Assignment 8
Due on 9th June, 11 p.m.

1. **Vaccine**: It is well established that vaccines are critical in fighting and eradicating infectious diseases in the world. ***Vaccine efficacy*** is one commonly used metric to evaluate how good a vaccine is. It is defined as:

$$\text{vaccine efficacy} = \frac{ARU - ARV}{ARU} \times 100\%$$

where

- ARU = attack rate of unvaccinated people, which is basically the proportion of people with the disease among unvaccinated people.
- ARV = attack rate of vaccinated people, which is the proportion of people with the disease among vaccinated people.

1.1) **(5 points)** **True or False**: read the following statements, and put a **T** in front of the statements if you think they are correct; put an **F** otherwise:

[ ] ARU is the risk of the disease in unvaccinated people

[ ] ARU is the odds of the disease in unvaccinated people

[ ] ARV is the risk of the disease in vaccinated people

[ ] ARV is the odds of the disease in vaccinated people

[ ] The ratio of ARV/ARU is the relative risk of the disease for the vaccinated people compared to the unvaccinated people

[ ] The ratio of ARV/ARU is the relative risk of the disease for the unvaccinated people compared to the vaccinated people

[ ] The ratio of ARV/ARU is the odds ratio of the disease for the vaccinated people compared to the unvaccinated people

[ ] The ratio of ARV/ARU is the odds ratio of the disease for the unvaccinated people compared to the vaccinated people

[ ] For a good vaccine, the ratio ARV/ARU should be small.

1.2) **(2.5 points)** We can easily see that:

$$\text{vaccine efficacy} = \left(1 - \frac{ARV}{ARU}\right) \times 100\%$$

Which of the following is true about $\frac{ARV}{ARU}$?

(A) It is the risk of the disease for all people.

(B) It is the relative risk of the disease for the vaccinated people compared to the unvaccinated people.

(C) It is the odds of the disease for all people.

(D) It is the odds ratio of the disease for the vaccinated people compared to the unvaccinated people.

In late 2019 and early 2020, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and the resulting coronavirus disease 2019 (Covid-19) started to spread the world and affect the lives of all people in a worldwide pandemic. In late 2020, BioNTech and Pfizer developed their vaccine (BNT162b2) against COVID-19. The following text is copied from their publication in the journal *The New England Journal of Medicine* about BNT162b2:

"... ..., of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; ... ..."

from **Polack *et al.* NEJM, 383;27**.

1.3) **(5 points)** Based on the information above, summarise the data by completing the contingency table below:

	BNT162b2	Placebo	**Total**
COVID-19			
Healthy			
Total			43,448

1.4) **(5 points)** Calculate the efficacy of the BNT162b2 vaccine based on the data.

1.5) **(15 points)** There are multiple ways of assessing whether the protection offered by the vaccine is statistically significant or not. Here, let's use the **chi-square test**. Write the null and alternative hypotheses, compute the expected table, the test statistic and the $p-$value.

1.6) **(2.5 points)** What assumptions do you need to check before you actually perform the **chi-square test**.

2. Read the paper **"Using Effect Size - or Why the P Value Is Not Enough"** by Sullivan and Feinn. Answering the following questions:

2.1) **(5 points)** During the lectures, we repeatedly used the example from the our medical school, where the authors used the **chi-square** test to figure out whether there is any relationship between ABO blood groups and COVID-19 susceptibility. What is the effect size if we look at the blood type A group in their test?

2.2) **(2.5 points)** Which of the following is true when we look at the blood type A group?

A) The p-value is large, and the effect size is large

B) The p-value is large, and the effect size is small

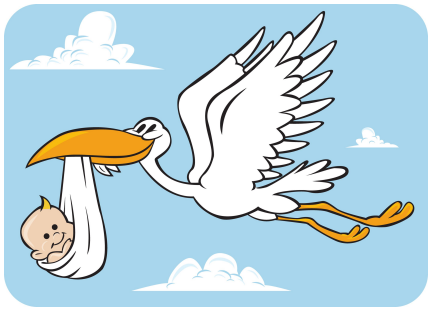
C) The p-value is small, and the effect size is large

D) The p-value is small, and the effect size is small

2.3) **(5 points)** Now, look at the first question in this assignment regarding Pfizer's BNT162b2 vaccine. What is the effect size of BNT162b2 and is it large, medium or small?

2.4) **(5 points)** Finally, in your own words, explain: what does effect size tell us, and why should we care about effect size ?

3. The white stork (*Ciconia ciconia*) is a large bird in the stork family, *Ciconiidae*. Its plumage is manly white, with black on the bird's wings. According to European folklore, the stork is responsible for bringing babies to new parents. The legend is very ancient, but was popularised by a 19th-century Hans Christian Andersen story called "**The Storks**".



The above text is from the "White Stork" entry of Wikipedia. The lore appears in many countries in Europe, North Africa and the Middle East. Read the paper "**Storks Deliver Babies ($p = 0.008$)**" by Robert Matthews and complete the following questions using **Table 1** from the paper:

3.1) Reproduce the results from the paper:

i. **(5 points)** Write down the equation of Pearson's Correlation Coefficient r (just choose one), and calcuate r between "**Storks (pairs)**" and "**Birth rate** ($10^3/yr$)" using the equation you choose.

ii. **(2.5 points)** Is there a positive or negative linear relationship between "**Storks (pairs)**" and "**Birth rate** $(10^3/yr)$"?

iii. **(5 points)** To check if there is a linear relationship, you can perform a statistical test. Write down the null and alternative hypotheses, calculate the test statistic and the $p-$value. Based on the $p-$value, do you reject the null hypothesis?

iv. **(7.5 points)** Perform a simple linear regression between "**Storks (pairs)**" and "**Birth rate** $(10^3/yr)$" using OLS. Write out the equations (choose one for each) for calculating the slope, the intercept and the r^2. Then calculate them based on the equations you choose.

v. **(2.5 points)** Compare your reults to those in the paper. Are they the same? If not, explain why.

3.2) **(10 points)** Repeat the analyses of (i), (ii), (iii) and (iv) from **3.1)** using "**Area** (km^2)" and "**Birth rate** $(10^3/yr)$" as variables.

3.3) **(10 points)** Explore the data on your own. Then in your own words, try to explain why there seems to be a positive relationship between "**Storks (pairs)**" and "**Birth rate** $(10^3/yr)$"

3.4) **(5 points)** Hopefully, through the above practice, we can agree that correlation does not necessarily indicate causation. Using your own words, explain what a confounding variable (or sometimes simply called a confounder) is.