STAT 536 Midterm

Name: ________________________

October 19, 2006

Formulae

$$\mathrm{Var}(\tilde{p}_u) = \frac{1}{2n}\left(p_u + P_{uu} - 2p_u^2\right)$$

$$\mathrm{Cov}(\tilde{p}_1, \tilde{p}_2) = \frac{1}{4n}\left(P_{12} - 4p_1p_2\right)$$

$$\mathrm{Var}(T) \approx n\sum_i \left(\left.\frac{\partial T}{\partial n_i}\right|_{E[n_i]}\right)^2 Q_i - n\left(\left.\frac{\partial T}{\partial n}\right|_{E[n_i]}\right)^2$$

$$\text{NR: } p_{t+1} = p_t - \frac{S_p}{I_p}$$

$$\hat{\phi}_J = n\hat{\phi} - \frac{n-1}{n}\sum_i \hat{\phi}_{(i)}$$

$$\mathrm{Var}(\hat{\phi})_J = \frac{n-1}{n}\sum_i \left(\hat{\phi}_{(i)} - \frac{1}{n}\sum_j \hat{\phi}_{(j)}\right)^2$$

For multinomial counts n_i and n_j with populaton proportions q_i and q_j:

$$\mathrm{Var}(n_i) = nq_i(1-q_i)$$

$$\mathrm{Cov}(n_i, n_j) = -nq_iq_j$$

$$I_p = -\left(\frac{\partial^2 \ln L(p)}{\partial p^2}\right) \text{ and } \mathrm{Var}(p) = \frac{1}{E[I_p]}$$

$$S_p = \frac{\partial \ln L(p)}{\partial p}$$

$$\mathrm{E}(\hat{D}_1) = D_1 - \frac{1}{2n}\left[p_1(1-p_1) + D_1\right]$$

$$\mathrm{Var}(\hat{D}_1) \approx \frac{1}{n}\left[p_1^2(1-p_1)^2 + (1-2p_1)^2D_1 - D_1^2\right]$$

$$\begin{aligned}\mathrm{Var}\left(\hat{D}_{uv}\right) &= \frac{1}{2n}\Bigg\{p_up_v\left[(1-p_u)(1-p_v) + p_up_v\right] \\ &\quad - \left[(1-p_u-p_v)^2 - 2(p_u-p_v)^2\right]D_{uv} \\ &\quad + \sum_{w\neq u,v}(p_u^2D_{vw} + p_v^2D_{uw}) - D_{uv}^2\Bigg\}\end{aligned}$$

$$\begin{aligned}\frac{v_{11}}{v_{12}} &= 2R'_{11}\frac{2R_{22}+1}{2R_{11}+1} \\ \frac{v_{22}}{v_{12}} &= 2R'_{22}\frac{2R_{11}+1}{2R_{22}+1} \\ w_{22} &= \frac{x^2}{x'}w_{11} + \frac{x}{x'} - x\end{aligned}$$

$$\mathrm{E}(\hat{D}_{AB}) = \frac{2n-1}{2n}D_{AB}$$

$$\begin{aligned}\mathrm{Var}(\hat{D}_{AB}) &= \frac{1}{2n}\big[p_A(1-p_A)p_B(1-p_B) \\ &\quad + (1-2p_A)(1-2p_B)D_{AB} - D_{AB}^2\big]\end{aligned}$$

$$\begin{aligned}\Delta_{AB} &= \frac{1}{n}\left(2n_{AABB} + n_{AABb} + n_{AaBB} + \frac{1}{2}n_{AaBb}\right) \\ &\quad - 2\tilde{p}_A\tilde{p}_B\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(\hat{\Delta}_{AB}) &= \frac{1}{n}\Big[(\pi_A + D_A)(\pi_B + D_B) + \frac{1}{2}\tau_A\tau_B\Delta_{AB} \\ &\quad + \tau_AD_{ABB} + \tau_BD_{AAB} + \Delta_{AABB}\Big]\end{aligned}$$

for $\pi_A = p_i(1-p_i), \tau_i = (1-2p_i), i = A, B$

$$\mathrm{E}(\hat{v}) = v\left[1 - \left(\frac{v}{v+1}\right)^n\right]$$

$$\mathrm{Var}(\hat{v}) \approx \frac{v(v+1)^2}{n}$$

Quantiles of Central ξ_1^2 with 1 degree of freedom

Left Tail Probability	Quantile
0.90	2.71
0.95	3.84
0.975	5.02

Quantiles of Standard Normal

Left Tail Probability	Quantile
0.90	1.28
0.95	1.64
0.975	1.96

1. (2.5 pts each) Short answers

(a) How many generations does it take to eliminate a dominant lethal mutation?

(b) What is the equilibrium allele frequency for a recessive allele who's forward mutation rate is μ and selection coefficient is t?

(c) Write an expression for the linkage disequilibrium coefficient D_{AB} in terms of population parameters.

(d) What is the name of the phenomena that describes the decline in overall fitness of a population due to mutation.

2. (5 pts each) Brief answers

(a) Describe two opposing forces on populations, how each force changes the population's genetics, and explain (in words) how they interact.

(b) What is the expected bias of relative (with respect to the heterozygote) viability estimator $\hat{v}_{11} = \frac{n_{11}}{n_{12}+1}$ if true relative viability is $v_{11} = 0.9$ and you sample $n = 8$ adults?

(c) Specify an equation summarizing the Whalund effect, and describe what it means in words.

(d) Compute a 95% confidence interval for the frequency of allele A if you observe the following genotype data $P_{AA} = 103, P_{AB} = 53, P_{BB} = 80$.

(e) Compute the confidence interval for $\hat{p}_A$ from the following bootstrap data (in rank order from smallest to largest):

1	0.500	6	0.517		91	0.587	96	0.593
2	0.506	7	0.519		92	0.587	97	0.600
3	0.513	8	0.523	$\cdots$	93	0.587	98	0.604
4	0.515	9	0.523		94	0.593	99	0.608
5	0.517	10	0.525		95	0.593	100	0.627

(f) Write the recursion relation for an island that receives a proportion m_c of immigrants from a continent with allele frequency p_c and exchanges a proportion m_i of immigrants with another island with allele frequency p_i.

3. (15 pts) Perform a test of linkage disequilibrium for the following haplotype data.

AB	Ab	aB	ab
6	8	31	15

4. (20 pts) *Epistasis* is a phenomena where genotypes or phenotypes at multiple loci interact to produce a combined phenotype. The importance of epistasis is only beginning to be properly appreciated. Here, you will consider an exact test for epistasis. Consider the following data for genotypes at two loci, A and B, each with two alleles. Compute the probability of this dataset assuming independence of genotypes at the two loci. Fill in the second table with a data set more extreme than the first one, but with the same marginal genotype counts. Compute its probability. Speculate on the results of an exact test for independence of genotypes.

	AA	Aa	aa
BB	9	0	0
Bb	0	3	1
bb	0	1	4

	AA	Aa	aa
BB			
Bb			
bb			

5. (25 pts) Consider a locus with two possible alleles A and B and corresponding allele frequencies p_A and p_B. There are three genotypes AA, AB, and BB and three distinguishable phenotypes A, AB, and B that correspond to each genotype. A gene is called *partially penetrant* if the same genotype does not always result in the same phenotype. Suppose that the AB genotype is penetrant with probability $e < 1$. Otherwise, the default phenotype is A. Assume both the AA and BB genotypes are fully penetrant.

(a) Describe a null hypothesis and statistical test that could be used to test for this kind of partial penetrance.

(b) Write all steps of an EM algorithm for obtaining maximum likelihood estimates of p_A, p_B, and e.