1. (a) The mean litter size is

$$p^2(2.66) + 2pq(2.17) + q^2(1.48)$$
$$= (0.18)^2(2.66) + 2(0.18)(0.82)(2.17) + (0.82)^2(1.48) = 1.722.$$

If we scale the population relative to the average of the mean litter sizes of the homozygotes, we get

$$a = 2.66 - \frac{2.66 + 1.48}{2} = 2.66 - 2.07 = 0.59, \quad \text{and} \quad d = 2.17 - 2.07 = 0.10,$$

so the mean litter size is

$$(0.18)^2(0.59) + 2(0.18)(0.82)(0.10) + (0.82)^2(-0.59) = -0.348.$$

(b) If the frequency of the B allele is doubled, then these means change to

$$(0.36)^2(2.66) + 2(0.36)(0.64)(2.17) + (0.64)^2(1.48) = 1.951,$$

for the unscaled population, and

$$(0.36)^2(0.59) + 2(0.36)(0.64)(0.10) + (0.64)^2(-0.59) = -0.119,$$

for the scaled population.

(c) When $p = 0.18$, the average allele effects are

$$\alpha_B = q[a + d(q - p)] = 0.82[0.59 + 0.10(0.82 - 0.18)] = 0.536, \quad \text{and}$$
$$\alpha_b = -p[a + d(q - p)] = -0.18[0.59 + 0.10(0.82 - 0.18)] = -0.118.$$

When $p = 0.36$, we get

$$\alpha_B = 0.64[0.59 + 0.10(0.82 - 0.18)] = 0.396, \quad \text{and}$$
$$\alpha_b = -0.36[0.59 + 0.10(0.82 - 0.18)] = -0.222.$$

(d) When $p = 0.18$, the average effect of allele substitution is

$$\alpha = \alpha_D - \alpha_d = 0.536 + 0.118 = 0.654.$$

When $p = 0.36$, we get

$$\alpha = 0.396 + 0.222 = 0.618.$$

2. (a) i. The data were fit to the following linear regression model:

$$P = \beta_0 + \beta_1 X_D + \epsilon,$$

where $X_D \in \{0, 1, 2\}$ is the number of copies of the D allele that each individual has. Recall that under this model we have

$$\beta_0 = \mu_G + 2\alpha_d, \quad \text{and} \quad \beta_1 = \alpha_D - \alpha_d.$$

Note that the error term ϵ incorporates both the dominance effect δ and any additional environmental effects.

This resulted in the following estimates:

$$\hat{\beta}_0 = 1.8029, \quad \text{and} \quad \hat{\beta}_1 = 0.3131,$$

so we have

$$\hat{\alpha} = \hat{\beta}_1 = 0.3131$$

$$\hat{\alpha}_d = \frac{\hat{\beta}_0 - \hat{\mu}_G}{2} = \frac{1.8029 - 1.9260}{2} = -0.0615, \quad \text{and}$$

$$\hat{\alpha}_D = \hat{\alpha} + \hat{\alpha}_d = 0.3131 - 0.0615 = 0.2516,$$

where $\hat{\mu}_G = 1.9260$ was estimated by the mean of the phenotype values in the sample.

ii. The estimated breeding values for each genotype are

$$
\begin{aligned}
DD: \quad & 2\hat{\alpha}_D = 2(0.2516) = 0.5032 \\
Dd: \quad & \hat{\alpha}_D + \hat{\alpha}_d = 0.2516 - 0.0615 = 0.1901 \\
dd: \quad & 2\hat{\alpha}_d = 2(-0.0615) = -0.1231.
\end{aligned}
$$

iii. The narrow sense heritability is the proportion of total phenotypic variance that can be explained by the additive genetic variance, which is simply the R^2 for the linear model fit to the data. Therefore, we have

$$\hat{h}^2 = R^2 = 0.7506.$$

(b) First, we estimate the phenotype means for each genotype as

$$
\begin{aligned}
DD: \quad & 2.3789 \\
Dd: \quad & 2.1284 \\
dd: \quad & 1.7999,
\end{aligned}
$$

so we have

$$a = 2.3789 - \frac{2.3789 + 1.7999}{2} = 2.3789 - 2.0894 = 0.2895, \quad \text{and}$$

$$d = 2.1284 - 2.0894 = 0.0390.$$

Therefore, we have

$$\hat{\alpha} = a + d(\hat{q} - \hat{p}) = 0.2895 + 0.0390(0.8035 - 0.1965) = 0.3132$$
$$\hat{\alpha}_D = \hat{q}\hat{\alpha} = 0.8035(0.3132) = 0.2517, \quad \text{and}$$
$$\hat{\alpha}_d = -\hat{p}\hat{\alpha} = -0.1965(0.3132) = -0.0615,$$

where $\hat{p} = 0.1965$ was estimated by the proportion of D alleles in the sample. These values are very close to the ones obtained through the regression model. I believe that their small differences are due to the proportions of genotypes in the sample being slightly out of Hardy-Weinberg equilibrium.

3. (a) First note that

$$X_1 \sim \text{Binomial}(2, p), \quad X_2 = 2 - X_1, \quad \text{and} \quad I_{X_1=1} \sim \text{Bernoulli}(2pq),$$

so we have

$$E(X_1) = 2p, \qquad \text{Var}(X_1) = 2pq,$$
$$E(I_{X_1=1}) = 2pq, \quad \text{Var}(I_{X_1=1}) = 2pq(1 - 2pq).$$

In addition, we have

$$\text{Cov}(X_1, I_{X_1=1}) = E(X_1 I_{X_1=1}) - E(X_1)E(I_{X_1=1})$$
$$= 2pq - (2p)(2pq) = 2pq(1 - 2p) = 2pq(q - p).$$

Therefore, for

$$Y = \mu_Y + \alpha_1 X_1 + \alpha_2 X_2 + \beta I_{X_1=1} + \epsilon$$
$$= \mu_Y + \alpha_1 X_1 + \alpha_2(2 - X_1) + \beta I_{X_1=1} + \epsilon$$
$$= \mu_Y + 2\alpha_2 + (\alpha_1 - \alpha_2)X_1 + \beta I_{X_1=1} + \epsilon,$$

we have

$$E(Y) = \mu_Y + 2\alpha_2 + (\alpha_1 - \alpha_2)(2p) + \beta(2pq) + 0$$
$$= \mu_Y + 2(p\alpha_1 + q\alpha_2 + pq\beta).$$

However, since we know that $E(Y) = \mu_Y$ by definition, this expectation imposes the constraint

$$p\alpha_1 + q\alpha_2 + pq\beta = 0.$$

We also have

$$\text{Var}(Y) = (\alpha_1 - \alpha_2)^2(2pq) + \beta^2[2pq(1 - 2pq)] + 2(\alpha_1 - \alpha_2)\beta[2pq(q - p)] + \sigma^2$$
$$= 2pq[(\alpha_1 - \alpha_2)^2 + (1 - 2pq)\beta^2 + 2(q - p)(\alpha_1 - \alpha_2)\beta] + \sigma^2.$$

(b) The broad sense heritability is the proportion of total phenotypic variance that can be explained by the total genetic variance, so under this model, we have

$$H^2 = 1 - \frac{\sigma^2}{\text{Var}(Y)} = 1 - \frac{\sigma^2}{2pq[(\alpha_1 - \alpha_2)^2 + (1 - 2pq)\beta^2 + 2(q - p)(\alpha_1 - \alpha2)\beta] + \sigma^2}.$$