STAT7203: Applied Probability and Statistics
Assignment 2

Due by 14:00 on Friday the 30^{th} of October, 2020. Submission via Blackboard.

The marks for each question are indicate by the number in square brackets. There are a total of 15 marks for this assignment.

1. Let $X_1, \ldots, X_n$ be a simple random sample where the X_i have a Poisson distribution with mean 1 and define $\bar{X} = n^{-1}(X_1 + X_2 + \cdots + X_n)$. Using Markov's inequality it can be shown that for any $a \in \mathbb{R}$ and any $\mathbf{t} > \mathbf{0}$ for which the expectations below exist

$$\mathbb{P}(\bar{X} \geq a) = \mathbb{P}(e^{\bar{X}t} \geq e^{at}) \leq \frac{\mathbb{E}(e^{t\bar{X}})}{e^{at}}.$$

Recall the moment generating function of a Poisson distribution with mean λ is $M(t) = \exp(\lambda(e^t - 1))$.

   (a) Define $\mathcal{H}(t; a) := e^{-at}\mathbb{E}(e^{t\bar{X}})$. For a fixed value of a, find the value t_a which mimimises $\mathcal{H}(t; a)$. [2 marks]

   (b) Let $n = 100$ and $a = 1.1$. Compute the bound on $\mathbb{P}(\bar{X} \geq 1.1)$ and compare this with the approximation from the central limit theorem. [2 marks]

2. A study conducted in the USA investigated the general public's perception of risk in participating in political activities online. Telephone numbers of 758 adults were randomly selected and respondents were asked the following question: If a person posted a message on an Internet news site about a national issue, how likely is it that the message would be edited in a way that changes its original meaning? The responses, split by age category are summarised below.

	Likely	Neutral	Unlikely
18 - 39 years old	221	46	113
Over 40 years old	201	30	147

   Based on the table above, is there any evidence of an association between age and the perceived likelihood of their message being edited in a way that changes its meaning? State the null and alternative hypotheses, compute the appropriate test statistic, and compute the p-value. What do you conclude? [2 marks]

3. The university is interested in academic dishonesty among students and so decides to interview a group of randomly selected students. For obvious reasons, a student maybe reluctant to answer honestly when asked if they have ever cheated on an exam. To give the students some anonymity, the following protocol is instituted. The interview asks the student to roll a die. Only the student sees the outcome of

the roll. Then the student picks one of two question cards with equal probability, independent of the outcome of the roll. The first card asks if they rolled a number **strictly greater than 2**. The second card asks if the student has ever cheated on an exam. Using this protocol 220 students were interviewed, of which 90 responded 'Yes'.

(a) Let p denote the probability that a randomly selected student has ever cheated on an exam. Give an expression for the probability that a student answers "Yes" in terms of p, assuming they answer honestly. [1 mark]

(b) Construct an unbiased estimator for p and give your estimate of p based on this data. [1 mark]

(c) Construct a 95% confidence interval for the proportion of students who would answer 'Yes' under this protocol. Transform this interval to construct a confidence interval for p.

[2 marks]

4. Training dogs with electric collars is controversial and regulations vary according to country. In France, where the use of electric collars is unregulated, a study was conducted to identify characteristics associated with the use of electric collars. Dog owners were asked to complete a questionnaire covering the use of electric collars, the dog's training and characteristics of the dog. Of the 320 dog owners who use electric collars 194 took their dog to formal obedience training while of the 873 dog owners who did not use electric collars 614 took their dog to formal obedience training.

   Does the data provide evidence that dog owners who use electric collars are less likely to take their dog to formal obedience training? State the null and alternative hypotheses and use an appropriate test statistic to determine the p-value. What do you conclude? [2 marks]

5. Trial is conducted to compare the effectiveness of two anti-congestant drugs. The effectiveness of the drug is measured by a patients PEF, the maximum rate at which they are able to exhale. The PEF is measured 30 minutes after the drug has been administerd. In the trial 40 patients where randomly assigned to drug A or drug B. The 18 patients who received drug A reported a sample average PEF of 460 (L/min) and a sample standard deviation of 32 (L/min). The 22 patients who received drug B reported a sample average PEF of 478 (L/min) and a sample standard deviation of 46 (L/min).

   (a) Construct a 90% confidence interval for the mean PEF of patients receiving drug B. [1 mark]

   (b) Does the data provide evidence of a difference in mean PEF between the two drugs? State the null and alternative hypotheses and use an appropriate test statistic to determine the p-value. What do you conclude? [2 marks]

Total [15 marks]