Student #__________________________ FULL NAME_______________________________________

Grade: /40

Population Genetics Midterm

BIO 3119 Fall 2016

Time: 80 min

Show your work and answers in spaces below each question. Use the back of the page if necessary. There are 6 questions in total and some have multiple parts. Faculty-approved calculators are permitted.

Good luck!

Cellular phones, unauthorized electronic devices or course notes (unless an open-book exam) are not allowed during this exam. Phones and devices must be turned off and put away in your bag. Do not keep them in your possession, such as in your pockets. If caught with such a device or document, the following may occur: you will be asked to leave immediately the exam and academic fraud allegations will be filed which may result in you obtaining a 0 (zero) for the exam.

By signing below, you acknowledge that you have ensured that you are complying with the above statement.

Il est interdit de se servir de téléphones cellulaires, de dispositifs électroniques non autorisés ou de notes de cours (à moins qu'il s'agisse d'un examen à livre ouvert). Les téléphones et les dispositifs doivent être fermés et rangés dans votre sac : vous ne pouvez pas les laisser dans vos poches ou sur votre personne. Sinon, on pourrait vous demander de quitter l'examen immédiatement et des allégations de fraude scolaire pourraient être déposées dont le résultat pourrait être un 0 (zéro) pour l'examen.

En apposant votre signature, vous reconnaissez vous être assuré de respecter l'énoncé ci-dessus.

____

Signature

Question 1. The loss of genetic variance due to drift can be a problem in perpetually small populations (e.g., those in captivity or in isolated reserves). In a population of size N = 10, on average the expected heterozygosity would be reduced to what proportion of its original value after 20 generations? [2 marks]

$H_t = H_0 e^{-t/2N}$, $H_t/H_0 = e^{-20/20}$, so $H_t/H_0 = e^{-1}$, $H_t/H_0 = 0.368$, so average heterozygosity is expected to be 36.8% of its initial value.

OR

$H_t = H_0(1 - 1/2N)^t$ so $H_t/H_0 = (1 - 1/20)^{20} = 0.358486$ so 35.9%.

Question 2. Porphyria is an autosomal dominant disease that can be imperfectly diagnosed by low levels of a marker enzyme (porphobilinogen deaminase; Motulsky 1995). Laboratory records indicate that 82% of individuals with the disease allele test positive for the marker, while 3.7% of individuals without the disease allele test positive. The incidence of the disease allele in the population is 10^{-4}. What is the probability that a person who tests positive for the marker actually has the disease? [5 marks]

Given:
$Pr(M|D) = 0.82$
$Pr(D) = 10^{-4}$
$Pr(M|d) = 0.037$

Calculate $Pr(M) = Pr(MD)*Pr(D) + Pr(M|d)*Pr(d) = 0.82*10^{-4} + (0.037*(1- 10^{-4})) = 0.037078$

So $Pr(D|M) = \{Pr(D)*Pr(M|D)\}/Pr(M) = \{(10^{-4})*0.82\}/0.037078 = 0.002212$ or 2.2×10^{-3}

Question 3. In birds, males are the homogametic sex (ZZ), while females are the heterogametic sex (ZW). This is referred to as the ZW system and is opposite to the XY system seen in mammals and insects, where females are homogametic (XX) and males are heterogametic (XY). The lutino (*l*) allele is a Z-linked feather colour mutation in parakeets that causes yellow feathers and is recessive to the common (i.e. wild-type) allele at this locus (+) that causes green feathers.

a) In one population of parakeets, 5% of males have yellow feathers. What percentage of females are expected to have yellow feathers and what assumption do you need to make in calculating this? [2 marks]

$p = \sqrt{p^2} = \sqrt{(0.05)} = 0.2236$ so approximately 22.4% of females are expected to be yellow.

You have to assume male genotypes are at HW in this population.

b) The numbers below come from a genetic study of males from another population of parakeets. Are genotypes at this locus at Hardy-Weinburg expectation in males? Show your calculations below by filling in the table, then state the degrees of freedom and, by referring to the table at the end of the exam, explicitly state whether the test results lead you to accept or reject Hardy-Weinburg equilibrium for males in this population. [8 marks]

Male genotype	**Observed #**	**HW expected frequency**	**HW expected counts**	**(O-E)^2/E**
++	277	0.808935261	273.4201183	0.046871286
+*l*	54	0.162744372	61.15976331	0.838168886
ll	7	0.010118693	3.420118343	3.747107963
TOTAL	338	1	338	4.632148135

p = ((2*277) + 54)/(338*2) = 0.8994; *q* = *1 – p* = ((2*7)+54)/(338*2) = 0.1006

df = 3-2 = 1; $0.01 < P < 0.05$ (i.e. Xcrit = 3.84 and Xobs > Xcrit)

Therefore we reject the null hypothesis that males in this population are at HWE at this locus.

Question 4. The table below gives genotype counts for two blood group loci for a sample of 810 people.

a) From these data, estimate the gametic disequilibrium (*D*) and *D'* for these two loci. Hint: use the frequencies of the gametes whose phase you know. [6 marks]

	BB	Bb	bb
AA	140	80	50
Aa	80	90	70
aa	30	70	200

Ignore Aa/Bb cell.

$N = N - N_{22} = 810 - 90 = 720$.

$g_1 = P_{AB} = (AABB + ½(AABb) + ½(AaBB))/N = 140 + ½(80) + ½(80)/720 = 0.31$

$g_2 = P_{Ab} = (50 + ½(80) + ½(70))/720 = 0.17$

$g_3 = P_{aB} = (30 + ½(80) + ½(70))/720 = 0.14$

$g_4 = P_{ab} = (200 + ½(70) + ½(70))/720 = 0.38$

$D = g_1 g_4 - g_2 g_3 = 0.31(0.38) - 0.17(0.14) = 0.094$

$D_{max} = \min(p_A * p_b, p_a * p_B)$

$p_A = P_{AB} + P_{Ab} = 0.31 + 0.17 = 0.48$

$p_a = 1 - p_A = 1 - 0.48 = 0.52$

$p_B = 0.31 + 0.14 = 0.45$

$p_b = 1 - 0.45 = 0.55$

$D_{max} = \min(0.48 \times 0.55, 0.52 \times 0.45) = \min(0.264, 0.234) = 0.234$

$D' = D/D_{max} = 0.094/0.234 = 0.4$

b) In one sentence, explain what the sign and magnitude of *D'* means [1 mark].

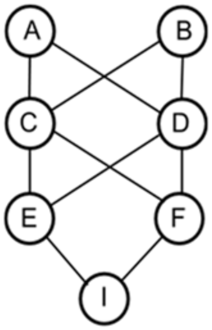
The positive D' tells us that there is an abundance of coupling gametes, and disequilibrium is at about 40% of the maximum possible disequilibrium.

c) Give one reason why the value of *D* should be considered an ESTIMATE of the actual gametic disequilibrium in this population and specify the assumption you must make for this estimate to be reasonable. [2 marks; BONUS mark for a 2nd reason and accompanying assumption]

It's an estimate because you are ignoring the double heterozygotes and you are thus assuming they are rare.

It's also an estimate because it is calculated using a sample of the actual population and you are assuming that the sample is random or otherwise representative of the population.

Question 5. What is the inbreeding coefficient (*f*) of individual I in the hypothetical pedigree shown below depicting two generations of brother-sister mating, assuming none of the common ancestors are themselves inbred? [6 marks]



ANSWER:

Note that along with the single paths through each of the common ancestors C and D, there are also two unique paths through common ancestor A (EC**A**DF and ED**A**CF) and two more through common ancestor B (EC**B**DF and ED**B**CF), yielding:

$$f_I = (½)^3(1 + f_C) + (½)^3(1 + f_D) + ½^5(1 + f_A) + ½^5(1 + f_B) + ½^5(1 + f_A) + ½^5(1 + f_B)$$

$$f_I = 1/8 + 1/8 + 4(1/32)$$

$$= 3/8$$

Question 6a) In four sentences or less, define inbreeding and describe its effects on allele vs. genotype frequencies. [3 marks]

Inbreeding is the mating between individuals that are more closely related than expected by chance. Inbreeding has no effect on allele frequencies. In terms of genotypes, it reduces the frequency of heterozygotes and increases the frequency of both homozygotes.

6b) In four sentences or less, define inbreeding depression and outline the two hypotheses for why it occurs. [5 marks]

Inbreeding depression is the reduction in population mean fitness that occurs with inbreeding. It can occur because recessive deleterious alleles, which are usually masked in heterozygotes, are now 'exposed' to selection as recessive homozygotes. Alternatively, it can occur if heterozygotes normally have higher fitness than homozygotes, but inbreeding causes a decrease in freq of heterozygotes, leading to lower population mean fitness.

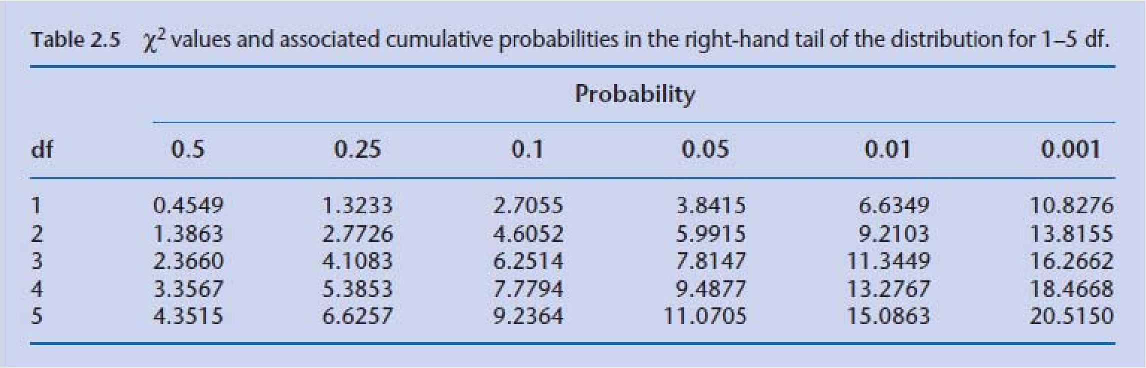


Table 2.5 χ^2 values and associated cumulative probabilities in the right-hand tail of the distribution for 1–5 df.

df	Probability					
	0.5	0.25	0.1	0.05	0.01	0.001
1	0.4549	1.3233	2.7055	3.8415	6.6349	10.8276
2	1.3863	2.7726	4.6052	5.9915	9.2103	13.8155
3	2.3660	4.1083	6.2514	7.8147	11.3449	16.2662
4	3.3567	5.3853	7.7794	9.4877	13.2767	18.4668
5	4.3515	6.6257	9.2364	11.0705	15.0863	20.5150

BIO 3119 Formula Sheet - Fall 2016 midterm exam

This sheet is provided so that you do not need to memorize complex formulas for the exam. It is not intended as a study guide - many important concepts and theories that we covered in class are not listed here. You will need to know how to apply these formulas in a variety of scenarios and **what they mean**. Some very basic formulas (e.g., how to go from genotype to allele frequency and vice versa) are not included. **A copy of this sheet will be provided with your midterm exam (do not bring your own).**

Probability Theory

Pr(A or B) = Pr(A) + Pr(B) – Pr(A and B); Pr(A and B) = Pr(A) x Pr(B) = Pr(A)Pr(B)

$$\Pr(A|B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}$$

Bayes' theorem: $\Pr(A|B) = \dfrac{\Pr(A) \times \Pr(B|A)}{\Pr(B)}$

Law to total probability: $\Pr(A) = \Pr(A \mid B) \times \Pr(B) + \Pr(A \mid C) \times \Pr(C)$

Binomial distribution: $\Pr(X = k) = \dbinom{n}{k} p^k (1-p)^{n-k} = \dfrac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$

Poisson distribution: $\Pr(X = k) = \dfrac{e^{-\mu} \mu^k}{k!}$

Quantifying genetic variance

$$P = \frac{\text{number of polymorphic loci}}{\text{total number of loci examined}} \qquad \overline{H}_{obs} = \frac{1}{n} \sum_{i=1}^{n} H_i$$

Hardy-Weinburg

$$p^2 + 2pq + q^2 = 1$$

$\chi^2 = \displaystyle\sum_{\text{all classes}} \frac{(O-E)^2}{E}$, df = # data classes - # parameters estimated from the data - 1

Linkage Disequilibrium

$$D = g_{AB} g_{ab} - g_{Ab} g_{aB}; \quad D' = \frac{D}{D_{\max}}; \quad D_t = (1-r) D_{t-1}; \quad D_t = (1-r)^t D_0$$

$D_{\max} = \min(p_A p_b, p_a p_B)$, if $D > 0$ $D_{\max} = \min(p_A p_B, p_a p_b)$, if $D < 0$

Disease Mapping

$$L = \frac{\Pr\left(\text{sibship} \mid r = \hat{r}\right)}{\Pr\left(\text{sibship} \mid r = 0.5\right)};\; LOD = \log(L);\;\; \Pr(D|M) = \frac{\Pr(D)\Pr(M|D)}{\Pr(M)}$$

Inbreeding and Fixation Index

$$f_I = \sum_{i=1}^{p} \left(\frac{1}{2}\right)^{n_i} (1 + f_{A_i})\,; \quad f = \frac{H_{\mathrm{exp}} - H_{\mathrm{f}}}{H_{\mathrm{exp}}}\,;\;\; F = \frac{H_{\mathrm{exp}} - H_{\mathrm{obs}}}{H_{\mathrm{exp}}}$$

freq(AA) = $P_{\mathrm{AA}} = p\,f + p^2(1 - f) = p^2 + pqf$
freq(Aa) = $P_{\mathrm{Aa}} = 2pq(1 - f) = 2pq - 2pqf$
freq(aa) = $P_{\mathrm{aa}} = q\,f + q^2(1 - f) = q^2 + pqf$

self-fertilization: $f_t = \frac{1}{2}(1 + f_{t-1})$; $H_t = \frac{1}{2}H_{t-1}$; $f_t = 1 - \frac{1}{2}^t(1 - f_0)$; $H_t = (\frac{1}{2})^t H_0$

self-fertilization & outcrossing: $f_{t+1} = \frac{1}{2} S(1 + f_t)$; $\hat{f} = \frac{S}{2 - S}$; $r_{effective} = r\left(1 - \frac{S}{2 - S}\right)$

Genetic Drift

$$E(p_{t+1}) = p_t\,; \quad V(p_{t+1}) = \frac{p_t(1 - p_t)}{2N}\,; \quad V_t = p_0 q_0\left[1 - \left(1 - \frac{1}{2N}\right)^t\right]$$

$$G_{t+1} = F_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F_t\,; \quad H_{t+1} = H_t\left(1 - \frac{1}{2N}\right)$$

$$H_t = H_0\left(1 - \frac{1}{2N}\right)^t \approx H_0 e^{-t/2N}\,; \quad F_t = 1 - \left(1 - \frac{1}{2N}\right)^t$$

$$\bar{t}_{fix} = -4N_e\left(\frac{1 - p}{p}\right)\ln(1 - p)\,; \quad \bar{t}_{loss} = -4N_e\left(\frac{p}{1 - p}\right)\ln(p)$$

$$\frac{1}{N_e} = \frac{1}{t}\left(\frac{1}{N_1} + \frac{1}{N_2} + \ldots + \frac{1}{N_t}\right); \quad N_e = \frac{4N_m N_f}{N_m + N_f}\,; \quad N_e = \frac{4N}{Var(k) + 2}$$

x-linked: $N_e = \frac{9N_m N_f}{4N_m + 2N_f}$; y-linked: $N_e = \frac{N_m}{2}$; mtDNA: $N_e = \frac{N_f}{2}$