**Assignment 3**

1. **Comparing two proportions:** We will use data reported in the randomized controlled trial (RCT) by Perkins et al. for this question. Relevant information is in the abstract and in the supplementary appendix section on sample size calculation (see text within red rectangle).
   * What were the intervention under study and control in this RCT? (0.5 marks)
   * What was the primary outcome of interest? (0.5 marks)
   * When planning the study, the authors made certain assumptions about the expected probability of the primary outcome in the two group. Depict their plans using a 2-by-2 table and show how they calculated the anticipated 95% confidence interval around the risk ratio. (4 marks)
   * Equivalent to the function power.t.test() seen earlier in the course, the function power.prop.test() serves to determine the power available for a given sample size. What was the power available for the planned sample size? (2 marks)
   * As the abstract shows, the observed probability of the outcome was much lower. What was the actual power available as a result? (2 marks)
   * Calculate the odds ratio and 95% confidence interval for the final result reported in the abstract. What is your interpretation of the odds ratio and its confidence interval? (4 marks)
   * Though the study was planned with the risk ratio in focus, the authors finally reported an odds ratio. Was this appropriate and why? (1 mark)
   * Carry out the hypothesis test for the null hypothesis that the odds ratio is 1 and obtain a p-value. What is your conclusion? (2 marks). Explain whether an exact test is necessary or if an approximate test is adequate (1 mark).
2. **Chi-square and Fisher exact tests:** This question is based on the publication by: Stub T et al. Complementary and conventional providers in cancer care: experience of communication with patients and steps to improve communication with other providers. BMC Complement Altern Med. 2017 Jun 8;17(1):301. Complimentary therapies included acupuncture, massage therapy and reflexology. Conventional providers include oncologists, nurses and family physicians.
   * Replicate the chi-square test of the association between Location and Type of provider (last comparison in Table 3). (4 marks)
   * Calculate the expected values for the contingency table between “How often do you ask your cancer patients if they use complimentary therapy (CT)?” and Type of provider (first comparison in Table 4). Is it appropriate to use the chi-square test? (2 marks)
   * For the previous question, replicate the Fisher exact test using R. How do you interpret the p-value? (2 marks)
3. **Non-parametric methods:** This question is based on the study by Corbett, K.S., Edwards, D.K., Leist, S.R. *et al.* SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586,**567–571 (2020). Take a look at Figure 2f.
   * What was the objective of this analysis? (1 mark)
   * Repeat the normal-approximation based Mann-Whitney (i.e. Wilcoxon rank sum) test in this figure using the dataset provided. State the null and alternative hypotheses, calculate the test statistic, report the p-value and state your conclusion assuming a Type I error level of 0.05. (4 marks)
   * Use the t.test() function in R to carry out a t-test instead. What do you conclude? (2 marks) Is the t-test appropriate for use in this setting? (1 mark)
   * Use the R program provided in your lecture to estimate an 80% confidence interval for the median antibody concentration in each of the two groups. (2 marks)