GENE613 - Exam 3

Daniel Osorio - dcosorioh@tamu.edu
Department of Veterinary Integrative Biosciences
Texas A&M University

1. A single locus with 2 alleles in HWE in each of four populations

GENOTYPE:	A_1A_1	A_1A_2	A_2A_2	$f(A_1)$
VALUE:	4	22	50	
Population I:	0.01	0.2	0.79	0.11
Population II:	0.79	0.2	0.01	0.89
Population II:	0	0	1	0
Population IV:	0	0.02	0.98	0.01

a) Two F1 populations are produced by mating individuals from population 1 to individuals in population 2 and separately for individuals from population 3 to individuals in population 4. Two F2 populations are produced by inter se matings within each F1 group (one F2 from the population 1 & 2 F1s and one F2 from the populations 3 & 4 F1s). Calculate the heterosis expressed in both F1 populations and both F2 populations.

```
> aValue <- 4 - mean (c(4, 50))
> dValue <- 22 - mean (c(4, 50))
> heterosisF1 <- function(p1, p2, a, d){
+    q1 <- 1 - p1
+    q2 <- 1 - p2
+    y <- p1 - p2
+    MP1 <- a * (p1 - q1) + 2 * p1 * q1 * d
+    MP2 <- a * (p2 - q2) + 2 * p2 * q2 * d
+    midP <- mean(c(MP1, MP2))
+    MF <- a * (p1 - q1 - y) + d * (2 * p1 * q1 + y * (p1 - q1))
+    HF <- MF - midP
+    return(HF)
+ }
> heterosisF1(p1 = 0.11, p2 = 0.89, a = aValue, d = dValue)

[1] -3.042

> heterosisF1(p1 = 0, p2 = 0.01, a = aValue, d = dValue)

[1] -5e-04

> heterosisF2 <- function (p1, p2, a, d){
+    0.5 * heterosisF1(p1, p2, a, d)
+ }
> heterosisF2(p1 = 0.11, p2 = 0.89, a = aValue, d = dValue)
```

```
[1] -1.521

> heterosisF2(p1 = 0, p2 = 0.01, a = aValue, d = dValue)

[1] -0.00025
```

b) Estimated heterosis values were -0.25 at 5 additional loci (none appear to interact with each other, nor with the locus above). What is total heterosis in both of the F1 populations? What do you expect it to be in the F2 populations?

```
> - 0.25 + (dValue * (0.11 - 0.89) ^ 2)

[1] -3.292

> - 0.25 + (dValue * (0 - 0.01) ^ 2)

[1] -0.2505

> 0.5 * (-0.25 + (dValue * (0.11 - 0.89) ^ 2))

[1] -1.646

> 0.5 * (-0.25 + (dValue * (0 - 0.01) ^ 2))

[1] -0.12525
```

2. The table below is for pigs from Suzuki et al. (2004) J. Anim. Sci. 82:994-999. Blood plasma concentrations of IGFI were measured in pigs at 8 wk of age (IGFI-8W) and 105 kg of body weight (IGFI-105KG). Estimates of heritability (h^2) and additive genetic correlation (r_a) for IGFI concentration, and additive genetic standard deviations (σ_a) at the two measured points are shown for traits in the far right column. Standard errors (SE) for each genetic parameter are in columns immediately to the right of h^2 and r_a estimates. Pigs in each generation are assigned to two equal (in number) groups based on their IGFI-8W value: the highest values are in group 1 and the lowest values are in group 2.

			IGFI-8W, ng/mL		IGFI-105KG, ng/mL		
Traits	h^2	SE	r_a	SE	r_a	SE	σ_a
Daily gain, g/d	0.49	0.03	0.26	0.08	0.22	0.07	56.49
Loin muscle area, cm^2	0.43	0.03	0.22	0.1	0.42	0.08	2.419695
Backfat thickness, cm	0.73	0.02	0.13	0.08	-0.02	0.07	0.333216
Intramuscular fat, %	0.39	0.03	0.32	0.1	0.26	0.09	0.8743
Tenderness, kdf/cm^2	0.39	0.04	-0.05	0.12	0.36	0.1	7.931147
Body weight at 8wk, kg	0.24	0.02	0.45	0.08	0.09	0.09	1.557875
Backfat thickness at 8 wk, cm	0.41	0.02	0.33	0.06	0	0.06	1.062919
Feed conversion ratio	0.35	0.04	0.2	0.08	-0.17	0.07	0.100573
IGFI-8W, ng/mL	0.23	0.03			0.73	0.08	19.1833
IGFI-105KG, ng/mL	0.26	0.03					22.69064

a) In a selection program in which only group 1 pigs are allowed to be parents, what would be the response to selection and the correlated response in each of the other traits?

$$R_k = ih\sigma_a$$

As the two groups are equal in number, which means each one has the 50 % of the individuals and an associate selection intensity of 0.8

```
> i <- 0.8
> h <- sqrt(c(0.49, 0.43, 0.73, 0.39, 0.39,
+             0.24, 0.41, 0.35, 0.23, 0.26))
> sigma_a <- c(56.49, 2.42, 0.33, 0.87, 7.93,
+              1.56, 1.06, 0.1, 19.18, 22.69)
> rSelection <- i * h * sigma_a
> names(rSelection) <- c("Daily gain", "Loin muscle area",
+                    "Backfat thickness (BT)",
+                    "Intramuscular fat", "Tenderness",
+                    "Body weight at 8wk", "BT at 8 wk",
+                    "Feed conversion ratio", "IGFI-8W",
+                    "IGFI-105KG")
> rSelection
           Daily gain       Loin muscle area Backfat thickness (BT)
           31.63440000             1.26952010             0.22556170
    Intramuscular fat             Tenderness     Body weight at 8wk
            0.43465186             3.96182673             0.61139264
            BT at 8 wk  Feed conversion ratio                IGFI-8W
            0.54298494             0.04732864             7.35872389
            IGFI-105KG
            9.25574022
```

$$CR_y = ih_x r_a \sigma_{a_y}$$

```
> r_a <- c(0.26, 0.22, 0.13, 0.32, -0.05, 0.45, 0.33, 0.2)
> sigma_ay <- c(56.49, 2.42, 0.33, 0.87, 7.93, 1.56, 1.06, 0.1)
> h_x <- sqrt(0.23)
> cResponse_IGFI8W <- i * h_x * r_a * sigma_ay
> names(cResponse_IGFI8W) <- c("Daily gain", "Loin muscle area",
+                    "Backfat thickness (BT)",
+                    "Intramuscular fat", "Tenderness",
+                    "Body weight at 8wk", "BT at 8 wk",
+                    "Feed conversion ratio")
> cResponse_IGFI8W
           Daily gain       Loin muscle area Backfat thickness (BT)
            5.63506367             0.20426406             0.01645929
    Intramuscular fat             Tenderness     Body weight at 8wk
            0.10681276            -0.15212378             0.26933390
            BT at 8 wk  Feed conversion ratio
            0.13420655             0.00767333
> h_x <- sqrt(0.26)
> r_a <- c(0.22, 0.42, -0.02, 0.26, 0.36, 0.09, 0, -0.17)
> cResponse_IGFI105KG <- i * h_x * r_a * sigma_ay
> names(cResponse_IGFI105KG) <- c("Daily gain", "Loin muscle area",
+                    "Backfat thickness (BT)",
+                    "Intramuscular fat", "Tenderness",
+                    "Body weight at 8wk", "BT at 8 wk",
+                    "Feed conversion ratio")
> cResponse_IGFI105KG
```

Daily gain	Loin muscle area	Backfat thickness (BT)
5.069567577	0.414611475	-0.002692282
Intramuscular fat	Tenderness	Body weight at 8wk
0.092271857	1.164534473	0.057272187
BT at 8 wk	Feed conversion ratio	
0.000000000	-0.006934667	

b) What is required to annualize response? *The generation interval (L) which is the time for generation replacement.* What would you project that to be for pigs? *As for pigs the L value is a small number, the annualize response to selection will be high.*

c) The additive genetic correlation between IGFI-105KG and Feed Conversion Ratio (this is the ratio of all feed consumed to total weight gained) is 0.17 ± 0.07. Interpret this value, including its SE. *-0.17 is the average degree (%) of association between the breeding value between IGFI-105KG and Feed Conversion Ratio and $\pm$ 0.07 is the standard deviation of the computation of this coefficient which means that the population coefficient of correlation between these two traits is more likely (64% of probability if we assume standardized normal distribution) to be between 0.10 and 0.24*

3. What parameters influence the accuracies of estimated breeding values? *If we consider accuracy as a bias in the calculation of the breeding value, this could come from the determination of the allele frequencies p and q from the sampled population, as well as from the determination of the parameters a and d. By other hand, if we consider accuracy as the precision, then the estimation of the breeding value is affected by the number of observations per individual as well as for the number of observations on relatives.*

4. Two lines of Hereford cattle that were homozygous for opposite alleles at two syntenic marker loci (Line 1 genotype MMqq; Line 2 genotype = mmQQ). Ten F1 bulls that were contemporaries to each other had semen samples collected and mass genotyped for these two markers. Gametic frequencies:

MQ	Mq	mQ	mq
0.05	0.44	0.4	0.1

Estimate the recombination rate r from these frequencies. Convert to map distance for the interval between the two loci in cM using Haldanes original map function and then his subsequent approximation. *If we consider that: $MQ = \frac{1-r}{2}$, $mq = \frac{1-r}{2}$, $Mq = \frac{r}{2}$, $mQ = \frac{r}{2}$ and $r = \frac{r}{2} + \frac{r}{2}$, then we have that:*

```
> r <- 0.5 * mean(abs((c(0.05, 0.1, 0.44, 0.4) -
+                      c(0.5, 0.5, 0, 0)) / 0.5))
> r

[1] 0.4225

> haldane <- function(r){(-0.5*log(1-2*r))/0.01}
> haldane(r)

[1] 93.21651

> haldaneR <- function(r){(0.7*r-0.15*log(1-2*r))/0.01}
> haldaneR(r)
```

```
[1] 57.53995
```

5. Describe the essential features of Hendersons Mixed Model Equations relative to prediction, and what made them different from previous linear regression models. *Hendersons Mixed Model Equations allow us to estimate the average breeding value of contemporary groups taking into account the performance of relatives in other contemporary groups. These association between relatives (correlated data) is the highest difference between this method and the linear regression model where each individual is assumed to be independent.*

6. Assume you have phenotypes for all of the individuals in this (extremely) old Angus pedigree. If phenotypes of the others were used to predict breeding values for the youngest animal in the pedigree, rank those from best to worst. *The ranking in order of how well a phenotype of a related individual would predict breeding values of Octavia is: Old Jock > Grey Breasted Jock = Favorite* Give the average inbreeding coefficient for this group.
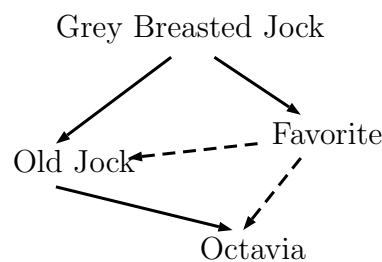

```
                Grey Breasted Jock Favorite Old Jock Octavia
Grey Breasted Jock              1.00     0.50     0.75   0.750
Favorite                        0.50     1.00     0.75   0.750
Old Jock                        0.75     0.75     1.25   1.250
Octavia                         0.75     0.75     1.25   1.625

F:

Grey Breasted Jock          Favorite          Old Jock          Octavia
           0.000             0.000             0.250            0.625

Average inbreeding coefficient:

[1] 0.21875
```

7. Assume in the context of experimentation with a simple null hypothesis of $\beta_1 = 0$ for m SNP loci in a GWAS, you set the probability of Type error I to 0.05 (you can control this) for each test. Assume that the probability of a Type II error is 0.2.

	Declared Non Significant	Declared Significant	Total
True Null Hypotheses	U	V	m_0
Non-True Null Hypotheses	T	S	$m - m_0$
	$m - R$	R	m

a) What will be the probabilities of V incorrect rejections of null hypotheses?
$\alpha = 0.05$

b) What is the probability of S correct rejections of null hypotheses?
$1 - \beta = 0.8$

c) What will be the corresponding FDR in that scenario?

$$Q = \frac{V}{R}, R > 0$$

$$Q = \frac{0{,}05}{0{,}8 + 0{,}05} = 0{,}06$$

8. Define σ_a^2 *The additive genetic variance is the deviation from the mean phenotype due to inheritance of a particular allele and this allele's relative (to the mean phenotype of the population) effect on phenotype.*

9. A Bonferroni correction of P values for multiple testing in a genome-wide association requires independence of tests, which is not the case. Why are such tests not independent? *Because not all the considered genetic markers (SNPs) are segregated in an independent manner, several ones segregate together due to the linkage disequilibrium (LD)*

10. The breeding objective in the below mouse population is to increase litter size.

MOUSE ID:	1	2	3	4	5	6	7	8	9	10
1st parity	11	9	13	10	9	8	10	11	10	13
2nd parity	10	12	12	10	7	6	12	9	12	12
3rd parity	12	12	11	11	7	10	9	11	10	14
4th parity	12	10	11	10	10	8	4	12	12	12

Using the above information, make a recommendation to remove 4 of the above females. $R = 0.57$. *I recomend to remove the mouses with ID 4, 5, 6 and 7 because they have the lowest ERPA values in the sample.*

```
        ID AVERAGE n    ERPA       ACC
 [1,]    1   0.875 4  0.49875 0.9172396
 [2,]    2   0.375 4  0.21375 0.9172396
 [3,]    3   1.375 4  0.78375 0.9172396
 [4,]    4  -0.125 4 -0.07125 0.9172396
 [5,]    5  -1.875 4 -1.06875 0.9172396
 [6,]    6  -2.375 4 -1.35375 0.9172396
 [7,]    7  -1.625 4 -0.92625 0.9172396
 [8,]    8   0.375 4  0.21375 0.9172396
 [9,]    9   0.625 4  0.35625 0.9172396
[10,]   10   2.375 4  1.35375 0.9172396
```

How good are the estimates of merit that you base your decision on? *All the computed values have a ACC of 0.917* If $V_P = 12.2$, what is estimate for V_{PA}? *As $R = \frac{V_{PA}}{V_P}$ then $V_{PA} = RV_P = 0{,}57 \times 12{,}2 = 6{,}954$*

11. For two syntenic loci with alleles A, a and B, b, give an example showing why the parameter space of D_{AB} must be: $max(-p_A p_B, -p_a p_b) \leq D_{AB} \leq min(p_A p_b, p_a p_B)$.
As D_{AB} is a portion of additional or missing expected genotype frequency, and the sum of all the genotypes in a population must be 1 with each genotype frequency in a range between 0 and 1.

	A	a	
B	$p_B p_A + D_{AB}$	$p_B p_a - D_{AB}$	p_B
b	$p_b p_A - D_{AB}$	$p_b p_a + D_{AB}$	p_b
	p_A	p_a	1

If we consider as an as example that: $p_A = 0{,}7, p_B = 0{,}6, p_a = 0{,}3$ *and* $p_b = 0{,}4$, *then the expected values will be:*

		A	a	
	B	$0{,}42 + D_{AB}$	$0{,}18 - D_{AB}$	p_B
	b	$0{,}28 - D_{AB}$	$0{,}12 + D_{AB}$	p_b
		p_A	p_a	1

And the parameter space for D_{AB} will be: $max\,(-0{,}42, \textbf{-0.12}) \leq D_{AB} \leq min\,(0{,}28, \textbf{0.18})$ otherwise the values of D will lead generate negative values of the genotype frequencies, out of the range of the expected values (which is not possible).