Chapter 1

1.1 Biochemical Studies
1.2 Units of Measurements
1.3 Weak Electrolytes
1.4 Buffer Solution
1.6 Quantitative Biochemical Measurements
1.7.1-1.7.2 Principle of Clinical Biochemical Analysis

Others:

- Receiver Operating Characteristic Curve
- Clinical Sensitivity and Specificity

SI Units

Unit	Symbol	SI equivalent
Avogadro constant	L or N_A	$6.022 \times 10^{23}\ mol^{-1}$
Faraday constant	F	$9.648 \times 10^{4}\ C\ mol^{-1}$
Planck constant	h	$6.626 \times 10^{-34}\ J\ s$
Universal or molar gas constant	R	$8.314\ J\ K^{-1}\ mol^{-1}$
Molar volume of an ideal gas at s.t.p.		$22.41\ dm^3\ mol^{-1}$
Velocity of light in a vacuum	c	$2.997 \times 10^{8}\ m\ s^{-1}$
Energy		
Calorie	cal	4.184 J
Erg	erg	10^{-7} J
Electron volt	eV	1.602×10^{-19} J
Pressure		
Atmosphere	atm	101 325 Pa
Bar	bar	10^5 Pa
Millimetres of Hg	mmHg	133.322 Pa
Temperature		
Centigrade	°C	$(t$°C + 273.15) K
Fahrenheit	°F	$(t$°F − 32)5/9 + 273.15 K
Length		
Ångström	Å	10^{-10} m
Inch	in.	0.0254 m
Mass		
Pound	lb	0.4536 kg

s.t.p., standard temperature and pressure.

Basic principles

- **Molarity** : Number of moles of the substances in 1 dm^3 of solution.

- **One mole**: equal to molecular mass of the substance

- **Molecular mass**:

  Da: *daltons*

  kDa: *Kilodaltons* =1000 Da

  M_r: *no unit*

  Relative molecular mass

  = the molecular mass of a substance relative to 1/12 of the atomic mass of the ^{12}C .

Units for Different Concentrations

Table 1.5 **Interconversion of mol, mmol and μmol in different volumes to give different concentrations**

Molar (M)	Millimolar (mM)	Micromolar (μM)
1 mol dm^{-3} 1 mol l^{-3}	1 mmol dm^{-3}	1 μmol dm^{-3}
1 mmol cm^{-3}	1 μmol cm^{-3}	1 nmol cm^{-3}
1 μmol mm^{-3}	1 nmol mm^{-3}	1 pmol mm^{-3}

Biological substances are most frequently found at relatively low concentrations and in *in vitro* model systems the volumes of stock solutions regularly used for experimental purposes are also small. The consequence is that experimental solutions are usually in the mmol dm^{-3}, μmol dm^{-3} and nmol dm^{-3} range rather than molar. Table 1.5 shows the interconversion of these units.

Ion Strengths

Reason of deviation:

Presence of electrolytes will result in electrostatic interaction with other ions and solvents

Total ion charge in solution

$$M = 1/2 * (c_1 z_1^2 + c_1 z_1^2 + \ldots + c_n z_n^2)$$

c_1, c_2, …c_n: concentrations of each ion in *molarity*

z_1, z_2, …z_n: charge on the individual ion

Example 2 CALCULATION OF IONIC STRENGTHS

Question

Calculate the ionic strength of (i) 0. 1 M NaCl, (ii) 0. 1 M NaCl + 0.05 M KNO_3 + 0.01 M Na_2SO_4.

Answer Ionic strength can be calculated using the equation $\mu = \frac{1}{2}\sum cz^2$.

(i) Calculating cz^2 for each ion:

$$Na^+ = 0.1 \times (+1)^2 = 0.1\,M$$

$$Cl^- = 0.1 \times (-1)^2 = 0.1\,M$$

Hence

$$\frac{1}{2}\sum cz^2 = 0.2/2 = 0.1\,M$$

(ii)

$$Na^+ = 0.1 \times (+1)^2 + 0.02 \times (+1)^2 = 0.12\,M$$

$$Cl^- = 0.1 \times (-1)^2 = 0.10\,M$$

$$K^+ = 0.05 \times (+1)^2 = 0.05\,M$$

$$NO_3^- = 0.05 \times (-1)^2 = 0.05\,M$$

$$SO_4^{2-} = 0.01 \times (-2)^2 = 0.04\,M$$

Hence

$$\frac{1}{2}\sum cz^2 = \frac{1}{2}(0.36) = 0.18\,M$$

Note 1: the unit of ionic strength is M.

Note 2: that for Na_2SO_4 $c = 0.02$, since there are $2Na^+$ per mole.

Note 3: that for a 1 M 1 : 1 electrolyte such as NaCl, the ionic strength is 1 M; for a 1 M 2 : 1 electrolyte such as $MgCl_2$, the ionic strength is 3 M, for a 1 M 2 : 2 electrolyte such as $MgSO_4$, the ionic strength is 4 M and for a 3 : 1 electrolyte such as $FeCl_3$, the ionic strength is 6 M.

Note 4: As the concentration and ionic strength increase, this type of calculation becomes progressively inaccurate owing to the importance of activity coefficients.

Activity and Activity Coefficients

Activity : the effective concentration in solution

A_x = [Concentration] γ_x

γ_x : Activity coefficient

- The coefficient establish the relationship between activity and concentration.
- It will decrease when the ionic strength increases (include concentration, charge and ion mobility)

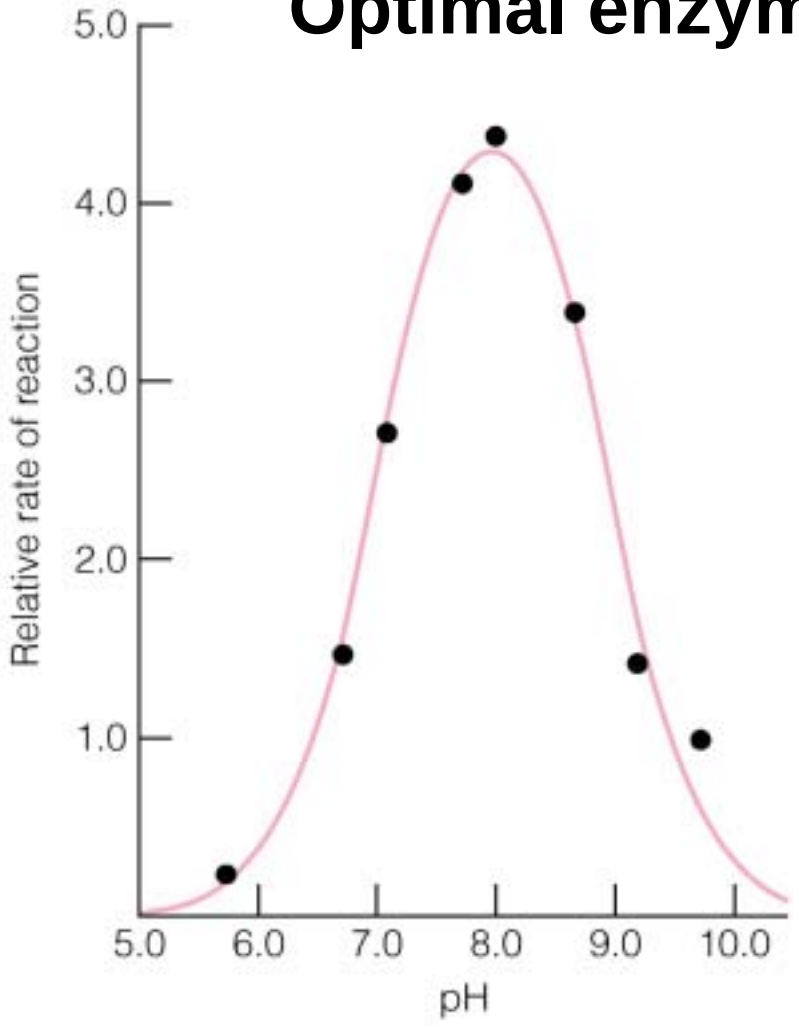
e.g. 0.001 M Mg^{2+} 0.872

Fe^{3+} 0.738

Except for very diluted solution, the effective concentrations are usually less than the actual concentration

Preparation of Buffer Solution

Optimal enzyme activity pH 8



α-Chymotrypsin: catalyzed cleavage of the C-N bond

Henderson-Hasselbalch Equation

For a weak acid, which dissociates as follows:

$$HA \rightarrow H^+ + A^-$$

$$\text{equilibrium constant} = K_{eq} = K_a = \frac{[H^+] \times [A^-]}{[HA]}$$

$$\log_{10}K_a = \log_{10}[H^+] + \log_{10}[A^-] - \log_{10}[HA]$$

$$-\log_{10}[H^+] = -\log_{10}K_a + \log_{10}[A^-] - \log_{10}[HA]$$

$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

$$pH = pK_a + \log_{10}\left(\frac{[\text{conjugate base}]}{[\text{conjugate acid}]}\right) = pK_a + \log_{10}\left(\frac{[\text{proton acceptor}]}{[\text{proton donor}]}\right)$$

Why is pKa useful?

$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

Perhaps it is useful to look at this in another way: if we consider the situation where the acid is one half dissociated, in other words where [A-] is equal to [HA], then, substituting in the Henderson-Hasselbalch Equation

pH = pKa + log10(1)

pH = pKa + 0

pH = pKa

This means that an acid is half dissociated when the pH of the solution is numerically equal to the pKa of the acid.

$$pH = pK_a + \log_{10}\left(\frac{[A^-]}{[HA]}\right)$$

$$HA \rightarrow H^+ + A^-$$

Acid	K_a		pK_a
Trichloroacetic	2×10^{-1}	$=10^{-0.7}$	0.7
Dichloroacetic	5×10^{-2}	$=10^{-1.3}$	1.3
Monochloroacetic	1.6×10^{-3}	$=10^{-2.8}$	2.8
Formic	2.1×10^{-4}	$=10^{-3.7}$	3.7
Benzoic	7.8×10^{-5}	$=10^{-4.1}$	4.1
Acetic	1.9×10^{-5}	$=10^{-4.7}$	4.7
H_2CO_3	2.9×10^{-7}	$=10^{-6.5}$	6.5
H_2S	5.8×10^{-8}	$=10^{-7.2}$	7.2
HCN	1.3×10^{-9}	$=10^{-8.9}$	8.9

Acids with the lowest pKa values are able to dissociate in solutions of low pH, i.e. even where the hydrogen ion concentration is high. Acids with higher pKa values dissociate only in solutions of high (more alkaline) pH.

Example 3 CALCULATION OF pH AND THE EXTENT OF IONISATION OF A WEAK ELECTROLYTE

Question

Calculate the pH of a 0.01 M solution of acetic acid and its fractional ionisation given that its K_a is 1.75×10^{-5}.

Answer

To calculate the pH we can write:

$$K_a = \frac{[\text{acetate}^-][\text{H}^+]}{[\text{acetic acid}]} = 1.75 \times 10^{-5}$$

Since acetate and hydrogen ions are produced in equal quantities, if x = the concentration of each then the concentration of unionised acetic acid remaining will be $0.01 - x$. Hence:

$$1.75 \times 10^{-5} = \frac{(x)(x)}{0.01 - x}$$

$$1.75 \times 10^{-7} - 1.75 \times 10^{-5}x = x^2$$

This can now be solved either by use of the quadratic formula or, more easily, by neglecting the x term since it is so small. Adopting the latter alternative gives:

$$x^2 = 1.75 \times 10^{-7}$$

hence

$$x = 4.18 \times 10^{-4}\text{M}$$

hence

$$\text{pH} = 3.38$$

Note that this solution has ignored the activity coefficients of the acetate and hydrogen ions. They are 0.90 and 0.91 respectively at 0.01 M and 25 °C. Inserting these values into the above expression and assuming that the activity coefficient of acetic acid is unity gives:

$$1.75 \times 10^{-5} = \frac{(x)0.90(x)0.91}{0.01 - x}$$

Solving this equation for x gives a value of 4.61×10^{-4}M, and hence a pH of 3.33. This illustrates the relatively small influence of activity coefficients in this case.

The fractional ionisation (α) of the acetic acid is defined as the fraction of the acetic acid that is in the form of acetate and is therefore given by the equation:

$$\alpha = \frac{[\text{acetate}]}{[\text{acetate}] + [\text{acetic acid}]}$$

$$= \frac{4.18 \times 10^{-4}}{4.18 \times 10^{-4} + 0.01 - 4.18 \times 10^{-4}}$$

$$= \frac{4.18 \times 10^{-4}}{0.01}$$

$$= 4.18 \times 10^{-2} \quad \text{or } 4.18\%$$

Thus the majority of the acetic acid is present as the unionised form. If the pH is increased above 3.33 the proportion of acetate present will increase in accordance with the Henderson–Hasselbalch equation.

Quantitative Biochemical Measurements

- What to study? **Model**
- How to study **Method**
- Is the results correct? **Performance**
- How to interpret results? **Report**

Quantitative Biochemical Measurements

- Analytical Considerations:

(I) Test Model :

in vivo v.s. ***in vitro***

Material: urine, serum/plasma/blood

Matrix v.s Analyte

Sampling v.s population

in vivo v.s. *in vitro*

In vivo: In a living cell or organism

In vitro: Biological or chemical work done in the test tube

(in glass)

Sampling v.s Population

Population: Representative portion of analyte

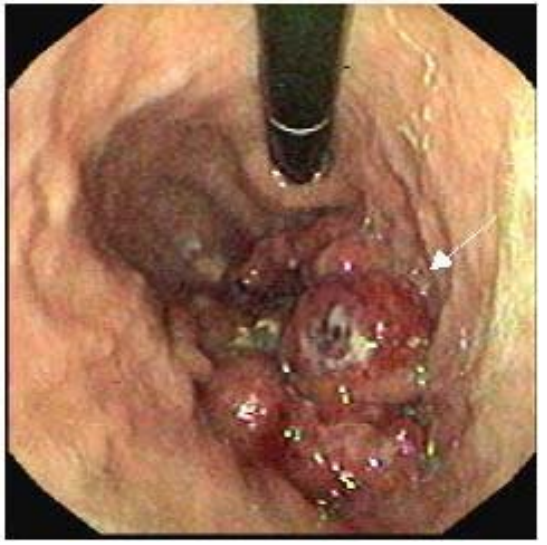
Heterogeneous v.s Homogeneous

Extraction Methods:

- Liquid extraction
- Solid-phase extraction
- Laser microdisection

(cancer cell)

- .....etc



Quantitative Biochemical Measurements

(II) Selection of Analytical Methods

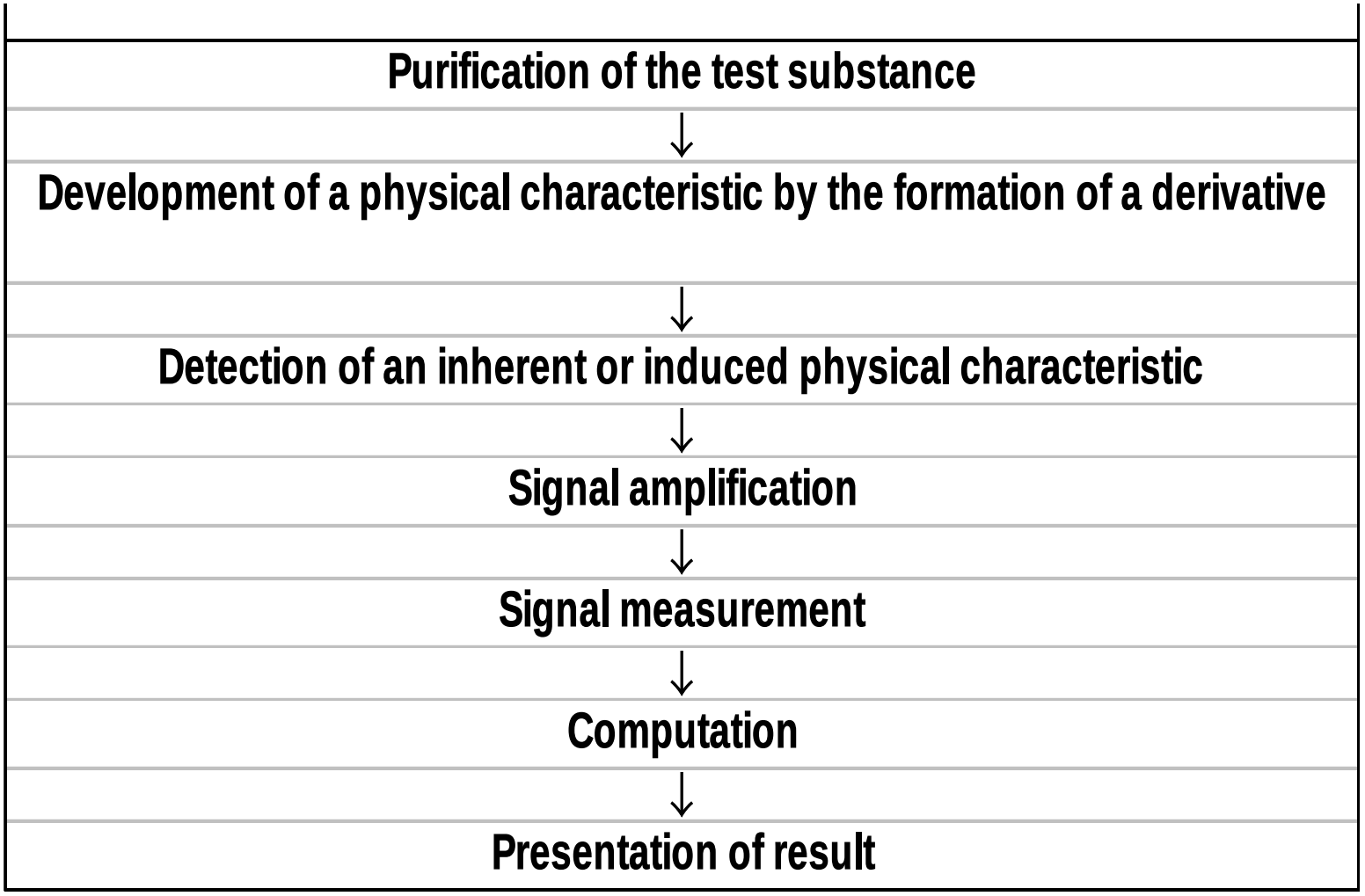
- Qualitative v.s Quantitative analysis
- Chemical and physical properties of analyte
- Precision, accuracy and detection limit
- Interference from matrix
- Cost and value
- Possible hazard and risk

NOTE

Physical Basis of Analytical Methods

Physical properties that can be measured with some degree of precision	Examples of properties used in the		
	Protein	Lead	Oxygen
Extensive			
Mass	+	+	
Volume			+
Mechanical			
Specific gravity	+		
Viscosity	+		
Surface tension	+		
Spectral			
Absorption	+	+	
Emission			
Fluorescence			
Turbidity	+		
Rotation			
Electrical			
Conductivity			
Cuurent/voltage			+
Half-cell potential			+
Nuclear			
Radioactivity	+		

Major manipulative steps in a generalized method of analysis



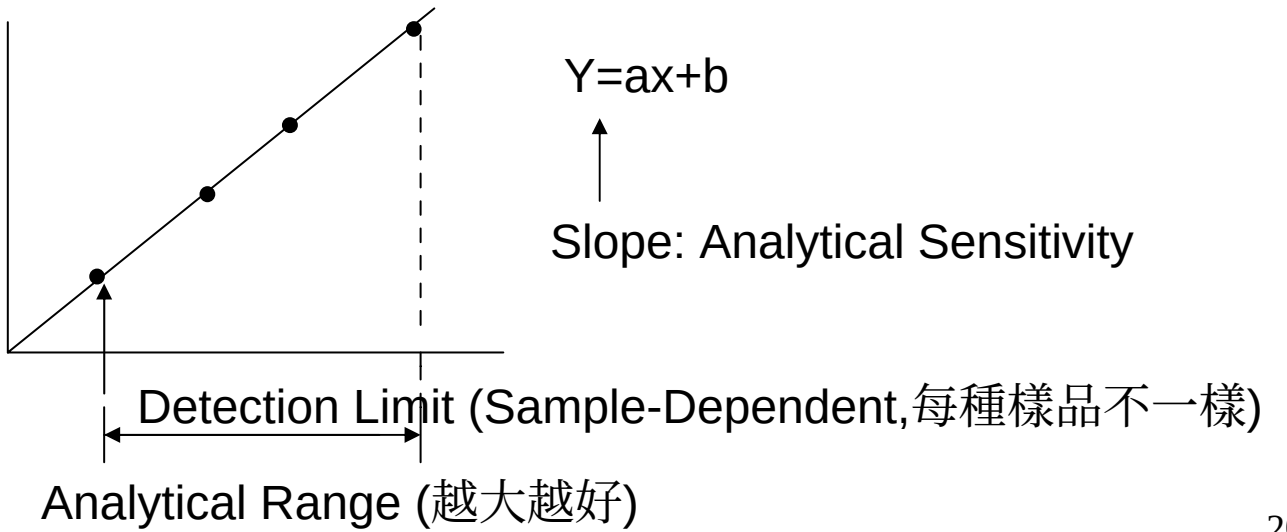
Basic Terminology in Analytical Science

Sensitivity

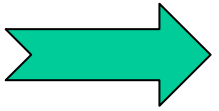
1.Detection limit: Smallest concentration/amount of analyte that can be detected.

2.Analytical sensitivity :The ability to measure the change in quality of analyte.

3.Analitical Range: The range of concentrations of analyte that can be measured reproducibility.



Specificity (Selectivity)

- The ability to detect ONLY the test analyte in the presence of other interfering substances.
- Loss of Specificity  false positive

  (resulting from interference in the sample)

Robustness :

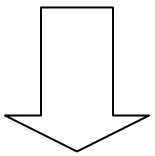
The ability of the method to give consistent result in the change of experimental parameters,

E.g: pH. temperature……etc

Quantitative Biochemical Measurements

(III) Experimental Errors

- Systematic error
- Random error



Standard Operation Procedures (SOP)

NOTE

Systematic Error

- Constant or proportional (**Bias)**
- Also called

Overestimation /underestimation

(1) **Analyst error**: pipette, calibration, solution preparation, method design

(2) **Instrumental error**: contamination of instrument, power fluctuation, variation in T, pH, electronic noise

(3) **Method error**: side reaction, incomplete reaction

Identification of Systematic Errors

- Blank sample
- Standard reference sample
- Alternative methods
- External quality assessment sample

Random Error

- Variable, either positive or negative
- also called

  Indeterminate error

(1) **Instrumental error**: random electric noise

Standard Operating Procedures (SOP)

Include:

- Quantity/quality of reagent
- Preparation of standard solution
- Calibration of instrument
- Methodology of actual analytical procedures

Assessment of performance of analytical method

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effector Memory T Cells, Early Metastasis, and Survival in Colorectal Cancer

Franck Pagès, M.D., Ph.D., Anne Berger, M.D., Ph.D., Matthieu Camus, M.D.,
Fatima Sanchez-Cabo, Ph.D., Anne Costes, B.S., Robert Molidor, Ph.D.,
Bernhard Mlecnik, M.Sc., Amos Kirilovsky, M.Sc., Malin Nilsson, B.S.,
Diane Damotte, M.D., Ph.D., Tchao Meatchi, M.D., Patrick Bruneval, M.D., Ph.D.,
Paul-Henri Cugnenc, M.D., Ph.D., Zlatko Trajanoski, Ph.D.,
Wolf-Herman Fridman, M.D., Ph.D., and Jérôme Galon, Ph.D.

(大腸直腸癌)

NEJM, 353, 2654-2666, 2005

Background

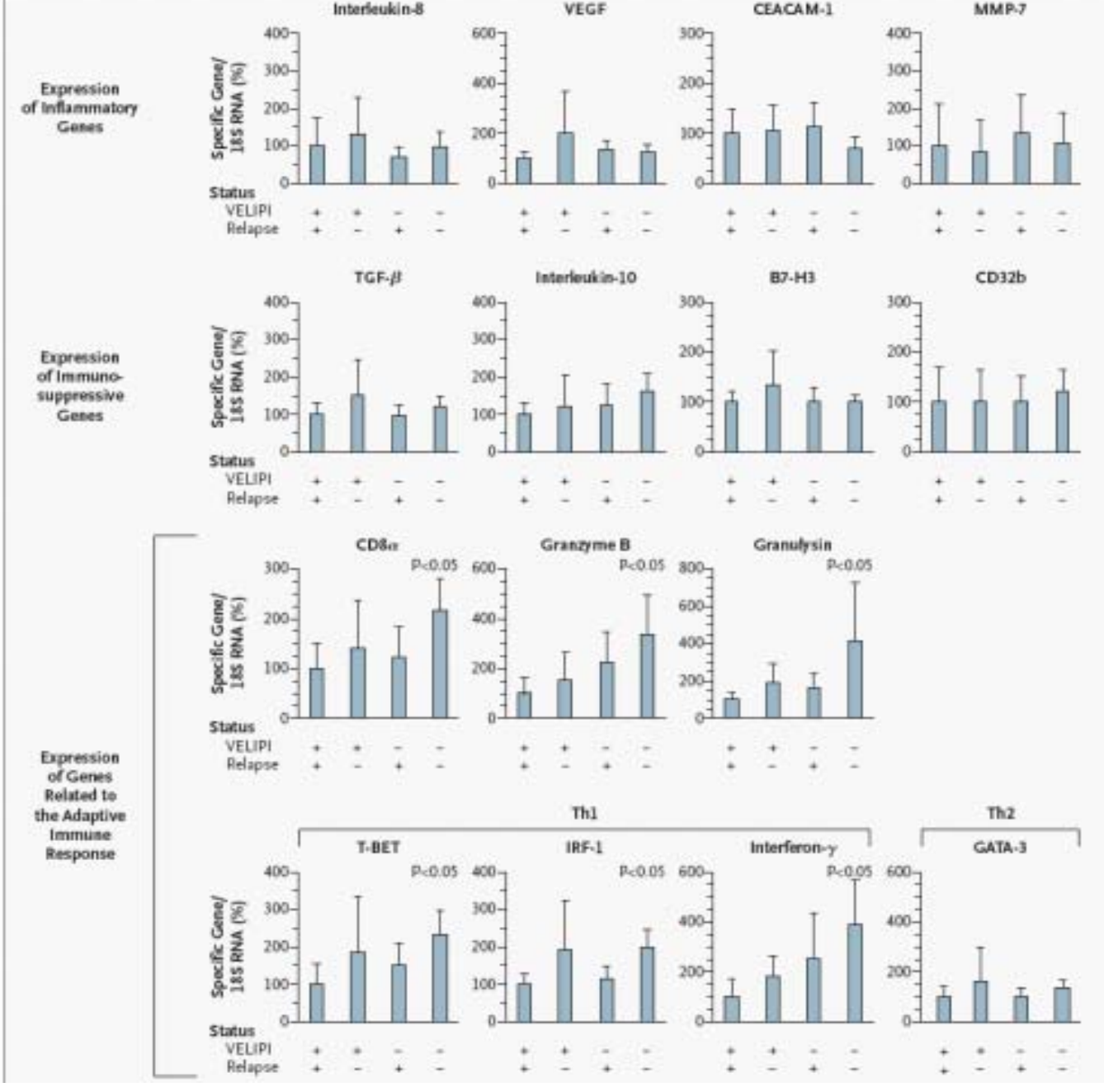
The role of tumor-infiltrating (浸潤) immune cells in the early metastatic invasion (轉移性侵犯) of colorectal cancer（直腸癌）is unknown.

Methods

We studied pathological signs of early metastatic invasion (venous emboli 靜脈栓塞 and lymphatic 淋巴 and perineural invasion(神經旁間隙) in 959 specimens of resected colorectal cancer. The local immune response within the tumor was studied by flow cytometry (39 tumors), low density-array real-time polymerase-chain-reaction assay (75 tumors), and tissue microarrays (415 tumors).

Table 1. Disease-free and Overall Survival among 959 Patients with Colorectal Cancer.

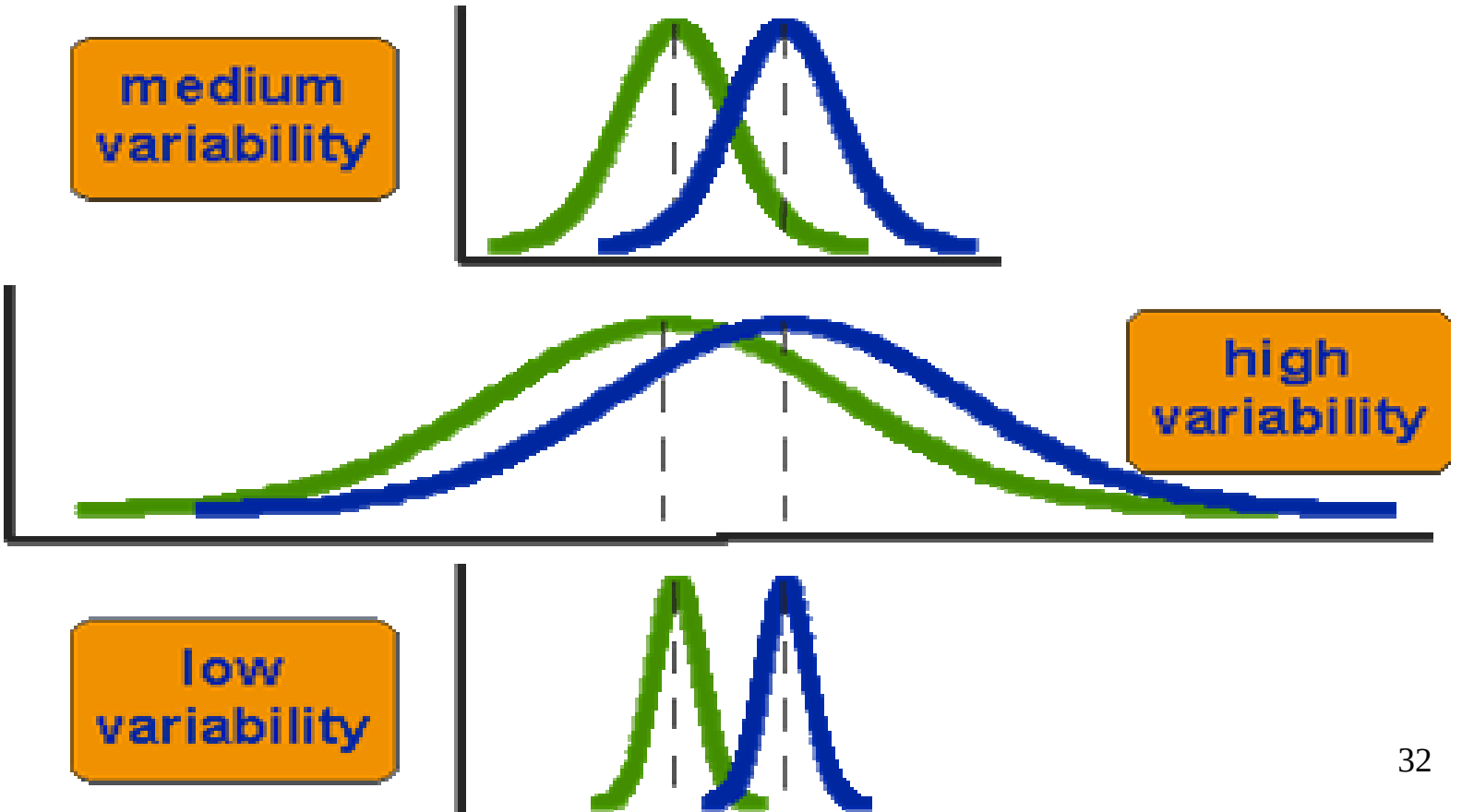
Characteristic	No. of Patients	Disease-free Survival			Overall Survival		
		5 yr	Median	P Value*	5 yr	Median	P Value*
		%	*mo*		%	*mo*	
Tumor (T) stage†				<0.001			<0.001
pTis	39	48.7	55.7		48.7	55.7	
pT1	54	42.6	52.2		44.4	53.8	
pT2	156	40.4	43.6		44.2	49.1	
pT3	502	23.7	16.5		26.7	25.8	
pT4	208	16.8	1.6		17.8	16.8	
Nodal (N) status				<0.001			<0.001
Negative	568	35.4	34.6		38.6	43.1	
Positive	384	15.1	4.3		16.7	16.9	
Nx‡	7						
Distant metastases (M)				<0.001			<0.001
None detected	747	34.5	32.6		37.6	41.1	
Present	212	0.5	0.1		0.9	12.3	
Dukes' stage				<0.001			<0.001
A	84	47.0	55.6		47.0	55.6	
B	438	37.2	39.2		41.1	46.8	
C	228	24.7	19.5		27.3	28.1	
D	209	0.5	0.1		1.0	12.1	
Sex				0.38			0.47
Male	494	25.9	16.4		28.5	29.4	
Female	465	28.2	19.3		30.5	27.3	
Location				0.20			0.14
Right side of colon	242	23.9	14.5		24.7	19.7	
Transverse colon	50	7.8	9.2		9.8	22.2	
Left side of colon	83	28.6	15.3		31.0	27.2	
Sigmoid colon	297	26.8	14.7		29.5	29.5	
Rectum	287	32.4	32.1		36.5	40.4	
Differentiation				0.26			0.09
Well	737	30.7	21.7		33.6	33.2	
Moderate	187	14.4	9.3		15.5	17.8	
Poor	35	17.1	2.6		17.1	11.6	
Mucinous (colloid) adenocarcinoma				0.087			0.27
No	766	28.2	19.5		30.9	30.9	
Yes	193	22.3	14.9		23.8	21.8	


Expression of Inflammatory Genes
Interleukin-8
VEGF
CEACAM-1
MMP-7
Expression of Immuno-suppressive Genes
TGF-β
Interleukin-10
B7-H3
CD32b
Expression of Genes Related to the Adaptive Immune Response
CD8α
Granzyme B
Granulysin
P<0.05
Th1
Th2
T-BET
IRF-1
Interferon-γ
GATA-3
Specific Gene/18S RNA (%)
Status
VELIPI
Relapse

How to interpret quantitative data?

How do we evaluate the data ?

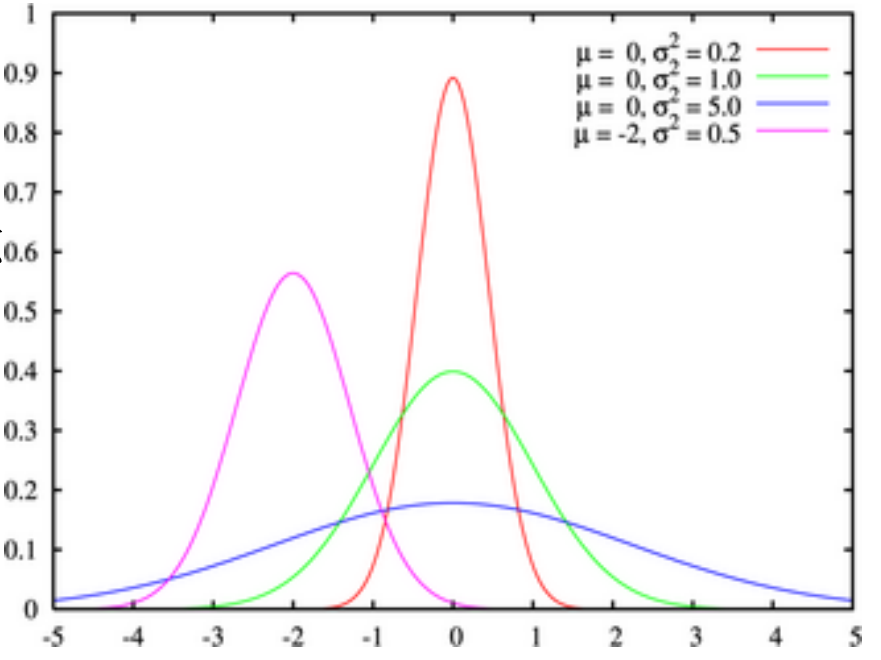
Normal control **52 54** **Cancer Patient**



Normal Distribution

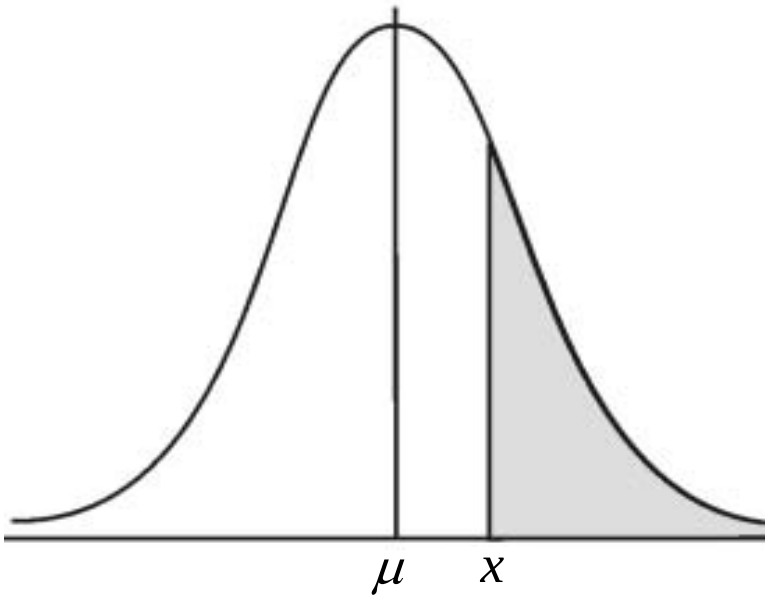
- 常態分佈、高斯分佈 (Gaussian Distribution)
- 隨機變數 x 為連續，樣本數 n 趨近無限大時，機率密度

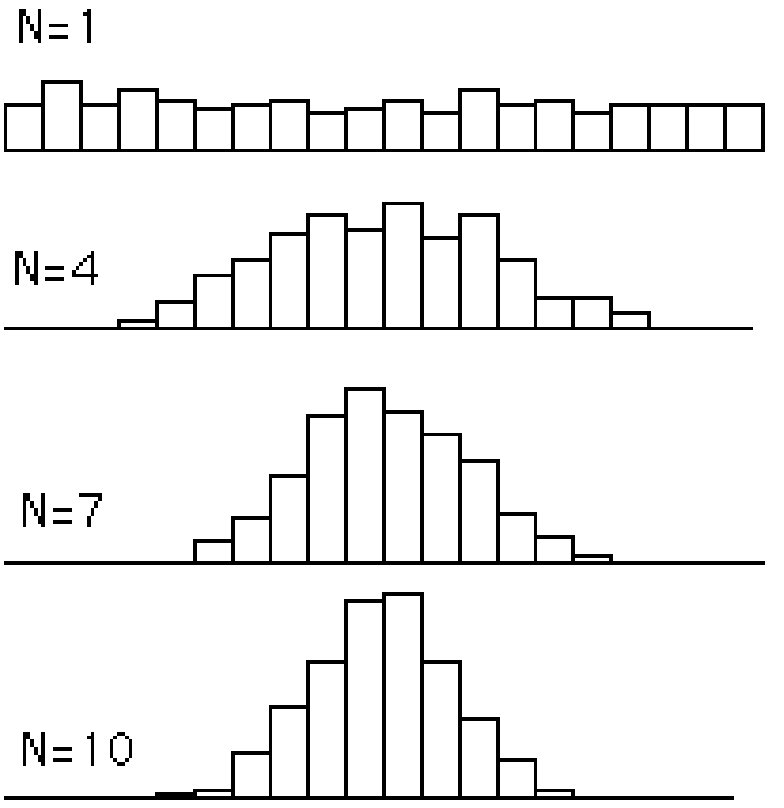
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$



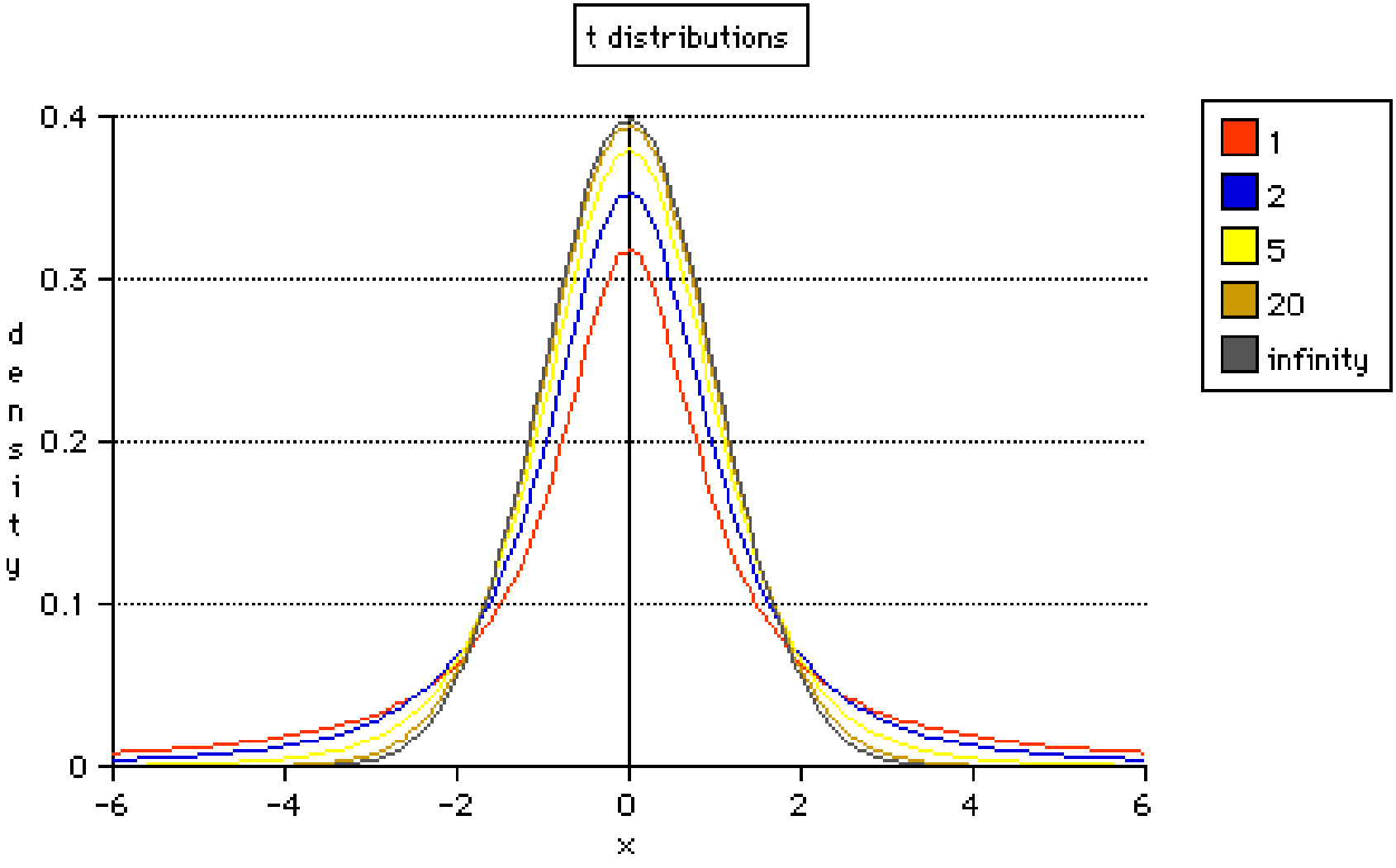
- 當隨機變數不連續、樣本數不足、事件發生機率不固定時不適用

- 樣本平均數抽樣之平均等於母群體平均數 μ
- 樣本平均數抽樣分佈的標準差為 $\sigma / \sqrt{n}$
- 若樣本數 n 夠大，則樣本平均數抽樣分佈接近常態分佈





Student's *t* Distribution

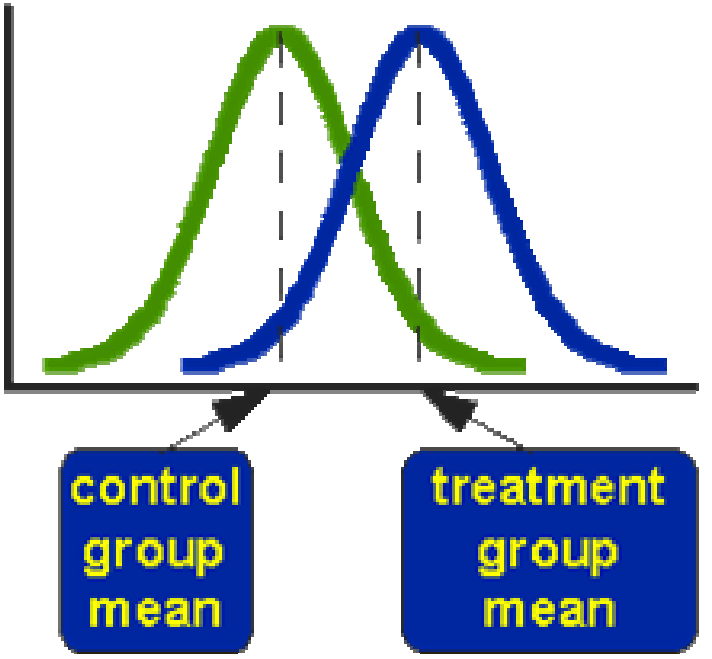


Normal or Patient?

A. Discrimination - Comparision of Data Groups
   1. 2 groups with equal variances
   2. 2 groups with unique variances
B. Receiving Operating Characteristic (ROC) curve
   1. 2 X 2 contingency table
   2. sensitivity & specificity
   3. plotting ROC curve
   4. uses of ROC curve

'Student's' t Test

The *t*-test compares the actual difference between two means in relation to the variation in the data



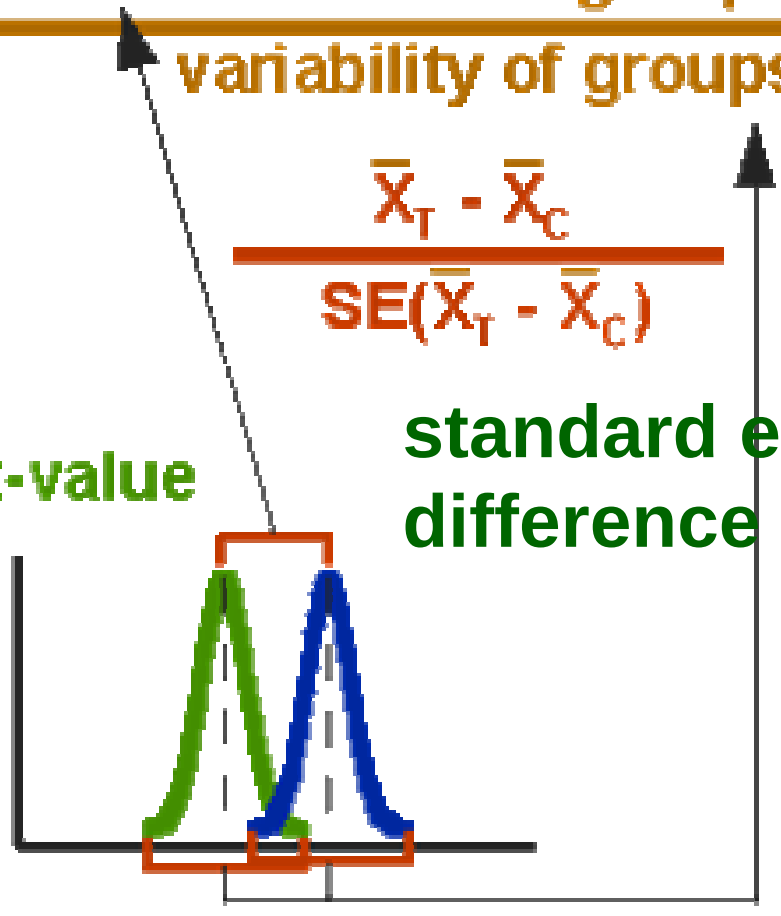
http://www.socialresearchmethods.net/kb/stat_t.htm

$$\frac{\text{signal}}{\text{noise}} = \frac{\text{difference between group means}}{\text{variability of groups}}$$

$$= \frac{\bar{X}_T - \bar{X}_C}{SE(\bar{X}_T - \bar{X}_C)}$$

$$= \text{t-value}$$

standard error of the difference



Student's *t* Distribution

- When *m* , *s* are unknown, Student's *t* distribution wil be applied

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

  - ***s*** standard deviation
  - degree of freedom, d.f. df = *n*-1
  - When df > 30, the distribution is approaching normal distribution

The top part of the ratio is just the difference between the two means or averages. The bottom part is a measure of the variability or dispersion of the scores.

- **One-sample t-test**: Compare the mean with a known standard solution
- **Unpaired t-test**: Comparing two competitive analytical methods
- **Paired t-test**: Comparing two competitive analytical methods

- The comparison tells us whether the "difference" is really difference within the confidence intervals.
  - When $t_{calc} > t_{table}$, the mean in two datasets are statistically different

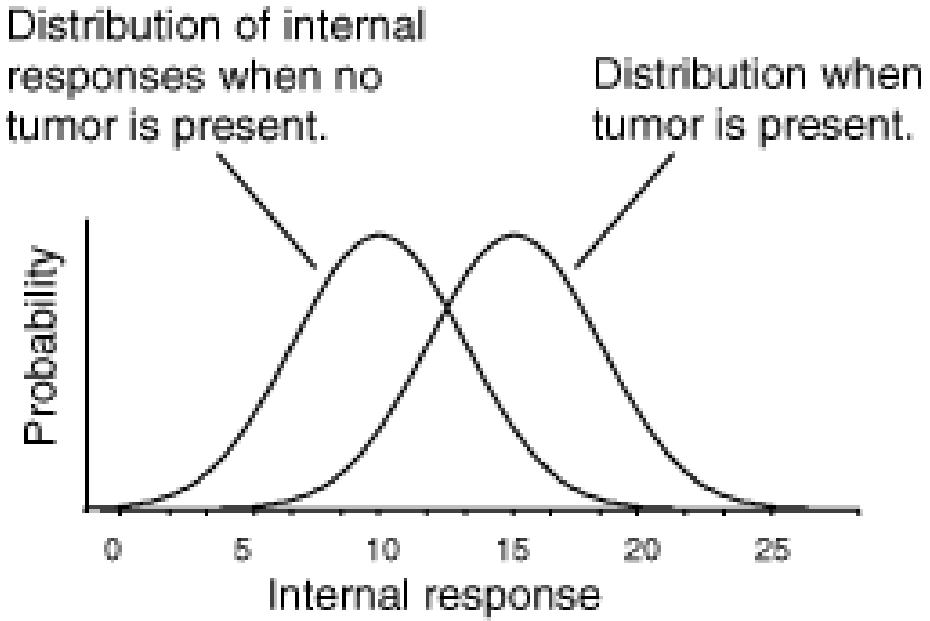
Sensitivity & Specificity

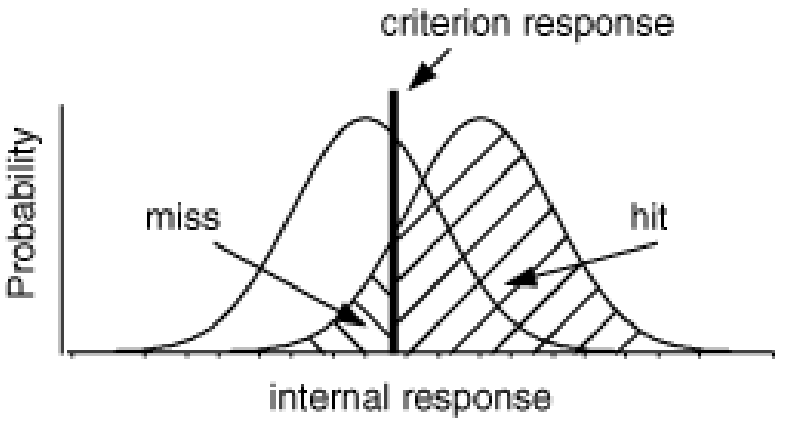
Sensitivity

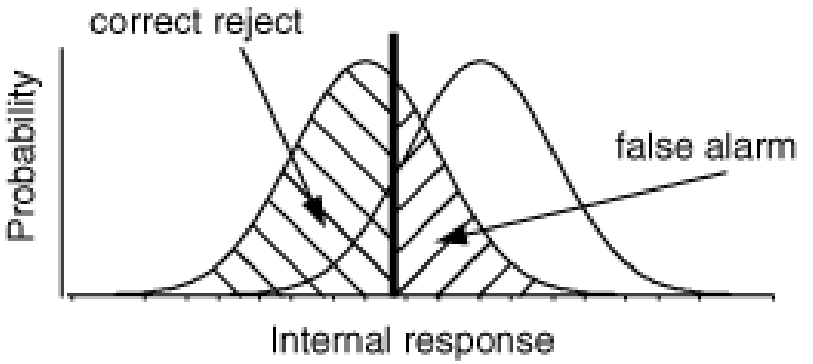
- 敏感度：已知為病患，檢驗結果為患病的機率
  - sensitivity = P(檢驗為患病｜病患) = $d / (b+d)$
  - true positive
  - (1-sensitivity) : false negative

Specificity

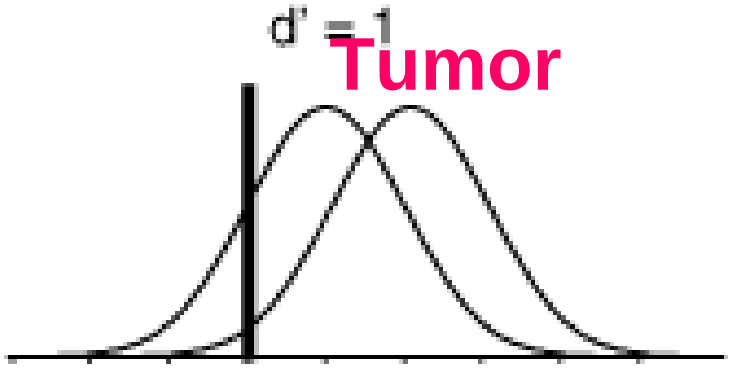
- 特異度：已知未患病，檢驗結果為正常的機率
  - specificity = P(檢驗為正常｜未患病) = $a / (a+c)$
  - true negative
  - (1-specificity) : false true


Distribution of internal responses when no tumor is present.
Distribution when tumor is present.
Probability
0
5
10
15
20
25
Internal response

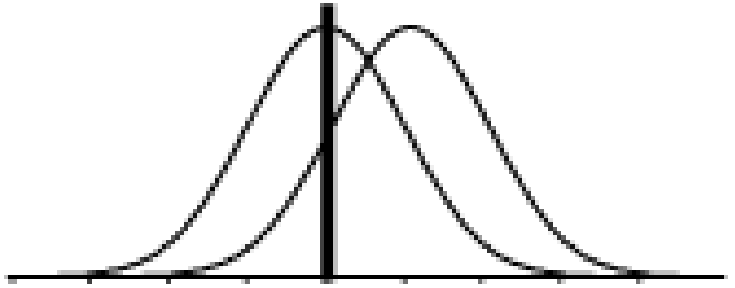

criterion response
Probability
miss
hit
internal response


correct reject
Probability
false alarm
Internal response

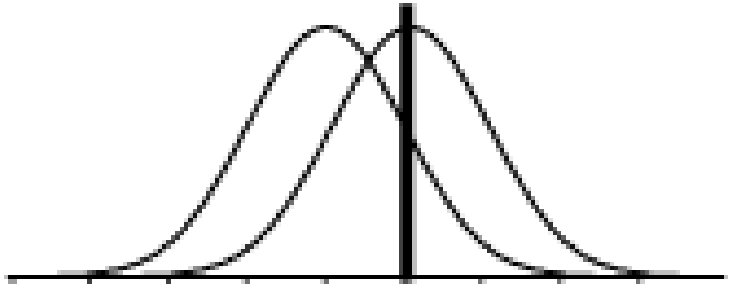
No tumor



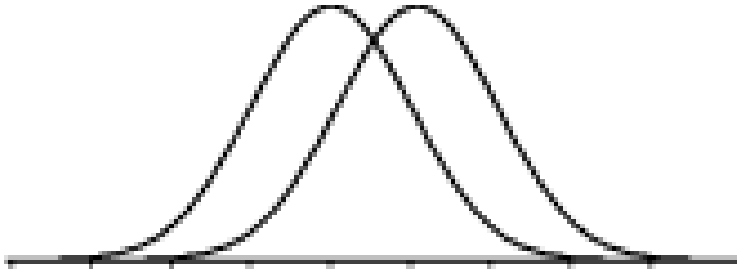
Hits = 97.5%
False alarms = 84%



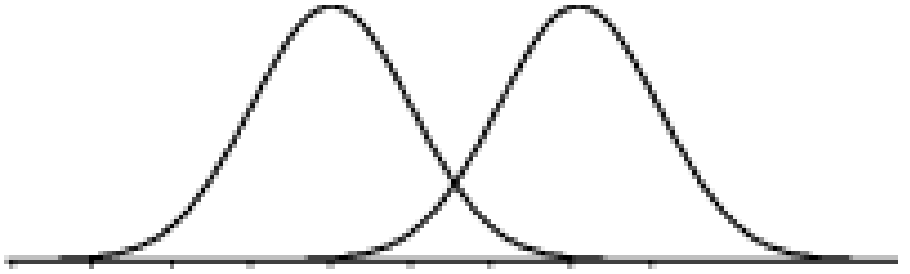
Hits = 84%
False alarms = 50%



Hits = 50%
False alarms = 16%



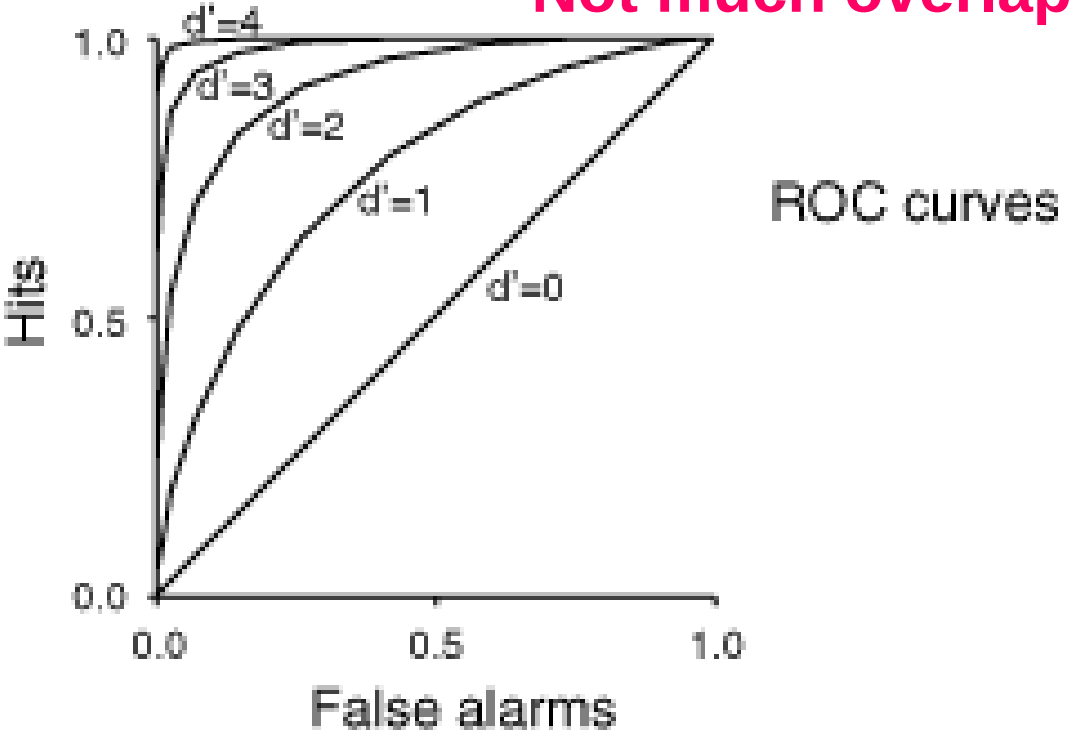
d' = 1 (lots of overlap)



d' = 3 (not much overlap)

High noise,
Lots of overlap

Low noise,
Not much overlap



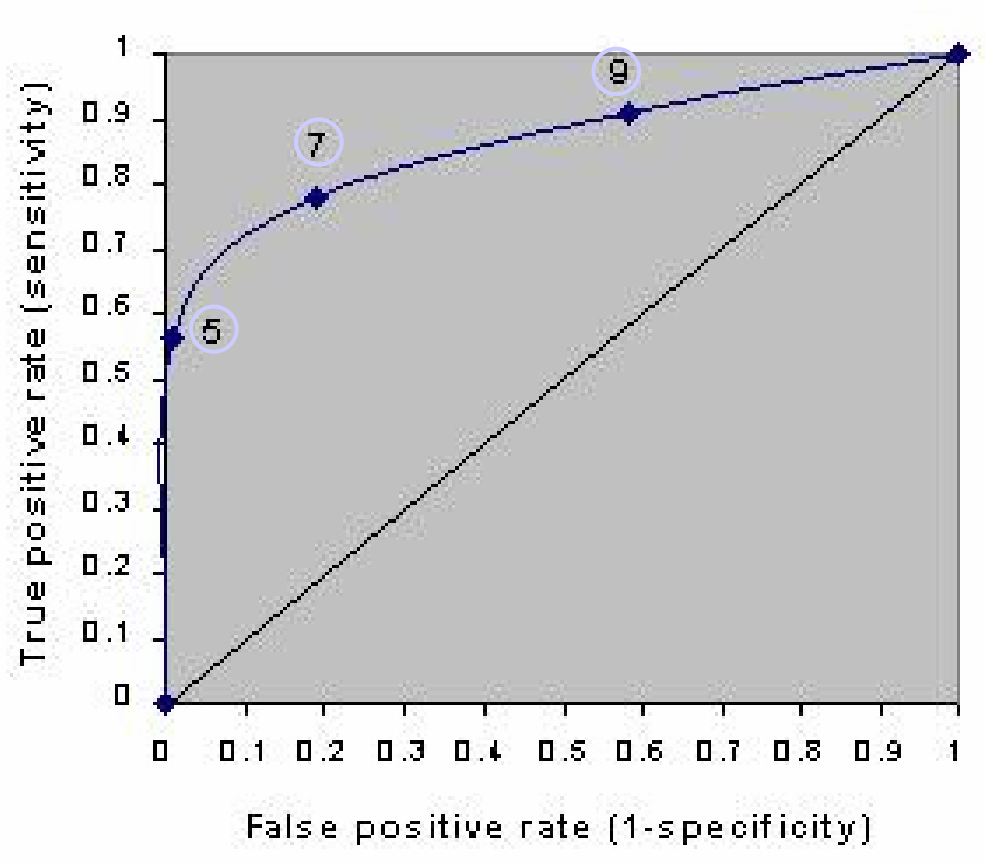
ROC curves

Plotting ROC Curve
Receiver Operating Characteristics Curve

- 縱軸：敏感度 (true positive) ← 患病且被正確判斷
- 橫軸：偽陽性率 (false positive) ← 未患病但被誤判為患病

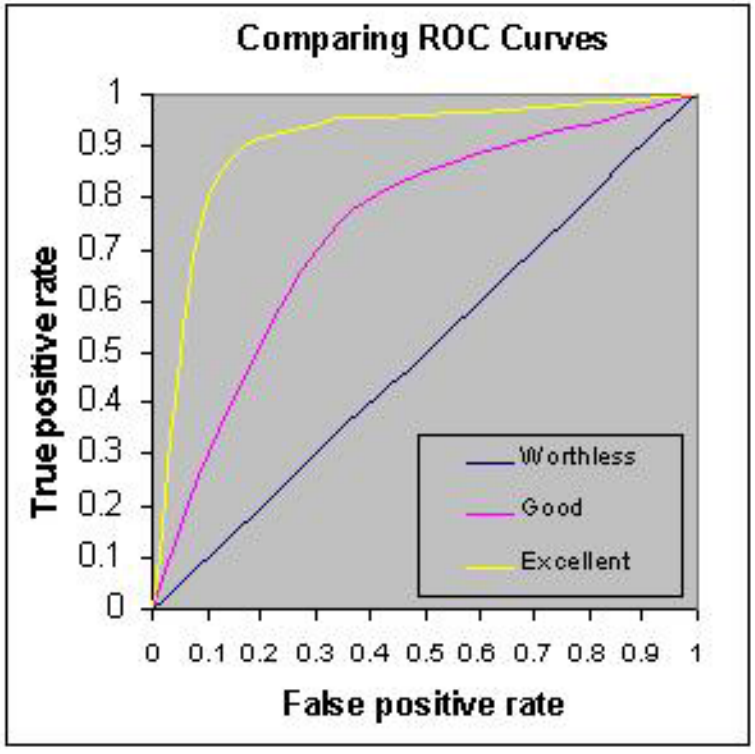
Cutpoint	True Positives	False Positives
5	0.56	0.01
7	0.78	0.19
9	0.91	0.58

不同判定標準

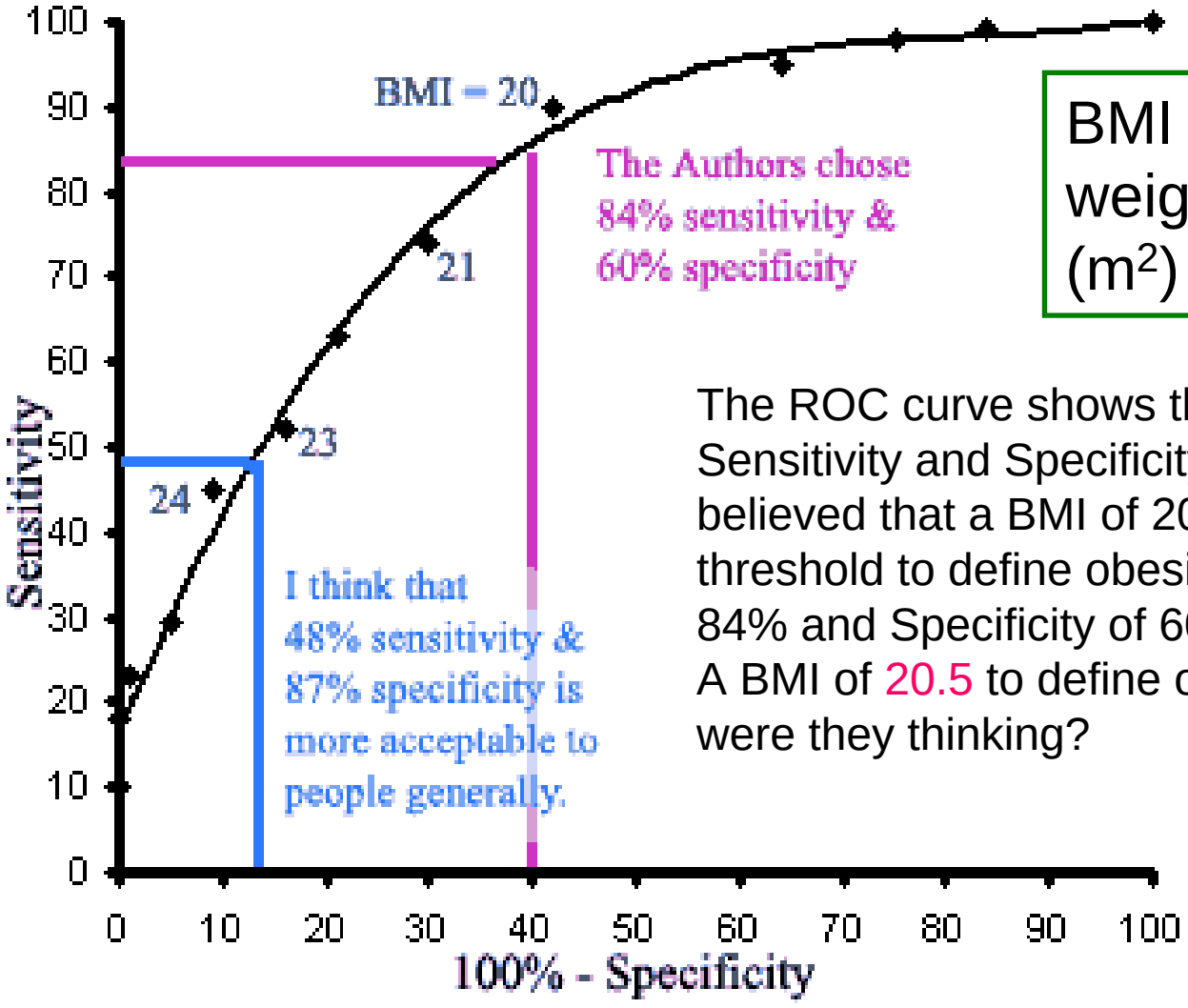


Uses of ROC curve

- Area under the curve：
  - 0.9 ~ 1.0: excellent
  - 0.8 ~ 0.9: good
  - 0.7 ~ 0.8: fair
  - 0.6 ~ 0.7: poor
  - 0.5 ~ 0.6: worthless



J Clin Epidemiol, Jul 1997;50(7):837-43



BMI :
weight (Kg)/Height (m^2)

The ROC curve shows the trade-offs between Sensitivity and Specificity. This article's Authors believed that a BMI of 20.5 was the optimum threshold to define obesity, with a Sensitivity of 84% and Specificity of 60%. Can you believe it? A BMI of 20.5 to define obesity **(肥胖)** ? What were they thinking?

Assessment of the Performance of a Method

(BMB 1.6.2)

Summary Statistics

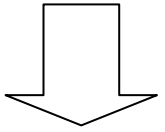
- **Measures of Central Tendency**
  - Mean, Median, Mode
- **Spread**
  - Range
  - Variance
  - Standard deviation
  - Stander error
- **Shape**

Precision v.s. Accuracy

Precision: (Imprecision, Variability, Reproducibility)

Reproducibility of analytical measurement

= closeness of replicate data

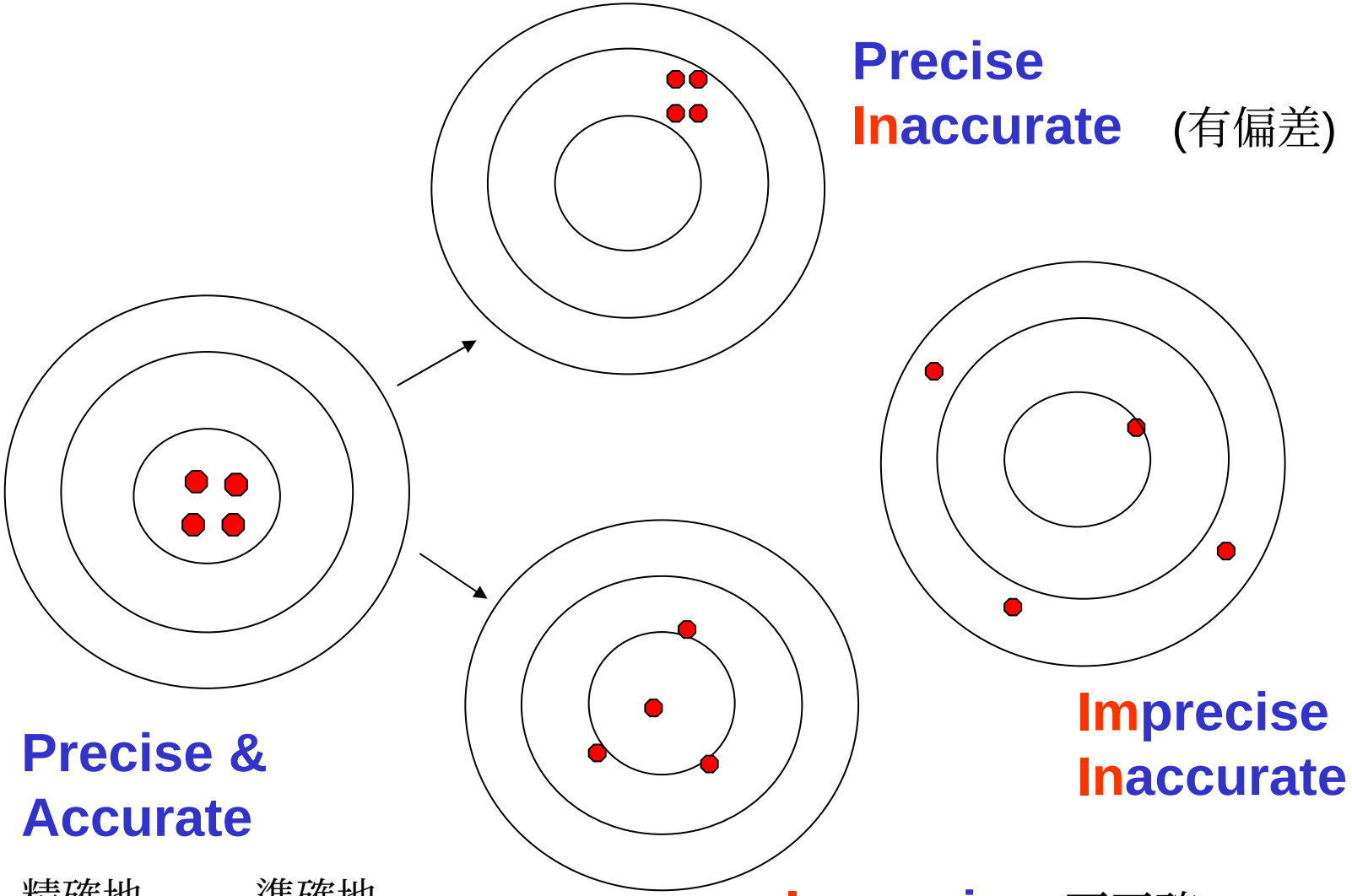


Within Batch Precision

Between Batch Precision

Accuracy (Bias, Inaccuracy)

Differences between “mean” and “true”value



Precise
Inaccurate (有偏差)

Precise & Accurate

精確地,
絲毫不差地

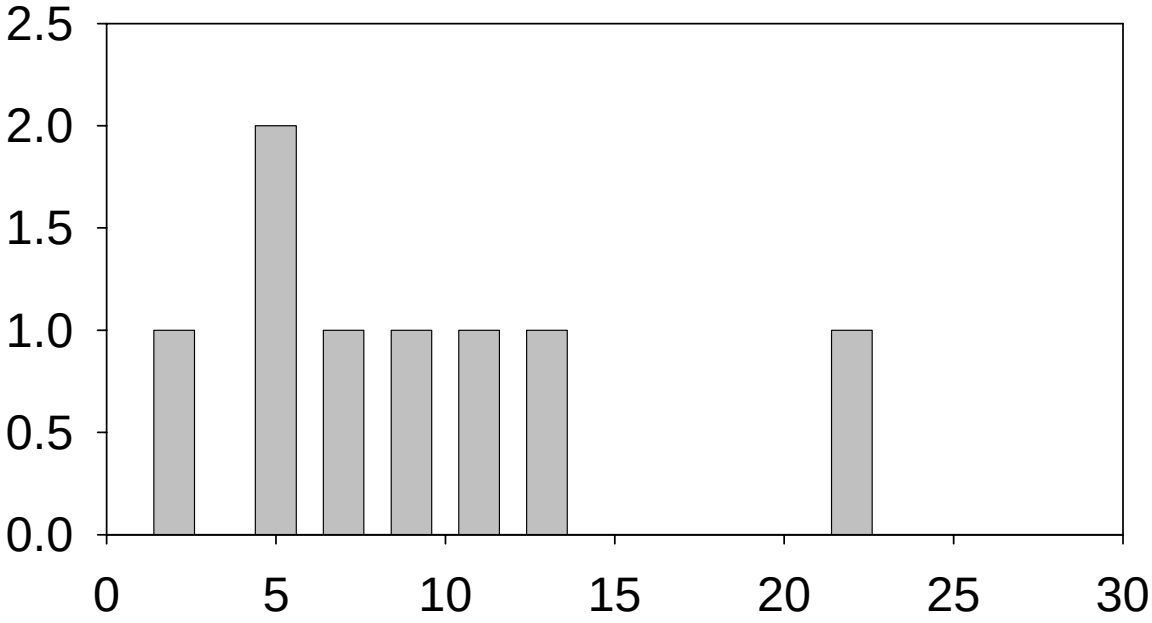
準確地,
正確無誤地

Imprecise Inaccurate

Imprecise (不正確)
Accurate

Measures of Central Tendency

- **Mode**
- **Median**
- **Mean**



e.g. 2, 5, 5, 7, 9, 11, 13, 22

mode = 5 (greatest frequency)
median= (7+9)/2=8 (
mean=(2+5+5+7+9+11+13+22)/8

Median=50%
Odd：中間數值
Even：中間兩數之平均

Spread -----Variance

- Variance (變異數)：

$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

- standard deviation (標準差)= gives the dispersion of numerical data around the mean value ：

$$s = \left(\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{1/2}$$

N-1: degree of freedom

= [Number of observation — 1]

Q: Why do we divide by (n-1) and not by (n)?

- Use of n as a divisor will give a sample standard deviation which tends to underestimate the population standard deviation, whereas the use of **(n-1)** gives what is known as an"unbiased estimator"
- Score deviates less from their own mean than from any other number. So, the calculation subtracting each score from the sample mean will be smaller than subtracting form the population mean------ ***underestimate*** the SD

(n-1)

Statistics for Analytical Chemists, by R. Caulcutt and R. Boddy

Spread -----Coefficient of Variance

Cofficient of variation (變異係數)：
Relative standard deviation

$$CV = \frac{s}{\bar{x}} \times 100\%$$

e.g. A: 2.00± 0.10 mM, CV=5.0%

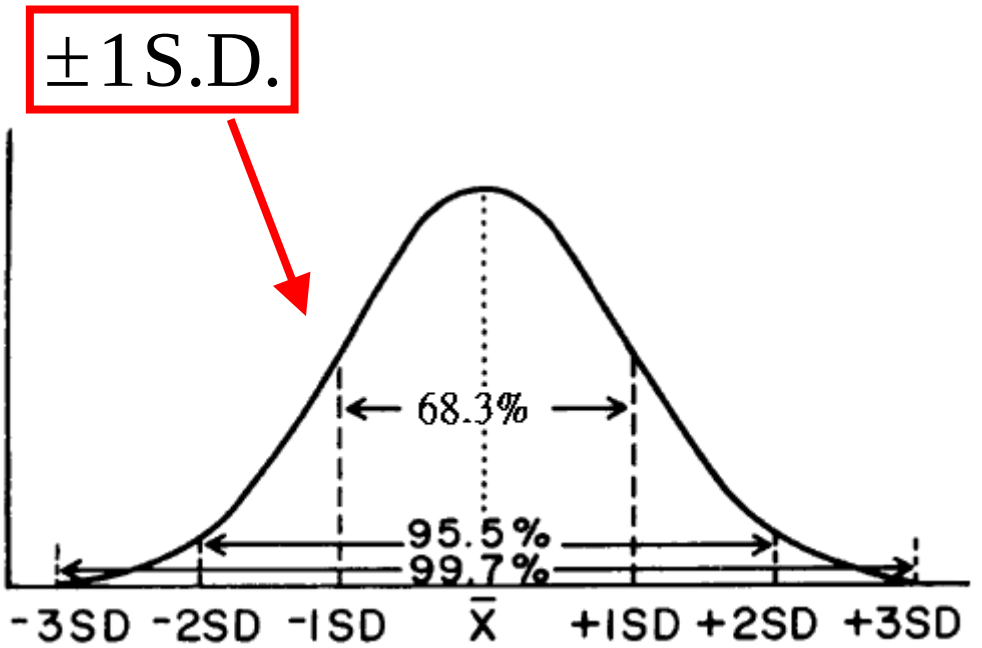
B: 8.00± 0.41 mM, CV=5.0%

Spread -----Coefficient of Variance

Define the spread or distribution of the data

有68.3%的data會落入 $\bar{X} \pm 1S.D.$ 的範圍內,

或看成單次實驗的data在 $\bar{X} \pm 1S.D.$ 的機率為68.3%.



Gaussian Distribution/Normal Distribution

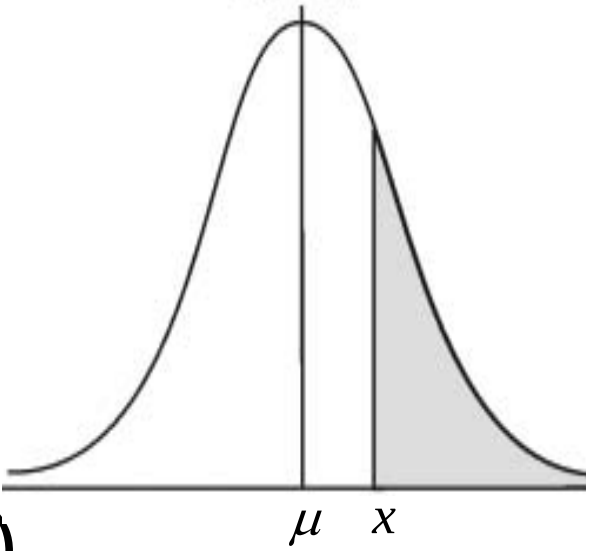
Spread -----Coefficient of Variance

- Possibility of occurrence

80% Rain

98% Sunny/Cloudy/Rainy

- *p*-value



P=0.05 (=95% confidence)

-----***Statistically significant***

Example 5

ASSESSMENT OF THE PRECISION OF AN ANALYTICAL DATA SET

Five measurements of the fasting serum glucose concentration were made on the same sample taken from a diabetic patient. The values obtained were 2.3, 2.5, 2.2, 2.6 and 2.5 mM. Calculate the precision of the data set.

Answer Precision is normally expressed either as one standard deviation of the mean or as the coefficient of variation of the mean. These statistical parameters therefore need to be calculated.

Mean

$$\bar{x} = \frac{2.2 + 2.3 + 2.5 + 2.5 + 2.6}{5} = 2.42 \text{ mM}$$

Standard deviation

Using both equations 1.12 and 1.13 to calculate the value of s:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i^2
2.2	−0.22	0.0484	4.84
2.3	−0.12	0.0144	5.29
2.5	+0.08	0.0064	6.25
2.5	+0.08	0.0064	6.25
2.6	+0.18	0.0324	6.75
Σx_i 12.1	Σ0.00	Σ0.1080	Σ29.39

Using equation 1.12

$$s = \sqrt{(0.108/4)} = 0.164 \text{ mM}$$

Using equation 1.13

$$s = \sqrt{\frac{29.39 - (12.1)^2/5}{4}} = \sqrt{\frac{29.39 - 29.28}{4}} = 0.166 \text{ mM}$$

Coefficient of variation

Using equation 1.9

$$\text{CV} = \frac{0.165 \times 100\%}{2.42}$$

$$= 6.82\%$$

Accuracy（Bias, Inaccuracy)

Differences between "mean" and "true"value

① 理論上，當**n**為無限大時，**"mean"**應趨近於 **"population mean** μ **"**

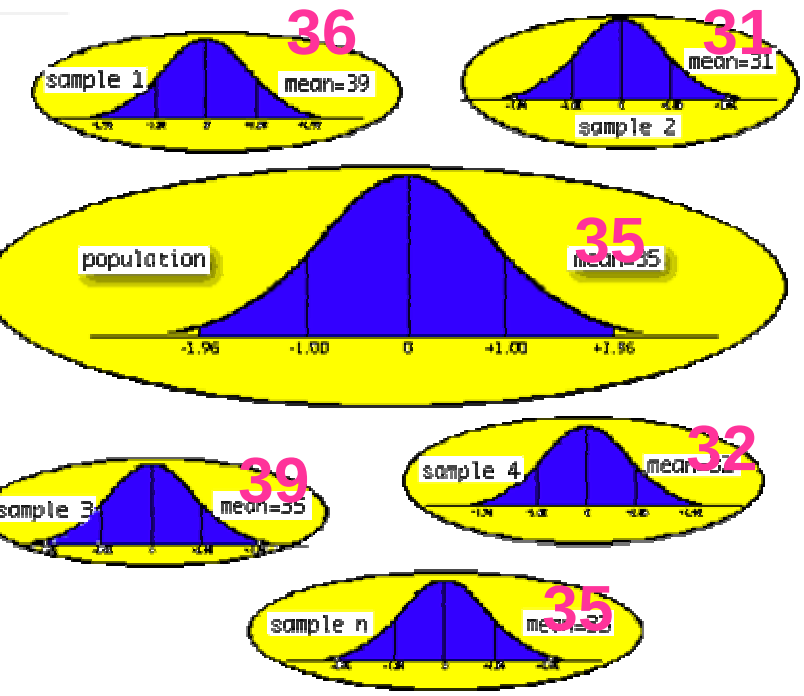
② 希望**uncertainty**趨近於**0,**
此時，**n**必須為無限大
(例：減為**1/2**→**n**為**4**倍**)**

$$s = \left(\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{1/2}$$

但在實驗設計上，**time/money-consuming**，不可行
如何判斷 **mean** 和正確數據的差異呢？

看另一個數值

Standard Error (of the mean)



S.E $\frac{SD}{\sqrt{N}}$ variability of mean

S.D variability of original data 即使知道**S.D**，也無法判斷實驗結果和真正數值之差距！

…..Why does the denominator read $N^{1/2}$ instead of just N? Because we are really dividing the variance, which is SD^2, by N, but we end up again with squared units, so we take the square root of everything…….

Can J Psychiatry, Vol 41, October 1996

Spread--Confidence Interval

- for normal distribution

$$P(-1.96 \le z \le 1.96) = 0.95 \text{, and } z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$\Rightarrow P(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$$

– 若由母群體隨機選取一百組樣本數為 n 的樣本，並使用這些樣本計算 100 組不同的信賴區間，約有 95 個區間會涵蓋真實的母體平均值 μ

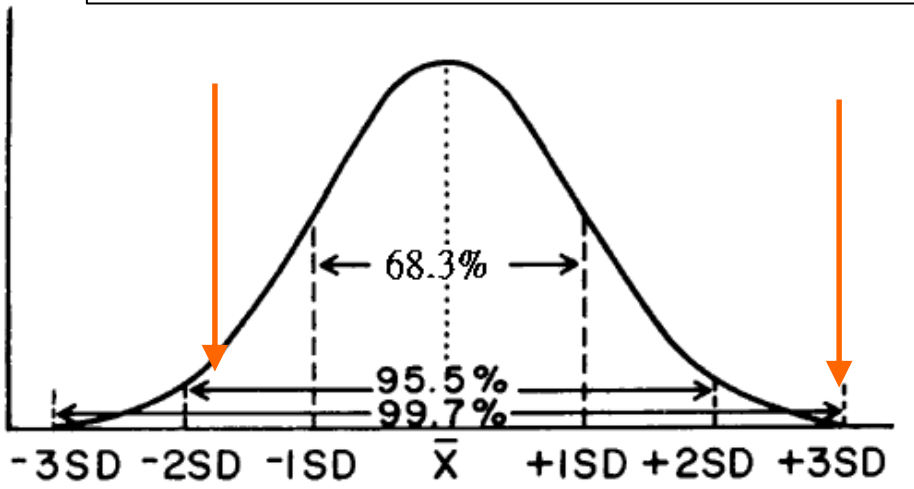
Spread---Confidence Interval

Gives a range of values about the sample mean within a given probability

$$\left[\bar{X} - (t)\cdot\frac{SD}{\sqrt{n}}\right] \le M \le \left[\bar{X} + (t)\cdot\frac{SD}{\sqrt{n}}\right]$$

Confidence Limit

t：student‘s factor (Table 1.9)



xample 6

ASSESSMENT OF THE ACCURACY OF AN ANALYTICAL DATA SET

Calculate the confidence intervals at the 50%, 95% and 99% confidence levels of the fasting serum glucose concentrations given in Example 5.

$$\left[\overline{X} - (t)\cdot\frac{SD}{\sqrt{n}}\right] \leq M \leq \left[\overline{X} + (t)\cdot\frac{SD}{\sqrt{n}}\right]$$

$$\text{Confidence interval} = 2.42 \pm \frac{(0.741)(0.16)}{\sqrt{5}}$$

$$= 2.42 \pm 0.05\ \text{mM}$$

For the 95% confidence level and the same number of degrees of freedom, $t = 2.776$, hence the confidence interval for the population mean is given by:

$$\text{Confidence interval} = 2.42 \pm \frac{(2.776)(0.16)}{\sqrt{5}}$$

$$= 2.42 \pm 0.20\ \text{mM}$$

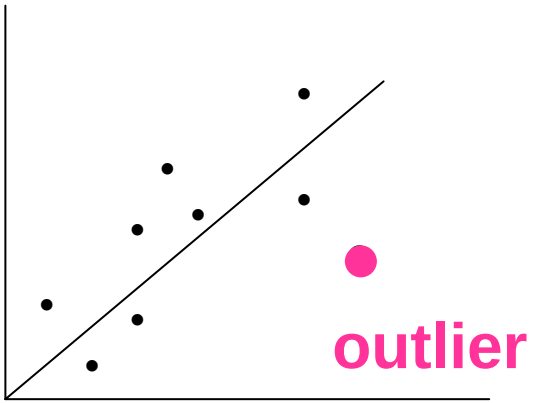
For the 99% confidence level and the same number of degrees of freedom, $t = 4.604$; hence the confidence interval for the population mean is given by:

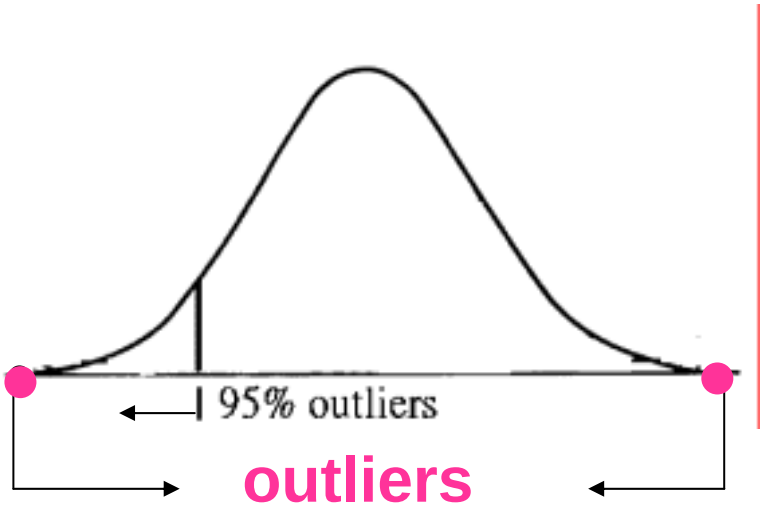
$$\text{Confidence interval} = 2.42 \pm \frac{(4.604)(0.16)}{\sqrt{5}}$$

$$= 2.42 \pm 0.33\ \text{mM}$$

Outlier

Rejection of outlier experimental data





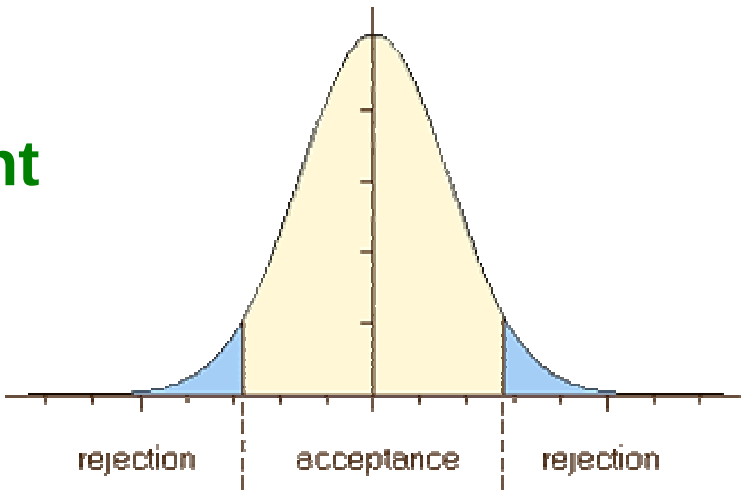
Q_{exp} (Dixon's Q-test)

Experimental rejection quotient

最接近的點

$$Q_{exp} = \frac{X_n - X_{n-1}}{X_n - X_1} = \frac{\text{gap}}{\text{range}}$$

相距最大的點



Outlier– Q values

Table.1.1	Values of Q for the rejection of outliers	
Number of observations		Q (95% conflidence)
4		0.83
5		0.72
6		0.62
7		0.57
8		0.52

$Q_{exp} < Q_{\text{Table 1.10}}$ ------ Accept the datapoint

$Q_{exp} > Q_{\text{Table 1.10}}$ ------ Reject the datapoint

xample 7

IDENTIFICATION OF AN OUTLIER EXPERIMENTAL RESULT

If the data set in Example 6 contained an addition value of 3.0 mM, could this value be regarded as an outlier point at the 95% confidence level?

From equation 1.16

$$Q_{exp} = \frac{3.0 - 2.6}{3.0 - 2.2} = \frac{0.4}{0.8} = 0.5$$

Using Table 1.11 for six data points, $Q_{table} = 0.62$.

Since Q_{exp} is smaller than Q_{table} the point should not be rejected as there is a more than 5% chance that it is part of the same data set as the other five values. It is easy to show that an additional data point of 3.3 rather than 3.0 mM would give a Q_{exp} of 0.64 and could be rejected.

- **One-sample t-test**：用以檢驗一組數據之平均值是否與某特定數值相等
- **Unpaired t-test**：用以檢驗兩組「獨立樣本」之數據平均值是否相等
- **Paired t-test**：用以檢驗兩組「配對樣本」之數據平均值是否相等

• 此處所指「相等」，非僅平均數本身，而是指數據是否位於包含由樣本本身分佈與各類型實驗誤差、自由度所形成之信賴區間（confidence intervals）內

  – 當$t_{calc} > t_{table}$，則兩組數據不相等