

# HARDY WEIBERG EQUILIBRIUM & BIOMETRY

DR. KOFI OWUSU-DAAKU

POPULATION GENETICS AND EVOLUTION

LECTURE V

# Hardy- Weinberg Law

- ▶ The Hardy-Weinberg Law is a basic concept in the population genetics developed independently in 1908 by the English mathematician, G H Hardy and German physician, W Weinberg.

► The Hardy-Weinberg law states that the gene frequencies and genotypic ratios in a natural breeding population remain constant from generation to generation if:

- (a) The population size is large.
- (b) There is no mutation.
- (c) There is no immigration or emigration.
- (d) Mating is random.
- (e) There is random reproductive success.

# Predictions of the Hardy-Weinberg Law

- ▶ 1. If the conditions are met, the population will be in genetic equilibrium, with two expected results:
- ▶ a. Allele frequencies do not change over generations, so the gene pool is not evolving at the locus under study.
- ▶ b. After one generation of random mating, genotypic frequencies will be  $p^2$ ,  $2pq$ , and  $q^2$ , and will stay constant in these proportions as long as the conditions above are met. This is Hardy-Weinberg equilibrium, which allows predictions to be made about genotypic frequencies.



- The relationship between gene frequency and genotype frequency can be described by The equation  $p^2 + 2pq + q^2 = 1$ , where  $p^2$  represents the frequency of dominant gene,  $2pq$  is the frequency of heterozygote genotype and  $q^2$  is the frequency of the homozygous recessive genotype.

- ▶ The Hardy-Weinberg law forms a theoretical base line for measuring evolutionary change as it predicts that there is no evolution.
- ▶ Whenever evolution occurs, the Hardy-Weinberg equilibrium is disturbed as a change in gene frequencies and indicates that evolution is in progress.

- ▶ The Hardy-Weinberg law is used to determine whether the number of harmful mutations in a population is increasing.
- ▶ In nature, the Hardy-Weinberg law is impossible.

# Derivation of the Hardy-Weinberg Law

- ▶ 1. Zygotes are formed by random combinations of alleles, in proportion to the abundance of that allele in the population (Figure 24.3).
- ▶ 2. When a population is in equilibrium, genotypic frequencies will be in the proportions  $p^2$ ,  $2pq$ , and  $q^2$ . This results from the expansion of the square of the allelic frequencies:  $(p+q)^2=p^2+2pq+q^2$ .
- ▶ 3. Mendelian principles acting on a population in equilibrium will work to maintain that equilibrium. Albinism is an example.



# **Mathematical proof that Gene Frequencies are Maintained from generation to generation**

# Mathematical proof -1

- ▶ In a population the relative frequencies of different alleles tend to be maintained constant from one generation to the next. This can be demonstrated mathematically and helps to explain why dominant traits do not automatically increase at the expense of recessives.

# Mathematical proof - 2

► consider one locus with two alleles A and a. If the frequency of the allele 'A' is  $p$  and the frequency of the allele 'a' is  $q$  then, since each individual must have one or other allele, the sum of these allele frequencies must be one or 100%. Therefore

►  $p + q = 1$

# Mathematical proof – 3 Table 1

		Paternal	gametes
		A	a
		(p)	(q)
	A	AA	Aa
Maternal gametes	(p)	(p <sup>2</sup> )	(pq)
	a	Aa	aa
	(q)	(pq)	(q <sup>2</sup> )



## Mathematical proof - 4

- ▶ In production of the next generation of the three types of paternal genotype may mate with each of the three types of maternal genotype (Table 2). Table 3 indicates the genotypes of the offspring for each mating type, and as can be seen the relative frequencies of each is unchanged and the population is said to be in genetic equilibrium. Although the actual numbers of individuals with each genotype may have increased, the relative proportions of each genotype (and allele) have remained constant (AA at  $p^2$ , Aa at  $2pq$  and aa at  $q^2$ ). This principle is called the Hardy Weinberg law.

# Mathematical proof - 5

Table 2		Paternal genotypes		
		AA	Aa	aa
		$(p^2)$	$(2pq)$	$(q^2)$
	AA	AA x AA	AA x Aa	AA x aa
Maternal genotype s	$(p^2)$	$(p^4)$	$(2p^3q)$	$(p^2q^2)$
	Aa	Aa x AA	Aa x Aa	Aa x aa
	$(2pq)$	$(2p^3q)$	$(4p^2q^2)$	$(2pq^3)$
	aa	aa x AA	aa x Aa	aa x aa
	$(q^2)$	$(p^2q^2)$	$(2pq^3)$	$(q^4)$

# Mathematical proof - 6

Table 3						
			Offspring			
Mating type	RESULTS	Frequency (from Table 2)	AA	Aa	aa	
AA x AA	AA,AA,AA,AA	$p^4$	$p^4$	-	-	
AA x Aa	AA,AA, Aa, Aa	$4p^3q$	$2p^3q$	$2p^3q$	-	
AA x aa	Aa, Aa, Aa, Aa	$2p^2q^2$	-	$2p^2q^2$	-	
Aa x Aa	AA,Aa, Aa, aa	$4p^2q^2$	$p^2q^2$	$2p^2q^2$	$p^2q^2$	
Aa x aa	Aa,Aa,aa,aa	$4pq^3$	-	$2pq^3$	$2pq^3$	
aa x aa	aa, aa,aa,aa	$q^4$	-	-	$q^4$	

# Mathematical proof - 7

▶ AA offspring =  $p^4 + 2p^3q + p^2q^2$

$$= p^2(p^2 + 2pq + q^2)$$

$$= p^2(p + q)^2$$

$$= p^2(1)^2 = p^2$$

▶ Aa offspring =  $2p^3q + 4p^2q^2 + 2pq^3$

$$= 2pq(p^2 + 2pq + q^2)$$

$$= 2pq$$

▶ aa offspring =  $p^2q^2 + 2pq^3 + q^4$

$$= q^2(p^2 + 2pq + q^2)$$

$$= q^2$$



## Extensions of the Hardy-Weinberg Law to Loci with More than Two Alleles

- ▶ 1. Often more than two alleles are possible at a given locus, and the frequencies of possible genotypes are still given by the square of the allelic frequencies.
- ▶ 2. If three alleles are present (e.g., alleles A, B, and O) with frequencies  $p$ ,  $q$ , and  $r$ , the frequencies of the genotypes at equilibrium will be:
  - ▶  $(p+q+r)^2 = p^2(AA) + 2pq(AB) + q^2(BB) + 2pr(AO) + 2qr(BO) + r^2(OO)$
- ▶ 3. Human blood groups is an example.

## Extensions of the Hardy-Weinberg Law to Sex-Linked Alleles

- ▶ 1. In species where sex is chromosomally determined, humans or *Drosophila* for example, females have two X chromosomes while males have only one. In females, Hardy-Weinberg frequencies are the same as for any other locus. In males, frequencies of the genotypes are the same as frequencies of the alleles in the population.
- ▶ 2. Because males receive their X chromosome from their mothers, the frequency of an X-linked allele will be the same as the frequency of that allele in their mothers. For females the frequency will be the average of both parents.

# Extensions of the Hardy-Weinberg Law to Sex-Linked Alleles

- ▶ 3. With random mating, the difference in allelic frequency between the sexes will be reduced by half in each generation . One generation after allelic frequencies become equal in males and females, the genotypes will be in Hardy-Weinberg proportions.

# Testing for Hardy-Weinberg Proportions

- ▶ 1. Data from real populations rarely match Hardy-Weinberg proportions. Use a chi-square test to check whether deviation is larger than expected by chance.
- ▶ 2. If the deviation is larger than expected, researchers begin to study which of the Hardy-Weinberg assumptions is being violated.



- ▶ If we know the frequency of each allele in a population, we can predict the genotypes and phenotypes we should see in that population. If the phenotypic frequencies in a population are not those predicted from the allele frequencies, the population is not in Hardy-Weinberg equilibrium, because an assumption has been violated.

- ▶ Either non-random mating or evolution is occurring.
- ▶ But as long as the Hardy-Weinberg assumptions are not violated, recessive alleles and dominant alleles do not change their frequencies over time.

Humans can inherit various alleles from the liver enzyme ADH (alcohol dehydrogenase), which breaks down ingested alcohol. People of Italian and Jewish descent commonly have a form of ADH that detoxifies very rapidly. People of northern European descent have forms of ADH that are moderately effective in alcohol breakdown, while people of Asian descent typically have ADH that is less effective at processing alcohol. Explain why researchers have been able to use this information to help trace the origin of Human use of alcoholic beverages.

## Extensions of the Hardy-Weinberg Law to Loci with More than Two Alleles

- ▶ 1. Often more than two alleles are possible at a given locus, and the frequencies of possible genotypes are still given by the square of the allelic frequencies.
- ▶ 2. If three alleles are present (e.g., alleles A, B, and O) with frequencies  $p$ ,  $q$ , and  $r$ , the frequencies of the genotypes at equilibrium will be:
  - ▶  $(p+q+r)^2 = p^2(AA) + 2pq(AB) + q^2(BB) + 2pr(AO) + 2qr(BO) + r^2(OO)$
- ▶ 3. Human blood groups is an example.



# Extensions of the Hardy-Weinberg Law to Sex-Linked Alleles

25

- ▶ 1. In species where sex is chromosomally determined, humans or *Drosophila* for example, females have two X chromosomes while males have only one. In females, Hardy-Weinberg frequencies are the same as for any other locus. In males, frequencies of the genotypes are the same as frequencies of the alleles in the population.
- ▶ 2. Because males receive their X chromosome from their mothers, the frequency of an X-linked allele will be the same as the frequency of that allele in their mothers. For females the frequency will be the average of both parents.
- ▶ 3. With random mating, the difference in allelic frequency between the sexes will be reduced by half in each generation. One generation after allelic frequencies become equal in males and females, the genotypes will be in Hardy-Weinberg proportions.

# Testing for Hardy-Weinberg Proportions

26

2/22/2017

- ▶ 1. Data from real populations rarely match Hardy-Weinberg proportions. Use a chi-square test to check whether deviation is larger than expected by chance.
- ▶ 2. If the deviation is larger than expected, researchers begin to study which of the Hardy-Weinberg assumptions is being violated.

# Using the Hardy-Weinberg Law to Estimate Allelic Frequencies

- ▶ 1. If one or more of the alleles is recessive, one can't distinguish between heterozygous and homozygous dominant individuals. Can use Hardy-Weinberg law to calculate the allele frequency based on information about the number of homozygous recessive individuals.



# Genetic Variation in Space and Time

28

2/22/2017

- ▶ 1. The genetic structure of populations can vary in space and time .
- ▶ 2. An allele frequency cline is a clear pattern of variation across a geographic transect, usually correlated with a physical feature such as temperature or rainfall .
- ▶ 3. Statistical tools are used to quantify spatial patterns of genetic variation. These are important in conservation biology.



# Genetic Variation in Natural Populations

- ▶ 1. Genetic variation is important in natural populations:
  - ▶ a. It determines the potential for evolutionary change and adaptation.
  - ▶ b. It provides clues about roles of various evolutionary processes.
  - ▶ c. It allows predictions about a population's chances for long-term survival.

# Measuring Genetic Variation at the Protein Level - 1

30

2/22/2017

- ▶ 1. Understanding genetic variation in a population was difficult before molecular biology because most phenotypes are the result of multiple genes, and there was no technique available to determine allele frequencies.
- ▶ 2. Protein electrophoresis (Lewontin and Hubby, 1966) separates proteins on the basis of size, charge, and conformation, and so often can separate the gene products of different alleles.

- ▶ 3. The amount of genetic variation within a population is usually measured by two parameters:
  - ▶ a. Proportion of polymorphic loci (those with more than one allele within a population). Proportion is calculated by dividing total of loci with more than one allele by total number of loci examined.
  - ▶ b. Heterozygosity (the proportion of an individual's loci that are heterozygous). Determine for individuals, and then average to obtain estimate of heterozygosity of the population.

## Measuring Genetic Variation at the Protein Level - 2

- ▶ 4. Proteins with similar sizes and charges will conform in gel electrophoresis, and so allele differences are likely to be underestimated. Even so, much more variation is seen at most loci than would be predicted by the classical model.
- ▶ 5. Kimura proposed the neutral-mutation model, saying that the combination of random mutations and chance fixation of alleles is responsible for some variations that are not functionally different and so are not acted upon by natural selection. However, DNA sequencing has undermined this model.



# Measuring Genetic Variation at the DNA Level

33

2/22/2017

- ▶ 1. PCR (polymerase chain reaction) allows amplification of DNA regions from many individuals. Fragments may be analyzed for size, restriction sites, and DNA sequence .
- ▶ 2. Restriction fragment length polymorphisms (RFLPs) are a quick way to map genes and gain an idea of how many DNA differences occur within a population . Limitation of RFLP analysis is that it only assesses variation in sites for particular restriction enzymes.

- ▶ 3. DNA sequence analysis shows that there is more variation and therefore many more alleles of most genes than previously believed. Different regions will have different levels of variation. Exons are less likely to vary than introns and flanking sequences, and changes within exons are often synonymous.
- ▶ 4. DNA length polymorphisms result from deletions and insertions of short stretches of nucleotides, especially in noncoding regions. Microsatellites or short tandem repeat polymorphisms (STRPs) are very simple repetitive sequences that occur different numbers of times in different individuals. Both are useful in determining genetic variability within a population .

# Forces That Change Gene Frequencies in Populations

- ▶ 1. Few populations are actually in Hardy-Weinberg equilibrium, and so their allele frequencies do change, and evolution occurs.

# Mutation

- ▶ 1. Usually a mutation converts one allelic form of a gene to another.
- ▶ 2. Mutations may be neutral, detrimental, or advantageous, depending on the environment. Environmental changes may favor alleles different from those previously favored.
- ▶ 3. The frequency of alleles in a population is determined by interaction of mutation rates and natural selection.



# Random Genetic Drift - 1

37

2/22/2017

- ▶ 1. Called “drift” for short. Results from random events in small populations (sampling error).
- ▶ 2. Effective population size is the number of adults contributing gametes to the next generation. Includes the number of breeding females plus the number of breeding males. (Remember that if, for example, one male contributes most of the gametes, his alleles will be present at a higher frequency in the next generation.)
- ▶ 3. The standard error of allelic frequency is a useful mathematical analysis for understanding the limits of allelic frequency.

# Random Genetic Drift - 2

- ▶ 4. Genetic bottlenecks and founder effects arise when populations expand from a small number of ancestors.
- ▶ a. Even though the population may become large, it will contain only alleles that:
  - ▶ i. Were present in the ancestors of this population.
  - ▶ ii. Developed since speciation by mutation of ancestral genes.
- ▶ b. Examples include:
  - ▶ i. The islanders of Tristan da Cunha.
  - ▶ ii. The Amish sect in the United States

# Random Genetic Drift - 3

39

- ▶ 5. **Effects of genetic drift:**
- ▶ a. Allelic frequencies will change over time, and may reach values of 0.0 or 1.0. When this occurs, the remaining allele is “fixed” in the population, and only mutation can change its frequency. This reduces the heterozygosity of the population, resulting in reduced genetic variation .

2/22/2017

- ▶ b. Individual populations will not necessarily drift in the same direction, and so genetic divergence can result. This may eventually result in speciation. Experimental evidence confirms that there is more variance in allelic frequency among small populations than among large ones.
- ▶ c. Neutral mutations are not subject to natural selection, and may be used to estimate the time elapsed since two species shared a common ancestor.



# *Balance Between Mutation and Random Genetic Drift*

41

2/22/2017

- ▶ 1. In a population, mutation adds variation, and random genetic drift removes variation. When these forces are combined, the infinite alleles model predicts that they will balance each other and a steady state of heterozygosity will result .

# Migration - 1

42

2/22/2017

- ▶ 1. While Hardy-Weinberg assumes no migration, many populations are not isolated and will exchange genes with other populations. Genetic migration is about gene movement, rather than actual movement of organisms, and is referred to as gene flow.
- ▶ 2. Gene flow has two major effects on a population:
  - ▶ a. May introduce new alleles to a population.
  - ▶ b. When migrants have different allelic frequency than recipient population, allelic frequencies will be altered in the recipient population.

# Migration - 2

- ▶ 3. Gene flow is diagrammed in Figure 24.15. Note that if gene flow continues, the differences in allelic frequencies between the populations will decrease (Figure 24.16).
- ▶ 4. The balance between drift and migration will determine whether two populations remain similar to one another. This is important in conservation biology, where fragmentation of habitats may prevent gene flow.

How does migration cause change in allele frequencies?



Allele freq. before migration:

*Did allele frequencies change?*

Allele freq. after migration:

*Is the population in  
Hardy-Weinberg equilibrium?*

$$\begin{aligned}
 p^2 &= (0.95)^2 = 0.9025 \\
 2pq &= 2(0.95)(0.05) = 0.0950 \\
 q^2 &= (0.05)^2 = 0.0025
 \end{aligned}$$

*What if migration continued  
over many generations?*



Migration makes population more similar

AlleleA1 simulation – one-way migration (gene flow)

Population 1 (“island”)



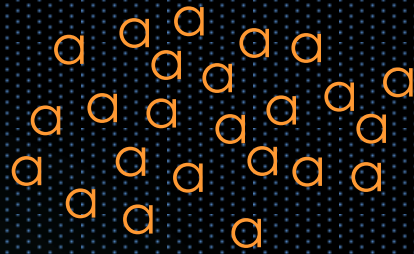
Population 2 (“mainland”)



Migration makes population more similar

Allele A1 simulation – one-way migration (gene flow)

Population 1 (“island”)

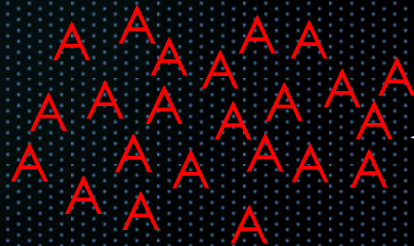


Population 2 (“mainland”)



Real life – gene flow can be one-way or two-way

Population 1



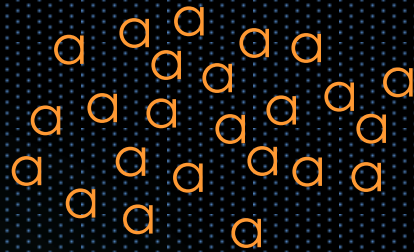
Population 2



Migration makes population more similar

Allele A1 simulation – one-way migration (gene flow)

Population 1 (“island”)

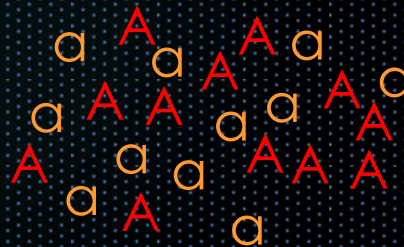


Population 2 (“mainland”)

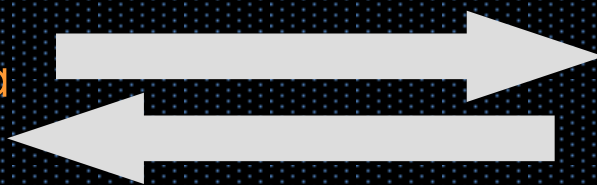
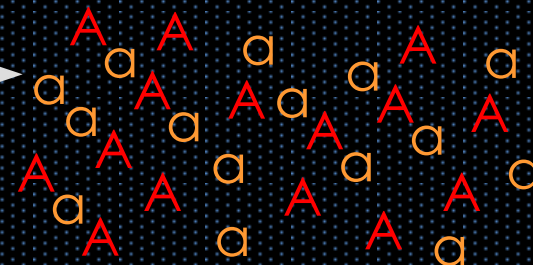


Real life – gene flow can be one-way or two-way

Population 1



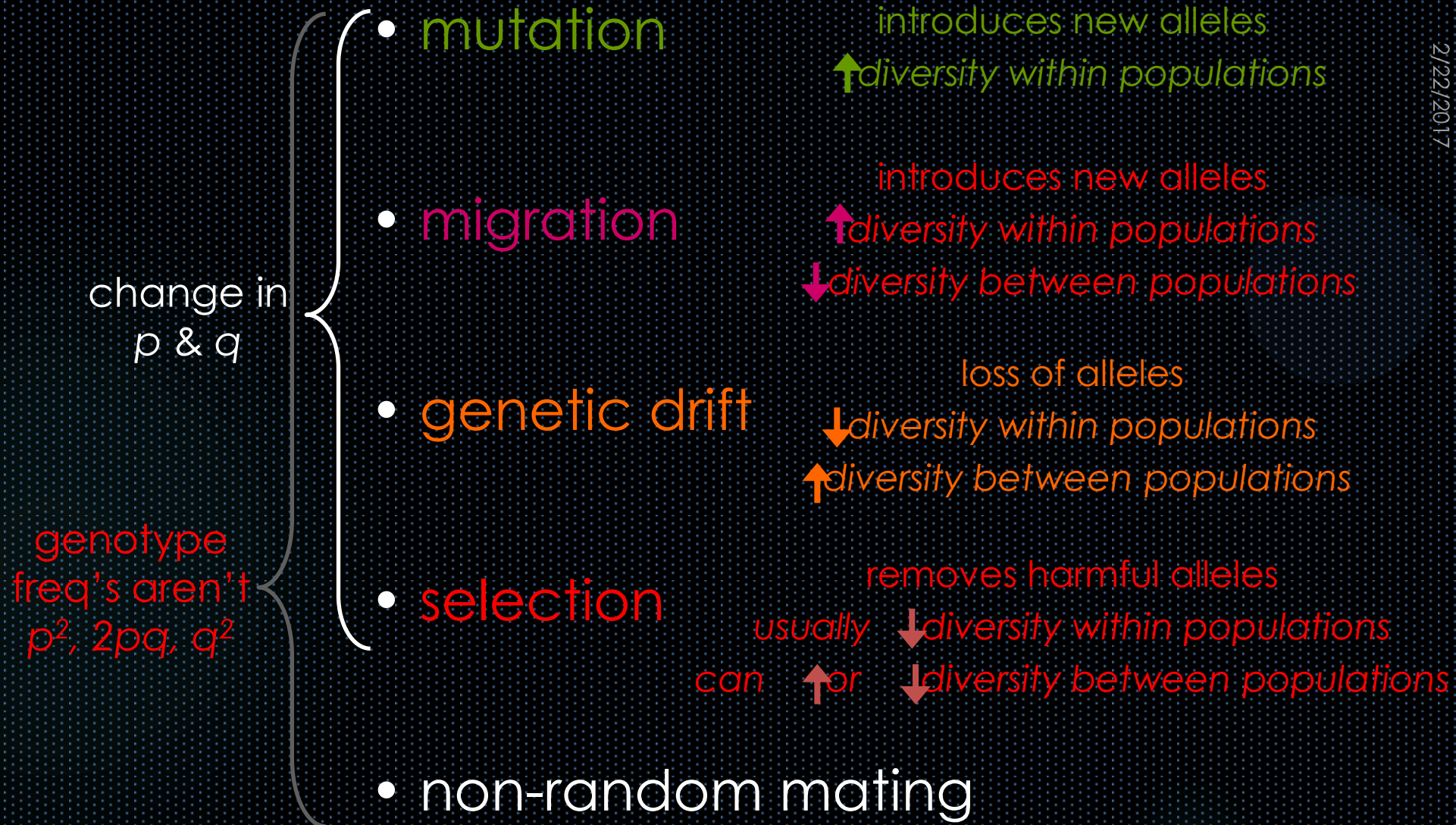
Population 2



# What can change population genetic structure?

48

2/22/2017





# Population genetic forces can interact

49

2/22/2017

- mutation

introduces new alleles  
↑diversity within populations

- migration

- genetic drift

- selection

removes harmful alleles  
usually ↓diversity within populations  
can ↑or ↓diversity between populations

- non-random mating

# Population genetic forces can interact

50

## mutation vs. selection

introduces new alleles  
↑ diversity within populations

removes harmful alleles  
↓ diversity within populations

### Mutation-selection balance

recurrent mutations offset removal by selection

more mutations  
per generation → higher frequency of mutant allele at equilibrium

stronger selection  
against mutant allele → lower frequency of mutant allele at equilibrium

2/22/2017

# Population genetic forces can interact

51

2/22/2017

- mutation

- migration

introduces new alleles  
↑ diversity within populations  
↓ diversity between populations

- genetic drift

- selection

removes harmful alleles  
usually ↓ diversity within populations  
can ↑ or ↓ diversity between populations

- non-random mating

# Population genetic forces can interact

52

## migration vs. selection



2/22/2017

### Balance between migration and selection

- input from migration offsets removal by selection
- homogenizing force of migration offset by diversifying force of selection

more migration per generation → higher frequency of migrant allele at equilibrium; populations become similar

stronger selection against migrant allele → lower frequency of migrant allele at equilibrium; populations remain distinct



# Population genetic forces can interact

53

2/22/2017

- mutation

- migration

introduces new alleles  
↑diversity within populations  
↓diversity between populations

- genetic drift

loss of alleles  
↓diversity within populations  
↑diversity between populations

- selection

- non-random mating

# Population genetic forces can interact

54

2/22/2017

## migration vs. drift

introduces new alleles  
↑diversity within populations  
↓diversity between populations

loss of alleles  
↓diversity within populations  
↑diversity between populations

### Balance between migration and drift

- input from migration offsets removal by drift
- homogenizing force of migration offset by diversifying force of drift

more migration  
per generation → higher frequency of migrant allele at equilibrium;  
populations become similar

smaller  
population size → drift affects frequency of migrant allele (usually lost);  
populations remain distinct

*Balance depends on population size*

# Population genetic forces can interact

55

2/22/2017

- mutation
- migration
- **genetic drift**
  - loss of alleles
  - ↓ *diversity within populations*
  - ↑ *diversity between populations*
- **selection**
  - removes harmful alleles
  - usually ↓ *diversity within populations*
  - can ↑ or ↓ *diversity between populations*
- non-random mating

# Population genetic forces can interact

56

## selection vs. drift

removes harmful alleles	loss of alleles
↓diversity within populations	↓diversity within populations
can ↑or ↓diversity between populations	↑diversity between populations

### Balance between selection and drift

- random force of drift can oppose selection against deleterious allele
- drift opposes adaptation

stronger selection against deleterious allele → lower frequency of deleterious allele at equilibrium; populations become adapted

smaller population size → drift affects frequency of deleterious allele (may be kept); populations drift

*Balance depends on population size*

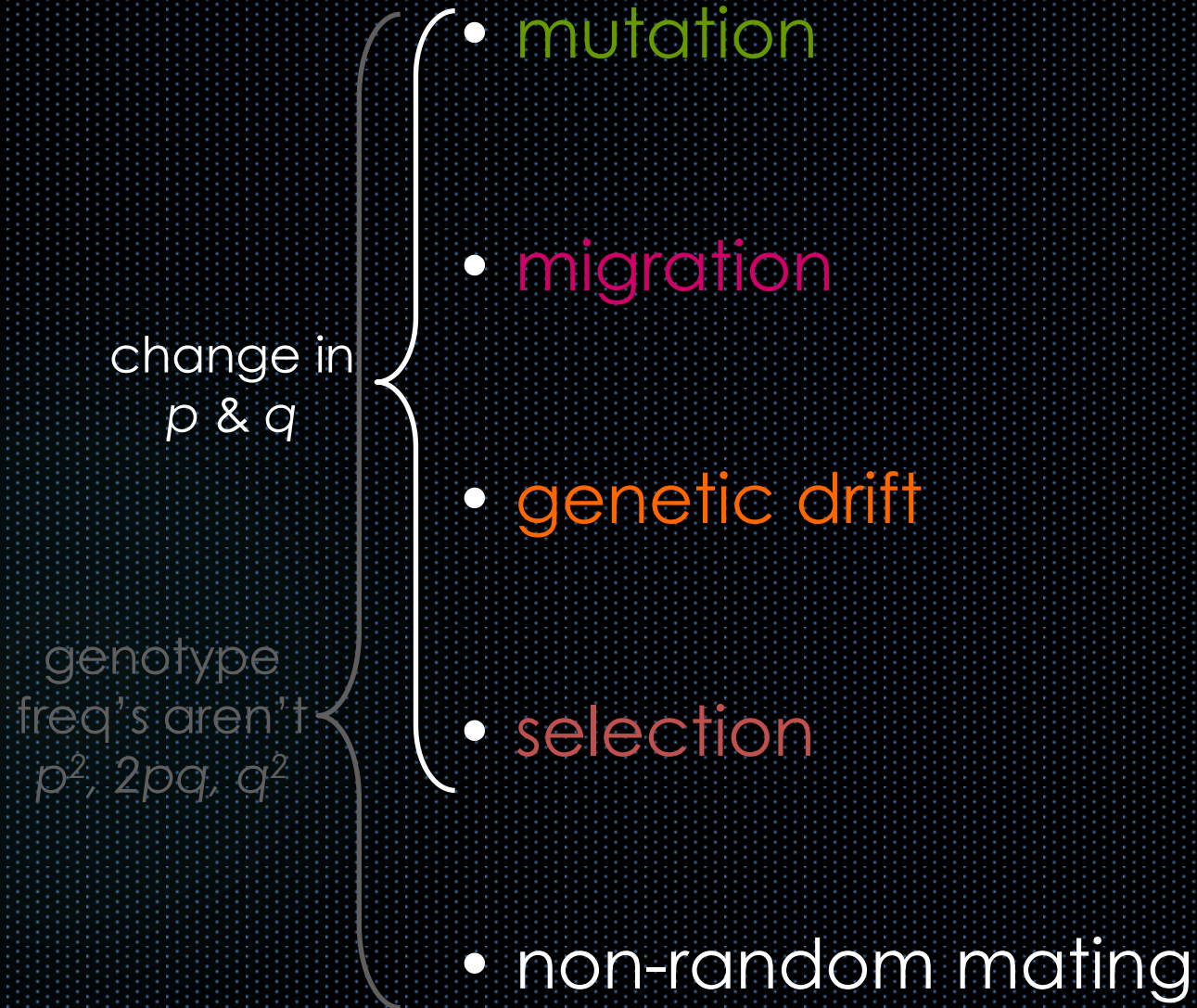
2/22/2017



# What can change population genetic structure?

57

2/22/2017



# Non-random mating can change population genetic structure

58

2/22/2017

## Non-random mating

- does *not* change allele frequencies
- does cause genotype frequencies to differ from  $p^2$ ,  $2pq$ ,  $q^2$

## Types of non-random mating

Assortative mating – mates are genetically similar

ex: *inbreeding* – mating between close relatives  
includes self-pollination

Disassortative mating – mates are genetically different

# Natural Selection -1

- ▶ 1. Adaptation is the process by which traits evolve that increase the organism's chances of surviving to reproduce. Adaptation is mainly the result of natural selection
- ▶ 2. Natural selection is the differential survival of genotypes, and the alleles that survive are more likely to be represented in the next generation. Over time, this increases the adaptation of organisms to their environment .

- ▶ 3. Darwinian fitness is the relative reproductive ability of a particular genotype. It involves both the number of offspring, and their relative fitness.
- ▶ 4. Natural selection may result in increasing or decreasing genetic variation, depending on environmental conditions. It can be calculated using the “table method”. The premise is that the contribution of each genotype to the next generation will be equal to the initial frequency of the genotype multiplied by its fitness.



# Natural Selection -2

61

- ▶ 5. Recessive traits often result in reduced fitness , and so there will be selection against homozygous recessives, decreasing the frequency of the recessive allele, but not eliminating it from the population, because as the allele becomes less frequent it is more likely to be found in a heterozygote, creating a protected polymorphism.
- ▶ 6. Some forms of selection result in maintaining genetic variation. If a heterozygote has higher fitness than either of the homozygotes, allelic frequencies will reach equilibrium and become stable. The relationship between sickle-cell anemia and resistance to malaria is a famous example

# Balance Between Mutation and Selection

62

2/22/2017

- ▶ 1. The balance between mutations and natural selection results in evolution.
- ▶ 2. When an allele becomes rare, its change in frequency with each generation becomes very small. At the same time, mutations occur that produce new alleles and increase the frequency. Eventually equilibrium will be obtained.
- ▶ 3. Selection is continually acting on dominant alleles, but can act only on recessive ones in homozygotes. Therefore, detrimental dominant alleles are generally less common than those that are recessive.

# Assortative Mating

63

2/22/2017

- ▶ 1. Individuals do not always mate randomly. When a particular phenotype is preferred in mates, allele frequencies will be affected.
- ▶ 2. Positive assortative mating occurs when individuals with similar phenotypes mate preferentially.
- ▶ 3. Negative assortative mating occurs when phenotypically dissimilar individuals mate preferentially.



# Inbreeding

- ▶ 1. Inbreeding is the preferential mating between close relatives.
- ▶ 2. Small populations will show this effect even if there is no tendency to select relatives, because even-chance matings are likely to involve relatives.
- ▶ 3. Self-fertilization is an extreme case of inbreeding seen in many plants and a few animals.
- ▶ 4. Inbreeding has results similar to genetic drift in a small population. Heterozygosity decreases and homozygosity increases. In large populations, inbreeding will result in constant allele frequencies even though homozygosity increases.



# Effects of Evolutionary Forces on the Genetic Structure of a Population - 1

65

2/22/2017

- ▶ **Changes in Allelic Frequency Within a Population**
- ▶ 1. Factors with potential to change allelic frequencies are:
  - ▶ a. Mutation, although it occurs at a slow rate and creates a relatively small change in allelic frequencies.
  - ▶ b. Migration, which may create major changes in allelic frequencies.
  - ▶ c. Genetic drift, which produces substantial changes in small populations.

- ▶ d. Selection, which alters frequencies and continues to act even when equilibrium has been reached.
- ▶ 2. Nonrandom mating affects genotypic frequencies in a population. Inbreeding increases homozygosity, and decreases fitness if deleterious recessive alleles are present

# Effects of Evolutionary Forces on the Genetic Structure of a Population - 2

67

2/22/2017

- ▶ **Genetic Divergence Among Populations**
- ▶ 1. Genetic drift can produce divergence among populations.
- ▶ 2. Migration will decrease divergence by encouraging gene flow between populations.
- ▶ 3. Natural selection can either increase divergence by favoring different alleles in different populations, or decrease divergence by selecting against certain alleles.
- ▶ 4. Nonrandom mating may contribute to the effects of other processes by altering the effective population size.

# Effects of Evolutionary Forces on the Genetic Structure of a Population - 3

- ▶ **Increases and Decreases in Genetic Variation Within Populations**
- ▶ 1. Effects are similar to those among populations (above):
  - ▶ a. Migration increases genetic variation.
  - ▶ b. Mutation also increases variation.
  - ▶ c. Genetic drift decreases variation due to loss of alleles.



- ▶ D. Inbreeding decreases variation.
- ▶ e. Natural selection may either increase or decrease genetic variation, depending on the environment and other circumstances.
- ▶ f. All of these effects combine within a population in complex ways to determine an overall pattern of genetic variation.

# The Role of Genetics in Conservation Biology

70

2/22/2017

- ▶ 1. Human activities are reducing the available habitat and altering the environment for many species, resulting in declining numbers and escalating rates of extinction. Intelligent conservation efforts must consider survival of gene pools, as well as survival of individuals.
- ▶ 2. Population viability analysis is used to determine how large a population needs to be to prevent extinction within a set period of time. An adequate gene pool is needed to ensure the potential for the population to evolve over time.

- ▶ 3. Inbreeding has occurred in zoos and game management programs, and is now a recognized concern in developing conservation strategies.
- ▶ 4. Until habitat destruction is addressed, genetic conservation efforts can only slow the depletion of gene pools and the loss of species that will inevitably result.

# BIOMETRY



# MEASURES OF CENTRAL TENDENCY

73

2/22/2017

- The term 'measures of central tendency' refers to the determination of mean, mode and median.
- The properties of large collected statistical data are difficult to understand without further treatment.
- ▶ The vast statistical data are condensed in such a way that the basic character of the data does not change

# Measures of Central Tendency Cont'd

74

- ▶ An average reduces the large number of data/observations to one figure.
- ▶ The average is a number indicating the central value of a group of observations.
- ▶ The average value of any characteristics is the one central value around which lie other observations.

11/22/2017

# Measures of Central Tendency Cont'd

75

- ▶ Thus, 'average' is a general term that describes the centre of observations.
- ▶ Three common types of average are mean, median and mode

2/22/2017

# Mean

76

2/22/2017

- ▶ Mean is the sum of all observations divided by the number of observations.
- ▶ It is the most common measure of the central tendency.
- ▶ It is the best known and most useful form of average.
- ▶ The method of calculation of arithmetic mean depends upon the nature of data available, which may be explained as follows:



# Mean cont'd

77

- (a) When the observations are small in size (i.e., in a series of individual observations):

$$\text{Mean or } \bar{X} = \frac{\sum x}{N}$$

where  $X$  = Data (values of variable) and  
 $N$  = Number of observations.

- (b) Calculation of arithmetic mean in a discrete series:

$$\text{Mean or } \bar{X} = \frac{\sum fx}{N} \text{ or } \frac{\sum fx}{\sum f}$$

where  $f$  = Frequency,  
 $x$  = The concerned variable and  
 $N$  or  $\sum f$  = Total number of observations.

(c) Calculations of arithmetic mean in a continuous series:

$$\text{Mean or } \bar{X} = \frac{\sum fm}{N}$$

where

$\sum fm$  = Total of the frequency of each class multiplied with the mid value of respective class and

$N$  = Total of the frequencies.

# Median

79

9/22/2017

- ▶ When all the observations of a variable are arranged in either ascending or descending order, the middle observation is known as median.
- ▶ Median is neither based on the total nor is it affected by the extreme values of variables.
- ▶ Median is a point, not a score or any particular measurement.

# Calculation of Median

(i) For individual observations:

Median or  $M = \text{Sum of the } N + 1/2\text{th item}$   
where  $N = \text{Number of items}$ .

(ii) For Discrete Series

For calculating median in a discrete series, frequency is made cumulative and then median is calculated on the basis of above formula.



# Calculation of Median cont'd

81

2/22/2017

(iii) For Continuous Series

After making the frequencies cumulative, the median item is found out as  $N/2$ th item and then the median is calculated as per the following formula:

# Calculation of Median cont'd

82

$$M = l_1 + \frac{i}{f} (m - c)$$

or

$$M = l_1 + \frac{\frac{N}{2} - c}{f} \times i$$

where

$M$  = Median,

$l_1$  = Lower limit of the median class,

$i$  = Class interval of the median class,

$f$  = Frequency of the median class,

$c$  = Cumulative frequency of the class preceding the median class and

$m$  =  $N/2$ th item.

# MODE

83

2/22/2017

- ▶ The mode may be defined as the observation with the highest frequency.
- ▶ This is a value that occurs most frequently in a statistical distribution.
- ▶ Normally, mode is frequently used for categorical data.
- ▶  $\text{Mode} - \text{Median} = 2 (\text{Median} - \text{Mean})$  or
- ▶  $\text{Mode} = \text{Mean} = 3 (\text{Median} - \text{Mean})$

## Calculation of Mode

- ▶ For individual series - After converting the data into discrete series, the modal item should be picked up as the most occurring value.
- ▶ For discrete series - Mode can be located simply by inspection of the series, i.e., the size having the highest frequency will be mode of that series.



# Calculation of Mode cont'd

85

- ▶ For continuous series - In a distribution of grouped data, the mode is estimated at the midpoint of the class interval having the greatest frequency.
- ▶ Mode can be calculated in a continuous series by the following formula

2/25/2017

# Calculation of Mode cont'd

86

$$MO = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

or

$$MO = \frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)} \times i$$

where

$l_1$  = Lower limit of the modal class,

$f_1$  = Frequency of the modal class,

$f_0$  = Frequency of the class preceding modal class,

$f_2$  = Frequency of the class following the modal class and

$i$  = Class interval of the modal class.

## DISTRIBUTION

- ▶ Frequency distribution is of two types, viz., observed frequency distribution and expected frequency distribution.
- ▶ Observed frequency distribution is prepared on the basis of actual data, whereas expected frequency distribution is a theoretical one.

# Distribution cont'd

88

2/22/2017

- ▶ Calculations of theoretical distribution are useful in many ways, such as to understand the risk and uncertainty in any event, helps in forecasting, serves as benchmarks for comparison, etc.



## Types of Distribution

- ▶ There are different types of theoretical frequency distribution, but the following three are of great importance:

### 1. Binomial Distribution

- ▶ It is also known as Bernoulli's distribution.
- ▶ It is identified by the number of the observations,  $n$ , and the probability of occurrence which is denoted by  $p$ .

# Binomial Distribution Cont'd

90

► The essential features of this distribution are as follows:

- (a) The number of trials is fixed.
- (b) There are two mutually exclusive possible outcomes of each trial.
- (c) The trials are independent

# Binomial Distribution Cont'd

91

2/22/2017

The binomial distribution is used when a researcher is interested in the occurrence of the events and not in its magnitude.

- This distribution is widely used in industries for quality control.

## 2. Poisson Distribution

92

- ▶ Poisson distribution was developed by the French mathematician, Simeon Denis Poisson (1837).
- ▶ It is a very useful probability distribution.
- ▶ Poisson distribution gives the idea of probability of rare events, i.e., the number of trials is very small and the probability of success is also very small.

2/28/2017



## 2. Poisson Distribution Cont'd

93

2/22/2017

- ▶ Poisson distribution is a discrete distribution with a single parameter, i.e., the mean of distribution.
- ▶ It is widely used in insurance, spread of diseases, physiology and genetics.

# 3. Normal Distribution

94

- ▶ The pattern of distribution of data that follows the bell-shaped curve is known as normal distribution.
- ▶ Normal distribution was used by mathematicians de Moivre and Laplace in the 1700s. German mathematician and physicist, Karl Gauss, used it to analyze astronomical data, hence it is also known as Gaussian distribution.

# 3. Normal Distribution cont'd

95

2/22/2017

- ▶ All normal distribution is symmetric.
- ▶ Normal distribution is the most useful theoretical distribution for continuous variables.
- ▶ The shape of normal distribution resembles the bell, so sometimes it is also referred to as the bell curve.
- ▶ It is the most frequently used of all probability distributions

- The general equation that describes normal distribution curve is as follows:

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

where

$N$  = Number of measures,

$Y$  = Frequency,

$\pi$  = 3.1416,

$\sigma$  = 2.7183,

$\sigma$  = Standard deviation of the distribution and

$x$  = Deviation of any unit of measurement from the mean.

- Biological distribution is generally assumed as normally distributed.



# CORRELATION

- ▶ Correlation is a statistical technique showing relationship between two variables.
- ▶ It is one of the most common and most useful statistics.
- ▶ The possible correlations range from +1 to -1.
- ▶ A zero correlation indicates that there is no relationship between variables.

# Correlation cont'd

98

2/22/2017

- ▶ A correlation of  $-1$  indicates that if one variable increases the other decreases, while a correlation of  $+1$  indicates that both variables move in the same direction.
- ▶ Further, it shows the closeness or degree of relationship between the variables.
- ▶ Correlation is also a marker of interdependence between two variables

# Types of Correlation

99

2/22/2017

- ▶ On the basis of nature of relationship between the variables, correlation may be of the following types:
  - (a) Positive or negative
  - (b) Simple, partial or multiple
  - (c) Linear or non-linear

# Degree of Correlation

100

2/22/2017

► On the basis of coefficient of correlation, the degree of correlation may be of the following types:

(a) Perfect , (b) Limited and (c) Absent

► The degree of relationship between two variables is the coefficient of correlation represented by the symbol 'r'.

► It is called Karl Pearson's coefficient of correlation and is most widely used.



# Calculation of Karl Pearson's Coefficient of Correlation

- For individual series

$$r_{xy} = \frac{\sum xy}{\sqrt{\sum x^2 \cdot \sum y^2}}$$

where

$r$  = Correlation coefficient,

$x$  = Deviation of X variables ( $X - \bar{X}$ ) and

$y$  = Deviation of Y variables ( $Y - \bar{Y}$ ).

$y$  = Deviation of  $y$  variables ( $y - \bar{y}$ ).

## Product-Moment Correlation

- When number of observation ( $N$ ) is small, their correlation can be calculated by product-moment method, according to the formula given below:

$$\frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left[ \sum x^2 - \frac{(\sum x)^2}{N} \right] \times \left[ \sum y^2 - \frac{(\sum y)^2}{N} \right]}}$$

$$\Sigma f dx dy - N \left( \frac{\Sigma f dx}{N} \right) \left( \frac{\Sigma f dy}{N} \right)$$

---


$$\sqrt{\left[ \frac{\Sigma f dx^2}{N} - \left( \frac{\Sigma f dx}{N} \right)^2 \right] \times \left[ \frac{\Sigma f dy^2}{N} - \left( \frac{\Sigma f dy}{N} \right)^2 \right]}$$

## REGRESSION ANALYSIS

- ▶ Regression analysis is the technique for the prediction of the relationship of a particular variable with another, on the basis of its relationship with a third variable.
- ▶ The variable to be estimated is called the dependent variable and the variable that provides the basis for estimation is called the independent variable.



## Regression Analysis cont'd

105

- ▶ In multiple regression, there are two or more independent variables and one dependent variable.
- ▶ In other words, from regression analysis, we can estimate the value of one variable from the given value of the other variable. For example, we can find out the expected weight of a fish from a given length.

2/22/2017

## Regression Analysis cont'd

- ▶ The relationship between the independent variable ( $X$ ) and the dependent variable ( $Y$ ) is expressed regression equation. The regression equation expresses the regression lines.

# Regression Analysis cont'd

107

- ▶ Since there are two regression lines, there are two regression equations.
- ▶ The regression equation  $X$  on  $Y$  shows the variation in the values of  $X$  for changes in  $Y$ .
- ▶ Likewise, regression equation  $Y$  on  $X$  describes the variation in the values of  $Y$  for changes in  $X$ .

2/22/2017

# Regression Analysis cont'd

108

2/22/2017

- ▶ Regression equation:
- ▶  $X = a + by$  (x on y)  
 $Y = a + bx$  (y on x)
- ▶ where 'a' is a constant (the point where regression line touches (Y-axis) and 'b' is also a constant call regression coefficient.



# Regression Analysis cont'd

109

2/22/2017

- ▶ The multiple regression equation shows the effect of a number of independent variables at the same time which may be written as follows:
- ▶  $Y_c = a + b_1x_1 + b_2x_2 + b_3x_3$  .
- ▶ Where  $Y_c$  = Value of dependent variable  
 $x_1, x_2, x_3, \dots$  = Independent variable  
 $b_1, b_2, b_3, \dots$  = Regression coefficient

# HYPOTHESIS TESTING AND TEST OF SIGNIFICANCE

110

- ▶ The test of significance is used by the researchers to determine whether the difference between calculated value and the hypothetical parameter is significant or not.
- ▶ It establishes whether there is relationship between variables or the observed values have been produced by the chance.

2/22/2017

# Hypothesis Testing And Test Of Significance Cont'd

111

- ▶ The phrase test of significance was coined by R A Fischer (1925).
- ▶ Every test of significance is associated with a basic concept known as the hypothesis.
- ▶ The hypothesis is basically a statement about the population parameters.
- ▶ It can be grouped into two types, viz., null hypothesis and alternative hypothesis

9/22/2017

# Hypothesis Testing And Test Of Significance Cont'd

112

3/22/2017

- ▶ Statistical inferences are drawn on the basis of information we get from the sample
- ▶ In other words, it is possible to make reasonable estimates from the sample data available.
- ▶ Even if we don't know about a population, we can get reliable information about it on the basis of random sample from that population



# Hypothesis Testing And Test Of Significance Cont'd

- The estimation deals with the methods by which population parameter/characteristics are estimated from sample information, whereas hypothesis testing deals with the process involved in the acceptance or non-acceptance of the assumption or a statement about the population parameter.

# Hypothesis Testing And Test Of Significance Cont'd

- ▶ Hypothesis testing enables us to verify whether or not such statements are in agreement with the available data.

## Null and Alternative Hypothesis

- ▶ The hypothesis to be tested is called 'Null Hypothesis' and is represented by  $H_0$ . This may be written as follows:
- ▶  $H_0: \mu - x = 0$  [where  $x$  = Sample mean and  $\mu$  = Population mean]
- ▶ From the above equation, it can be concluded that there is no difference between the population mean and sample mean.

# Null and Alternative Hypothesis

cont'd

- ▶ Null hypothesis must be tested.
- ▶ To test the null hypothesis, there is an alternative hypothesis represented as  $H_1$ .
- ▶ If this alternative hypothesis is correct, the null hypothesis is rejected.



## Errors in Testing of Hypothesis

► Since the acceptance or rejection of null hypothesis ( $H_0$ ) depends on sample study, there is every chance of error. The error may be:

1. Type I ( $\alpha$ ) error - to reject null hypothesis when it is true.

2. Type II ( $\beta$ ) error - to accept null hypothesis when it is false.

# Level of Significance

118

2/22/2017

- ▶ The probability of committing a  $\alpha$ -error is called level of significance.
- ▶ 5 per cent (0.05) and 1 per cent (0.01) are the most commonly used levels of significance.
- ▶ 5 per cent level of significance shows that out of 100 times, there is a probability that 5 times correct  $H_0$  will be rejected.

# Test of Significance

- ▶ An assessment of significance of difference between parameters of different samples is known as the test of significance.
- ▶ Such a test gives an idea whether observed differences between two samples are significant or have occurred due to chance.

# STANDARD ERROR OF MEAN

- ▶ The standard deviation of the sample means is called the standard error of mean.

- The standard deviation of the sample means is called the standard error of mean.

$$\text{S.E. } \bar{X} = \sqrt{\frac{\sum(\bar{x} - \mu)^2}{N - 1}}$$

where  $\mu$  = Mean of the sample mean and  $X$  = Sample mean.

- When the standard deviation of a population is known,

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



# Standard Error Of Mean (Sex)

121

## Cont'd

- ▶ A small value of standard error of mean is a clear indication of the fact that the various values of  $X$  are close to each other and average difference between these  $X$ s and  $\mu$  is small.
- ▶ As the sample size increases, the standard error of mean becomes smaller. At the same time, on increasing the sample size, various sample means become more uniform.

# Standard Error Of Mean (Sex)

## Cont'd

122

2/22/2017

- ▶ Standard error is useful in testing a given hypothesis.
- ▶ It gives an idea about unreliability of a sample.

# STANDARD ERROR OF STANDARD DEVIATION

123

2/22/2017

- ▶ Standard deviation of different samples of the same population varies.
- ▶ So the standard error of standard deviation can be calculated to test the significance.
- ▶ SE of standard deviation data can be calculated as follows:

# STANDARD ERROR OF STANDARD DEVIATION

124

2/22/2017

- SE of standard deviation data can be calculated as follows:

$$SE \sigma = \frac{SD}{\sqrt{2N}}$$

For grouped data, it can be calculated by the following formula:

$$SE \sigma = \frac{SD}{\sqrt{2\Sigma f}}$$



# STUDENT t-TEST

125

2/22/2017

- ▶ Student t-test is a small sample test.
- ▶ Student t-test was developed by W S Gosset (1908).
- ▶ Gosset published his work in pseudonym 'Student' in 1908.
- ▶ It is the most common statistical technique used to test the hypothesis based on difference between sample means.

It is also called t-ratio because it is a ratio of difference between two means.

$$t = \frac{\bar{X} - \mu}{S / \sqrt{N}}$$

where

$\bar{X}$  = mean of the sample,

$S$  = Standard deviation of the sample and

$N$  = Number of observations in the sample.

## STUDENT t-TEST Cont'd

- ▶ A conclusion based on t-test is good if the distribution is normal or near normal and samples are chosen randomly.
- ▶ Fisher's table gives the highest obtainable values of 't' under different probabilities, with (P ) in decimal fractions corresponding to the degrees of freedom.

## STUDENT t-TEST Cont'd

128

2/22/2017

- ▶ Probability of occurrence of any calculated value of 't' is determined by comparing it with the value given in the table.
- ▶ If the calculated 't' value exceeds the value given in the table, it is said to be significant.



# Application of the t - Test

129

- (a) Student 't' test for single mean is used to test a hypothesis on specific value of the population mean.
- (b) Student t-test is used to test the difference between the means of two samples.
- (c) The paired t-test is applied when the two samples are dependent.
- (d) A t-test is used to test the significance of an observed correlation coefficient.
- (e) A t-test is used for testing significance of regression coefficient.

2/29/2017

# CHI-SQUARE ( $\chi^2$ ) TEST

130

2/22/2017

- ▶ Chi-square test is the most commonly used method for comparing frequencies.
- ▶ It is a statistical test that is used to measure difference between an observed data with the data we would expect according to a given hypothesis.
- ▶ Chi-square is calculated on the basis of frequencies in a sample.

# CHI-SQUARE ( $\chi^2$ ) TEST CONT'D

131

- ▶ It is used as a test of significance when the data are in forms of frequencies or percentages or proportions.
- ▶ It is one of the simplest and widely used non-parametric tests in statistical analysis.
- ▶ Chi-square test compares the observed value with the expected value and find out how far the differences between the two values can be attributed to fluctuations of simple sampling.

2/22/2017

# CHI-SQUARE ( $\chi^2$ ) TEST CONT'D

132

- ▶ The Chi-square test was developed by Prof. A R Fischer (1870) and it was further developed by Karl .
- ▶ Pearson (1906) in its present form.
- ▶ The following are the essentials to apply  $\chi^2$  test:
  - ▶ Random sample
  - ▶ Qualitative data



# CHI-SQUARE ( $\chi^2$ ) TEST CONT'D

133

- ▶ Lowest expected frequency not less than 5

- Chi-square can be calculated by the following formula :

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where

O = Observed number of trails and

- ▶ E = Expected number of trails.

# CHI-SQUARE ( $\chi^2$ ) TEST CONT'D

134

2/22/2017

- ▶ In  $\chi^2$  test, the number of degrees of freedom is equal to the number of classes minus one.
- ▶ The value of  $\chi^2$  depends on the degrees of freedom.
- ▶  $\chi^2$  test is also applied as a test of goodness of fit as it shows the closeness of observed and expected frequency.

# Characteristics of $\chi^2$ test

135

2/22/2017

- (a) It is based on frequencies.
- (b) It is non-negative.
- (c) It is highly skewed.
- (d) It is based on degrees of freedom.

# Characteristics of $\chi^2$ test

136

2/22/2017

(e) With the change in degree of freedom, a new chi-square distribution is created.

(f) The shape of chi-square distribution does not depend on the size of sample. It may depend upon the number of categories.



## Uses

137

2/22/2017

(a) A chi-square test is used as a test of homogeneity. It is a test which is used to determine whether several populations are similar or equal or homogenous in some characteristics.

(b) Chi-square test is used as test of independence. With the help of chi-square test, one can be able to know whether two attributes are associated or not.

## Uses Cont'd

138

2/22/2017

(c) Chi-square test as a test of goodness of fit is used to determine whether the sample data are in consistent with the hypothesized data.

# ANALYSIS OF VARIANCE

139

2/22/2017

- ▶ To test the hypothesis whether the means of several samples have significant differences or not, a method called analysis of variance is used.
- ▶ This method is based on the comparison of variances estimated from various sources.
- ▶ The analysis of variance is based on the following assumptions:
  - (a) Populations are normally distributed.

# ANALYSIS OF VARIANCE CONT'D

- ▶ (b) Populations from which the samples have been taken have means ( $\mu_1, \mu_2, \mu_3$  etc.) and variances ( $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 \dots$ )
- ▶ (c) Samples have been randomly selected.
- ▶ F-test is widely used in the analysis of variance and is calculated as follows:



# ANALYSIS OF VARIANCE CONT'D

141

2/22/2017

►  $F = \text{variance between samples} / \text{variance within samples}$

The analysis of variance is mainly of the following two types:

- (i) One-way analysis of variance
- (ii) Two-way analysis of variance

# I. One-way Analysis of Variance

- ▶ Here, analysis of variance observations are grouped on the basis of single criterion, i.e., the influence of only one factor is considered.
- ▶ In this type of analysis of variance, samples have been taken from normal populations with common variance.

## II. Two-way Analysis of Variance

143

2/22/2017

- ▶ Here we have to take consideration of the influence of two factors.
- ▶ The data are grouped according to the two different factors