Kód předmětu: Bi8980

MASARYKOVA UNIVERZITA

Protein expression and purification

II. Calculation in the molecular biosciences

Lubomír Janda, Blanka Pekárová and Radka Dopitová

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

2.1. The key mathematical tools

2.1.1. Estimation of the results of calculations

Relating the number of amino acids in the polypeptide chain to molecular mass.

1 Dalton (Da) is equal to 1/12 the mass of the ^{12}C isotope of carbon (1.66×10^{-24} g)

1 amino-acid contributes about 110 (112) Da to the protein's mass.

A: A protein 260 aa long = ? kDa

B: A protein of 40 kDa = ? amino acid ? nt

Strategy A: (260 x 100) + (260 x 10) = 28,600 Da + (2 x 260) = **29,120 Da**

Strategy B: 40,000/100 = 400 aa – 10% (40 aa) = **360 aa**

360 aa x 3 = 900 + 180 = **1,080 nt**

The molar concentration of a solution.

Estimate the molarity of a 3.5 mg/ml solution of BSA, whose molecular mass is 66,000 Da (66 kDa).

Strategy: $3.5/66{,}000 = 3.5 \times 10^{-5}/0.66 \sim 5 \times 10^{-5} = 0.00005$ M = 0.05 mM = **50 µM** (control 5 x 0.66 = 3.3)

2.1. The key mathematical tools

2.1.2. Significant figures

•If the molecular mass of a protein is quoted as 30 kDa, this represents only one significant figure (mass is between 25 and 35 kDa)

- •MS might give an answer of 34.503 kDa (5 significant figures).
- •If mobility on protein electrophoresis compared with standard proteins of known molecular mass gives us no more than two significant figures, then molecular mass also must be quoted to two significant figures, i.e. 35 kDa.

A calculator gives the result of a calculation as 4,623.708 Da. Express this result to 1, 2, 3, and 4 significant figures

Answer: The results are

1. 5,000 Da
2. 4,600 Da
3. 4,620 Da
4. 4,624 Da

2.1. The key mathematical tools

2.1.3. Logarithms

The logarithm of a number ***n*** is the power to which the reference base number (usually 10) must be raised to give ***n***.

Thus, $10^2 = 100$, so log 100 = 2; similarly log 100 000 = log 10^5= 5

$10^a \times 10^b = 10^{a+b}$, so log (a x b) = log a + log b

$10^a/10^b = 10^{a-b}$, so log (a/b) = log a – log b

$a^2 = a \times a$, so log (a^2) = log a + log a = 2log a; in general log (a^n) = n log a

An alternative reference base number for logarithms is the Euler number (***e***, equal to 2.71828…)
Logarithms to base *e* are known as natural logarithms and generally denoted by ***ln.*** ***ln*** 10 = 2.303, so in general ***ln*** x= 2.303 ***log*** x

<u>Key properties of logarithms</u>
- The log of 1 = 0 (this is because $10^0 = 1$).
- The log of number between 0 and 1 is negative (pH).
- The log of a number greater than 1 is positive.
- Negative numbers do not have logarithms.

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.1. Acid-base behaviour and the pH scale

Acidity is quantitatively defined by the concentration of protons (H^+ ions) present in a solution.

Stomach	0.03 M		(30 mM)
Duodenum	0.00000001 M	(1 x 10^{-8} M)	(10 nM)
Lysosome	0.00003 M	(3 x 10^{-5} M)	(30 μM)

pH is defined by the equation: pH = - log [H^+]

Stomach (0.03 M)	pH = -log (0.03) = -(-1.52) = **1.52**
Duodenum (0.00000001 M)	pH = -log (1 x 10^{-8} M) = -(-8) = **8**
Lysosome (0.00003 M)	pH = -log (3 x 10^{-5} M) = -(-4.52) = **4.52**

Henderson-Hasselbalch equation **$pH = pK_a + \log([A^-]/[HA]$ (p91)**

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.2. Variation of reaction rates with temperature

The rates of reactions increase dramatically with temperature as a greater proportion of the reactants possess the energy necessary to surmount the activation energy barrier for reaction to occur.

$$k = Ae^{-E_a/RT}$$

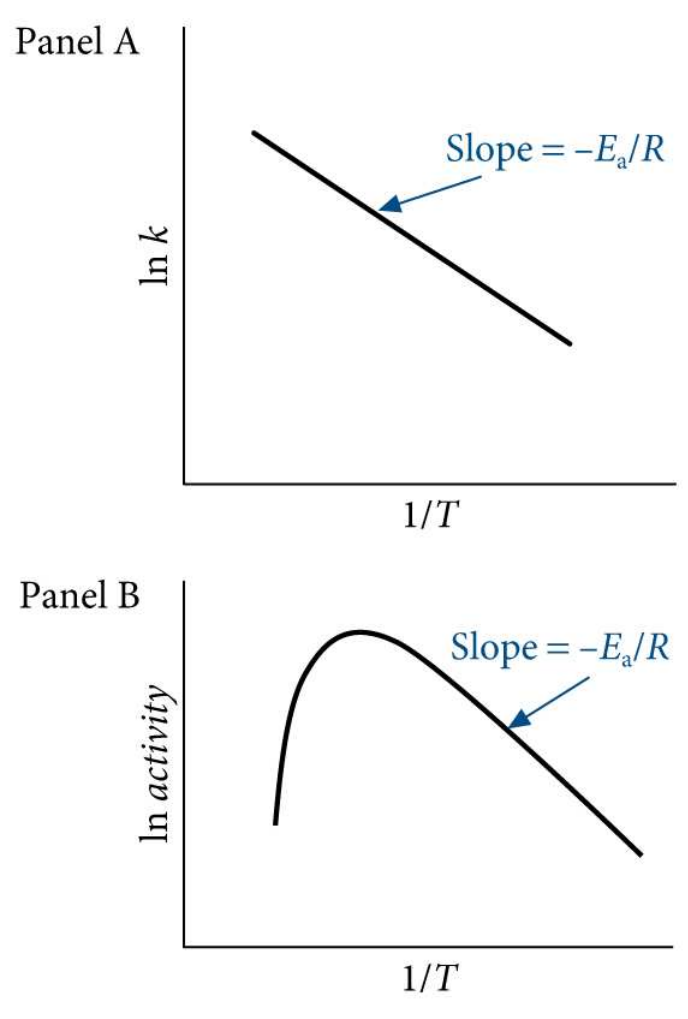
k	rate constant for the reaction
A	pre-exponential factor (related to the frequency of successful collisions between reacting molecules)
E_a	activation energy for the reaction
R	gas constant (8.31 J/K mol)
T	temperature in degrees Kelvin

$$\ln k = \ln A - (E_a/RT)$$

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.2. Variation of reaction rates with temperature



Arhenius plots show the effect of temperature on reaction rate.

$$\ln k = \ln A - (E_a/RT)$$

Typical plot for an enzyme-catalyzed reaction.

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.2. Variation of reaction rates with temperature

Table: The variation of lactate dehydrogenase activity with temperature

Temperature (°C)	5	15	25	35	45	55	65
Activity (µmol/min.mg)	52.5	108	212	401	488	270	40.3

Explain the dependence of activity on temperature and determine the activation energy for the enzyme-catalyzed reaction.

The data can be plotted according to the Arrhenius equation to give a straight line

$$\ln k = \ln A - (E_a/RT)$$

T - temperature K = °C + 273

Slope of the graph = -5,810 K = $-E_a/R$

E_a = 5,810 x 8.31 J/mol = 48.3 kJ/mol

At 45°C, the rate is less than predicted on the basis of this straight line.

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.3. First-order processes and bacterial growth

$$[A]_t - [A]_0 e^{-kt}$$

- The decay of radioactive isotopes
- The decrease in the concentration of drugs in the blood
- Bacterial growth

$[A]_t$ concentration at time t
$[A]_0$ concentration at time zero
k is the rate constant for the reaction

Input 100 cells		**generation time 30 min**
after	30 min	200 cells
	60 min	400 cells
	10 h	1.0486×10^8 (2^{20} x 100 cells)
	20 h	1.0995×10^{14} (2^{40} x 100 cells)

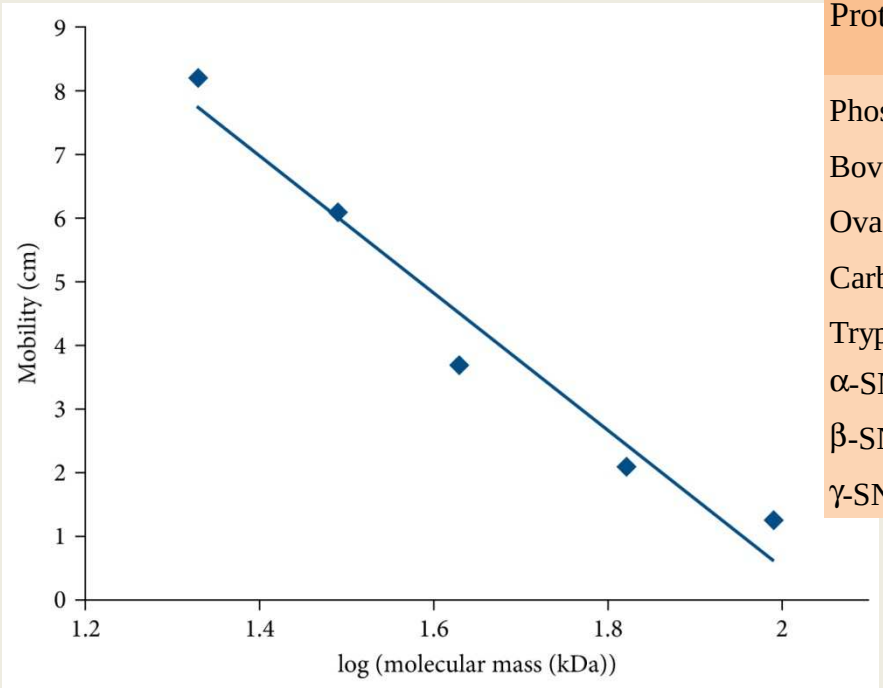
Q: The number of bacterial cells in a culture increases from 1.2×10^5 to 5.8×10^5 over 120 min. What is the generation time for the bacteria under these conditions?
ST: The two data points can be used to calculate the slope of the plot of log against time.
S: The slope of the plot is 0.684/120 = 0.0057 min^{-1}. Thus $0.301/t_{1/2} = 0.0057$
From which $t_{1/2}$ = 0.301/0.0057 min = 52.8 min

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.4. Molecular mass calibration graphs

•Molecular masses of proteins are often estimated by the techniques of gel filtration and SDS-PAGE.
•Two graphs may be constructed: log molecular mass vs. elution volume and log molecular mass vs. mobility.



Protein	Mobility (cm)	Molecular mass (kDa)	Log (molecular mass)
Phosphorylase b	1.25	97.00	1.99
Bovine serum albumin	2.10	66.00	1.82
Ovalbumin	3.70	43.00	1.63
Carbonic anhydrase	6.10	29.00	1.46
Trypsin inhibitor	8.20	21.50	1.33
α-SNAP	5.30	35.00	1.54
β-SNAP	5.20	36.00	1.56
γ-SNAP	4.85	39.00	1.59

Please solve the problem. 5 points

The molecular mass of lysoyzme is 14,300 Da, and its density is 1.4 mg/ml. Avogadro's number is 6.02×10^{23} mol^{-1}. The volume of a sphere of radius **r** is given by $(4\pi/3)r^3$.

Q*ue*stion 1: Assuming that the lysozyme molecule can be regarded as spherical, calculate its radius.

Answer: The mass of 1 mol lysozyme is 14,300 g. Hence the mass of 1 molecule = $14,300/(6.02 \times 10^{23})$ g = 2.375×10^{-20} ml (cm^3). So $(4\pi/3)r^3 = 1.697 \times 10^{-20}$, hence $r3 = 4.05 \times 10^{-21}$, therefore $r = 1.59 \times 10^{-7}$ cm, or 1.59 nm. Thus, the radius of a lysozyme molecule is 1.59 nm.

Please solve the problem. 5 points

The concentration of the protein required for a nuclear magnetic resonance experiment is 0.5 mM.

Q*ue*stion 2: If the molecular mass of the protein is 27.5 kDa, what concentration is required in terms of mg/ml? If the sample volume is 0.4 ml, what mass of protein is required?

Answer: A 1 M solution of protein contains 27,500 mg/ml. Thus a 0.5 mM solution contains $0.5 \times 10^{-3} \times 27{,}500$ mg/ml = 13.75 mg/ml. In 0.4 ml, there would be 13.75 x 0.4 mg, i.e. 5.5 mg.

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.5. Spectrophotometry

Beer–Lambert law $A = \varepsilon \times c \times l$

A is the absorbance at a particular wavelength (without units; merely a ratio):

$$A = \log (I_0/I_t)$$

I_0 is the intensity of the incident light (light striking the cuvette).

I_t is the intenisty of the transmitted light (the light leaving the cuvette).

ε is the absorption coefficient (degree of absorption) (M^{-1} cm^{-1}).

l is the path length of the cuvette (cm).

c is the concentration of the solution (M).

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.5. Spectrophotometry

Q: What percentage of the incident light is transmitted after passage through solutions which have absorbance values of (a) 0.05, (b) 0.3, (c) 1.0, and (d) 2.0

ST: This is an application of the equation: $\mathbf{A = \log (I_0/I_t)}$

10^x of 0.05 = 1.122 10^x of 0.3 = 1.995

10^x of 1 = 10 10^x of 2 = 100

S: The value of I_0/I_t at each value of A is multiplied by 100 to give the percent of the incident light that is transmitted

100/1.122 = 89.1% 100/1.995 = 50.1%

100/10 = 10% 100/100 = 1%

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.5. Spectrophotometry

Beer–Lambert law $A = \varepsilon \times c \times l$

A is the absorbance at a particular wavelength (without units; merely a ratio).
ε is the absorption coefficient (degree of absorption) (M^{-1} cm^{-1}).
l is the path length of the cuvette (cm).
c is the concentration of the solution (M).

- The logarithmic nature of A is not always readily appreciated in a digital display of the absorbance.
- However, it is particularly important to note that you should always aim to measure absorbance values in the **range 0.1–1.0.** Despite what the manufacturers of the instrument may claim, absorbance values significantly above 1.0 are not usually reliable.

Q: The absorption coefficient of NADH at 340 nm is 6,220 $M^{-1}cm^{-1}$. You have made up a 2.5 mM stock solution of NADH. How would you check its concentration spectrophotometrically?

ST: The absorbance to be measured should be in the range for reliable measurements (0.1–1.0).

S: 2.5 mM solution in 1 cm cuvette will have absorbance at 340 nM of

$$\mathbf{6{,}220 \times 2.5 \times 10^{-3} = 15.6}$$

50-fold diluted solution will have absorbance of 15.6/50 = **0.312**.

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.6. Energy changes and equilibrium constants of reactions

The standard free energy change in a reaction (ΔG^0) is related to the equilibrium constant for the reaction (K_{eq}):

$$\Delta G^0 = -RT \ln K_{eq}$$

R is the gas constant (8.31 J K^{-1} mol^{-1})
T is the absolute temperature.
$\mathbf{K_{eq}}$ will change logarithmically with changes in ΔG^0.

$$\Delta G^0 = \Delta H^0 - T\Delta S^0$$

ΔH is the change of enthalpy (change in heat content).
ΔS is the change in entropy (change of disorder of a system).
ΔG is the change in (Gibbs) free energy.
T is the absolute temperature (degrees Kelvin, K).

Standard state of a substance refers to: solution of 1 M of the solute
pressure of 1 atm at the temperature in question

2.1. The key mathematical tools

2.1.3. Logarithms

2.1.3.6. Energy changes and equilibrium constants of reactions

$$\Delta G = \Delta H - T\Delta S$$

Q: For the combustion of glucose ($C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O$), the values of ΔG^0 and ΔH^0 at 25°C are -2,872 and -2,822 kJ/mol, respectively. Calculate the value of ΔS^0 and comment on its sign.

ST: This is an application of above equation using the standard state values of the thermodynamic parameters.

S: We can rearrange the equation to give $\Delta S^0 = (\Delta H^0 - \Delta G^0)/ T$. Putting T = 298 K, we find that ΔS^0 = (-2,822 + 2,872)/298 kJ/K.mol = 0.168 kJ/K.mol, or 168 J/K.mol. The positive value of ΔS^0 reflects the greater number of molecules of products (12) compared with reactants (7). In this reaction, the enthalpy term makes the dominant contribution to the free energy change.

2.1. The key mathematical tools

2.1.4. Reciprocals

The reciprocal of a number is 1 divided by that number. Use the 1/x button on the calculator to calculate reciprocals and explore this function in:

- The Lineweaver-Burk plot of enzyme kinetic data and the subsequent calculation of the parameters K_m and V_{max}.
- Calculating the K_m or K_d from the slope of an Eadie-Hofstee or a Scatchard plot, respectively.
- Interconverting association and disocciation constant for binding processes.
- In the Arrhenius plot where the x-axis of the plot is 1/T.

Check that you have mastered the key concepts at the start of this section by attempting the following questions.

Q1: From a graph, $1/V_{max}$ is found to be 0.0235 min/μM. What is the value of V_{max}?

Q2: From the same graph, $-1/K_m$ is found to be -0.0065 μM^{-1}. What is the value of K_m?

Q3: The value of K_a for a binding process is 4.53×10^4 M^{-1}. What is the value of K_d, given that $K_a=1/K_d$?

S1: V_{max} = 42.6 μM/min **S2**: K_m = 154 μM **S3**: K_d = 22.08 μM

2.1. The key mathematical tools

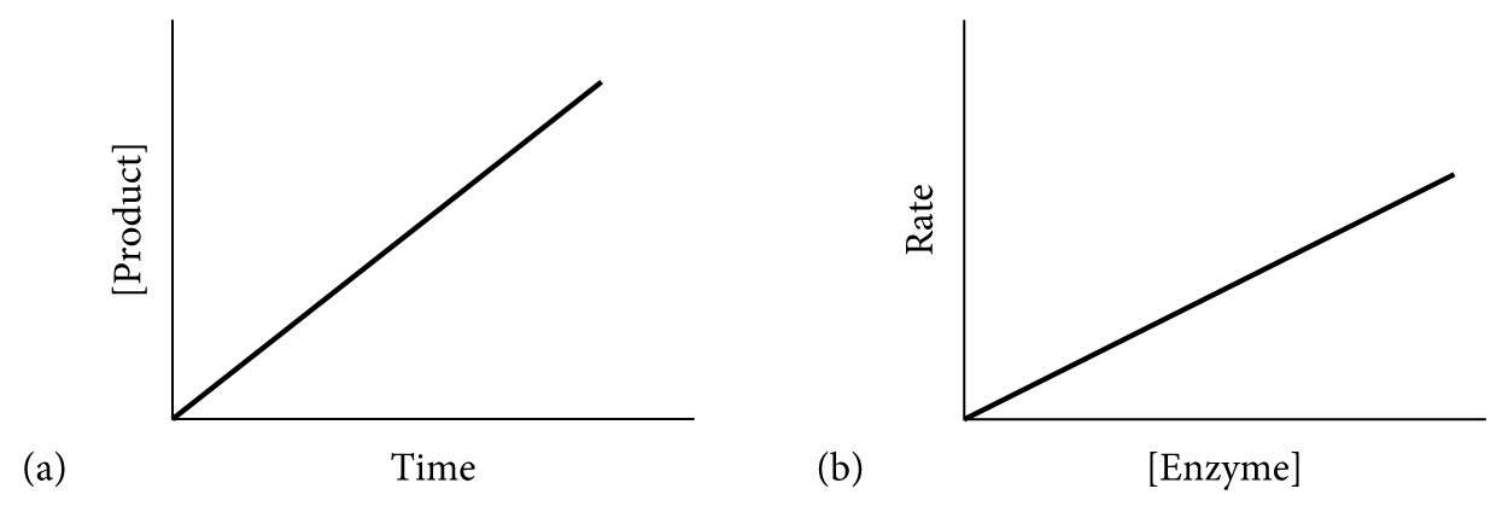
2.1.5. Testing hypotheses

2.1.5.1. Dependent and independent variables

In a graph, the convention is that the x-axis (abscissa) is used to plot the variable that the experimenter varies. This is the independent variable.
The y-axis (ordinate) is used to plot the quantity that is then observed. This is the dependent variable.

$$y = mx + c$$

m is the slope (gradient) of the line.
c is the intercept of the line on the y-axis.



2.1. The key mathematical tools

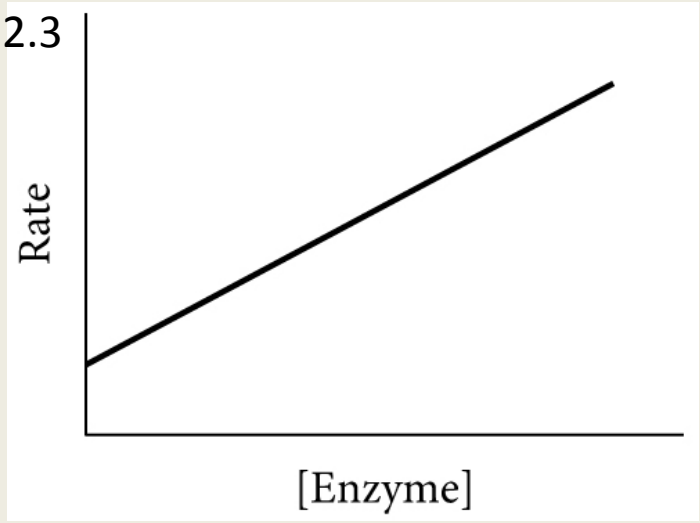
2.1.5. Testing hypotheses

2.1.5.1. Dependent and independent variables

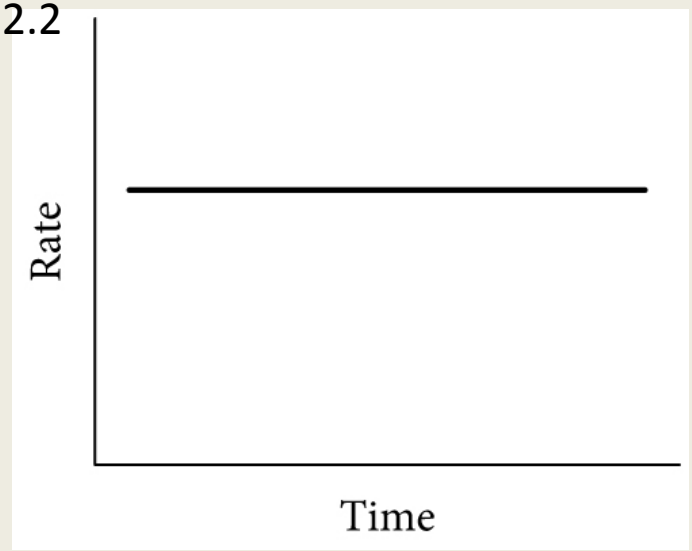
$$y = mx + c$$

m is the slope (gradient) of the line.
c is the intercept of the line on the y-axis.

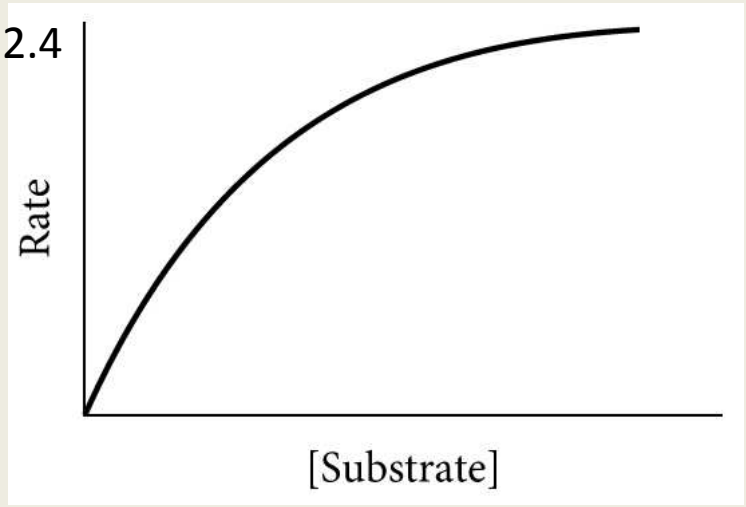
2.3



2.2



2.4



2.1. The key mathematical tools

2.1.5. Testing hypotheses

2.1.5.2. Rearranging equations

Q: Transform equation $v = \frac{V_{max} \cdot [S]}{K_m + [S]}$ so as to give the Hanes-Woolf equation for which the plot is **[S]/v** versus **[S]**.

ST: The approach is to rearrange the equation so as to be able to separate terms in the **y = mx + c** form.

S: By multiplying both sides of equation $v = \frac{V_{max} \cdot [S]}{K_m + [S]}$ **(Km + [S])**, we obtain:

$$vK_m + v[S] = V_{max}[S]$$

Rearranging terms: $V_{max}[S] = v[S] + vK_m$

Dividing each term on both sides of equation by V_{max} $\quad [S] = \frac{v[S]}{V_{max}} + \frac{vK_m}{V_{max}}$

Dividing each term by **v**: $\quad \frac{[S]}{v} = \frac{[S]}{V_{max}} + \frac{K_m}{V_{max}}$ **Hanes-Woolf equation**

Because K_m and V_{max} are constants, this equation is of the form $y = mx + c$, with $y =$ **[S]/v** and **x = [S].** A plot of **[S]/v** versus **[S]** will be a straight line with an intercept on the y-axis of K_m/V_{max} and a slope of $1/V_{max}$. The intercept on the x-axis is $-K_m$.

Please solve the problem. 5 points

The molecular mass of lysozyme is 14.3 kDa.

Question 3: What is the molar absorption coefficient of the protein at 280 nm, given that the A_{280} of a 1 mg/ml solution in a cuvette of 1 cm path length is 2.65?

Answer: A 1 M solution of lysozyme would be 14,300 mg/ml. Using the Beer-Lambert law, the A_{280} of this solution in a cuvette with a 1-cm path length is 2.65 x 14,300 = 37,900. Thus, the absorption coefficient of lysozyme is 37,900 M^{-1} cm^{-1}.

Please solve the problem. 5 points

The Tris buffer system has a pK_a of 8.1. During a reaction, 5 mM H^+ ions are formed.

Question 4: What are the concentrations of the $TrisH^+$ and Tris forms at pH 8.3, if the total concentration of Tris species is 50 mM? What is the new pH after adding 5 mM H^+ ions?

Answer: From the Henderson-Hasselbalch equation, the ratio $[Tris]/[TrisH^+] = 10^{0.2} = 1.58$. Thus, [Tris] = 61.2 mM and $[TrisH^+]$ = 38.8 mM. Upon addition of 5 mM H^+, [Tris] will be reduced by 5 mM to 56.2 mM and $[TrisH^+]$ will be raised by 5 mM to 43.8 mM. Application of the Henderson-Hasselbalch equation shows that the new pH would be 8.21. (Note: If no buffer had been present, 5 mM H^+ would give a pH of 2.3.)

2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.1. Distributions of variables

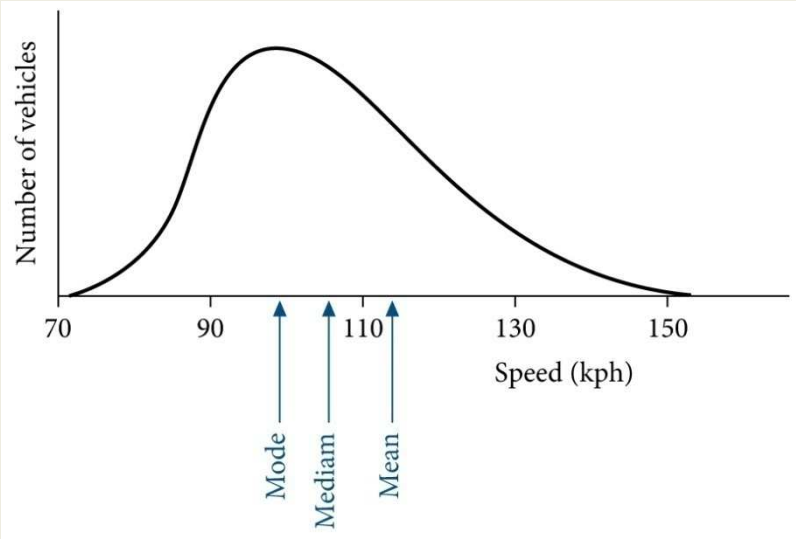
$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots\ldots\ldots x_n}{n} = \frac{\sum(x)}{n}$$

arithmetic mean ($\bar{x}$)

10.2, 10.7, 11.0, 11.1, 11.5, 11.6, 11.9, 12.2 median 11.3

The mode is valuable in describing a distribution of variables which cannot be ranked and distribution which might show two peaks.

mode

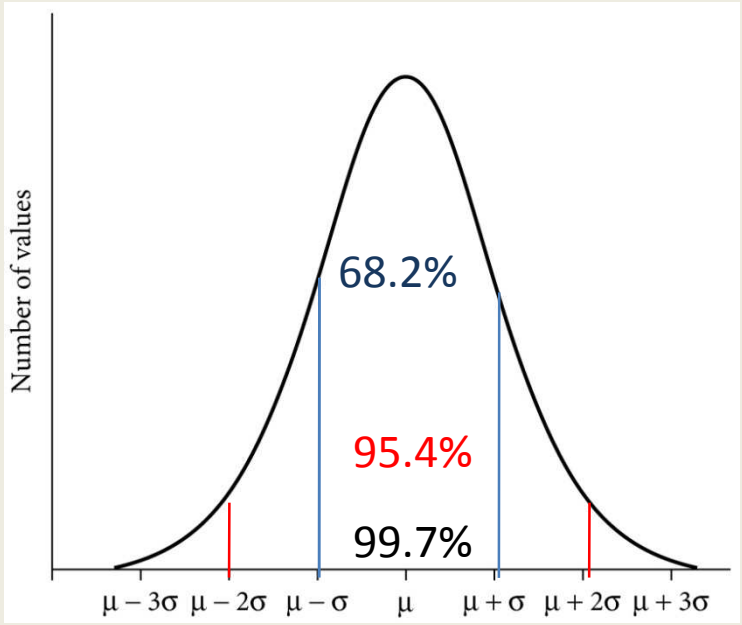


The *mode* is the most frequently occurring measurement in a data set or distribution.

2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.2. The normal distribution



The population mean is defined by:

$$\mu = \frac{x_1 + x_2 + x_3 + \ldots\ldots\ldots x_n}{n} = \frac{\sum(x)}{n}$$

Where, x_1, x_2, …… are the individual values of the property and **n** is the number of values in the population.

$$\sigma\ (SD) = \sqrt{\frac{\sum(x-\mu)^2)}{n}}$$

x is an individual value of the property.
μ is the population mean.
n is the number of values in the population.

$$\sigma\ (SD) = \sqrt{\frac{\sum(x-\bar{x})^2)}{n-1}}$$

Introduction of the term **n-1** rather than **n** into the expression for standard deviation

In terms of experimentally derived values, this would indicate the degree of confidence we had in stating the value of the given parameter.

2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.2. The normal distribution

$$(\mathrm{SEM}) = \frac{\mathrm{SD}}{\sqrt{n}}$$

Clearly, the larger the sample size *n*, the smaller the value of the *SEM*.

Q: The operation of a pipette was checked by repeatedly dispensing and weighing volumes of water. The volume on the pipette was set at 1 ml, and the following volumes (mL) were dispensed in succession: 0.932, 0.927, 0.948, 0.937, 0.918, 0.929, 0.940, and 0.942. What is the mean and standard deviation of these values?

ST: We use equation $\bar{x} = \frac{x_1 + x_2 \cdot + x_3 + \ldots\ldots\ldots x_n}{n} = \frac{\Sigma(x)}{n}$ and $\sigma\,(\mathrm{SD}) = \sqrt{\frac{\Sigma(x-\bar{x})^2)}{n-1}}$

S: The mean value is 0.934 ml and the standard deviation is 0.0096 ml. From the properties of the normal distribution, 99.7% of the values would be within the range 0.905 to 0.963 ml, which is significantly different from the nominal value of 1.000 ml. Thus we can conclude that the pipette is **precise**, but it is not **accurate**. If the experiment had given a mean of 1.002 mL with a standard deviation of 0.0096 mL the pipette would be **precise** and **accurate.**

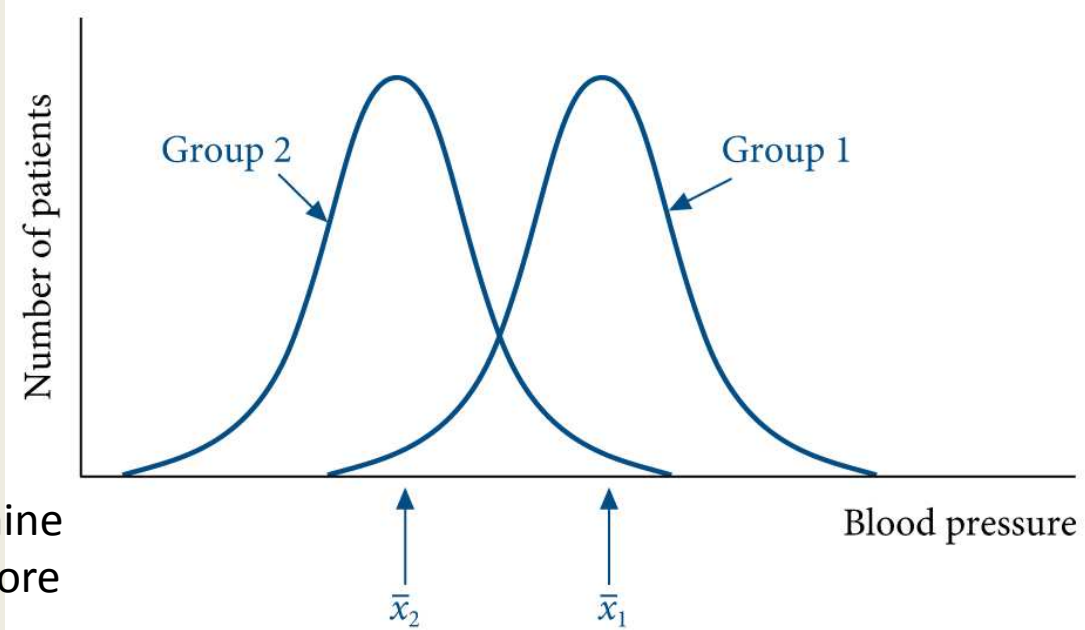
2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.3. Testing the difference between two means

$$(SE_d) = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}$$

$$t = \frac{(x - \mu)\sqrt{n}}{SD}$$



Student's t function – used to determine the significance of the differences (more appropriate for smaller sample sizes).

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{SE_d}$$

|| means 'irrespective of sign'.

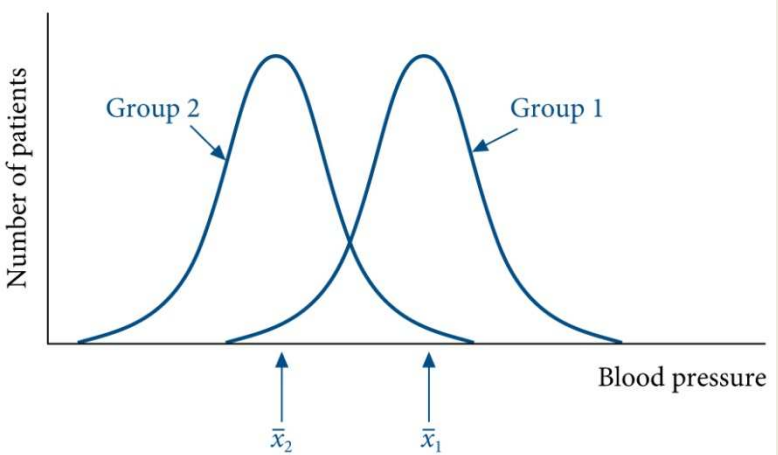
The probability (p) that the two sample means are identical can be deduced from the properties of the t function for the appropriate number of degrees of freedom (equal to $n_1 + n_2 - 2$).

2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.3. Testing the difference between two means

$$(SE_d) = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}} \qquad t = \frac{|\bar{x}_1 - \bar{x}_2|}{SE_d}$$



Q: In a small-scale drug trial, the diastolic blood pressures of the drug and placebo were 122.5 and 110.3 mm, respectively. There were 20 patients in each group. The standard deviations for the two groups were 20.5 and 18.1 mm, respectively. Do the data show (at >95% confidence) that the drug has an effect on the blood pressure?

ST: We calculate the standard error of the difference (equation above), and from that the value of t.

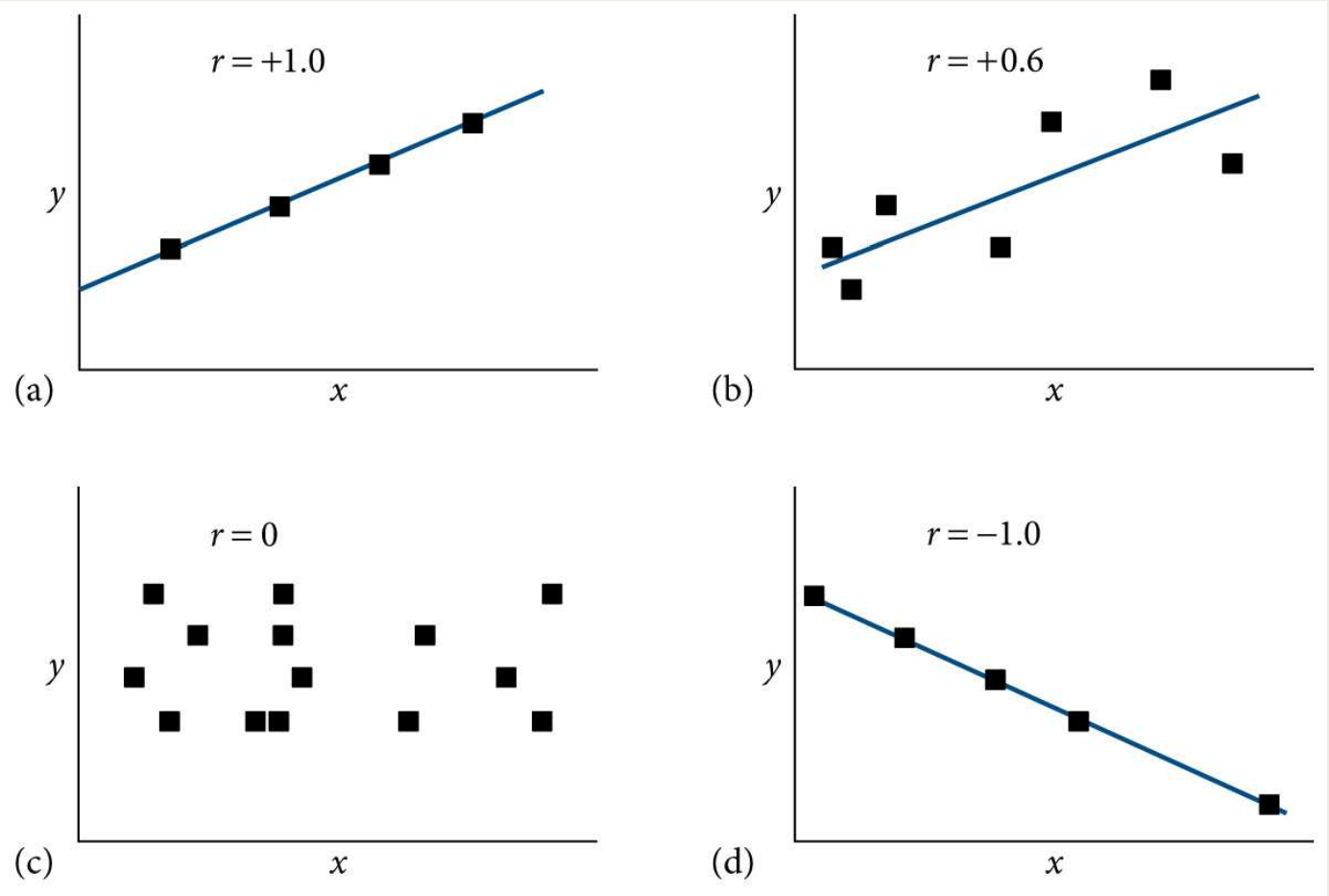
S: The value of SEd = 6.11 mm. The value of $\bar{x}_1 - \bar{x}_2$ = 12.2 mm. Hence t = 1.997. Reference to the appendix shows that t is below the entry value (2.02) for 95% confidence. Hence, we cannot reject the null hypothesis and must conclude that the drug has not been shown to have an effect. Since t is quite close to the entry value, it would probably be worthwhile extending the test to include more patients.

2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.4. The correlation coefficient and linear regression

The term “correlation” refers to how strongly two variables are related.



2.1. The key mathematical tools

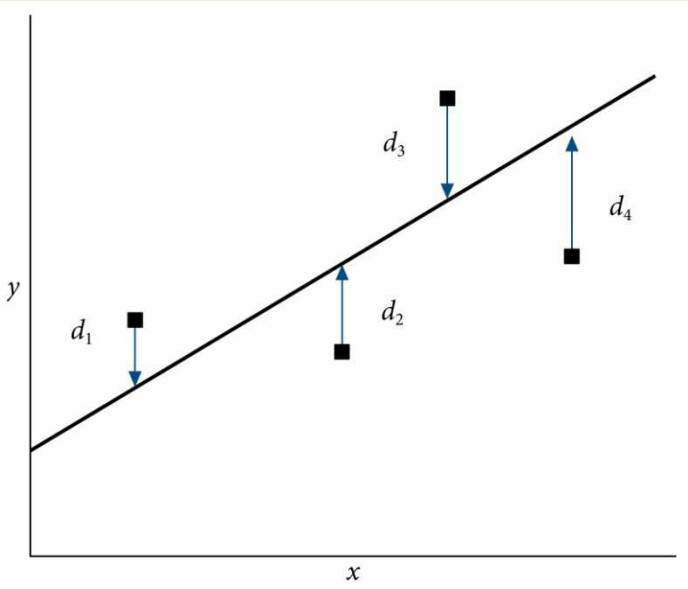
2.1.6. Some basic statistics

2.1.6.4. The correlation coefficient and linear regression

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

95% confidence level
r ≥ 0.754 (n=5)
r ≥ 0.576 (n=10)

r = +1, perfect positive correlation
r = -1, perfect negative correlation



The determination of the best straight line is known as linear regression.
Least-squares method is the most widely used method.

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2}}$$

$$\bar{y} = m\bar{x} + c$$

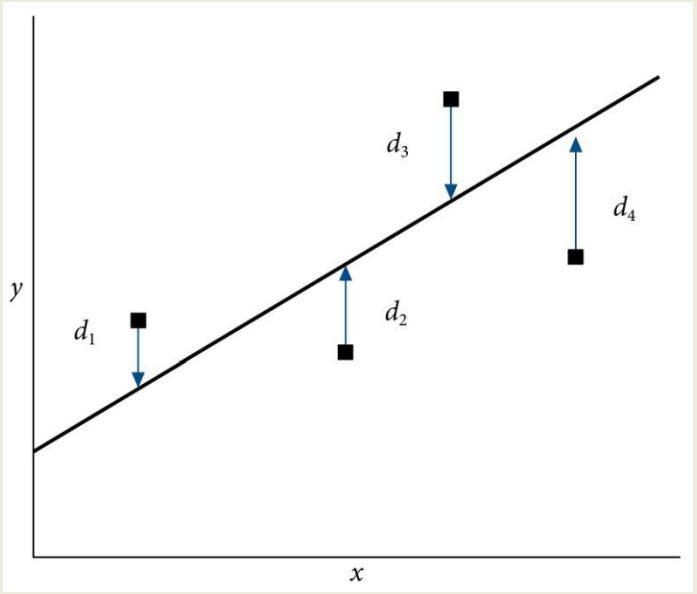
The value of the y-axis intercept, **c**, can be calculated from the equation above.

2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.4. The correlation coefficient and linear regression

$$s_m = \sqrt{\left(\frac{1}{n-2}\right)\left(\frac{n\sum(y^2) - (\sum(y))^2}{n\sum(x^2) - (\sum(x))^2} - m^2\right)}$$



s_m = standard errors in the estimates of the slope
s_c = y-axis intercept

$$s_c = s_m \sqrt{\frac{\sum(x)^2}{n}}$$

2.1. The key mathematical tools

2.1.6. Some basic statistics

$$\frac{[S]}{v} = \frac{[S]}{V_{max}} + \frac{K_m}{V_{max}}$$

2.1.6.4. The correlation coefficient and linear regression

Q: The rate (**v**, in units of **μM/min**) of an enzyme-catalyzed reaction was studied as a function of substrate concentration (**[S]**, in units of **μM**). The data was analyzed by the Hanes-Woolf plot in which **[S]/v** is plotted against **[S]**. The following values were obtained:

[S]/v	13.2	18.0	18.2	22.9	25.3	27.0	30.7
[S]	10	20	30	40	50	60	70

Calculate the correlation coefficient for the plot of **[S]/v** against **[S],** and use linear regression to calculate the best straight line.

ST: $$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2\ \Sigma(y-\bar{y})^2}} \qquad m = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}} \qquad \bar{y} = m\bar{x} + c$$

S: [S]/v is designated as y and **[S]** as x. The values of y and x are 22.19 and 40, respectively. The values of **∑ (y-y)²** and **∑ (x-x)²** are 220.03 and 2,800 respectively. The value of **∑(x-x)(y-y)** is 776. From this, using the first equation **r** = 0.9886; this is highly significant correlation (**p < 0.001** for 5 degrees of freedom, i.e. the number of (x , y) data points (7) -2). The slope (**m**) and y-axis intercept (**c**) of the least-squares line are 0.227 and 11.1, respectively. Further analysis using the equations for s_m and s_c above gives 0.029 and 1.3, respectively.

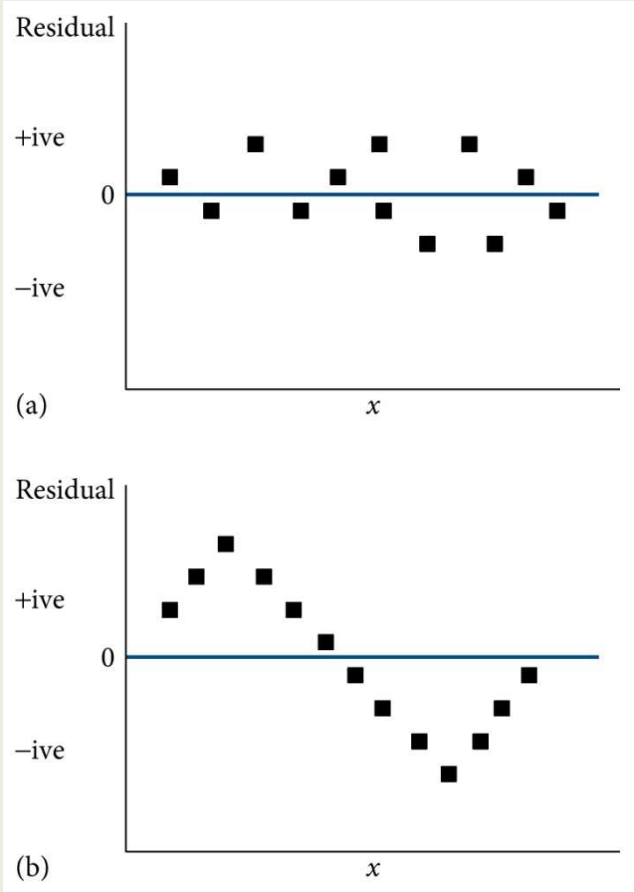
2.1. The key mathematical tools

2.1.6. Some basic statistics

2.1.6.5. Non-linear regression

NRMSD – the normalized root mean square deviation
y_{obs} and y_{cal} are the observed and calculated values of y, at each specified value of x.

$$\mathbf{NRMSD} = \sqrt{\frac{\sum(y_{obs} - y_{cal})^2}{\sum(y_{obs})^2}}$$



Please solve the problem. 5 points

The following values of blood cholesterol (mM) were found in a sample of eight healthy females: 4.3, 4.1, 5.8, 5.0, 3.9, 5.5, 5.2, and 4.9.

Question 5: What is the median, mean and standard deviation of these values? What is the 95% confidence limit for the population mean (μ).

Answer: The values are: median, 4.95 mM; mean, 4.84 mM; standard deviation, 0.68 mM; 95% confidence limit for μ, 4.37–5.31 mM.

Please solve the problem. 5 points

A group of 21 healthy students undertook a glucose tolerance test in which they fasted overnight and then ingested 75 g glucose. Before taking the glucose, the blood glucose levels of the group had a mean value of 4.73 mM (SD = 0.48 mM). 30 min after taking the glucose, the blood glucose levels had a mean value of 7.95 mM (SD = 1.63 mM). After another 90 min, the levels had a mean value of 5.25 mM (SD = 1.53 mM).

Question 6: Are the levels at 30 and 120 min significantly different from that at the start?

Answer: Comparing 30 min and start values, t = 8.68; this gives $p < 0.01$, i.e. the null hypothesis can be rejected with >99% confidence. Comparing 120 min and start values, t = 1.49; this gives a **p** value between 0.2 and 0.1 ($0.2 > p > 0.1$); i.e. the null hypothesis cannot be rejected with at least 95% confidence. Thus, the 30 min value is significantly higher than the start value, but the 120 values is not.

Please solve the problem. 5 points

A Hanes-Woolf plot used to analyze a set of enzyme kinetic data obtained at eight values of substrate concentration showed a correlation coefficient (**r**) of 0.669.

Question 7: What would you recommend to the investigator who produced the data?

Answer: The value of **r** is below the value required for 95% confidence of a positive correlation. It would not, therefore, be appropriate to use the plot to try to obtain reliable values of the kinetic parameters (K_m and V_{max}) for the enzyme. It would be sensible to try to improve the experimental technique and to obtain more data points.