Question 1
Wet lab research: Characteristics of rat liver alcohol dehydrogenase - Dr. G. Bosman (10 points)

Alcohol dehydrogenases (ADHs) are ubiquitous in higher organisms. By catalyzing the first step in alcohol oxidation, they act as an important detoxification mechanism. The specificities of ADHs for various alcohols is of fundamental interest, because they catalyze not only the oxidation of ethanol, but also of a large variety of other alcohols.

A. In an experimental determination of the specificity of rat liver ADH, the following values were obtained (Table 1.1). Use these values to draw a graph that shows the relationship between velocity and substrate concentration for each substrate. (4 pts)

alcohol	**Vrel**	**Km (mM)**	
ethanol	1.0	0.64 ± 0.22	
1-butanol	1.30 ± 0.10	0.14 ± 0.03	
2-propanol	0.40 ± 0.02	36 ± 7	

Table 1.1 Kinetic constants for oxidation of alcohols by isolated rat liver alcohol dehydrogenase. Vrel is the Vmax relative to the Vmax of ethanol as a substrate.

One graph with three lines of different shapes (1-butanol higher and steeper than ethanol; 2-propanol lower and much less steep than ethanol).

B. What is the most likely explanation for the differences in Vmax and Km between 1-butanol and ethanol? (3 pts)

The structure of 1-butanol provides a better fit for 1-butanol than ethanol in the binding site of ADH (lower Km = higher affinity), and has a configuration that makes it easier to be oxidized = to lower the activation more (higher Vmax)

C. In an experiment 2-propanol is added to a reaction mixture containing ADH, NAD and ethanol (final concentration of both alcohols 1 mM). What is the effect of this addition on the initial rate of NADH production? Provide the reasons for your answer. (2 pts)

This would have no measurable/significant effect, since the Km of 2-propanol is much higher than 1 mM. This means that the affinity of ADH for 2-propanol is much lower than for ethanol, and thereby the competition with ethanol for the binding site. Also, the Vmax for propanol is considerably lower than for ethanol, so any oxidation of propanol will hardly contribute to the formation of NADH.

D. While working in a laboratory, you may need to wear gloves. Describe one situation in which the wearing of gloves is an absolute requirement; explain why. (1 pt)

When working with toxic substances (for own protection), possibility to contaminate cultures with (other) microbes, breakdown of biomolecules by skin enzymes (protein, DNA, RNA), etc (for outcome experiment)

Question 2
Modelling multiple sclerosis – dr. T. Oostendorp
(15 points)

In multiple sclerosis, periods of inflammation occur during which axons are damaged. These axons may subsequently either recover of become lost. In 2019, Montolío et al. introduced a model for this process[1]. Figure 2.1 depictures their model. denotes the number of healthy axons, the number of damaged axons, and the number of lost axons.

1 Montolío et al: A mathematical model to predict the evolution of retinal nerve fiber layer thinning in multiple sclerosis patients. *Computers in Biology and Medicine* 111 (2019) 103357.

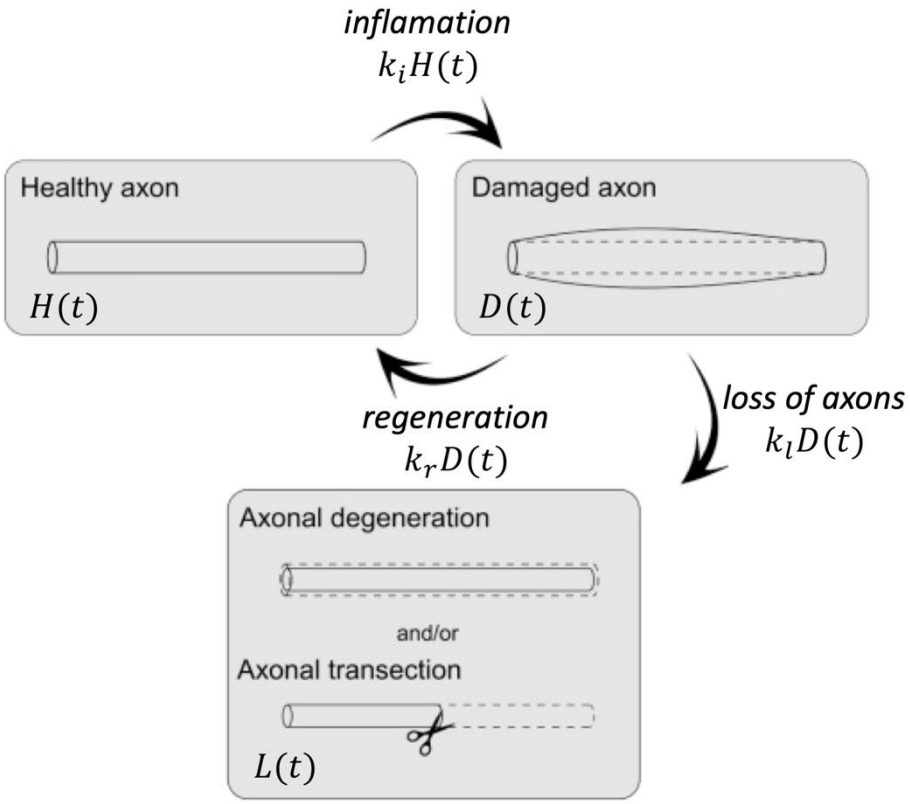


Figure 2.1 Model for multiple sclerosis.

In this model, the differential equation that describes the change in number of damaged axons is

$$\frac{d}{dt}D(t) = k_i H(t) - (k_r + k_l)D(t)$$

A. Explain why the differential equation on the previous page describes the change in number of damaged axons. (4 pt)

We start with change = in – out.

The change we are talking about is the change per time of the number of damaged axons = $\frac{d}{dt}D(t)$

In = the number of new damaged axons per time = $k_i H(t)$

Out = the number of axons per time that are not damaged anymore (i.e. regenerated or lost) = $k_r D(t) + k_l D(t)$

This yields:

$$\frac{d}{dt}D(t) = k_i H(t) - (k_r + k_l)D(t)$$

The differential equation that describes the change in number of healthy axons is

$$\frac{d}{dt}H(t) = k_r D(t) - k_i H(t)$$

If the processes of axonal degeneration and transection could be stopped, an equilibrium in the number of healthy and damaged axons would occur after time during a period of inflammation.

B. What would be the ratio $D(t)/H(t)$ in this equilibrium? (4 pts)

In equilibrium, we have $\frac{d}{dt}H(t) = 0$ (you arrive at the same answer from $\frac{d}{dt}D(t) = 0$)

If no there is no degeneration or transection, the differential equation is

$$\frac{d}{dt}H(t) = k_r D(t) - k_i H(t)$$

$$0 = k_r D(t) - k_i H(t)$$

$$k_r D(t) = k_i H(t)$$

$$D(t)/H(t) = k_i/k_r$$

Unfortunately, axonal degeneration and transection cannot be stopped. As a consequence, there will be no equilibrium with a non-zero number of healthy or damaged axons.

C. Explain why there will not be an equilibrium (3 pt).

There are two acceptable answer:

- From figure 2.1:
  Figure 2.1 shows that part of the damaged axons become irrevocably lost. This goes on while there are damaged axons. So, every period of inflammation will lead to loss of axons until there are no axons left (or rather: until so many axons have been lost that the patient dies).

- From the differential equations:
  For an equilibrium to exist, both $\frac{d}{dt}H(t)$ and $\frac{d}{dt}D(t)$ have to be zero. This implies
  $\frac{d}{dt}D(t) = k_i H(t) - (k_r + k_l)D(t) = 0$, and
  $\frac{d}{dt}H(t) = -k_i H(t) + k_r D(t) = 0$. One can rewrite these equations to
  $k_i H(t) = (k_r + k_l)D(t)$ and
  $k_i H(t) = k_r D(t)$.
  This is only true if $k_l = 0$

In an MS patient, periods of inflammation are interspersed by periods without inflammation. Consider a period without inflammation starting at $t = 0$.

D. Show that, during this period, the number of damaged axons is given by $D(t) = D_0 e^{-at}$, with D_0 the number of damaged axons at the start of the period and $a = k_r + k_l$. (4 pt)

In a period without inflammation we have

$\frac{d}{dt}D(t) = -(k_r + k_l)D(t)$

We have to show that $D(t) = D_0 e^{-at}$ is a solution to this equation. Substituting that into the left side of the equation give:

$\frac{d}{dt}(D_0 e^{-at}) = -a\, D_0 e^{-at} = -(k_r + k_l)\, D_0 e^{-at}$

The same for the right side:

$= -(k_r + k_l)D(t) = -(k_r + k_l)\, D_0 e^{-at}$

The left and the right side match, so $D(t) = D_0 e^{-at}$ is a solution to the equation.

If we substitute $t = 0$, we find $D(0) = D_0 e^0 = D_0$, so D_0 is the number of damaged axons at $t = 0$

Question 3
Recording nerve activity – dr. T. Oostendorp
(10 points)

In multiple sclerosis patients, axons lose myelin. As a result, the propagation of the action potential through the axons is delayed. In patients suspected for MS the propagation speed is recorded using the setup op figure 3.1.

At the wrist, the nerve is stimulated electrically, and at the thumb the ElectroNeuroGram (ENG) is recorded. Figure 3.2 shows a typical ENG recording.

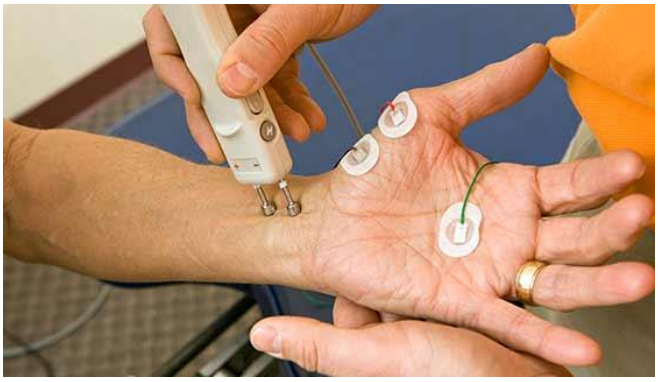


Figure 3.1 Recording nerve propagation

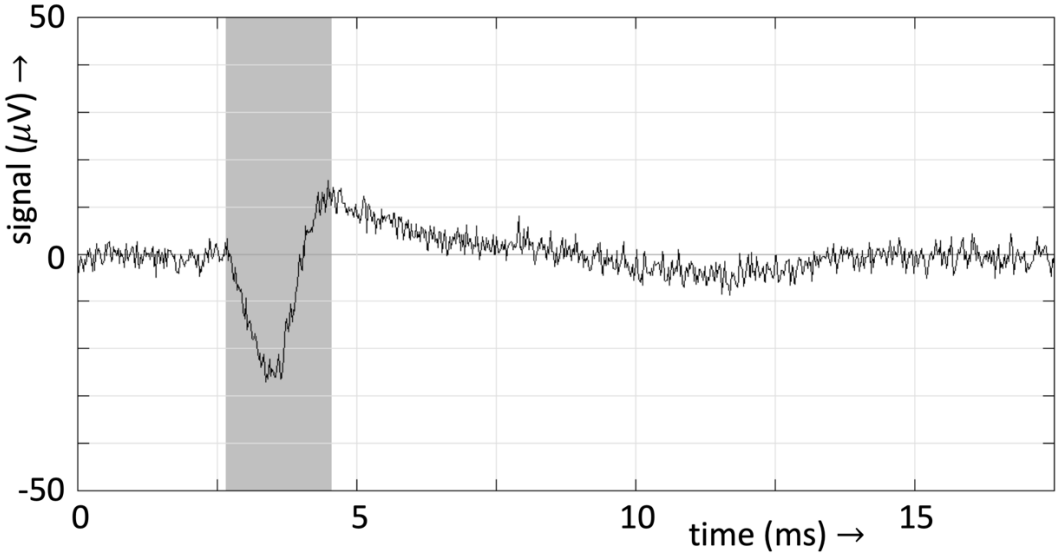


Figure 3.2 ENG recording. The selected time interval shows the Compound Action Potential (CAP): the effect of the action potential in the axons of the nerve passing between the recording electrodes.

Figure 3.3 shows the spectrum of an almost noise free ENG.

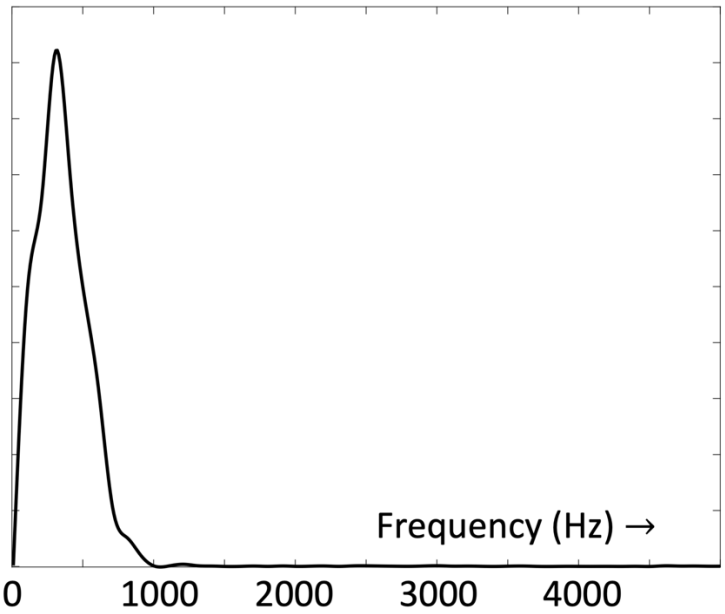


Figure 3.3 The spectrum of an almost noise free4 ENG recording (obtained by averaging 1000 individual recordings).

A. The strongest frequency in the spectrum is about 300 Hz. Explain why that is (more or less) consistent with the recording of figure 3.2. (3 pt)

The CAP looks a bit like halve a period of a sine function. The duration of the CAP is about 2 ms. This means the period of the corresponding sine function is 4 ms, which corresponds to a frequency of 1/0.004 = 250 Hz.

B. What sample rate should be used to record the ENG so that the CAP is clearly visible? Explain your answer. (2 pts)

The maximum frequency in the spectrum is 1000 Hz. At least 10 samples are necessary to render the highest frequency well. Consequently, the sample rate should at least be 10.000 Hz.

C. As figure 3.2 shows, actual ENG recordings contain quite some noise. What kind of filter should be used to improve signal-to-noise ratio? Explain your answer. (3 pt)

The noise that is visible in figure 3.2 has clearly a higher frequency that the CAP. Consequently, a low-pass filter should be used.

D. What should be the cut-off frequency of that filter? Explain your answer. (2 pt)

The filter should suppress the noise as much as possible, but should also leave the signal untouched. The highest frequency in the signal (the CAP) is about 1000 Hz, so a cut-off frequency of 1000 Hz is optimal.

Question 4
Epidemiology - dr. F. de Vegt
(20 points)

Use '<u>Abdollahpour</u> et al. *Lifestyle factors and multiple sclerosis: A population-based incident case-control study- abstract and table 1'* for question 4 and 5.

A. In their abstract, Abdollahpour and colleagues say that "smoking is an established risk factor of multiple sclerosis". Explain in epidemiological terms what this statement about smoking as a risk factor means. (3 pts)

Answer – Two issues to be mentioned: disease likelihood and comparison. The risk of developing MS in smokers is higher than the risk in non-smokers. So the incidence of MS is higher in smokers compared to non-smokers. The RR and OR will be higher than 1.

B. In this Iranian study, 547 incident cases and 1057 general population controls were included. What is, in this study, the meaning of 'cases', the adjective 'incident', and 'general population controls'? How might 'general population controls' have been collected? (4 pts)

Answer – Cases, people diagnosed with multiple sclerosis. Incident, newly diagnosed MS patients, not prevalent cases, i.e. patients being diagnosed years earlier, and treated. General population controls are people sampled from the general population, not having MS in principle. The latter might be enrolled with help of administrative registers, municipalities, etc.

C. The authors reported that lifetime drug abuse was associated with multiple sclerosis: OR of ever vs never use: 2.93 and 95% CI: 1.83-4.70. How should these numerical results be interpreted? (3 pts)

Answer – OR = 2.93 means that people with drug abuse have three-fold risk of developing MS compared to people unexposed to drugs. The 95% confidence interval, (1.83 – 4.70), appears to be fairly wide, indicating that the data are compatible with, still, a vast range of possible values: two-fold to four-fold elevated risk, meaning that drug abuse might double or even quadruple MS-risk. No increased risk seems to be implausible.

D. Is gender a risk factor of multiple sclerosis? Of the 574 cases, 401 (73%) were female; of the 1057 controls, the figures were 544 (51%). Fill out the two-by-two table below (2 pts).

	MS	No MS	total
female	401	544	945
male	173	513	686
total	574	1057	1631

E. Use the two-by-two table (see D1) to calculate the odds ratio of multiple sclerosis and gender, and give a proper interpretation of the result. (4 pts)

Answer – OR = (401/173) : (544/513) = 2.19. Alternative calculation, OR = (73%/27%) : (56%/44%) = 2.12. Interpretation: women obtain a two-fold risk of MS compared with men (and not the other way around, men having higher risk).

F. Instead of a case-control study, researchers could have designed a cohort study to investigate the causal relation between drug abuse and the development of multiple sclerosis. How does a cohort study look like for this research question? Which association measures can be calculated? (4 pts)

Answer – Assemble a cohort of people not having MS, and measure drug use and abuse at the start of the follow-up. Then follow for a couple of years to see who will develop MS. Data analysis is calculating cumulative incidence of MS in sub-cohorts of

yes/no drug abuse. Association measures which can be calculated are the AR (attributable risk) [also called RD, Risk Difference] and the RR (relative risk).

Question 5
Statistics– dr. R. Donders
(15 points)

Use '<u>Abdollahpour</u> et al. *Lifestyle factors and multiple sclerosis: A population-based incident case-control study-abstract and table 1'* for question 4 and 5.

A. One of the variables studied is the amount of physical activity, expressed in Metabolic Equivalent of Task (MET) per week (a positive number). Physical activity is presented as a continuous variable in Table 1 with an average of 3199.4 MET/week and a standard deviation of 4222 MET/week for the 1057 participating general population controls. Explain why it is clear that this variable is likely not normally distributed in the general population.(3 pt)

If physical activity would be normally distributed approximately 95% of all values should be between the population mean minus 2 times the standard deviation and the population mean plus 2 times the standard deviation. Since the standard deviation is larger than the mean, this would imply that negative numbers should occur, which is impossible as physical activity is a positive number by definition.

B. What is the sampling distribution of the sample mean of physical activity? (3 pts)

Even though psychical activity is heavily skewed, the large sample size together with the Central Limit Theorem sort of guarantee that the sampling distribution will be a normal distribution.

C. Assume that the population mean and standard deviation are equal to the sample estimates. What would be the standard error of the sample mean? (4 pts)

The standard error of a sample mean is equal to $\sigma/\sqrt{n}$. In this case σ is equal to 4222 and n is equal to 1057. So the standard error is $4222/\sqrt{1057} = 129.9$ MET/week.

D. The measurements of physical activity are based on recollections (memory) of the participants. This implies that these measurements might contain a larger measurement error when compared to a situation where instant physical activity would be recorded. Suppose that the fact that these measurements are based on recollection would only imply a larger random measurement error. Would this cause problems when comparing MS cases with the controls with respect to physical activity? (3 pts)

When you want to compare the cases to the controls you really want to compare the population mean for the cases to the population mean for the controls. Each individual measurement can be seen as a measurement of the population mean Viewed in this manner all deviations of individuals from the population mean can be viewed as random measurement error. By increasing the sample size we can make the measurement error of the average as small as we want to (See Duplo, triplo, … measurements). When recollection would cause an increase in random measurement error, it would only imply that on average the individuals are further away from the true value/population mean. This problem can be overcome by increasing the sample size so this would not cause problems that cannot be overcome.

E. Suppose that recollection would lead to a systematic measurement error. Would this cause problems when comparing MS cases with the controls with respect to physical activity? (2 pts)

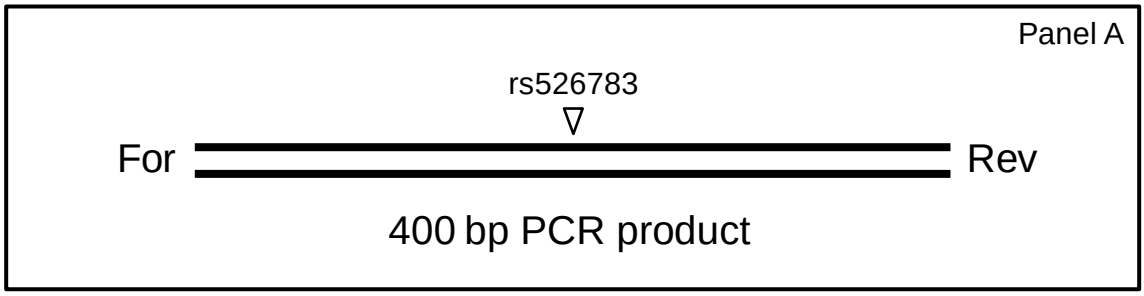
That depends on the way the systematic measurement error would work out. Suppose that recollection would lead to a systematic overestimation of physical activity. As long as this effect is identical (within probability limits) in both populations, there is no problem when comparing both groups. However, when for MS cases the recollections would be nearer to the truth (so a smaller overestimation) than for the controls, problems might arise!

Genetic Lab Practice – Dr. Diederik de Bruijn

(10 points)

Introduction:
The occurrence of red hair in the human population is mainly caused by genetic factors, that are generally inherited in an autosomal recessive fashion. Literature data has shown that the rs526783 variant may be one of these genetic factors. This variant is an indel, in this case an insertion of 2 bases (GC) into the promoter of a gene that is involved in pigmentation. Allele 1 (A1) is the ancestral sequence (without the insertion), allele 2 (A2) is the variant sequence (with the insertion). It has been reported that the variant allele is associated with red (and reddish) hair color.

Your aim is to verify the association between rs526783 and red hair in a (randomly selected) cohort of 400 people from the English population, who were genotyped with PCR and Sanger sequencing. In short, PCR products of 400 basepairs (with rs526783 in the middle) were generated with a forward (For) and a reverse (Rev) primer (panel A). The central part of this PCR product, encompassing fifty basepairs of double stranded DNA (with rs526783 in the middle) is shown in panel B. Hydrogen bonds between the bases of both strands are depicted with single vertical lines ("|"). The rs526783 variant is an insertion of 2 bases (GC) between the T and C nucleotide at the location that is marked in panel B.



A. A PCR product from a person who is heterozygous (A1/A2) for the rs526783 variant was sequenced with the Rev primer. Describe what you should see in the sequencing results. In your answer, show the variant itself and a minimum of six bases on either side of it. If needed, you can use a drawing to illustrate your answer. (2 pts)

Answer: TGGCTG(G/A)(C/C)(A/T)(C/C)(T/A)(C/T)(A/C)(T/A), last part is double sequence, mixed between with/without insertion.

B. In light of your previous answer, give the most important reason to explain why it is necessary to sequence all samples in your cohort from both sides (with the For and Rev primer). (1 pt)

Sequencing from both sides is the best way to be sure about the sequence in heterozygotes. The sequence can only be read unambiguously until the place of the insertion.

The distribution of rs526783 genotypes in your cohort, grouped according to hair color, is shown in Box 1.

Box1	A1/A1	A1/A2	A2/A2	Total
Red	2	12	5	19
Red-brown	1	9	4	14
Red-blonde	3	8	1	12
Blonde	118	41	3	162
Brown/Black	126	65	2	193
Total	250	135	15	400

C. Use a calculation to explain why the variant allele (A2) should not be classified as a mutation in this population. (1 pt)

Number of A2 alleles: 2x15 + 135 = 165
Total number of alleles: 2x400 = 800
A2 allele frequency = 165/800 = 0,20625. This is higher than 1%, so not a mutation

As you have learned during the Genetic Lab Practice, genotype-phenotype correlations can be interpreted with the help of a 2x2 table (see Box 2). In order to do that, you will need to divide the phenotypes and the genotypes in two groups each.

D. Use information from the introduction to divide all hair phenotypes into two (logical) groups and name these groups in the column headers of box 2. Use one sentence to explain why you chose this division. (2 pt)

Reddish (Red; Red blonde and Red brown) in one group, non-Reddish (Brown and Blonde) in the other group. Reason: the introduction mentions that red and reddish hair may be associated with the variant allele.

E. Use information from the introduction to divide all rs526783 genotypes into two (logical) groups and name these groups in the row headers of box 2. Use one sentence to explain why you chose this division. (2 pt)

A1/A1 and A1/A2 in one group; A2/A2 in the other group. This division is based on the recessive inheritance of the variant allele that is mentioned in the introduction.

F. Fill the remaining cells of box 2 with the (summed up) genotype numbers from box 1, make sure that you end up with a total of 400. Explain whether or not there is evidence for a genotype-phenotype correlation in these data. (2 pt)

Box 2	Reddish	Non-Reddish
A2/A2	*5 + 4 + 1 = 10*	*3 + 2 = 5*
A1/A1 + A1/A2	*2+1+3+12+9+8=35*	*118+126+41+65=350*

Answer: Yes, the ratio reddish/non-reddish >1 for A2/A2, but <1 for the other genotypes