - $\hat{p}$：Sample proportion = x/n = 成功次數/總試驗次數

- **Theorem：Sampling Distribution of $\hat{p}$**

When the sample size **n** is large, the sampling distribution of $\boldsymbol{\hat{p}}$ is approximately normal with **mean** $p$ and **standard deviation** $\sqrt{\frac{pq}{n}}$.

$$\boxed{\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)}$$

# Sampling Distribution of the Sample Proportion

- **Example 4 :**
  - A production line at a manufacturing company produces **10%** defective items. If a sample of n=64 items is taken, what is the probability that the **sample defective rate** is **less** than 8%?

# Sampling Distribution of Sum

- **Theorem 6.10 on page 241**
  - **Sampling Distribution of Sum of Normal Random Variables**

Let $a_1, a_2, \cdots, a_n$ be constants and let $Y_1, Y_2, \cdots, Y_n$ be $n$ normally distributed random variables with $E(Y_i) = \mu_i$, $V(Y_i) = \sigma_i^2$, and $Cov(Y_i, Y_j) = \sigma_{ij}$ $(i = 1, 2, \cdots, n)$. Then the sampling distribution of a linear combination of the normal random variables

$$\ell = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n$$

possesses a normal density function with mean and variance

$$E(\ell) = \mu = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n$$

and

$$\begin{aligned} V(\ell) = {} & a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2 \\ & + 2a_1 a_2 \sigma_{12} + 2a_1 a_3 \sigma_{13} + \cdots + 2a_1 a_n \sigma_{1n} \\ & + 2a_2 a_3 \sigma_{23} + \cdots + 2a_2 a_n \sigma_{2n} \\ & + \cdots + 2a_{n-1} a_n \sigma_{n-1,n} \end{aligned}$$

# Sampling Distribution of Sum

- **Example 6.20 on page 241**

Suppose you select independent random samples from two normal populations, $n_1$ observations from population 1 and $n_2$ observations from population 2. If the means and variances for populations 1 and 2 are $(\mu_1, \sigma_1^2)$ and $(\mu_2, \sigma_2^2)$, respectively, and if $\overline{Y}_1$ and $\overline{Y}_2$ are the corresponding sample means, find the distribution of the difference $(\overline{Y}_1 - \overline{Y}_2)$.

# Solution

The means and variances of the sample means are

$$E(\bar{Y}_i) = \mu_i \quad \text{and} \quad V(\bar{Y}_i) = \frac{\sigma_i^2}{n_i} \quad (i = 1, 2)$$

Then, $\ell = \bar{Y}_1 - \bar{Y}_2$ is a linear function of two normally distributed random variables, $\bar{y}_1$ and $\bar{y}_2$. According to Theorem 6.10, $\ell$ will be normally distributed with

$$\boxed{E(\ell) = \mu_\ell = E(\bar{Y}_1) - E(\bar{Y}_2) = \mu_1 - \mu_2}$$

$$V(\ell) = \sigma_\ell^2 = (1)^2 V(\bar{Y}_1) + (-1)^2 V(\bar{Y}_2) + 2(1)(-1)Cov(\bar{Y}_1, \bar{Y}_2)$$

But, since the samples were independently selected, $\bar{Y}_1$ and $\bar{Y}_2$ are independent and $Cov(\bar{Y}_1, \bar{Y}_2) = 0$. Therefore,

$$\boxed{V(\ell) = \sigma_1^2 / n_1 + \sigma_2^2 / n_2}$$

We have shown that $(\bar{Y}_1 - \bar{Y}_2)$ is a normally distributed random variable with mean $(\mu_1 - \mu_2)$ and variance $(\sigma_1^2 / n_1 + \sigma_2^2 / n_2)$.

# Sampling Distribution of the Sample Mean

## 2) What is the Sampling Distribution of the Sample Mean, $\bar{X}$ (When σ is unknown) ？

- **Theorem**
  - If $\bar{X}$ is the mean of a random sample of size **n** taken from a **normal** population having the mean μ and the variance $\sigma^2$ , then the sample statistic

$$\boxed{T = \frac{\bar{X} - \mu}{s/\sqrt{n}}}$$

has a **T-distribution** with **degrees of freedom**(d.f. 自由度) ν= n-1.

Note: **T-distribution** is also called "**Student's T-distribution**"

# Student’s T Distribution

## • What is “Student’s” T-Distribution ?

The probability distribution of **T** statistic was first published in 1908 in a paper by **W.S. Gosset**. At the time, Gosset was employed by an Irish brewery (釀酒場) that disallowed publication of research by members of its staff. To circumvent this restriction, he published his work secretly under the name “**Student**”. Consequently, the distribution of **T** is usually called the **Student’s T-distribution**, or simply the **T**-distribution.

**William Sealy Gosset**

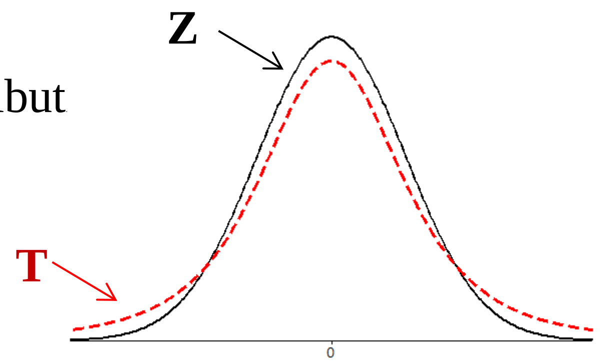
*Student* in 1908

| | |
|---|---|
| **Born** | June 13, 1876<br>Canterbury, Kent, England |
| **Died** | October 16, 1937 (aged 61)<br>Beaconsfield, Buckinghamshire, England |

source:
http://en.wikipedia.org/wiki/William_Sealy_Gosset

# Student's T Distribution

- **Properties of T-Distribution**
  - The **T**-distribution is very much like a **Z**-distribut




- **Comparison of T-distribution and Z-distribution**

  1) Both are **symmetric**, bell-shaped.
  2) Both have a **mean** of 0 .
  3) T is **more variable** than Z in repeated sampling. (There is more area in the tails of the T-distribution, and the Z-distribution is higher in the middle).
  4) As the number of **d.f.** increases (i.e., as **n** increases ) without limit, the **T**-distribution approaches **Z**-distribution.

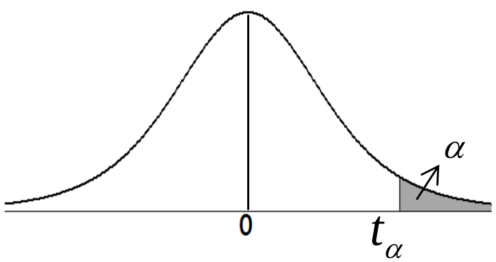
# Degree of Freedom

- **What is the Degree of Freedom?**
  - We use the **degrees of freedom** as a measure of **sample information**.
  - For example, we say that the **T** statistic has degrees of freedom ***n*-1**.
- **Why？**
  - There are **n *degrees of freedom*** or **independent pieces** of information in the random sample of size ***n*** from the normal distribution.
  - In calculating **T** = $\frac{\bar{X}-\mu}{s/\sqrt{n}}$, we **do not know σ** and need to use the sample data to estimate **σ**. When the data (the values in the sample) are used to compute the mean $\bar{\mathbf{X}}$ for obtaining $S^2 = \sum_{i=1}^{n}(x_i - \bar{X})^2/(n-1)$, there is **1** less degree of freedom in the information used to estimate $\boldsymbol{\sigma}^2$.

# Student's T Distribution

- **T-distribution table**
  - **TABLE 7 in the Appendix B** of the textbook **(page 910**) gives the value of $t_\alpha$ which locates an area of $\alpha$ in the **upper** tail of the T-distribution for various values of $\alpha$ and for **d.f.** ranging from **1** to $\infty$ .

- **Note**
  - When d.f. $\geqq 29$ or $\mathbf{n} \geqq 30$, **Z**-distribution is very close to **T**-distribution.




| $\nu$ | $t_{.100}$ | $t_{.050}$ | $t_{.025}$ | $t_{.010}$ | $t_{.005}$ | $t_{.001}$ | $t_{.0005}$ |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.076 | 31.821 | 63.657 | 318.310 | 636.620 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.326 | 31.598 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.213 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.767 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.659 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 | 3.646 |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551 |
| 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| 120 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 | 3.373 |
| $\infty$ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 |

# Student's T Distribution

- **Example 1 :**
  - Find $t_{\alpha}$ and $t_{\alpha/2}$ when $\alpha$=0.05 and n=6.
- **Solution :**

| ν | $t_{.100}$ | $t_{.050}$ | $t_{.025}$ | $t_{.010}$ | $t_{.005}$ | $t_{.001}$ | $t_{.0005}$ |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.076 | 31.821 | 63.657 | 318.310 | 636.620 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.326 | 31.598 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.213 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |

  - ANS: $t_{.05,\,5}=$ 2.015 , $t_{.025,\,5}=$ 2.571

# Student's T Distribution

- **Example 2 :**
  - Find $t_\alpha$ and $t_{\alpha/2}$ when α=0.01 and n=20.
- **Solution :**

| ν | $t_{.100}$ | $t_{.050}$ | $t_{.025}$ | $t_{.010}$ | $t_{.005}$ | $t_{.001}$ | $t_{.0005}$ |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.076 | 31.821 | 63.657 | 318.310 | 636.620 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.326 | 31.598 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.213 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| … | | | | … | | | |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |

  - ANS: $t_{.01,\,19}= 2.539$ , $t_{.005,\,19}= 2.861$

# Student's T Distribution

- **Example 3 :**
  - Find $t_{\alpha}$ and $t_{\alpha/2}$ when $\alpha$=0.10 and n=42.
- **Solution :**

| ν | $t_{.100}$ | $t_{.050}$ | $t_{.025}$ | $t_{.010}$ | $t_{.005}$ | $t_{.001}$ | $t_{.0005}$ |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.076 | 31.821 | 63.657 | 318.310 | 636.620 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.326 | 31.598 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.213 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| … | | | | … | | | |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551 |
| 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| 120 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 | 3.373 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 |

  - ANS: $t_{.10,\infty}$= 1.282 , $t_{.05,\infty}$= 1.645

# The Sampling Distributions Related to the Normal Distribution

# $\chi^2$ -Distribution

## 1. $\chi^2$ -Distribution

– If $\mathbf{s^2}$ is the variance of a random sample of size **n** taken from a ***Normal*** population having the variance $\boldsymbol{\sigma^2}$, then

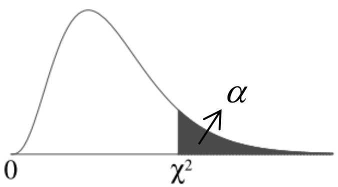
$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

has a (Greek letter, **Chi**) distribution with the **d.f.** = ν = n-1

– **TABLE 8 on pages 911-912** of the Appendix gives the value of $\chi^2_\alpha$ which locates an area of **α** in the **upper** tail of the $\chi^2$ -distribution for various values of **α** and **d.f**.

# $\chi^2$ -Distribution Table

- **Example 1 :**
  - If n= 20, use TABLE 8 (p.911) to determine $\chi^2_{0.05}$ = ?




| $df$ | $\chi^2_{.995}$ | $\chi^2_{.990}$ | $\chi^2_{.975}$ | $\chi^2_{.950}$ | $\chi^2_{.900}$ | $\chi^2_{.100}$ | $\chi^2_{.050}$ | $\chi^2_{.025}$ | $\chi^2_{.010}$ | $\chi^2_{.005}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.000 | 0.000 | 0.001 | 0.004 | 0.016 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 4.605 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 7.779 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.833 | 15.086 | 16.750 |
| 6 | 0.676 | 0.872 | 1.237 | 1.635 | 2.204 | 10.645 | 12.592 | 14.449 | 16.812 | 18.548 |
| 7 | 0.989 | 1.239 | 1.690 | 2.167 | 2.833 | 12.017 | 14.067 | 16.013 | 18.475 | 20.278 |
| 8 | 1.344 | 1.646 | 2.180 | 2.733 | 3.490 | 13.362 | 15.507 | 17.535 | 20.090 | 21.955 |
| 9 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 14.684 | 16.919 | 19.023 | 21.666 | 23.589 |
| 10 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 15.987 | 18.307 | 20.483 | 23.209 | 25.188 |
| 11 | 2.603 | 3.053 | 3.816 | 4.575 | 5.578 | 17.275 | 19.675 | 21.920 | 24.725 | 26.757 |
| 12 | 3.074 | 3.571 | 4.404 | 5.226 | 6.304 | 18.549 | 21.026 | 23.337 | 26.217 | 28.300 |
| 13 | 3.565 | 4.107 | 5.009 | 5.892 | 7.042 | 19.812 | 22.362 | 24.736 | 27.688 | 29.819 |
| 14 | 4.075 | 4.660 | 5.629 | 6.571 | 7.790 | 21.064 | 23.685 | 26.119 | 29.141 | 31.319 |
| 15 | 4.601 | 5.229 | 6.262 | 7.261 | 8.547 | 22.307 | 24.996 | 27.488 | 30.578 | 32.801 |
| 16 | 5.142 | 5.812 | 6.908 | 7.962 | 9.312 | 23.542 | 26.296 | 28.845 | 32.000 | 34.267 |
| 17 | 5.697 | 6.408 | 7.564 | 8.672 | 10.085 | 24.769 | 27.587 | 30.191 | 33.409 | 35.718 |
| 18 | 6.265 | 7.015 | 8.231 | 9.390 | 10.865 | 25.989 | 28.869 | 31.526 | 34.805 | 37.156 |
| 19 | 6.844 | 7.633 | 8.907 | 10.117 | 11.651 | 27.204 | 30.144 | 32.852 | 36.191 | 38.582 |
| 20 | 7.434 | 8.260 | 9.591 | 10.851 | 12.443 | 28.412 | 31.410 | 34.170 | 37.566 | 39.997 |

# $\chi^2$ -Distribution

- **Example 2:** See Example 6.24 on page 249
  - Consider a cannery that produces 8-ounce cans of processed corn. Quality control engineers have determined that process is operating properly when the true variation $\sigma^2$ of the fill amount per can is less than 0.0025. A random sample of n=10 cans is selected from a day's production, and the fill amount (in ounces) recorded for each. Of interest is the sample variance, $S^2$. If, in fact, $\sigma^2$ =0.001, find the probability that $S^2$ exceeds 0.0025. Assume that the fill amounts are normally distributed.

# $\chi^2$ -Distribution

- **Solution (1/2)**

We want to calculate $P(S^2 > 0.0025)$. Assume the sample of 10 fill amount is selected from a normal distribution. Theorem 6.11 states that the statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

has a chi-square probability distribution with $\nu=(n-1)$ degrees of freedom. Consequently, the probability we seek can be written

$$P(S^2 > 0.0025) = P\left[\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)(0.0025)}{\sigma^2}\right] = P\left[\chi^2 > \frac{(n-1)(0.0025)}{\sigma^2}\right]$$

Substituting $n=10$ and $\sigma^2=0.001$, we have

$$P(S^2 > 0.0025) = P\left[\chi^2 > \frac{9(0.0025)}{0.001}\right] = P(\chi^2 > 22.5)$$

# $\chi^2$ -Distribution

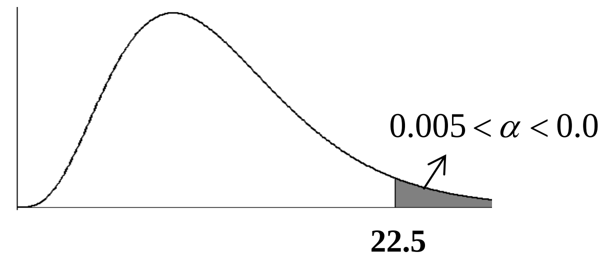
- **Solution (2/2)**

Then, we want to find the probability **α** such that $\chi^2_\alpha > 22.5$

for n=10 (ν =9) , we find that

$\chi^2_{0.01} = 21.666$ and $\chi^2_{0.005} = 23.589$

i.e., $0.005 < P(\chi^2 > 22.5) < 0.01$




| df | $\chi^2_{.995}$ | $\chi^2_{.990}$ | $\chi^2_{.975}$ | $\chi^2_{.950}$ | $\chi^2_{.900}$ | $\chi^2_{.100}$ | $\chi^2_{.050}$ | $\chi^2_{.025}$ | $\chi^2_{.010}$ | $\chi^2_{.005}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.000 | 0.000 | 0.001 | 0.004 | 0.016 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 4.605 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 7.779 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.833 | 15.086 | 16.750 |
| 6 | 0.676 | 0.872 | 1.237 | 1.635 | 2.204 | 10.645 | 12.592 | 14.449 | 16.812 | 18.548 |
| 7 | 0.989 | 1.239 | 1.690 | 2.167 | 2.833 | 12.017 | 14.067 | 16.013 | 18.475 | 20.278 |
| 8 | 1.344 | 1.646 | 2.180 | 2.733 | 3.490 | 13.362 | 15.507 | 17.535 | 20.090 | 21.955 |
| 9 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 14.684 | 16.919 | 19.023 | 21.666 | 23.589 |
| 10 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 15.987 | 18.307 | 20.483 | 23.209 | 25.188 |

Thus, the probability that the variance of the sample fill amounts exceeds 0.0025 is small (between 0.005 and 0.01) when the true population variance $\sigma^2$ equals 0.001.

# T-Distribution

## 2. T-Distribution

– Definition: Let Z be a **standard Normal** random variable and $\chi^2$ be a Chi-square random variable with degrees of freedom ν, then

$$T = \frac{Z}{\sqrt{\chi^2 / \nu}}$$

has a **T** distribution with ν numerator d.f. (分子自由度) and $\nu_2$ denominator d.f. (分母自由度)

- **Theorem :**

– If $\overline{X}$ is the mean of a random sample of size **n** taken from a **normal** population having the mean μ and the variance $\sigma^2$, then the sample statistic

$$T = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

has a **T-distribution** with **degrees of freedom (d.f.) (**自由度) ν = n-1.

# T-Distribution

- **Example 3:** See Example 6.25 on page 251
  - Suppose the random variables $\bar{Y}$ and $S^2$ are the mean and variance of a random sample of n observations from a normally distributed population with mean μ and variance $\sigma^2$. It can be shown that $\bar{Y}$ and $S^2$ are statistically independent when the sampled population has a normal distribution. Use this result to show that

$$T = \frac{\bar{Y} - \mu}{s / \sqrt{n}}$$

possesses a T distribution with $\nu=(n-1)$ degrees of freedom.

# T-Distribution

- **Solution**

We know from Theorem 6.10 that $\bar{Y}$ is normally distributed with μ and variance $\sigma^2/n$. Therefore,

$$Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$$

is a standard normal random variable. We also know from Theorem 6.11 that

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

is a $\chi^2$ random variable with ν=(n-1) degree of freedom. Then, using Definition 6.15 and the information that $\bar{Y}$ and $S^2$ are independent, we conclude that

$$T = \frac{Z}{\sqrt{\chi^2 / v}} = \frac{\dfrac{\bar{Y} - \mu}{\sigma / \sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2} \Big/ (n-1)}} = \frac{\bar{Y} - \mu}{S / \sqrt{n}}$$

has a Student's T distribution with ν=(n-1) degrees of freedom.

# F-Distribution

## 3. F-Distribution

– Definition: Let $\chi_1^2$ and $\chi_2^2$ be two **independent chi-square** random variables with $\nu_1$ and $\nu_2$ degrees of freedom, respectively, then

$$F = \frac{\chi_1^2 / \nu_1}{\chi_2^2 / \nu_2}$$

has a **F** distribution with $\nu_1$ numerator d.f. (分子自由度) and $\nu_2$ denominator d.f. (分母自由度)

- **Theorem :**

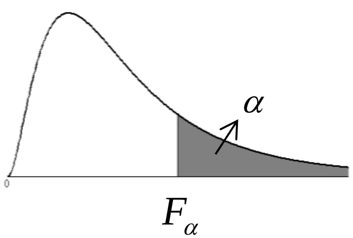
– If $s_1^2$ and $s_2^2$ are the variances of a random sample of size $n_1$ and $n_2$ taken from two normal population having the same variances, then

$$F = \frac{s_1^2}{s_2^2}$$

has a **F** distribution with **d. f.** = $(\nu_1, \nu_2) = (n_1 - 1, n_2 - 1)$.

# F-Distribution Table

- **F-Distribution Table**
  - TABLE 9 on pages 913-916 of the Appendix gives the value of $F_\alpha$ which locates an area of α in the **upper** tail of the F-distribution for various values of **α** and **d.f.**
- **Example**: If $n_1 = 7$, $n_2 = 13$, use TABLE 9 to determine $F_{.01} = ?$




| Denominator Degrees of Freedom ($\nu_2$) \ Numerator Degrees of Freedom ($\nu_1$) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 4052.2 | 4999.5 | 5403.4 | 5624.6 | 5763.7 | 5859.0 | 5928.4 | 5981.1 |
| 2 | 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 |
| 3 | 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 |
| 4 | 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 |
| 5 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 |
| 6 | 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 |
| 7 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 |
| 8 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 |
| 9 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 |
| 10 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 |
| 11 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 |
| 13 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 |
| 14 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 |

# 本單元結束