ANSC 446 / IB 416 Population Genetics
Exam 1, Sept. 25, 2009

Name _____

(5 pages) Please round decimals to 4 significant digits. Show your work or describe your logic to earn partial credit for incomplete answers. Indicate your answers.

(9) 1. Coat color in horses is determined by multiple alleles. A completely Black horse (C1 black horse with black mane and tail) is dominant to a Bay horse (C2 brown horse with black legs, mane and tail) and a Mahogany Bay. A Bay is dominant to a Mahogany Bay (C3 brown horse with black roots, legs, mane, and tail). Your sample has 2000 horses (Black, Bay, and Mahogany Bay).

Color	Observed Number
Black	720
Bay	560
Mahogany Bay	720

Estimate the allele frequencies for C1, C2, and C3.

Hint: draw a unit square

P(C3) = square root of 720/2000 = **0.6**

P(C2) = (square root of 1280/2000) – P(C3) = 0.8-0.6 = **0.2**

P(C1) = 1 – P(C2) – P(C3) = 1 – 0.6 – 0.2 = **0.2**

(8) 2. Give the best definition for the following terms (2 points each):

Locus: The position on a chromosome occupied by a particular gene or other DNA sequence

Hemizygous: Having a single copy of a gene in a haplo-diploid organism or chromosome, e.g., X-linked gene in a human male

Bonferroni correction: when testing n hypotheses on a set of data, using for each individual hypothesis a statistical significance level of 1/n times what it would be if only one hypothesis were tested

Autosome: a chromosome that is not a sex chromosome; in humans all nuclear chromosomes except the X and Y chromosomes are autosomes; and in humans the autosomes are diploid.

(15) 3. Assume that the following mtDNA sequences were found in four different individuals sampled from a population.

```
GATGGAGACTTTAGT
GTTGCAGATTTAAGA
GATCGAGACTTTAGT
GTTGCAGATTTAAGA
```

(3) a) How many sites are segregating? 6

(3) b) What proportion of nucleotide sites differ between the third and fourth sequences? 6/15 = 0.4

(3) c) How many transversions are present between the third and fourth sequences? 5

(6) d) Estimate the population nucleotide diversity from this sample.

Three unique sequences: A (sequence 1), B (sequences 2, 4), C (sequence 3)
Frequencies P(A) = 0.25, P(B) = 0.5, P(C) = 0.25
πij for comparisons: A-B 5/15 = 0.3333, A-C 1/15 = 0.06666, B-C 6/15 = 0.4
π-hat (estimate) = $(N/(N-1)) \Sigma (p_i X p_j) X \pi_{ij}$
= 4/3 X [(2 X 0.25 X 0.5 X 0.3333) + (2 X 0.25 X 0.25 X 0.06666) + (2 X 0.5 X 0.25 X 0.4)]
= 4/3 X (.08333 + 0.008333 + 0.1)] = **0.2555**

(6) 4. A population of caracals was sampled to determine the weight of adult males. A normal distribution was present in which the arithmetic mean weight of the sampled caracals was 20 kg, with standard deviation of 5 kg.

Hint: draw a bell shaped curve

(3) a) What proportion of adult male caracals would be expected to weigh 29.8 kg or less?

29.8 is 9.8 kg above the mean. 9.8/5 = 1.96 standard deviations; 95% of values fall within 1.96 sd's of the mean, only 5% are more extreme than this, with half above and half below. So 0.025 of adult male caracals would be $\geq$ 29.8 kg, and the proportion weighing less is: 1 – 0.025 = **0.975 or 97.5%**

(3) b) Among 200 adult male caracals, how many individuals would be expected to weigh between 15 and 20 kg?

This range of one standard deviation (5 kg) above and below the mean in a normal distribution includes 68% of samples; 0.68 X 200 = 136. One half of these are below the mean: 136/2 = **68** adult male caracals

(16) 5. Red-green color blindness in humans is an X-linked recessive trait that affects 7% of males in the United States. Assuming Hardy Weinberg Equilibrium:

(3) a) What would be the expected frequency of the recessive allele in **males**?

In males P(trait) = P(X_a) = 0.07

(3) b) What would be the expected frequency of carriers among **females**?

Female heterozygotes are carriers, with P(het) = 2 X P(X_a) X P(X_A)
= 2 X(0.07) X (1-0.07) = 2 X 0.07 X 0.93 = 0.1302

(3) c) Alice is a daughter born to a father with normal vision and a mother who is a carrier. Using just this information, what is the probability that Alice is also a carrier?

Hint: draw Punnett square. P(carrier) = 0.5

(7) d) As an adult, Alice has two sons who both have normal vision. Using information regarding her ancestors and descendants, what is the (posterior or Bayesian) probability that Alice is a carrier?

Define: event A: X_AX_a event A^C: X_AX_A outcome B: 2 sons X_A

P((X_AX_a) given 2 sons with normal vision) =

$$\frac{(P(2\ X_A \text{ sons}) \text{ given } X_AX_a) * P(X_AX_a)}{[(P(2\ X_A \text{ sons}) \text{ given } X_AX_a) * P(X_AX_a)] + [(P(2\ X_A \text{ sons}) \text{ given } X_AX_A) \times P(X_AX_A)}$$

$$= \frac{(.5)^2 * 0.5}{((.5)^2 * 0.5) + (1 * 0.5)}$$

= 0.125 / (0.125 + 0.5) = **0.2**

(6) 6. Four babies were born in a hospital on the same night, and their blood groups were later found to be O, A, B and AB. The four pairs of biological parents were:

AB and B – baby AB
B and B – baby B
A and O – baby A
O and O – baby O

Assign the four babies to their correct parents. Indicated next to parents

(17) 7. A survey of MN blood type frequencies was conducted using samples from 400 Navaho in New Mexico. The phenotypic results were:

296 M
48 MN
56 N

(3) a. What is the frequency of the M allele?

P(M allele) = (296 + (48/2)) / 400 = **0.8**

(3) b. What is the frequency of the N allele?

P(N allele) = (56 + (48/2)) / 400 = **0.2**

(6) c. What are the expected genotypic frequencies under Hardy-Weinberg equilibrium?

MM = 0.8^2 = **0.64**

MN = 2 X 0.9 X 0.1 = **0.32**

NN = 0.2^2 = **0.04**

(5) d. Using a chi-square ($\chi 2$) test (with one degree of freedom), are the observed genotypes in the sampling consistent with Hardy-Weinberg equilibrium?

Expected: MM 0.64 X 400 = 256, MN = 0.32 X 400 = 128, NN = 0.04 X 400 = 16

Σ ((O-E)^2/E) = [(296-256)^2/256] + [(48-128)^2/128] + [(56-16)^2/16]

= 6.25 + 50 + 100 = 156.25

Since 156.25 >> 3.84, we conclude that the population significantly differs from Hardy Weinberg expectations

Potentially useful chi square critical values.

Degrees of freedom	P value = .05
1	3.84
2	5.99
3	7.81
4	9.49

(16) 8. Two populations of leopards were sampled and found to have the following allele frequencies for two SNP sites:

	Site 1		Site 2	
	G	C	G	A
60 African leopards	.05	.95	.30	.70
140 Indian leopards	.35	.65	.45	.55

(4) a) Estimate the mean allele frequency of G at Site 1.

Answer $\bar{p} = \frac{60(.05) + 140(.35)}{200} = 52/200 = .26$

(8) b) Calculate genetic identity and its three components for Site 1.

Answer: $Jxy = \sum_{i=1}^{n} P_{ix} \bullet P_{iy} = (.05)(.35) + (.95)(.65) = \mathbf{.635}$

$$Jx = \sum_{i=1}^{n} P^2_{ix} = (.05)^2 + (.95)^2 = \mathbf{.905}$$

$$Jy = \sum_{i=1}^{n} P^2_{iy} = (.35)^2 + (.65)^2 = \mathbf{.545}$$

$$I = \frac{Jxy}{(Jx\ Jy)^{1/2}} = \frac{.635}{((.905)(.545))^{1/2}} = \frac{.635}{.7023} = \mathbf{.9042}$$

(4) c) Estimate Nei's standard genetic distance between the two leopard populations at Site 1.

Answer: $D = - \ln (I) = - \ln (.9042) = \mathbf{.1007}$

(2) 9. When is it most important to use an exact test rather than a chi-square test?

When sample sizes are small

(5) 10. Assume the probability of a child being born a boy is 0.52. What is the probability that a family would have two boys and two girls?

$(4!/(2!2!))X(0.52)^2X(0.48)^2 = ((4X3X2)/(2X2))(.2704)(.2304) = \mathbf{0.3738}$ or 37.38%