Fundamental Mathematics and Statistics for Health Data Science

Assessment 1:

*** Released on 5 Nov 2021, due on 3 Dec 2021 at 12 noon (UK time) ***

You have been provided with data of the blood pressure of all patients over 30 years old attending a drop-in service over the course of a month (one visit per patient). The columns provide data on the patient's identification number ('patient'), sex ('sex'), age group ('agegrp'), week of visit (WeekVisit (1-4)), blood pressure ('bp') and whether they were referred to another service or not ('Referral' (1=Referred, 0=Not referred)).

Questions

1. Produce appropriate initial summaries of each of the variables and comment on the summaries. Include discussion on whether blood pressure appears to follow a normal distribution. (20%)

2. Plot the probability mass function of X from $Bin(n,p)$ with n = the sample size of the data set, and p = 0.40. Repeat this with p = 0.55. Discuss how well you feel each binomial distribution matches the 'Referral' variable. (12%)

3. Calculate the 95% confidence interval for the true population mean blood pressure (bp) (using the mean, standard deviation and n of our sample size). Does the confidence interval contain a bp of 146? How would you interpret this? (8%)

4. Test if the proportion of male patients is 0.5. (10%)

5. Calculate odds ratio of association between sex and referral. How would you interpret it? Is there statistical evidence of an association between sex and referral? (15%)

6. Is there statistical evidence of a blood pressure difference between men and women? (15%)

7. Is there statistical evidence of a blood pressure difference between age groups? (20%)

Please specify the hypotheses clearly for each of the tests, and include your R commands with the answers.