

Variations in Plants and Animals

Purpose and expected outcomes

1. Define quantitative genetics and distinguish it from population genetics.
2. Distinguish between qualitative traits and quantitative traits.
3. Discuss polygenic inheritance
4. Discuss the variance components of quantitative traits.
5. Discuss the concept of heritability of traits.
6. Discuss selection and define the breeders' equation.



Quantitative genetic methods are being used to improve and predict racing speed in thoroughbred horses.

What is quantitative genetics?

- ▶ **Population genetics** deals with Mendel's laws and other genetic principles as they affect entire populations of organisms.
- ❖ Population genetics does not assign a genotypic or numerical value to each of the individuals (genotypes) in the population.
- ❖ **Quantitative genetics**: individual genotypes are unidentified, and the traits of individuals are measured.
- ❖ Genotypic values are assigned to genotypes in the population.
- ❖ Quantitative genetics emphasizes the role of selection in controlled populations of known ancestry.
- ❖ Population genetics is basic to quantitative genetics.

Quantitative genetics Continued...

- ❖ Phenotype, a product of the genotype and the environment.
- ❖ The genotypic array depends on mating systems and genetic linkage relationships, as well as on allelic frequencies, which in turn are impacted by mutation, migration, random drift, and selection.
- ❖ Assumptions about the mating system, allelic frequency altering forces, and the environment.

Assumptions of quantitative genetic analysis

- **Reference population defined:** Allelic and genotypic frequencies can only be defined with respect to a specified population.
- **Absence of linkage:** Trait (phenotype) observed is not affected by autosomal linkage genes.
- **Presence of diploid Mendelian inheritance:** diploid in which genes segregate and assort independently.
- **Absence of selection during the formation of inbred lines:** no selection occur when inbred lines are crossed.
- **No inbreeding of the reference population:** the inbreeding coefficient of the reference population is zero

Quantitative traits

- Most traits encountered in plant breeding are **quantitatively inherited**.
- **Many genes** control such traits, each contributing a **small effect** to the overall phenotypic expression of a trait.
- Variation in quantitative trait expression is **continuous**.
- The traits that exhibit continuous variations are also called **metric traits**.

Qualitative genetics versus quantitative genetics

Qualitative genetics:

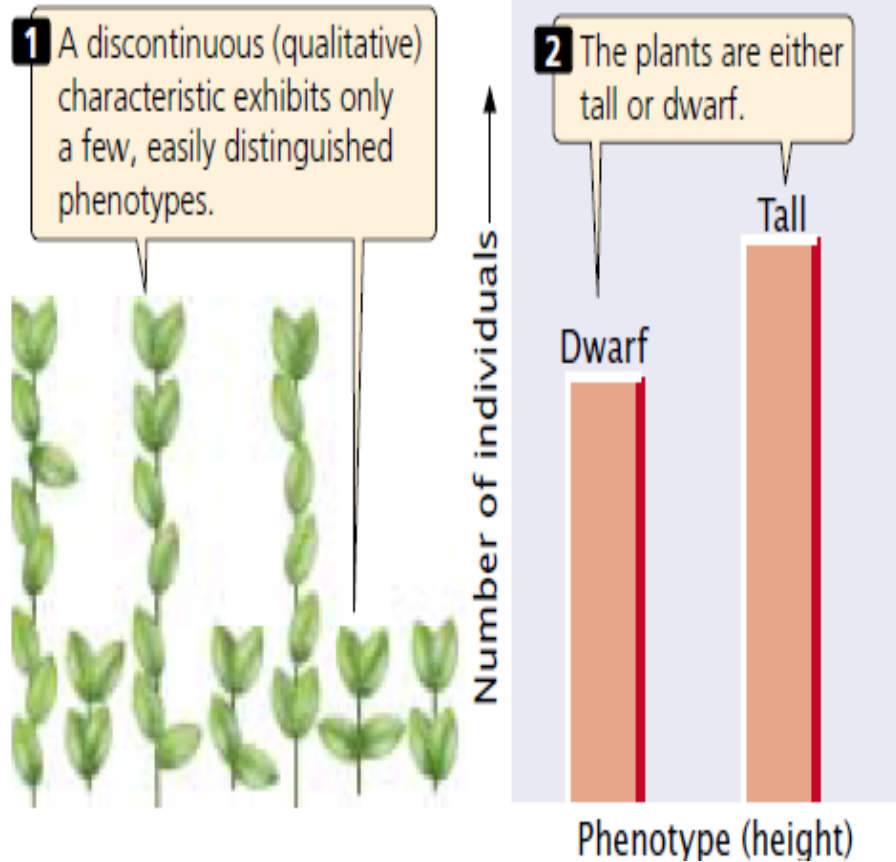
- ❖ Traits that have Mendelian inheritance and can be described according to kind and, can be unambiguously categorized.
- ❖ Traits provide discrete (discontinuous) phenotypic variation.
- ❖ Effects of single genes are readily detectable.
- ❖ Concerned with individual matings and their progenies.
- ❖ **Statistical** analysis is quite straightforward, and is based on counts and ratios

Qual. genetics versus quan. genetics Continued...

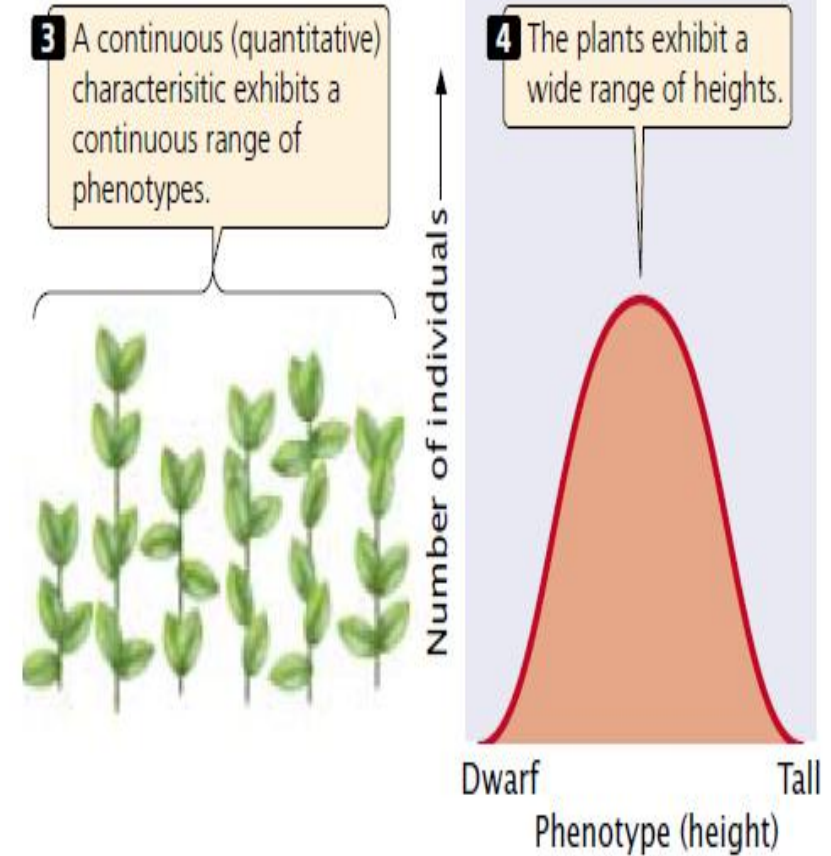
Quantitative genetics:

- ❖ Traits are described in terms of the degree of expression of the trait, rather than the kind.
- Traits produce phenotypic variation that spans the full spectrum (continuous).
- Single gene effects are not discernible.
- Traits are under polygenic control.
- Genes with small indistinguishable effects.
- Concerned with a population of individuals that may comprise a diversity of mating kinds.
- Quantitative genetic analysis provides estimates of population parameters (attributes of the population from which the sample was obtained).

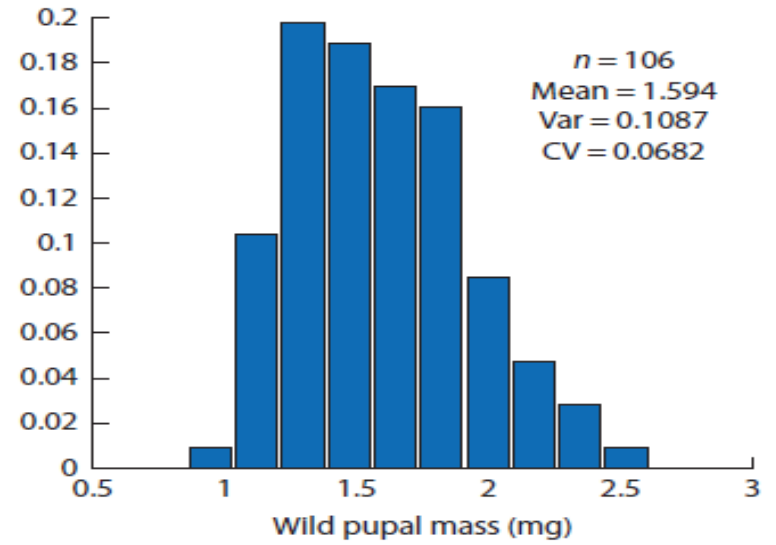
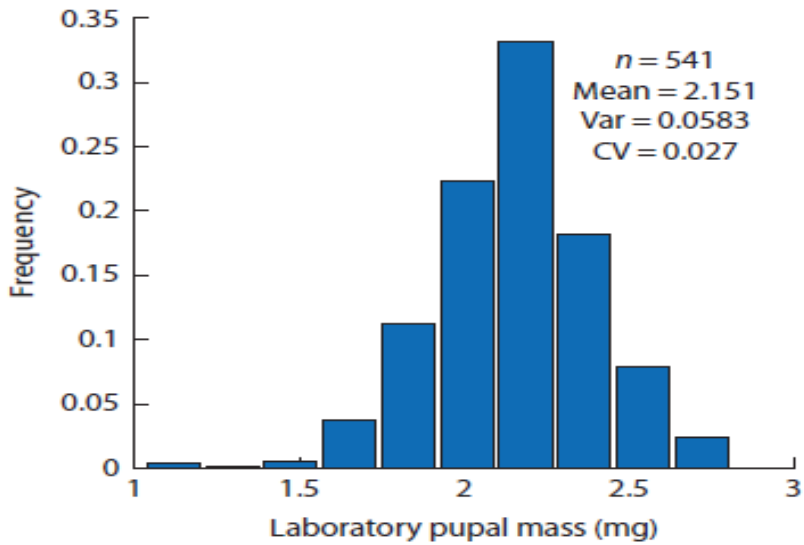
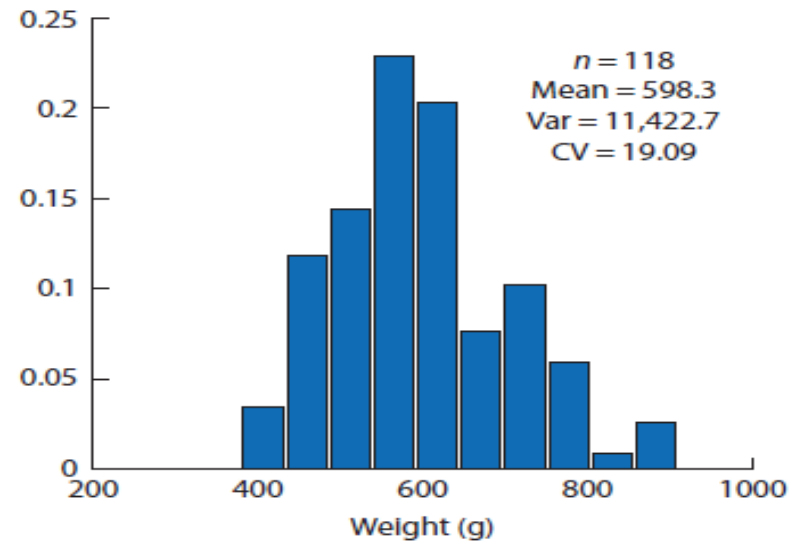
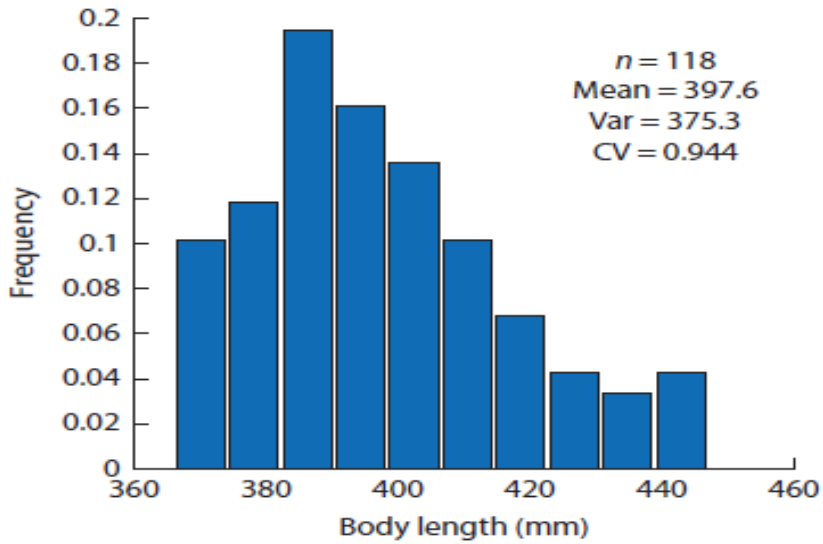
(a) Discontinuous characteristic



(b) Continuous characteristic



Discontinuous and continuous characteristics differ in the number of phenotypes exhibited



Examples of Quantitative traits

The Relation Between Genotype and Phenotype

- Discontinuous characteristics, there is a relatively straightforward relation between genotype and phenotype.
- Each genotype produces a single phenotype, and most phenotypes are encoded by a single genotype.
- Dominance and epistasis may allow two or three genotypes to produce the same phenotype, but the relation remains relatively simple.
- For quantitative characteristics, the relation between genotype and phenotype is often more complex.
- If the characteristic is polygenic, many different genotypes are possible, several of which may produce the same phenotype.

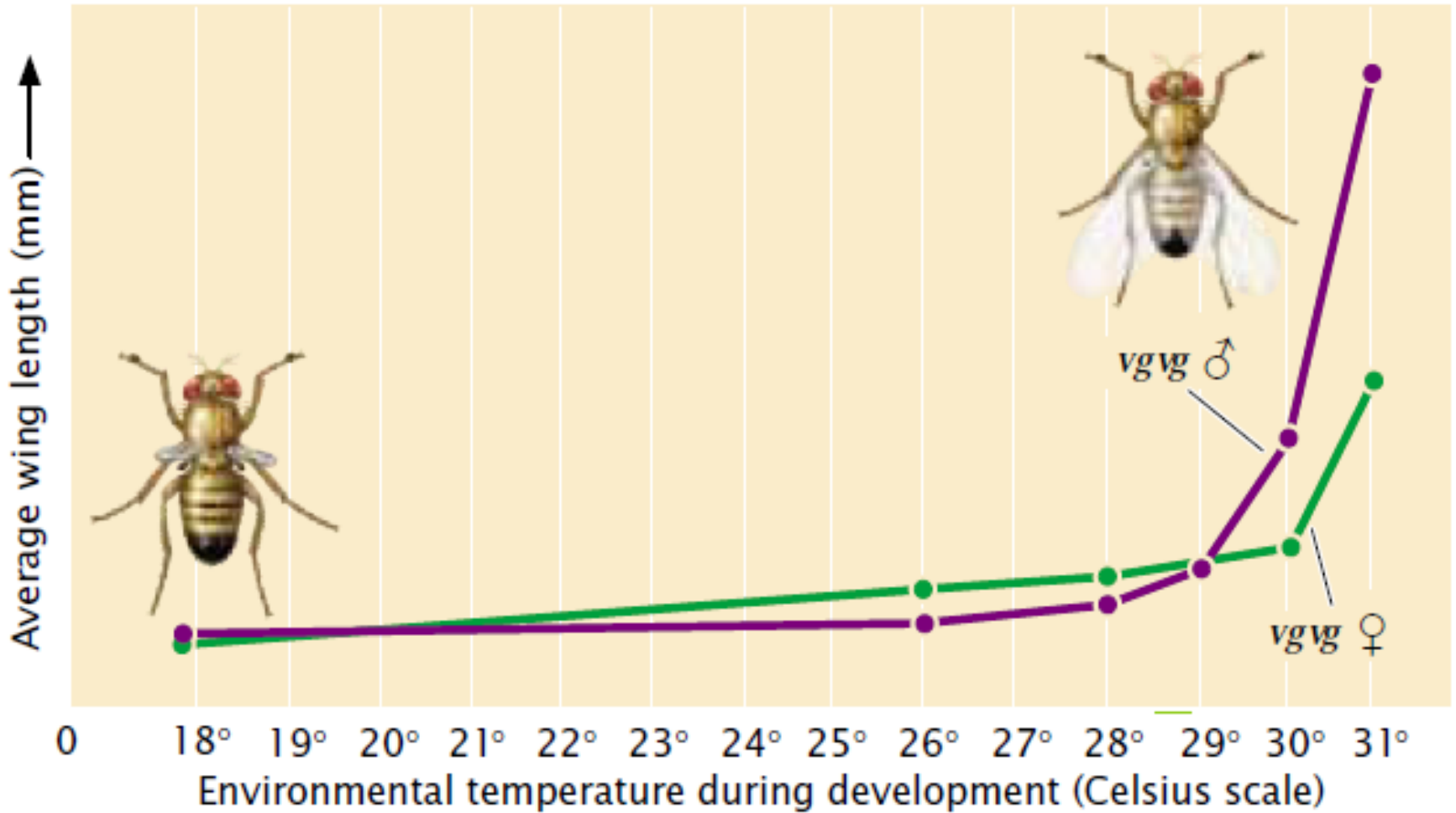
Genotype and Phenotype continued.....

- Consider a plant whose height is determined by three loci (A , B , and C), each of which has two alleles.
- Assume that one allele at each locus (A^+ , B^+ , and C^+) encodes a plant hormone that causes the plant to grow 1 cm above its baseline height of 10 cm.
- The second allele at each locus (A^- , B^- , and C^-) encodes no plant hormone and does not contribute to additional height.
- Considering only the two alleles at a single locus, 3 genotypes are possible ($A^+ A^+$, $A^+ A^-$, and $A^- A^-$).

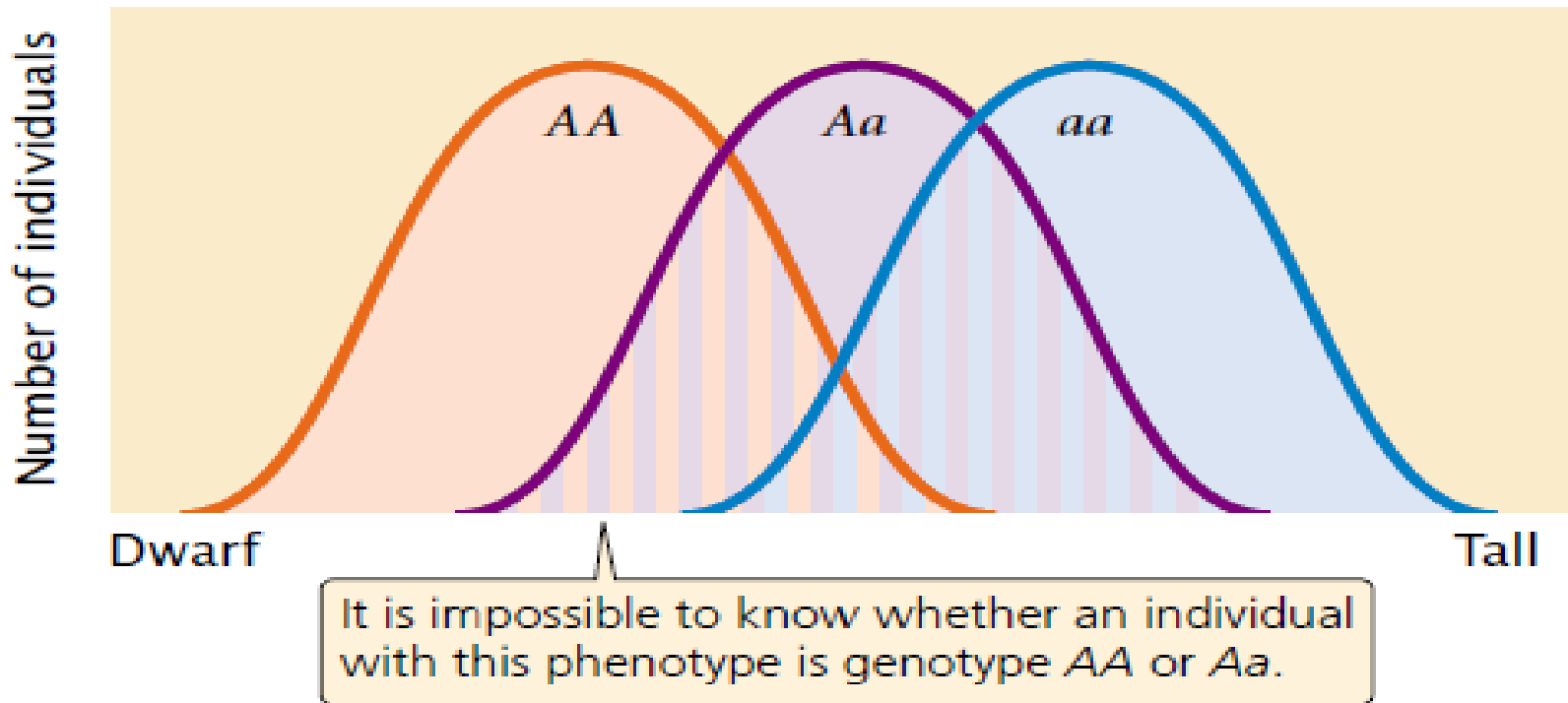
Hypothetical example of plant height determined by pairs of alleles at each of three loci

| Genotype | Doses of Plant Hormone | Height (cm) |
|----------------------|------------------------|-------------|
| $A^-A^-B^-B^-C^-C^-$ | 0 | 10 |
| $A^+A^-B^-B^-C^-C^-$ | 1 | 11 |
| $A^-A^-B^+B^-C^-C^-$ | | |
| $A^-A^-B^-B^+C^-C^-$ | | |
| $A^+A^+B^-B^-C^-C^-$ | 2 | 12 |
| $A^-A^-B^+B^+C^-C^-$ | | |
| $A^-A^-B^-B^-C^+C^+$ | | |
| $A^+A^-B^+B^-C^-C^-$ | | |
| $A^+A^-B^-B^-C^+C^-$ | | |
| $A^-A^-B^+B^-C^+C^-$ | | |
| $A^+A^+B^+B^-C^-C^-$ | 3 | 13 |
| $A^+A^+B^-B^-C^+C^-$ | | |
| $A^+A^-B^+B^+C^-C^-$ | | |
| $A^-A^-B^+B^+C^+C^-$ | | |
| $A^+A^-B^-B^-C^+C^+$ | | |
| $A^-A^-B^+B^-C^+C^+$ | | |
| $A^+A^-B^+B^-C^+C^-$ | | |
| $A^+A^+B^+B^+C^-C^-$ | 4 | 14 |
| $A^+A^+B^+B^-C^+C^-$ | | |
| $A^+A^-B^+B^+C^+C^-$ | | |
| $A^-A^-B^+B^+C^+C^+$ | | |
| $A^+A^+B^-B^-C^+C^+$ | | |
| $A^-A^-B^+B^-C^+C^+$ | | |
| $A^+A^+B^+B^+C^+C^-$ | 5 | 15 |
| $A^+A^-B^+B^+C^+C^+$ | | |
| $A^+A^+B^+B^-C^+C^+$ | | |
| $A^+A^+B^+B^+C^+C^+$ | 6 | 16 |

Note: Each allele contributes 1 cm in height above a baseline of 10 cm.



Norm of reaction is the range of phenotypes produced by a genotype in different environments.

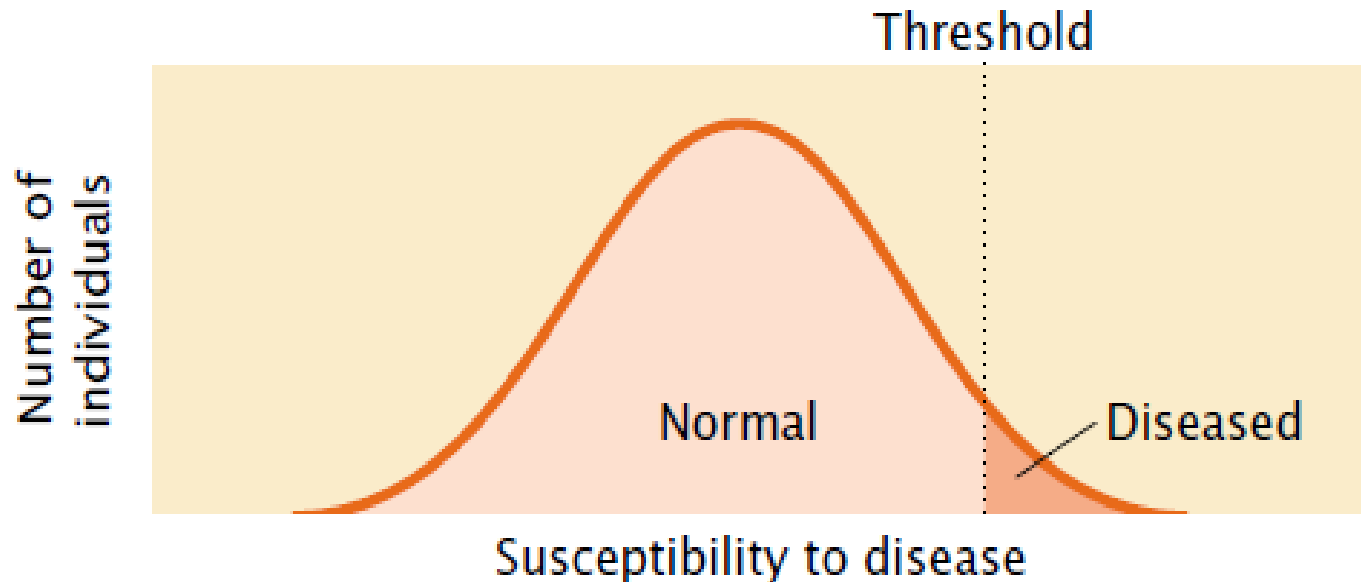


The phenotypic ranges of different genotypes may overlap, making it difficult to know whether individuals differ in phenotype because of genetic or environmental differences

Types of Quantitative Characteristics

Meristic characteristics:

- Are measured in whole numbers.
- An example is litter size: a female mouse may have 4, 5, or 6 pups but not 4.13 pups.
- Has a limited number of distinct phenotypes, but the underlying determination of the characteristic may still be quantitative.



Threshold characteristic:

- which is simply present or absent.
- Although threshold characteristics exhibit only two phenotypes, they are considered quantitative because they, too, are determined by multiple genetic and environmental factors.

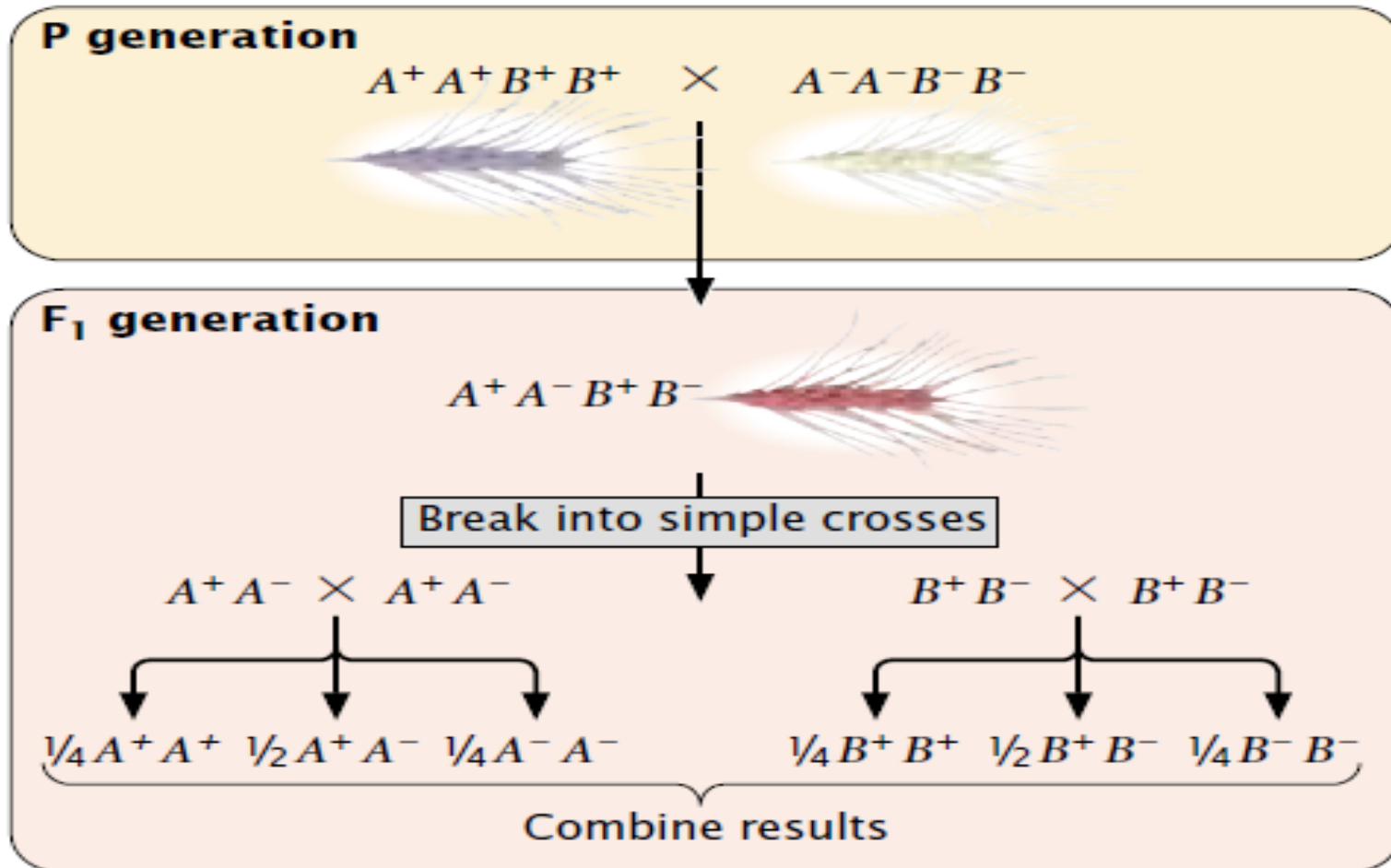
Polygenic Inheritance

- ❖ Ronald Fisher (1918) demonstrated that the inheritance of quantitative characteristics could indeed be explained by the cumulative effects of many genes, each following Mendel's rules.

Kernel Color in Wheat

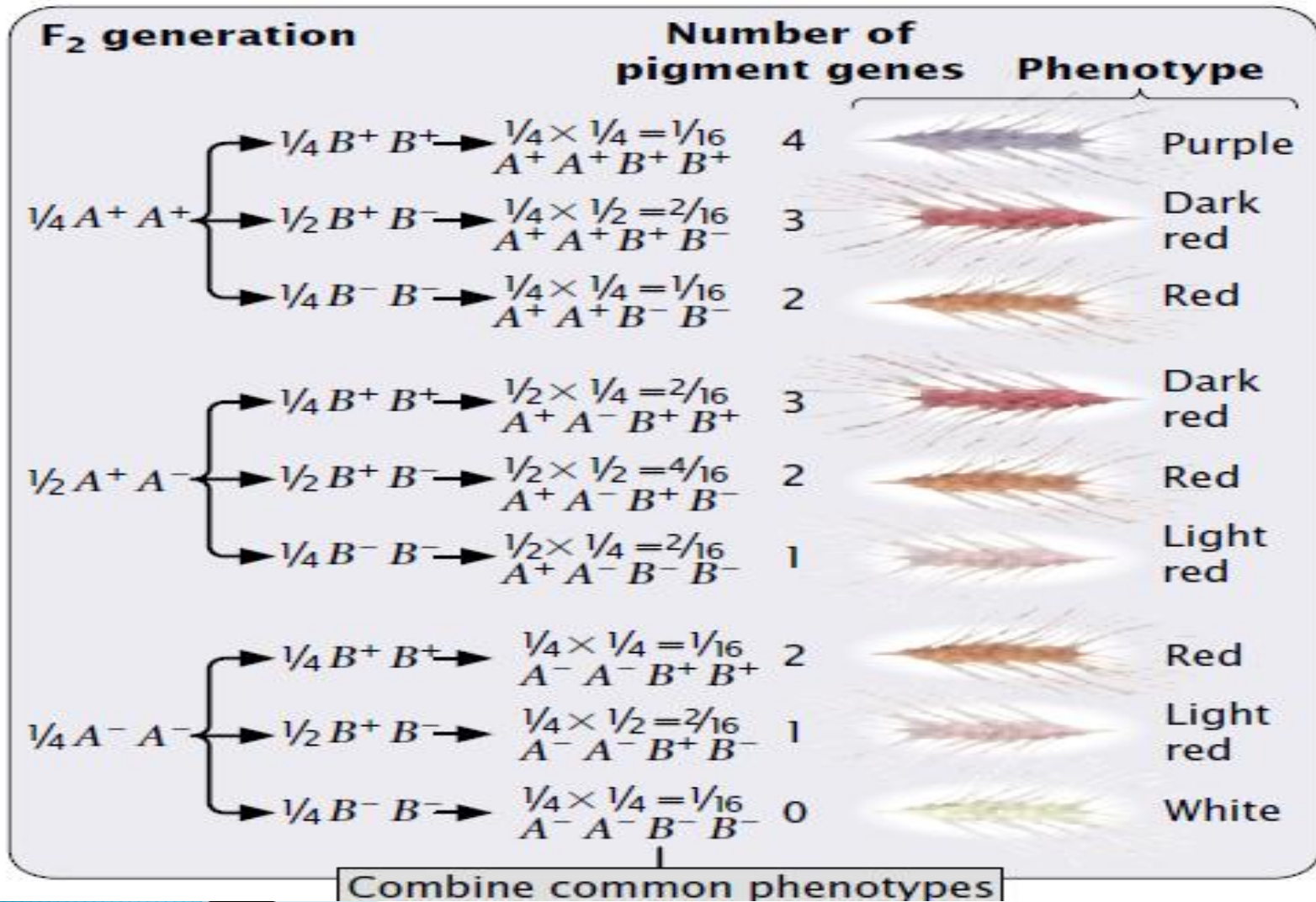
Nilsson-Ehle proposed that there were two alleles at each locus: one that produced red pigment and another that produced no pigment.

| Genotype | Doses of pigment | Phenotype |
|----------------|------------------|-----------|
| $A^+A^+B^+B^+$ | 4 | purple |
| $A^+A^+B^+B^-$ | 3 | dark red |
| $A^+A^-B^+B^+$ | | |
| $A^+A^+B^-B^-$ | 2 | red |
| $A^-A^-B^+B^+$ | | |
| $A^+A^-B^+B^-$ | | |
| $A^+A^-B^-B^-$ | 1 | light red |
| $A^-A^-B^+B^-$ | | |
| $A^-A^-B^-B^-$ | 0 | white |








Nilsson-Ehle demonstrated that kernel color in wheat is inherited according to Mendelian principles.

Nilsson-Ehle experiment continued...

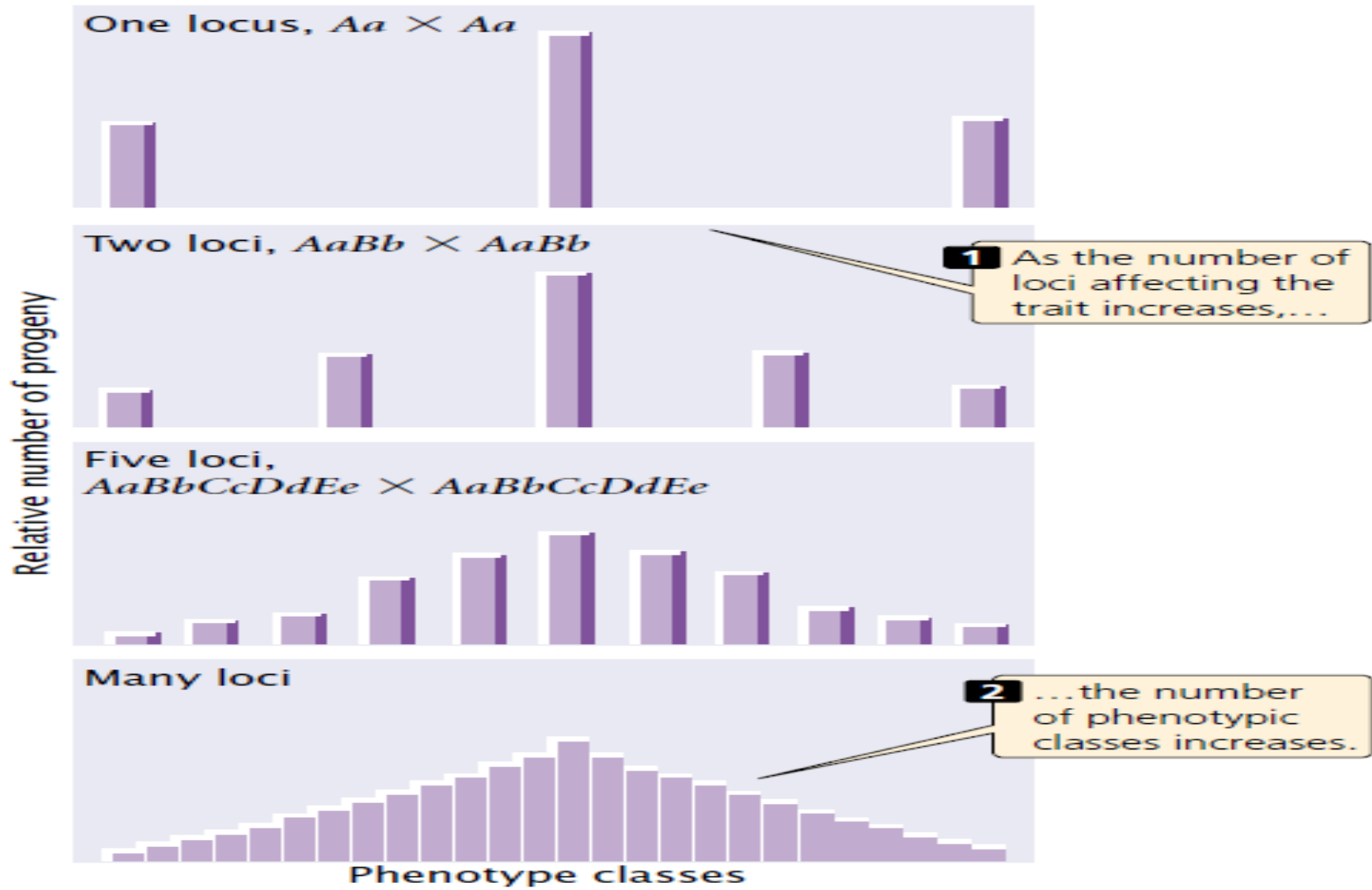


Nilsson-Ehle experiment continued...

Combine common phenotypes

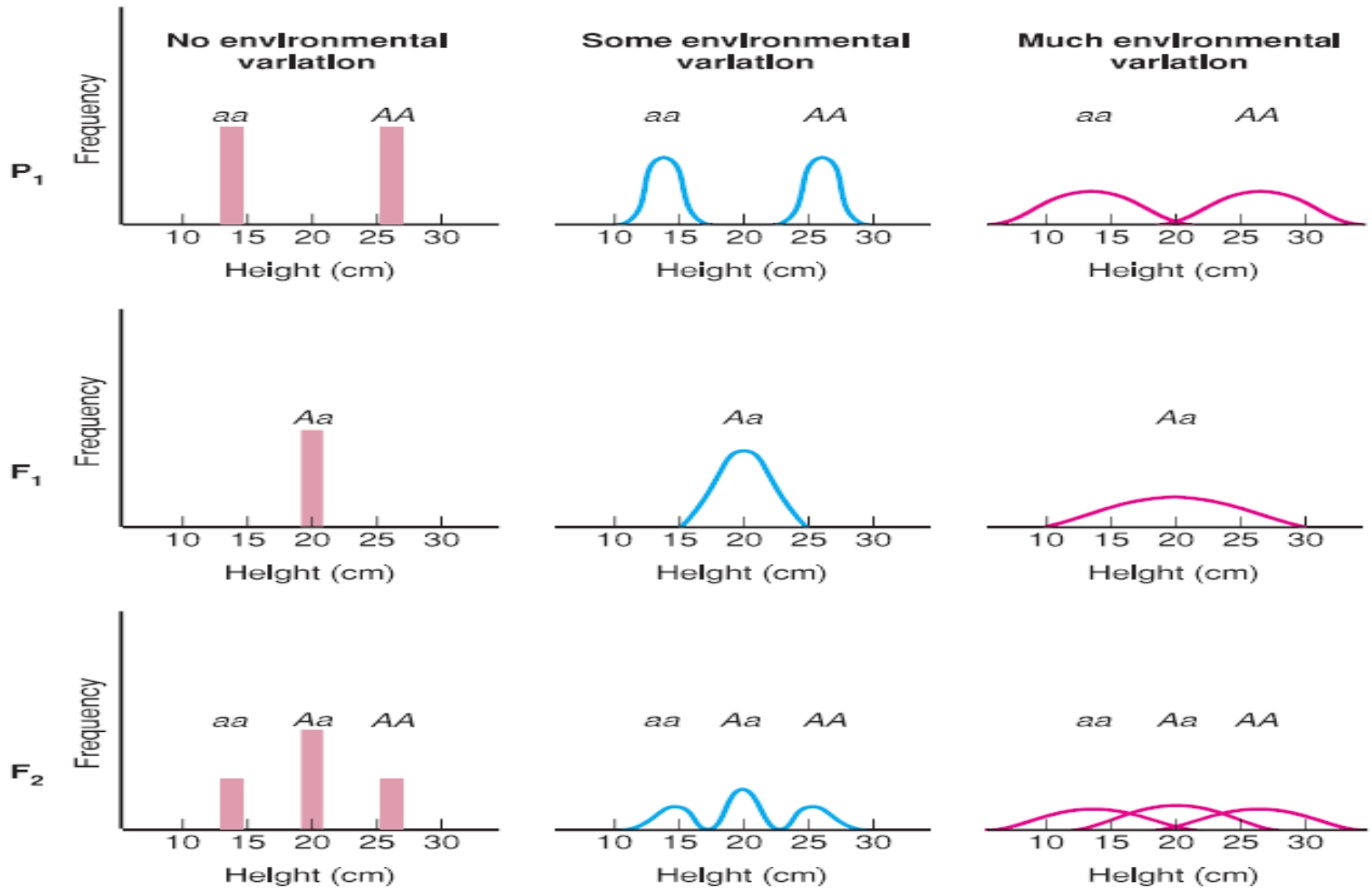
| F₂ ratio | Frequency | Number of pigment genes | Phenotype |
|----------------------------|------------------|--------------------------------|--|
| | $\frac{1}{16}$ | 4 |  Purple |
| | $\frac{4}{16}$ | 3 |  Dark red |
| | $\frac{6}{16}$ | 2 |  Red |
| | $\frac{4}{16}$ | 1 |  Light red |
| | $\frac{1}{16}$ | 0 |  White |

Conclusion: Polygenic characteristics are inherited according to Mendel's principles.



The results of crossing individuals heterozygous for different numbers of loci affecting a characteristic

➤ The principles that determine the inheritance of quantitative characteristics are the same as the principles that determine the inheritance of discontinuous characteristics, but more genes take part in the determination of quantitative characteristics.

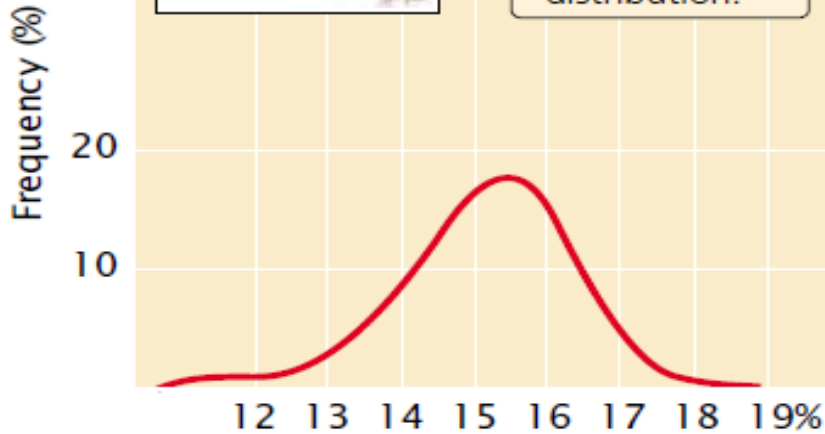


The influence of environment on phenotypic distribution.

(a) Sugar beet percentage of sucrose



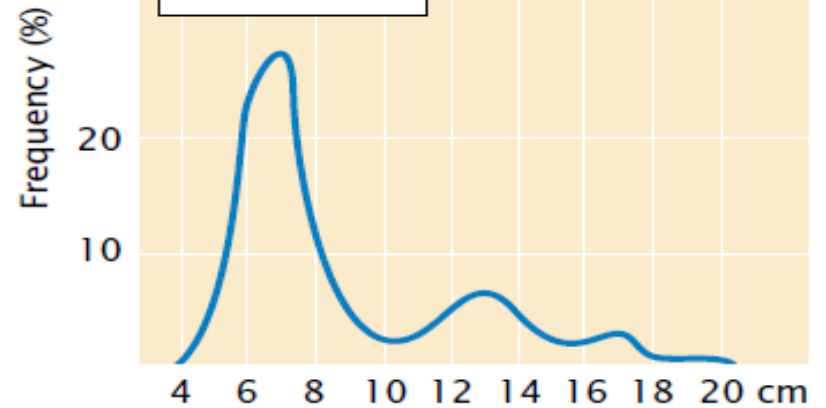
1 This type of symmetrical (bell-shaped) distribution is called a normal distribution.



(b) Squash fruit length



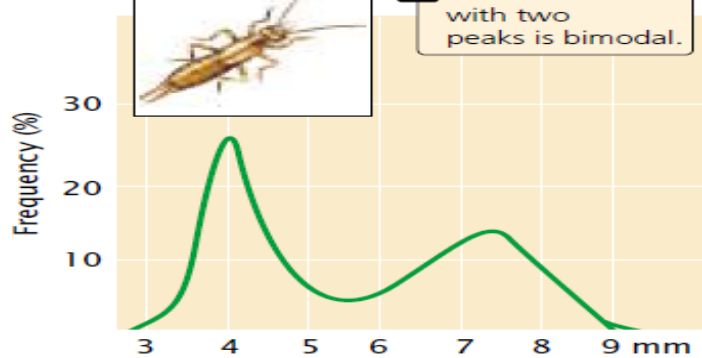
2 The distribution of fruit length among the F₂ progeny is skewed to the right.



(c) Earwig forceps length



3 A distribution with two peaks is bimodal.



Distributions of phenotypes may assume several different shapes

Common statistical methods in breeding

Purpose and expected outcomes

1. Describe the role of statistics in breeding.
2. Discuss the measures of central tendency.
3. Discuss the measures of dispersion.
4. Discuss the measures of association.
5. Discuss the method of analysis of variance.

Role of statistics in breeding

- To obtain a descriptive summary of the sample
- To provide a means of statistical inference.
- Comparison

Statistical hypothesis

Null hypothesis

H0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (i.e., all genotype means are equal)

Alternate hypothesis,

H1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$ (i.e., all genotype means are not equal)

- The H0 is accepted (i.e., automatically reject H1), or rejected (accept H0), at a chosen level of **statistical significance, α** (e.g., $\alpha = 0.01$ or 0.05 ; acknowledging that 1% or 5% of the time you could be mistaken in your conclusion).

Type I error: a false rejecting a true null hypothesis.

Type II error: incorrectly failing to accept a false null hypothesis.

Principles of experimental design

- Experimental designs are statistical procedures for arranging experimental units (experimental design) such that experimental error is minimized.

Replication

- ❖ Replication is the number of times a treatment is repeated in a study.

Randomization

- ❖ This is the principle of equal opportunity whereby treatment allocation to experimental units is made without bias.

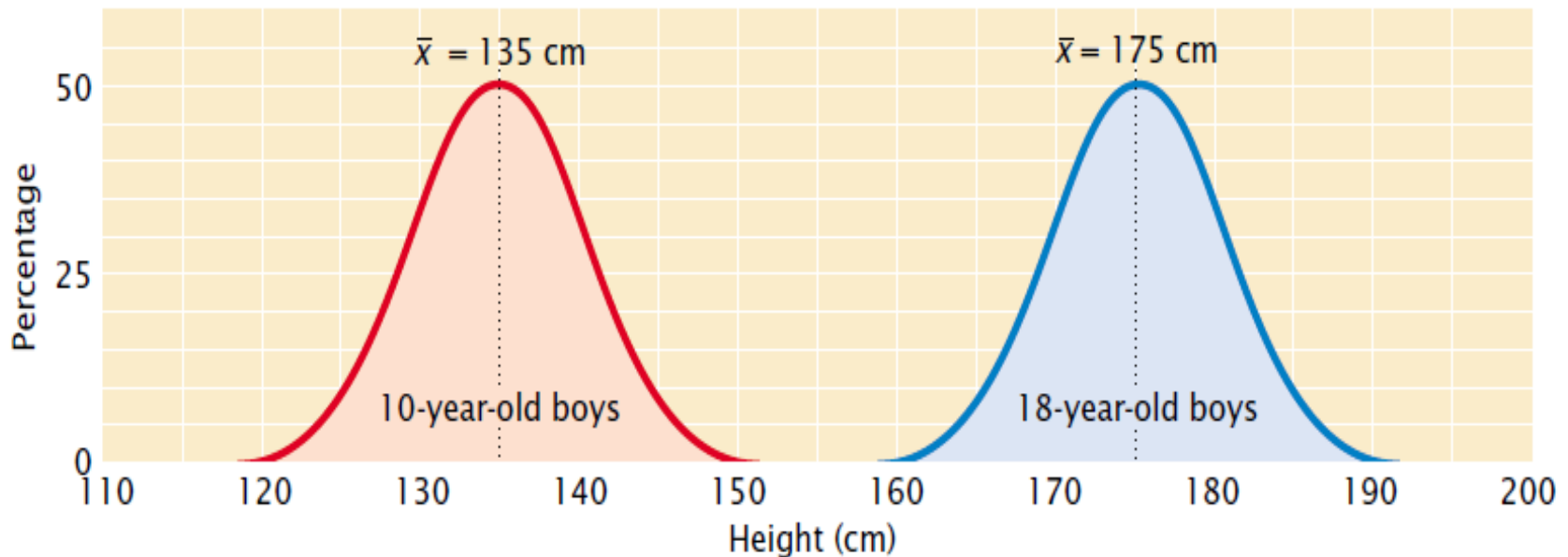
Local control

- ❖ Variation within groups is kept to a minimum, while enhancing variation between groups.

Measures of central tendency

Sample mean, $\bar{X} = \Sigma X/n$

Population mean, $\mu = \Sigma X/N$



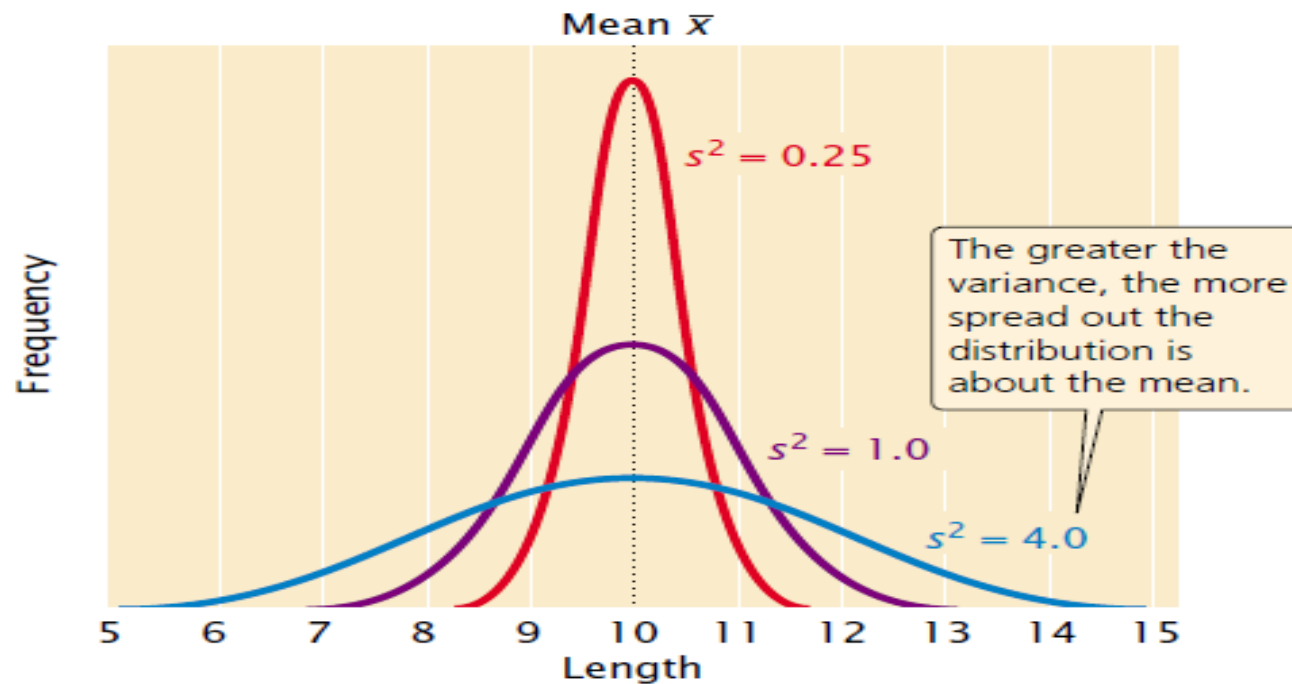
The mean provides information about the center of a distribution.

Measures of dispersion

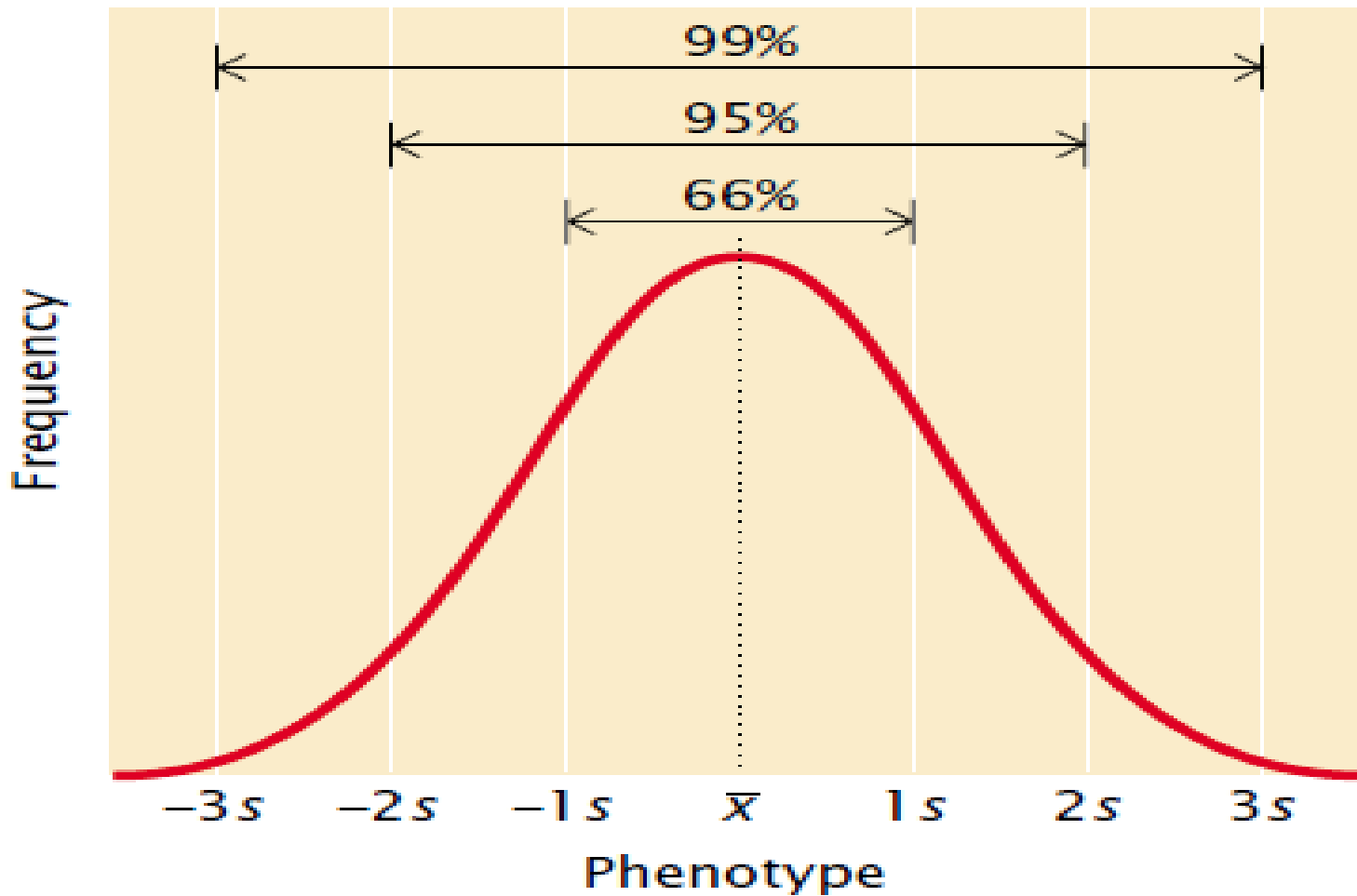
The population variance: $\sigma^2 = [\Sigma(X1 - \mu)^2]/N$

The sample variance: $s^2 = [\Sigma(X - \bar{X})^2]/(n - 1)$

Standard deviation: $s = \sqrt{s^2}$



The variance provides information about the variability of a group of phenotypes.



The proportions of a normal distribution occupied by plus or minus one, two, and three standard deviations from the mean.

Variance (σ^2) The sum of the squared deviations from the mean divided by one less than the sample size.

Standard deviation (σ) The square root of the variance; the average deviation from the mean for a single observation. Quantifies the range of values around the mean seen in a sample.

Standard error of an average (SE) The product of the standard deviation and the square root of the sample size divided by the sample size; how far the true population average (a parameter) may be from the sample average (a parameter estimate) by chance.

Exercise: The following table lists yearly amounts (in hundreds of pounds) of milk produced by 10 two-year-old Jersey cows. Calculate the mean, variance, and standard deviation of milk production for this sample of 10 cows.

**Annual milk production
(hundreds of pounds)**

60

74

58

61

56

55

54

57

65

42

Solution

$$\bar{x} = \sum_i x_i / n. \quad \bar{x} = (582/10) = 58.20$$

| Annual milk production (hundreds of pounds) x | Variance $x_i - \bar{x}$ | Standard deviation $(x_i - \bar{x})^2$ |
|---|-----------------------------|--|
| 60 | 1.80 | 3.24 |
| 74 | 15.80 | 249.64 |
| 58 | -0.20 | 0.04 |
| 61 | 2.80 | 7.84 |
| 56 | -2.20 | 4.84 |
| 55 | -3.20 | 10.24 |
| 54 | -4.20 | 17.64 |
| 57 | -1.20 | 1.44 |
| 65 | 6.80 | 46.24 |
| 42 | -16.20 | 262.44 |

$$\sum (x_i - \bar{x})^2 = 603.60$$

The variance is therefore:

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{603.60}{9} = 67.07$$

The standard deviation is the square root of the variance:

$$s_x = \sqrt{s_x^2} = \sqrt{67.07} = 8.19.$$

Simple linear correlation

- The population correlation coefficient (ρ) is given by:

$$\rho = \sigma^2 XY / \sqrt{(\sigma X^2 \times \sigma Y^2)}$$

- where $\sigma X^2 = \text{variance of } X$, $\sigma Y^2 = \text{variance of } Y$, and
- $\sigma^2 XY = \text{covariance of } X \text{ and } Y$.
- *The sample covariance is called the*
- ***Pearson correlation coefficient (r) and is calculated as:***

- $r = s^2 XY / \sqrt{(s^2 X \times s^2 Y)}$

$$\text{COV}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad r = \frac{\text{COV}_{xy}}{S_x S_y}$$

- A correlation coefficient can theoretically range from -1 to +1.
- A positive value indicates that there is a direct association between the variables as one variable increases; the other variable also tends to increase.
- A positive correlation exists for human height and weight: tall people tend to weigh more.
- A negative correlation coefficient indicates that there is an inverse relation between the two variables; as one variable increases, the other tends to decrease.

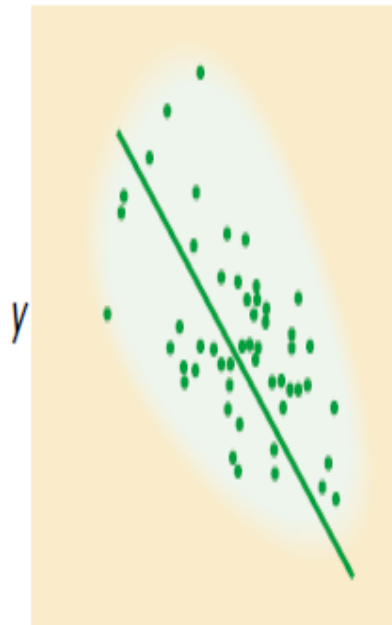
- An $r = -1$ or $+1$ indicates a perfect correlation between the variables, meaning that a change in x is always accompanied by a proportional change in y .
- r close to -1 or close to $+1$ indicate a strong association between the variables—a change in x is almost always associated with a proportional increase in y .
- An r closer to 0 indicates a weak correlation—a change in x is associated with a change in y but not always.
- An r of 0 indicates that there is no association between variables.
- An r can be computed for two variables measured for the same individual, such as height (x) and weight (y).
- An r can also be computed for a single variable measured for pairs of individuals.

(a) $r = .7$



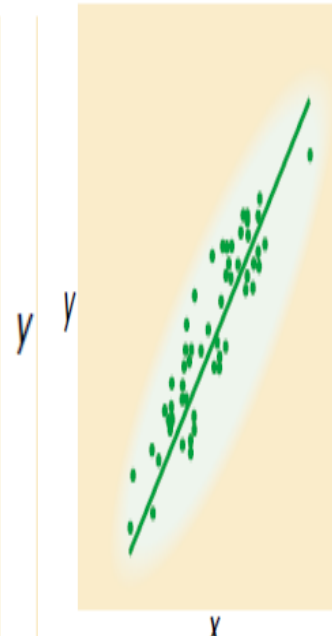
A positive correlation indicates that there is a direct association between variables.

(b) $r = -.7$



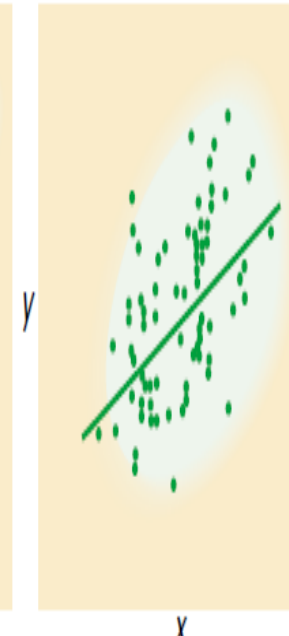
A negative correlation indicates that there is an inverse association between variables.

(c) $r = .9$



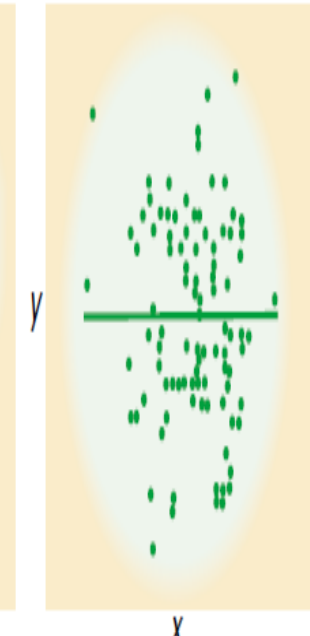
A strong positive correlation.

(d) $r = .3$



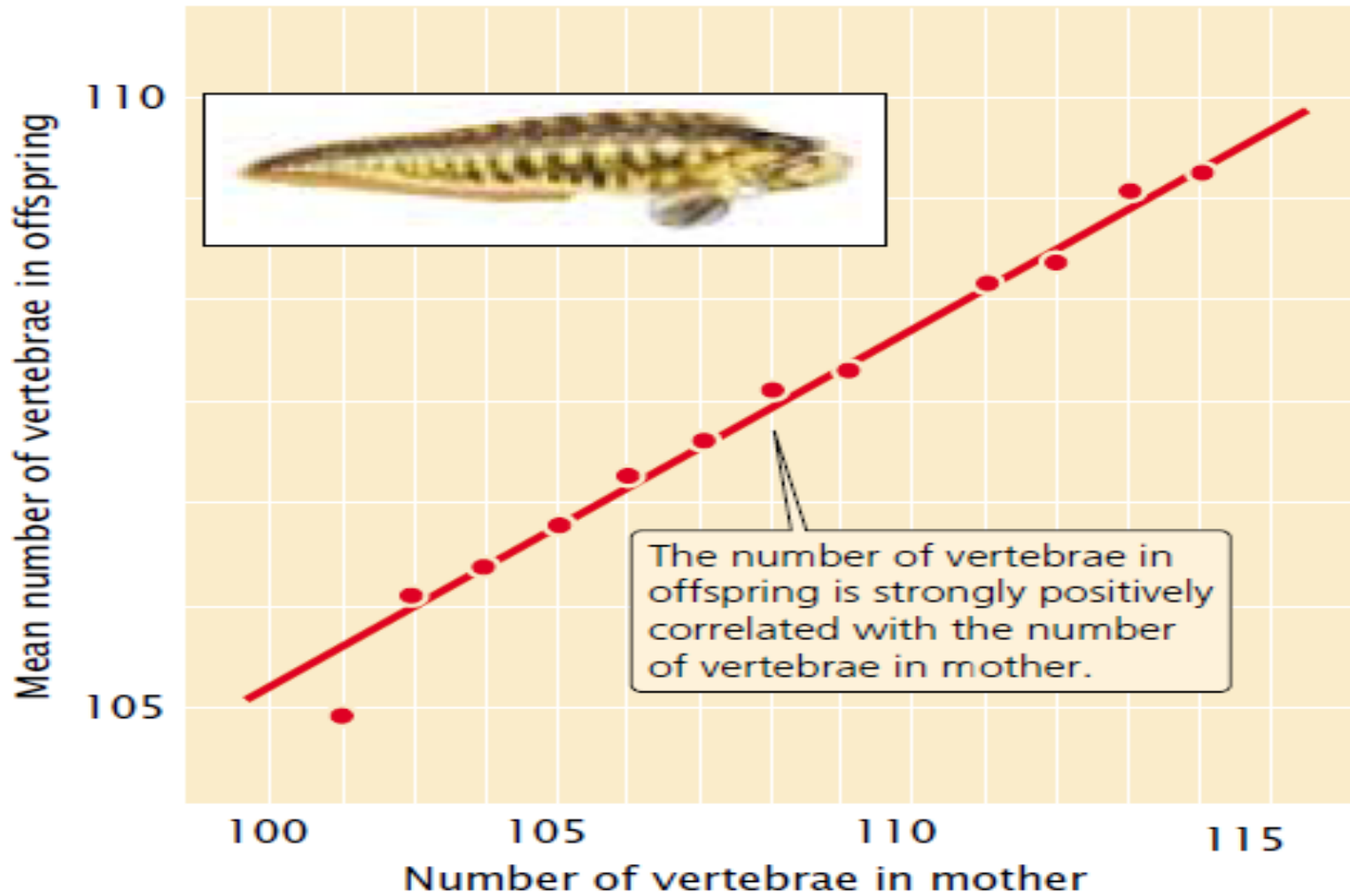
A weak positive correlation.

(e) $r = 0$



A correlation of zero indicates that there is no association between variables.

The correlation coefficient describes the relation between two or more variables.



A correlation coefficient can be computed for a single variable measured for pairs of individuals

Simple linear regression

- Simple linear regression has the mathematical form of the equation of a straight line:

$$Y = a + bX$$

- Where $Y =$ *dependent variable*, $X =$ *independent variable*, $b =$ *slope of the regression line*, and $a =$ *intercept on the y axis*.

Exercise: Body weights of 11 female fish and the numbers of eggs that they produce are given in the following table. What are the correlation coefficient and the regression coefficient for body weight and egg number in these 11 fish?

Weight (mg)

x

14

17

24

25

27

33

34

37

40

41

42

Eggs (thousands)

y

61

37

65

69

54

93

87

89

100

90

97

Solution

$$\bar{x} = \frac{\sum x_i}{n} = \frac{334}{11} = 30.36$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{842}{11} = 76.55$$

$$\text{COV}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{1743.84}{10} = 174.38$$

$$s_x^2 = \frac{(x_i - \bar{x})^2}{n - 1} = \frac{932.55}{10} = 93.26$$

Solution continued...

| A | B | C | D | E | F | G |
|-------------|-----------------|---------------------|------------------|-----------------------|---------------------|----------------------------------|
| Weight (mg) | | | Eggs (thousands) | | | |
| x | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | \bar{y} | $\bar{y}_i - \bar{y}$ | $(y_i - \bar{y})^2$ | $(x_i - \bar{x})(y_i - \bar{y})$ |
| 14 | -16.36 | 267.65 | 61 | -15.55 | 241.80 | 254.40 |
| 17 | -13.36 | 178.49 | 37 | -39.55 | 1564.20 | 528.39 |
| 24 | -6.36 | 40.45 | 65 | -11.55 | 133.40 | 73.46 |
| 25 | -5.36 | 28.73 | 69 | -7.55 | 57.00 | 40.47 |
| 27 | -3.36 | 11.29 | 54 | -22.55 | 508.50 | 75.77 |
| 33 | 2.64 | 6.97 | 93 | 16.45 | 270.60 | 43.43 |
| 34 | 3.64 | 13.25 | 87 | 10.45 | 109.20 | 38.04 |
| 37 | 6.64 | 44.09 | 89 | 12.45 | 155.00 | 82.67 |
| 40 | 9.64 | 92.93 | 100 | 23.45 | 549.90 | 226.06 |
| 41 | 10.64 | 113.21 | 90 | 13.45 | 180.90 | 143.11 |
| 42 | 11.64 | 135.49 | 97 | 20.45 | 418.20 | 238.04 |

$$\Sigma x_i = 334$$

$$\Sigma(x - \bar{x})^2 = 932.55$$

$$\Sigma y_i = 842$$

$$\Sigma(y - \bar{y})^2 = 4188.70$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 1743.84$$

Solution continued

$$s_x = \sqrt{s_x^2} = \sqrt{93.26} = 9.66$$

$$s_y^2 = \frac{(y_i - \bar{y})^2}{n - 1} = \frac{4188.70}{10} = 418.87$$

$$s_y = \sqrt{s_y^2} = \sqrt{418.87} = 20.47$$

Correlation coefficient:

$$r = \frac{\text{COV}_{xy}}{s_x s_y} = \frac{174.38}{9.66 \times 20.47} = 0.88$$

Regression coefficient:

$$b = \frac{\text{COV}_{xy}}{s_x^2} = \frac{174.38}{93.26} = 1.87$$

Chi-square test

- The chi-square (χ^2) test is used by plant breeders to test hypotheses related to categorical data such as would be collected from inheritance studies.
- It is defined mathematically as:

$$\chi^2 = \Sigma[(fo - fe)^2]/fe$$

In F_2 population of garden pea, 145 plants with red flower and 55 with white flower were observed. Test the validity of 3 :1 expected segregation ratio of flower colour.

Solution

For calculation of χ^2 value, the data is tabulated in the following manner.

| Segregation classes | Frequencies | | O - E | χ^2 value (O - E) ² /E |
|---------------------|--------------|-------------|-------|---|
| | Observed (O) | Expcted (E) | | |
| Red flower | 145 | 150 | -5 | 0.166 |
| White flower | 55 | 50 | 5 | 0.500 |
| Total | 200 | 20 | | 0.666 |

Calculated value of χ^2 = 0.666

Table value of χ^2 at 1 degree of freedom = 3.841

Conclusion

The calculated value of χ^2 is smaller than the table value hence the assumed 3 : 1 segregation ratio holds good.

T-test

- The *t*-test is used to make inferences about population means.
- A breeder may wish to compare the yields of two cultivars, for example.
- Assuming the sample observations are drawn at random, the two population variances are equal, and the populations from which the samples normal distribution.
- Null hypothesis
 - ▶ $H_0: \mu_1 = \mu_2$ (no difference between the two means)
- The alternative hypothesis
 - ▶ $H_1: \mu_1 \neq \mu_2$ (the two populations are not equal)

Test statistics

$$t = [\bar{X}_1 - \bar{X}_2] / s_p \sqrt{[1/n_1 + 1/n_2]}$$

where:

$$s_p = \sqrt{\{[(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / n_1 + n_2 - 2\}}$$

= pooled variance

and \bar{X}_1 and \bar{X}_2 are the means of samples 1 and 2, respectively.

Solution

Example A plant breeder wishes to compare the seed size of two navy bean cultivars, A and B. Samples are drawn and the 100 seed weight obtained. The following data were compiled:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

| | Cultivar | |
|-----------|----------|--------------------|
| | A | B |
| n | 10 | 8 |
| \bar{X} | 21.2 | 19.5 (g/100 seeds) |
| s | 1.3 | 1.1 (g/100 seeds) |

where:

$$t = \frac{[\bar{X}_1 - \bar{X}_2] / s_p \sqrt{[1/n_1 + 1/n_2]}}{s_p} = \frac{[(10 - 1)(1.3)^2 + (8 - 1)(1.1)^2] / (10 + 8 - 2)}{[9(1.69) + 7(1.21)] / 16}$$
$$= \frac{[15.21 + 8.47] / 16}{23.68 / 16}$$
$$= 1.48$$
$$t = 10 - 8$$
$$= 2 / (1.48 \times 0.47)$$
$$= 0.70$$
$$= 2 / 0.70$$
$$t(\text{calculated}) = 2.857$$

at $\alpha_{0.05}$:

$$df = 10 + 8 - 2 = 16$$

$$t(\text{tabulated}) = 1.746$$

Since calculated t exceeds tabulated t , we declare a significant difference between the two cultivars for seed size (measured as 100 seed weight).

Analysis of variance(ANOVA)

- Breeder needs to compare more than two cultivars.
- In yield trials, several advanced genotypes are evaluated at different locations and in different years.
- The *t*-test is not applicable in this circumstance but its extension, the **ANOVA**, is used instead.
- ANOVA allows the breeder to analyze measurements that depend on several kinds of effects, and which operate simultaneously, in order to decide which kinds of effects are important, and to estimate these effects.
- ANOVA is used to obtain and partition the total variation in a data set according to the sources of variation present and then to determine which ones are important.

- ‘On average’ = mean value.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k = \mu$$

$$H_a : \mu\text{'s are heterogeneous}$$

| Source of variation | df | SS | MS | <i>F</i> |
|---------------------|---------|-----------|---------------------------|----------------|
| Treatment | $k - 1$ | SS_{Tr} | $MS_{Tr} = SS_{Tr}/k - 1$ | MS_{Tr}/MS_E |
| Error | $N - k$ | SS_E | $MS_E = SS_E/N - k$ | |
| Total | $N - 1$ | SS_T | | |

$$s_1^2 = \frac{\sum_{i=1}^{n_1} (X_{i1} - \bar{X}_1)^2}{n_1 - 1}$$

⋮

$$s_k^2 = \frac{\sum_{i=1}^{n_k} (X_{ik} - \bar{X}_k)^2}{n_k - 1}$$

$$s_w^2 = \frac{n_1 s_1^2 + n_2 s_2^2 + \dots + n_k s_k^2}{N}$$

where $\sum_{j=1}^k n_j = N$

$$s_b^2 = \frac{n_1 (\bar{X}_1 - \bar{\bar{X}})^2 + n_2 (\bar{X}_2 - \bar{\bar{X}})^2 + \dots + n_k (\bar{X}_k - \bar{\bar{X}})^2}{N \cdot (k-1)} = \frac{\sum_{j=1}^k n_j (\bar{X}_j - \bar{\bar{X}})^2}{N \cdot (k-1)}$$

$$s_b^2 = MS_b = \frac{SS_b}{df_b} = \frac{SS_b}{k-1}$$

$$s_w^2 = MS_w = \frac{SS_w}{df_w} = \frac{SS_w}{N-k}$$

$$\hat{F} = \frac{s_b^2}{s_w^2} \sim F_{[v_1, v_2]}$$

where $v_1 = k-1$

$$v_2 = \sum_{j=1}^k (n_j - 1) = N - k$$

- *Example:*

| | Sample 1 | Sample 2 |
|-----------|----------|----------|
| | 5.34 | 7.46 |
| | 3.08 | 7.11 |
| | 3.46 | 6.07 |
| | 4.72 | 6.47 |
| | 5.65 | 6.05 |
| | 4.15 | 6.56 |
| | 6.20 | 3.76 |
| | | 5.58 |
| N: | 7 | 8 |
| Mean: | 4.66 | 6.13 |
| Variance: | 1.34 | 1.28 |

$$n_1 = 7, n_2 = 8, N = 15$$

$$\bar{X}_1 = 4.66, \bar{X}_2 = 6.13, \bar{\bar{X}} = 5.44$$

$$s_1^2 = 1.34, s_2^2 = 1.28, s_p^2 = 1.31$$

$$s_w^2 = 1.31, s_b^2 = 8.09$$

$$\hat{F} = \frac{s_b^2}{s_w^2} = \frac{8.09}{1.31} = 6.19 \sim F_{[1,13]}$$

$$p = 0.027$$

- Results usually presented in the form of an “ANOVA table”:

| Source of variation | Sum of squares | Degrees of freedom | Mean square (variance) | F |
|---------------------|----------------|--------------------|------------------------|-------------|
| Factor | SS_b | $k-1$ | ms_b | ms_b/ms_w |
| Error | SS_w | $N-k$ | ms_w | |
| Total | SS_t | $N-1$ | | |

- Based on the previous example:

| Source of variation | Sum of squares | Degrees of freedom | Mean square (variance) | F |
|---------------------|----------------|--------------------|------------------------|------|
| Factor | 8.09 | 1 | 8.09 | 6.19 |
| Error | 17.00 | 13 | 1.31 | |
| Total | 25.09 | 14 | | |