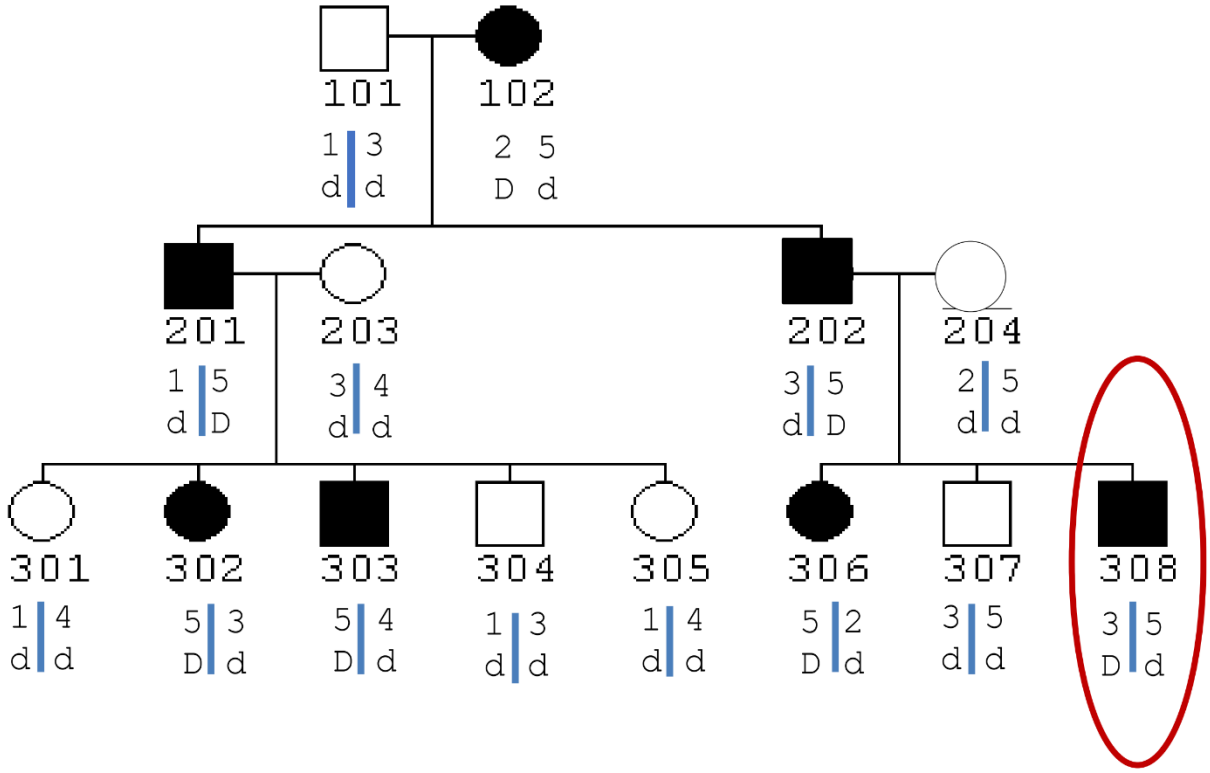
A researcher has collected data from a family with multiple members affected with a rare dominant disease. A prior study reported linkage of this disease to a marker on chromosome 5. They genotyped the previously linked marker, and recorded pedigree information and genotypes in the file HW2_family.xlsx.

Question 1

Part A

Draw the family, labeling everyone by their ID number and indicating each member's genotypes for the genetic marker. Infer the disease locus genotypes for each family member, assuming a very rare disease with full penetrance and a phenocopy rate of 0. Ignore the possibility that an individual may carry two copies of the disease allele (for a rare disease, the probability is very low). Determine the disease-marker haplotypes for all individuals. Identify which of the grandchildren are recombinants and which are non-recombinants.



Haplotypes do not have a line in between if the phases aren't known. Seven grandchildren (301, 302, 303, 304, 305, 306, 307) are non-recombinants. Only one grandchild, individual 308, is a recombinant.

Part B

Compute by hand the LOD score for θ=0.2 for linkage between the disease locus and the genetic marker assuming that the family is segregating a rare autosomal dominant disease with full penetrance and 0 phenocopy rate, using only the recombination information available in the grandchildren that you determined in Part A. Which marker allele is segregating with the disease?

Allele 5 of the marker is segregating with the disease.

H_0: $\theta_1 = \frac{1}{2}$, The disease locus and marker locus aren't linked. They are segregating independently.
H_A: $\theta_1 < \frac{1}{2}$, The disease locus and marker locus are close together on the same chromosome.

$$LOD(\theta_1) = log_{10}\left(\frac{LOD(\theta = \theta_1)}{LOD(\theta = 0.5)}\right) = log_{10}\left(\frac{{\theta_1}^R(1-\theta_1)^N}{0.5^{R+N}}\right)$$

R = 1 recombinant
N = 7 non-recombinants

$$LOD(0.2) = log_{10}\left(\frac{0.2^1(1-0.2)^7}{0.5^8}\right) = 1.0309$$

Part C

How many additional recombinant children in addition to the recombinant and non-recombinant children already observed would we need to be present in this pedigree in order to exclude linkage at θ=0.1? Remember, you need a LOD score ≤-2 to exclude linkage.

$$log_{10}\left(\frac{0.1^{1+n}(1-0.1)^{7+n}}{0.5^{8+n}}\right) \leq -2$$

$$n = 5$$

5 more recombinant grandchildren in addition to those present in the pedigree are needed to produce a LOD score of less than -2 in order to exclude linkage.

Question 2

Use the data file HW2_siblings.csv to answer questions 2 and 3.

famid:	family ID
id1:	ID of first child
id2:	ID of second child
ibd:	Number of alleles identical-by-descent between child 1 and child 2 at marker of interest
trait1:	trait value of first child
trait2:	trait value of second child
resid1:	adjusted trait value of first child
resid2:	adjusted trait value of second child
disease1:	disease status (1=unaffected; 2=affected) of first child
disease2:	disease status (1=unaffected; 2=affected) of second child

Part A

Demine the number of affected sibling pairs sharing 0, 1 and 2 alleles IBD. You will need to restrict your analysis to siblings who both are affected (disease1=2 and disease2=2).

0 Alleles	1 Allele	2 Alleles
68 pairs	156 pairs	115 pairs

Part B

Compute the Goodness of fit test of linkage using your favorite software and/or hand computations. Compute the LOD score.

A chi-squared goodness-of-fit test was used to test whether the disease locus and marker locus were linked. The chi-squared statistic with 2 degrees of freedom was 15.1829 and the resulting one-sided p-value was 0.00025. With p-value being less than the α=0.05 significance level and LOD score being greater than 3, the null hypothesis of the two loci segregating independently was rejected. There is evidence that the disease locus and marker locus are close together on the same chromosome.

$$LOD\ score = \frac{15.1829}{2ln(10)} = 3.2969$$

Question 4

Assess linkage between the quantitative trait and the marker for which IBD information is provided using:

a. The adjusted trait value (resid1 and resid2), corresponding to the trait value adjusted for significant covariates, and the original Haseman-Elston regression. You will need to compute the squared trait difference for the adjusted residuals (resid1 and resid2).
b. The adjusted trait value (resid1 and resid2) and Haseman-Elston revisited regression. You will need to compute the product of the mean centered trait values. This will require that you know the overall mean value for this trait.

Hand in a table that shows the regression coefficient, T statistic, degrees of freedom and one-sided p-value for the models evaluated in parts a and b and your assessment of whether there is linkage. Summarize your findings and conclusions. Do you believe there is linkage to this region?

Part A

Using the Haseman-Eston approach, a linear regression analysis was used to test whether the disease locus and marker locus were linked. The slope was -0.0746, the t-statistic was -2.1615, and resulting one-sided p-value was 0.0154. The negative slope and significant p-value indicates linkage, but the LOD score is within -2 to 3, so the results are inconclusive. There is evidence that the disease locus and marker locus are close together on the same chromosome, but there is not enough power in this sample to conclude linkage.

Part B

Using the revisited Haseman-Eston approach, a linear regression analysis was used to test whether the disease locus and marker locus were linked. The slope was 0.0444, the t-statistic was 4.5529, one-sided p-value was 2.9404×10^{-6}, and resulting LOD score was 5.2884. With a positive slope, p-value being less than the α=0.05 significance level, and LOD score being greater than 3, the null hypothesis of the two loci segregating independently was rejected. There is evidence that the disease locus and marker locus are close together on the same chromosome.

	Haseman-Eston Approach	**Haseman-Eston Revisited**
regression coefficient	-0.0746	0.0444
t-statistic	-2.1615	4.935
degrees of freedom	1098 degrees of freedom	1098 degrees of freedom
one-sided p-value	0.0154	4.6274×10^{-7}
LOD score	1.0145	5.2884
conclusion	inconclusive	linkage