

---

## MATH 4330 — Assignment 1

**Instructor:** Kevin McGregor

**Due Date:** Friday Sept 23rd, 2022, 11:59 PM

---

**Instructions:** Hand in all questions to Crowdmark by the deadline. For written questions, you may upload a scan or photo of your work; however, I encourage you to use  $\text{\LaTeX}$  (though it's not a requirement). **Show your work, and make sure that all work you submit is your own.**

**Question 1:** A researcher wants to study whether an existing drug causes a particular kind of skin reaction. She collects data from an existing medical database; she samples individuals who showed the skin reaction as well as individuals without a reaction. She then examines how many individuals had taken the drug of interest.

- (a) **[2 points]** What kind of study is this? Be as specific as possible, and explain your answer.
- (b) **[2 points]** Suppose the researcher finds that those who took the drug had a higher chance of developing a skin reaction. Do you believe this result? Can you think of other specific factors that might affect the conclusion? Explain your answers.

**Question 2:** **[2 points]** A researcher runs a randomized experiment where study participants are randomized to be given either a drug or a placebo. Another researcher wants to perform an additional study on this same study group. He asks participants whether they get more or less than 3 hours of exercise per week. He concludes that individuals with more than 3 hours of exercise per week have lower blood pressure than those who get less than 3 hours, on average. Are you worried about confounders in this conclusion? Explain.

**Question 3:** We want to study the genetics of colour distribution in a population of unicorns. Suppose that in unicorns there is a single gene that determines colour; there are two variants of the gene, one called  $A$  and the other called  $a$ . Each unicorn has two copies of the gene (one from each parent). The combination of the gene variants for copy 1 and copy 2 of the gene determines colour as follows:

Copy 1	Copy 2	Colour
$A$	$A$	Red
$A$	$a$	Pink
$a$	$A$	Pink
$a$	$a$	White

We want to determine if “random mating” is happening in this unicorn population. This would mean that the probability that a newly born unicorn inherits variant  $A$  or  $a$  with probability equal to the prevalence of each variant in the overall population.

- (a) **[3 points]** Let  $p$  be the proportion of the  $A$  variant in the population, and  $q$  be the proportion of the variant  $a$ , so that  $q = 1 - p$ . Under random mating, each newborn unicorn inherits  $A$  with probability  $p$  and  $a$  with probability  $q$  and the two copies inherited in each unicorn are independent of one another. Calculate the expected proportions of unicorn colours under random mating.
- (b) **[5 points]** Suppose you know that  $p = 0.75$  and  $q = 0.25$ , and that you collect the following data from a sample of unicorns:

Colour	Number of unicorns
Red	45
Pink	49
White	12

Use the appropriate statistical test to determine whether the random mating assumption holds in this population. **Write out your calculations for this part; using R for this part will not result in credit.**

### Question 1

A researcher wants to study whether an existing drug causes a particular kind of skin reaction. She collects data from an existing medical database; she samples individuals who showed the skin reaction as well as individuals without a reaction. She then examines how many individuals had taken the drug of interest.

- (a) What kind of study is this? Be as specific as possible, and explain your answer.
- (b) Suppose the researcher finds that those who took the drug had a higher chance of developing a skin reaction. Do you believe this result? Can you think of other specific factors that might affect the conclusion? Explain your answers.

### Solution

- (a) This study would be considered a retrospective study. The reason would be that the researcher is looking into an already existing database system, hence she is looking into the past for her subjects. Based on this medical database system, she is sampling individuals who already showed a skin reaction vs. individuals that did not have a reaction.
- (b) It is possible to have this result. However there might be certain factors that can affect the conclusion by the researcher. For example, seasonality, where exposure to high levels of sunlight can cause skin reaction, food allergies can cause skin reactions, and finally prior infections and diseases could have caused similar skin reaction. All of these have no relation to the drug that was used by the patients, however they were still able to get skin reactions though not from the drug. This can cause some problems with the way the experiment was held. However depending on the sample size, these may be outlier situations, and the drug does indeed cause a development of skin reaction.

## Question 2

A researcher runs a randomized experiment where study participants are randomized to be given either a drug or a placebo. Another researcher wants to perform an additional study on this same study group. He asks participants whether they get more or less than 3 hours of exercise per week. He concludes that individuals with more than 3 hours of exercise per week have lower blood pressure than those who get less than 3 hours, on average. Are you worried about confounders in this conclusion? Explain.

## Solution

Yes there is a worry about confounders in this conclusion. The first confounder that can be thought off is age of the participants. Age is a factor because older people may have less time to exercise due to work, family etc. and with age, high blood pressure is quite common than younger people. Because of the age confounder, it may seem that less hours of exercise will cause higher blood pressure, however it could just be the age of those individuals that may be causing the high blood pressure.

### Question 3

Question 3: We want to study the genetics of colour distribution in a population of unicorns. Suppose that in unicorns there is a single gene that determines colour; there are two variants of the gene, one called  $A$  and the other called  $a$ . Each unicorn has two copies of the gene (one from each parent). The combination of the gene variants for copy 1 and copy 2 of the gene determines colour as follows:

Copy 1	Copy 2	Colour
$A$	$A$	Red
$A$	$a$	Pink
$a$	$A$	Pink
$a$	$a$	White

We want to determine if “random mating” is happening in this unicorn population. This would mean that the probability that a newly born unicorn inherits variant  $A$  or  $a$  with probability equal to the prevalence of each variant in the overall population.

- Let  $p$  be the proportion of the  $A$  variant in the population, and  $q$  be the proportion of the variant  $a$ , so that  $q = 1 - p$ . Under random mating, each newborn unicorn inherits  $A$  with probability  $p$  and  $a$  with probability  $q$  and the two copies inherited in each unicorn are independent of one another. Calculate the expected proportions of unicorn colours under random mating.
- Suppose you know that  $p = 0.75$  and  $q = 0.25$ , and that you collect the following data from a sample of unicorns:

Colour	Number of unicorns
Red	45
Pink	49
White	12

Use the appropriate statistical test to determine whether the random mating assumption holds in this population. **Write out your calculations for this part; using R for this part will not result in credit.**

### Solution

(a)

$$\begin{aligned} E(\text{Red Unicorn Colour}) &= \text{proportion of } A \times \text{proportion of } A \\ &= (p \times p) \\ &= p^2 \end{aligned}$$

$$\begin{aligned} E(\text{Pink Unicorn Colour}) &= \text{proportion of } a \times \text{proportion of } A \\ &= (q \times p) \\ &= (qp) \\ &= p(1 - p) \end{aligned}$$

$$\begin{aligned} E(\text{White Unicorn Colour}) &= \text{proportion of } a \times \text{proportion of } a \\ &= (q \times q) \\ &= q^2 \\ &= (1 - p)^2 \end{aligned}$$

(b) Let  $n = 45 + 49 + 12 = 106$ .

$$\begin{aligned} E(\text{Red Unicorn colour}) &= (0.75)^2 \times 106 \\ &= 59.625 \end{aligned}$$

$$\begin{aligned} E(\text{Pink Unicorn Colour}) &= (0.75) \times (0.25) \times (106) \\ &= 19.875 \end{aligned}$$

$$\begin{aligned} E(\text{White Unicorn Colour}) &= (0.25)^2 \times 106 \\ &= 59.625 \end{aligned}$$

$$\begin{aligned} \chi^2 &= \frac{(45 - 59.625)^2}{59.625} + \frac{(49 - 19.875)^2}{19.875} + \frac{(12 - 59.625)^2}{59.625} \\ &= 84.3074 \end{aligned}$$

 $H_0$ : Random mating does occur in the unicorn population $H_a$ : Random mating does not occurLet  $\alpha = 0.05$  and  $df = 3 - 1 = 2$

Using R to calculate the test statistic, we get that the p-value is  $4.9304 \times 10^{-19}$  which is well below 0.05. Hence we reject  $H_0$  and say that there is no evidence to show that random mating does occur in the unicorn population.

# Question 1

**Question 1:** A researcher wants to study whether an existing drug causes a particular kind of skin reaction. She collects data from an existing medical database; she samples individuals who showed the skin reaction as well as individuals without a reaction. She then examines how many individuals had taken the drug of interest.

- (a) [2 points] What kind of study is this? Be as specific as possible, and explain your answer.
- (b) [2 points] Suppose the researcher finds that those who took the drug had a higher chance of developing a skin reaction. Do you believe this result? Can you think of other specific factors that might affect the conclusion? Explain your answers.

A) This study would be considered a retrospective study. The reason would be that the researcher is looking into an already existing database system, hence she is looking into the past for her subjects. Based on this medical database system, she is sampling individuals who already showed a skin reaction vs individuals that did not have a reaction.

B) It is possible to have this result. However there might be certain factors that can affect the conclusion by the researcher. For example, seasonality, where exposure to high levels of sunlight can cause skin reaction, food allergies can cause skin reactions, and finally, prior infections and diseases could have caused similar skin reaction. All of these have no relation to the drug that was used by the patients, however they were still able to get skin reactions though not from the drug. This can cause some problems with the way the experiment was held. However depending on the sample size, these may be outlier situations, and the drug does indeed cause a development of skin reaction.



## Question 2

**Question 2:** [2 points] A researcher runs a randomized experiment where study participants are randomized to be given either a drug or a placebo. Another researcher wants to perform an additional study on this same study group. He asks participants whether they get more or less than 3 hours of exercise per week. He concludes that individuals with more than 3 hours of exercise per week have lower blood pressure than those who get less than 3 hours, on average. Are you worried about confounders in this conclusion? Explain.

Yes there is a worry about confounders in this conclusion. The first confounder that can be thought off is age of the participants. Age is a factor because older people may have less time to exercise due to work, family etc. and with age, high blood pressure is quite common than younger people. Because of the age confounder, it may seem that less hours of exercise will cause higher blood pressure, however it could just be the age of those individuals that may be causing the high blood pressure.

## Question 3

**Question 3:** We want to study the genetics of colour distribution in a population of unicorns. Suppose that in unicorns there is a single gene that determines colour; there are two variants of the gene, one called  $A$  and the other called  $a$ . Each unicorn has two copies of the gene (one from each parent). The combination of the gene variants for copy 1 and copy 2 of the gene determines colour as follows:

Copy 1	Copy 2	Colour
$A$	$A$	Red
$A$	$a$	Pink
$a$	$A$	Pink
$a$	$a$	White

We want to determine if “random mating” is happening in this unicorn population. This would mean that the probability that a newly born unicorn inherits variant  $A$  or  $a$  with probability equal to the prevalence of each variant in the overall population.

- (a) [3 points] Let  $p$  be the proportion of the  $A$  variant in the population, and  $q$  be the proportion of the variant  $a$ , so that  $q = 1 - p$ . Under random mating, each newborn unicorn inherits  $A$  with probability  $p$  and  $a$  with probability  $q$  and the two copies inherited in each unicorn are independent of one another. Calculate the expected proportions of unicorn colours under random mating.
- (b) [5 points] Suppose you know that  $p = 0.75$  and  $q = 0.25$ , and that you collect the following data from a sample of unicorns:

Colour	Number of unicorns
Red	45
Pink	49
White	12

Use the appropriate statistical test to determine whether the random mating assumption holds in this population. **Write out your calculations for this part; using R for this part will not result in credit.**

$$\begin{aligned}
 a) \ E(\text{Red Unicorn Colour}) &= \text{proportion of } A \times \text{proportion of } A \\
 &= (p \times p) \\
 &= p^2
 \end{aligned}$$

$$\begin{aligned}
 E(\text{Pink Unicorn Colour}) &= \text{proportion of } a \times \text{proportion of } A \\
 &= (q \times p) \\
 &= (qp) = p(1-p)
 \end{aligned}$$

$$\begin{aligned}
 E(\text{White Unicorn Colour}) &= \text{proportion of } a \times \text{proportion of } a \\
 &= (q \times q) \\
 &= q^2 = (1-p)^2
 \end{aligned}$$

$$\begin{aligned}
 b) \ E(\text{Red Unicorn colour}) &= (0.75)^2 \cdot 106 \\
 &= 59.625
 \end{aligned}$$

$$\begin{aligned}
 \text{let } n &= 45 + 49 + 12 \\
 &= 106
 \end{aligned}$$

$$E(\text{Pink Unicorn colour}) = (0.75)(0.25)(106) \\ = 19.875$$

$$E(\text{White Unicorn colour}) = (0.25)^2 \times 106 \\ = 6.625$$

$$\chi^2 = \frac{(45 - 59.625)^2}{59.625} + \frac{(49 - 19.875)^2}{19.875} + \frac{(12 - 6.625)^2}{6.625} \\ = 50.6281$$

$H_0$ : Random mating does occur in the unicorn pop<sup>n</sup>

$H_a$ : Random mating does not occur

Let  $\alpha = 0.05$  and  $df = 3 - 1 = 2$

Using R to calculate the test statistic, we get that the p-value is  $1.01449 \times 10^{-11}$

which is well below 0.05. Hence we reject  $H_0$  and say that there is no evidence to show that random mating does occur in the unicorn pop<sup>n</sup>.