STAT0003 Further Probability and Statistics

Ge Li (Sunny)

March 15 2023

2

Contents

Preface

This document is used to document my assignments and weekly exercises of STAT0003.

Week 1

Exercise Sheet 1

1.1 Question 1

You and your friend are presented with 7 boxes. 3 of the boxes contain delicious chocolates while other 4 are filled with deadly mushrooms (disguised as delicious chocolates). Both of you will need to select a box and eat what is inside, but you need to decide which of you will go first.

Question 1.1 In order to increase your chance of eating delicious chocolates, should you choose first or second?
Solution. Let A denote the event of getting chocolate in the first trial.
Let B denote the event of getting chocolate in the second trial.

$$P(A) = \frac{3}{7}.$$

According to the law of total probabilities:

$$P(B) = P(B \mid A)P(A) + P(B \mid A')P(A')$$

$$\therefore P(B) = \frac{2}{6} \cdot \frac{3}{7} + \frac{3}{6} \cdot \frac{4}{7} = \frac{3}{7}$$

$$\therefore P(A) = P(B)$$

As a result, the chance of eating delicious chocolates remains the same regardless of choosing to go first or second.

Question 1.2 If you choose first, what is the probability that you eat delicious chocolates, given that your friend also does?
Solution.
$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{\frac{2}{6} \cdot \frac{3}{7}}{\frac{3}{7}} = \frac{1}{3}.$$

Question 1.3 How would your answer to (i) change if there were three of you: yourself and two friends? Answer this without doing any mathematics or algebra, perhaps by referring to examples from this module.
Solution. The answer would not change. This is the same question as the one discussed duringthe lecture where we tried to calculate the probability of the third ball drawn from an urn with 3 black balls and 4 red balls being a black ball. Essentially, the probability always remains the same regardless of the order.

1.2   Question 2

A class consists of 12 girls and 16 boys. Jane is one of the girls and Jonathan is one of the boys.

Question 2.1 In how many ways can we make a group of 2 girls and 3 boys?

$$\text{Number of ways} = \binom{16}{3} \cdot \binom{12}{2} = 36960.$$

Question 2.2 What is the probability that Jane is not in that group but Jonathan is?

$$P(\text{Jonathan without Jane}) = \frac{\binom{15}{2} \cdot \binom{11}{2}}{\text{Number of ways}} = \frac{5}{32}.$$

Question 2.3 What is the probability that Jonathan is and Jane is not a group of 5 students selected randomly out of the 28?
Solution. Let E_1 denote the event that Jonathan is but Jane is not a group of 5 students randomly selected out of the 28.

$$P(E_1) = \frac{\binom{26}{4}}{\binom{28}{5}} = \frac{115}{756}.$$

Question 2.4 What is the probability that there are at least 2 boys and at least 1 girl in a randomly selected group of 5 students?
Solution. Let E_2 denote the event that there are at least 2 boys and at least 1 girl in a randomly selected group of 5 students.

$$P(E_2) = \frac{\binom{16}{2} \cdot \binom{12}{3} + \binom{16}{3} \cdot \binom{12}{2} + \binom{16}{4} \cdot \binom{12}{1}}{\binom{28}{5}} = \frac{85200}{98280} \approx 0.8669.$$

Week 2

Exercise Sheet 2

2.1 Question 1

Let X be a random variable with expectation μ and variance σ^2. Find the expectation and variance of the random variable $Y = \frac{X-\mu}{\sigma}$.

Solution.

$$\textbf{Expectation:} \qquad \mathbb{E}(Y) = \frac{\mathbb{E}(X) - \mu}{\sigma} = 0.$$

$$\textbf{Variance:} \qquad Var(Y) = \frac{\sigma^2}{\sigma^2} = 1.$$

2.2 Question 2

Let X be a discrete random variable with $\mathbb{E}(X^2) = 0$.

Question 2.1 Show that $P(X = 0) = 1$.

Solution. Since we are given that

$$\mathbb{E}(X^2) = \sum_{\text{all } x} x^2 P(X = x) = 0,$$

and since both x^2 and $P(X = x)$ are non-negative values, we can deduce that $X = 0$ is the only possible circumstance. As a result, $P(X = 0) = 1$.

Question 2.2 What is the $Var(X)$?

Solution. We need to calculate the expected value of X first,

$$\mathbb{E}(X) = 0.$$

We are given that $\mathbb{E}(X^2) = 0$. As a result,

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 0.$$

2.3   Question 3

The random variable, X, has the following probability mass function

$$p(x) = \frac{c}{x(x+1)} \quad (x = 1, 2, 3, ...)$$

Question 3.1 Find the value of the constant c.
Solution.

$$p(x) = \frac{c}{x(x+1)} = \frac{c}{x} - \frac{c}{x+1}$$

Since we know that probability mass function (p(x)) has the property that $\sum p(x) = 1$,

$$\sum_{x=1}^{\infty} p(x) = \lim_{n \to \infty} \sum_{x=1}^{n} \left(\frac{c}{x} - \frac{c}{x+1} \right) = \lim_{n \to \infty} \left(\frac{c}{1} - \frac{c}{2} + \frac{c}{2} + ... - \frac{c}{n} + \frac{c}{n} - \frac{c}{n+1} \right) = 1$$

$$\therefore \lim_{n \to \infty} c \cdot \frac{n}{n+1} = 1$$
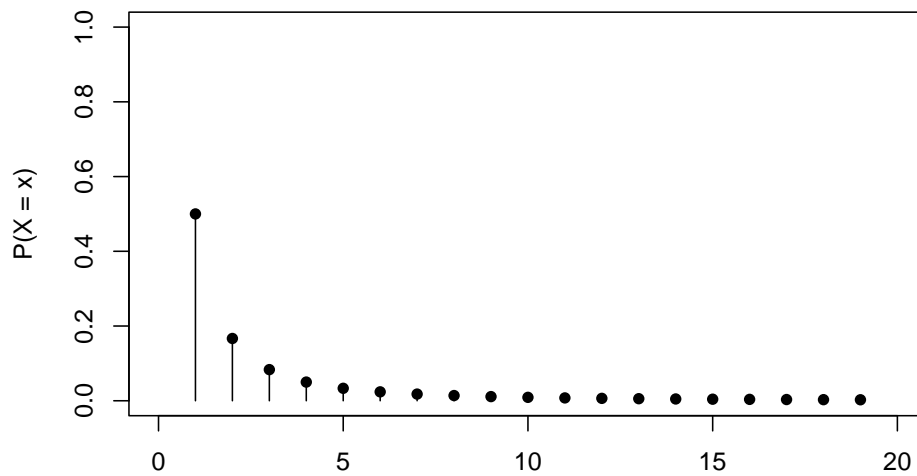
$$\therefore c = 1$$

Question 3.2 Find the cumulative distribution function of X and sketch both the probability mass function and the cumulative density function.
Solution.

$$F_X(x) = P(X \le x) = \sum_{1}^{x} \frac{1}{x} - \frac{1}{x+1} = \frac{x}{x+1}$$

$$F_X(x) = P(X \le x) = \begin{cases} 0 & -\infty < x < 1 \\ \frac{x}{x+1} & 1 \le x < \infty \end{cases}$$

Probability Mass Function

Cumulative Density Function
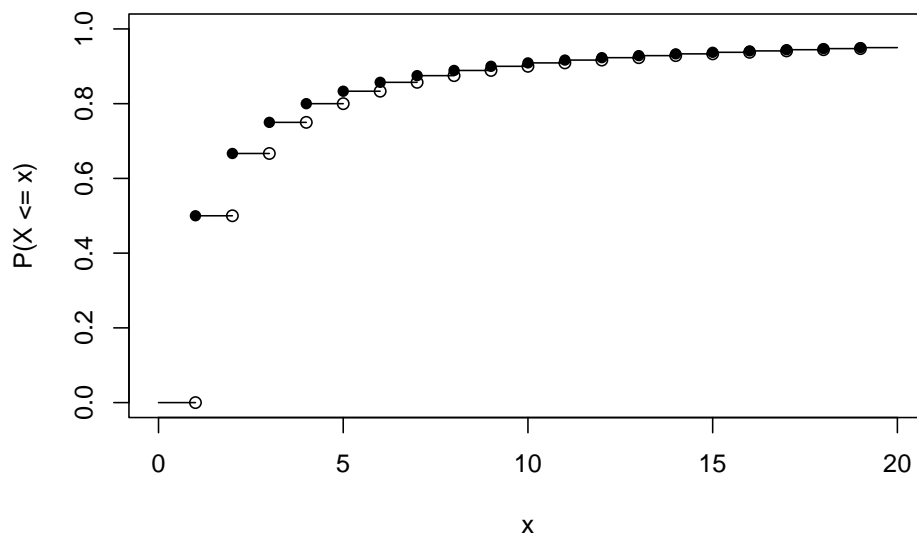
Question 3.3 Calculate $P(X \geq 50)$ and $P(X \geq 50 \mid X \geq 40)$.
Solution.

$$P(X \geq 50) = 1 - P(X < 50) - 1 - P(X \leq 49) = 1 - \frac{49}{50} = \frac{1}{50}$$

$$P(X \geq 50 \mid X \geq 40) = \frac{1/50}{1/40} = \frac{4}{5}$$

Week 3

Exercise Sheet 3

3.1 Question 1

A geological study indicates that an exploratory oil well drilled in a certain region should strike oil with probability 0.3, independently of other wells drilled. Find the probability that the first strike of oil comes on the third well drilled. Find the probability that the third strike of oil comes on the fifth well drilled.
Solution. Let X denote the number of well drilled until the first strike of oil comes, since then X follows the distribution

$$X \sim \operatorname{geom}(0.3).$$

Therefore, the probability that the first strike of oil comes on the third well drilled is
$$P(X = 3) = 0.3 \cdot (1 - 0.3)^2 = 0.147.$$

Let M denote the event that the third strike of oil comes on the fifth well drilled. The probability of striking oil successfully in the fifth well is 0.3, while the probability of having 2 wells striking oil successfully in the first 4 wells follows a binomial distribution $\sim \operatorname{bin}(4, 0.3)$. Therefore, the probability that the third strike of oil comes on the fifth well drilled is

$$P(M) = \binom{4}{2} \cdot 0.3^2 \cdot 0.7^2 \times 0.3 = 0.07938.$$

3.2 Question 2

An exam paper consists of 10 multiple choice questions, each offering 4 choices of which only 1 (randomly assigned) is correct. If a candidate chooses her answers completely at random, what is the probability of each of the following, specifying the distribution used in each case.

Question 2.1 She gets at least 8 questions right.
Solution. Let X denote the number of questions the candidate gets correct, as a result, X follows

$$X \sim \text{bin}(10, 0.25).$$

Therefore, the probability the candidate gets at least 8 questions right is

$$P(X \geq 8) = \sum_{i=8}^{10} P(X = i) = \sum_{i=8}^{10} \binom{10}{i} \cdot 0.25^i \cdot 0.75^{10-i} = 4.158 \cdot 10^{-4}.$$

Question 2.2 The last of the 10 questions is the 8th one she gets rights.
Solution. Let E denote the event that the last question being the eighth one the candidate gets right. The probability of the 10th question being correct is 0.25. The probability of the candidate getting 7 questions among the first 9 questions correct is $P(X = 7)$, where in this particular case X follows the distribution

$$X \sim \text{bin}(9, 0.25).$$

Therefore, the final probability

$$P(E) = 0.25 \cdot P(X = 7) = 0.25 \times \binom{9}{7} \cdot 0.25^7 \cdot 0.75^2 = 3.090 \cdot 10^{-4}.$$

Question 2.3 In 6 such exams, she gets at least 8 questions right in at most one exam.
Solution. Let A denote the number of exams the candidate gets at least 8 questions correct among the 6 exams, A follows the distribution

$$A \sim \text{bin}(6, 4.158 \cdot 10^{-4}).$$

As a result,

$$P(X \leq 1) = \sum_{i=0}^{1} \binom{6}{i} \cdot (4.158 \cdot 10^{-4})^i \cdot (1 - 4.158 \cdot 10^{-4})^{6-i} = 0.9999974.$$

Question 2.4 She is told that in one exam, 6 of the answers (at random) are given by the same unknown choice. What is the probability that in her first 4 answers the candidate chooses 3 from that set of 6.
Solution. The event that the candidate chooses 3 from that set of 6 answers in her first 4 answers follows the hypergeometric distribution $\sim \text{Hypergeometric}(10, 3, 6)$.

$$P = \frac{\binom{4}{3} \cdot \binom{6}{3}}{\binom{10}{6}} = 0.3810.$$

3.3 Question 3

Let $X \sim \text{Poi}(\mu)$. Prove that $\mathbb{E}(X) = \mathbb{Var}(X) = \mu$.

$$\mathbb{E}(X) = \sum_{\text{all x}} x \cdot \frac{e^{-\mu} \cdot \mu^x}{x!} = \mu e^{-\mu} \sum_{\text{all x}} \frac{\mu^{x-1}}{(x-1)!}$$

Since we can identify that $\sum_{\text{all x}} \frac{\mu^{x-1}}{(x-1)!}$ is the expansion of e^μ,

$$\mathbb{E}(X) = \mu e^{-\mu} \cdot e^\mu = \mu.$$

In order to calculate the variance $\mathbb{Var}(X)$, we need to first calculate the expectation of X^2,

$$\mathbb{E}(X^2) = \sum_{\text{all x}} x^2 \cdot \frac{e^{-\mu} \cdot \mu^x}{x!}$$

$$= \mu e^{-\mu} \cdot \sum_{\text{all x}} \frac{\mu^{x-1}}{(x-1)!} \cdot [(x-1)+1]$$

$$= \mu e^{-\mu} \cdot (\mu \sum_{\text{all x}} \frac{\mu^{x-2}}{(x-2)!} + \sum_{\text{all x}} \frac{\mu^{x-1}}{(x-1)!})$$

$$= \mu e^{-\mu} \cdot (\mu e^\mu + e^\mu)$$

$$= \mu^2 + \mu.$$

Therefore, the variance of X is

$$\mathbb{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mu^2 + \mu - \mu^2 = \mu$$

As a result, we have proved that $\mathbb{E}(X) = \mathbb{Var}(X) = \mu$.

Week 4

Exercise Sheet 4

4.1 Question 1

Trains arrive at a station with a mean rate of 3 trains/hour: assume that they are arriving from a Poisson Process. Compute the following probabilities (making sure that you clearly define the random variables you are using in each case). You may only use R or a calculator for the exponential function (not for any probability distributions):

Question 1.1 If we record the number of trains for 3 hours, there are more than 5.
Solution. Let X denote the number of trains arriving in 3 hours.

$$X \sim \text{pois}(9)$$

$$P(X > 5) = 1 - \sum_{n=0}^{5} \frac{e^{-9} \cdot 9^n}{n!} = 0.8843$$

Question 1.2 If we go to the station and wait for the first train to arrive, then we will not wait more than 10 minutes.
Solution. Let T denote the time passengers wait for the train to arrive.

$$T \sim \exp(3)$$

$$P(T \leq 1/6) = 1 - (1 - e^{-3(1/6)}) = 0.3934$$

Question 1.3 If we go to the station and take the third train that comes, we will not wait more than 30 minutes. Evaluate the integral by repeated integration by parts.
Solution.

$$T \sim \text{gamma}(\text{shape} = 3, \text{rate} = 3/2)$$

17

$$P(T \leq 1) = \int_0^1 \frac{(3/2)^3 x^2 e^{-3x/2}}{2!} \, dx = \frac{27}{16} \int_0^1 x^2 e^{-3x/2} \, dx$$

$$= \frac{27}{16} \left[x^2 (-2/3) e^{-3x/2} \right]_0^1 - \frac{27}{16} \int_0^1 (-\frac{4}{3} x e^{-3x/2}) \, dx$$

$$= \frac{27}{16} (-2 e^{-3/2}/3 - \frac{40}{27} e^{-3/2} + \frac{16}{27})$$

$$\approx 0.1912$$

We can confirm this answer by using *pgamma(1,3,3/2)*. These are not sensible assumptions, considering the fact that trains don't arrive at stations at randomly chosen time, and the time of arrivals of each train may not necessarily be independent.

4.2 Question 2

The probability density function of X, the lifetime in hours of a certain type of electronic device is given by

$$f(x) = \begin{cases} 10/x^2 & x > c \\ 0 & \text{otherwise} \end{cases}$$

Find c and $P(X > 15)$. What is the probability that out of 5 such devices at least 4 will function for at least 15 hours (assume the lifetimes of the devices are independent)?

Solution.

$$F_X(x) = \int_{x=c}^{\infty} \frac{10}{x^2} \, dx = 1$$

$$\therefore \left[-\frac{10}{x} \right]_c^{\infty} = 1$$

$$\therefore c = 10$$

$$P(X > 15) = \int_{15}^{\infty} \frac{10}{x^2} \, dx = \left[-\frac{10}{x} \right]_{15}^{\infty} = \frac{2}{3}$$

Let M denote the number of the electronic devices that function well for at least 15 hours.

$$M \sim \text{binom}(5, \frac{2}{3})$$

$$P(M \geq 4) = \binom{5}{4} \left(\frac{2}{3}\right)^4 \left(\frac{1}{3}\right)^1 + \binom{5}{5} \left(\frac{2}{3}\right)^5 = 0.4609$$

4.3 Question 3

Question 3.1 The mean of a Gamma(α, λ) distribution.
Solution. Since we are given that

$$\int_0^\infty \frac{x^{\alpha-1}e^{-\lambda x}\lambda^\alpha}{\Gamma(\alpha)}\,dx = 1.$$

$$
\begin{aligned}
\mathbb{E}(X) &= \int_0^\infty x \cdot \frac{x^{\alpha-1}e^{-\lambda x}\lambda^\alpha}{\Gamma(\alpha)}\,dx \\
&= \int_0^\infty \frac{x^\alpha e^{-\lambda x}\lambda^{\alpha+1}}{\lambda \cdot \Gamma(\alpha)}\,dx \\
&= \frac{\alpha}{\lambda}\int_0^\infty \frac{x^\alpha e^{-\lambda x}\lambda^{\alpha+1}}{\Gamma(\alpha+1)}\,dx \\
&= \frac{\alpha}{\lambda}
\end{aligned}
$$

Question 3.2 Find the result of $\int_0^\infty x^2 e^{-cx}\,dx$.
Solution.

$$\Gamma(3) = \int_0^\infty x^2 e^{-cx}\,dx$$

We use substitution, let $u = cx$, then $dx = \frac{1}{c}du$.

$$\int_0^\infty x^2 e^{-cx}\,dx = \frac{1}{c^3}\int_0^\infty u^2 e^{-u}\,du = \frac{2}{c^3}$$

Question 3.3 Derive the Mean and Variance of continuous r.v. X, whose pdf
is $\frac{a}{2}e^{-a|x|}$ over the real line.
Solution.

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^\infty \frac{ax}{2}e^{-a|x|}\,dx \\
&= \frac{1}{2}\left(\int_{-\infty}^0 axe^{-ax}\,dx + \int_0^\infty axe^{-ax}\,dx\right) \\
&= 0
\end{aligned}
$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} \frac{ax^2}{2} e^{-a|x|} \, dx$$

$$= \int_{0}^{\infty} ax^2 e^{-ax} \, dx$$

$$= a \int_{0}^{\infty} x^2 e^{-ax} \, dx$$

$$= a \cdot \frac{2}{a^3} \text{ (according to (b))}$$

$$= \frac{2}{a^2}$$

As a result, the variance is

$$\mathbb{V}\text{ar} = \frac{2}{a^2}.$$

Week 5

Exercise Sheet 5

5.1 Question 1

The annual rainfall in inches in a certain region is modelled by a normal random variable with mean 40 and variance 16 What is the probability that it will take at least 5 years until a year occurs having a rainfall of over 45 inches? What assumptions are you making?

Solution. Assumption: annual rainfall of each year is independent.

Let X denote the annual rainfall in inches in that certain region.

$$X \sim \text{norm}(40, 16)$$

Therefore, the probability of a year having a rainfall of over 45 inches is

$$P(X > 45) = 1 - P(X \leq 45) = 1 - \Phi\left(\frac{45 - 40}{4}\right) = 1 - 0.8944 = 0.1056.$$

Alternatively, we can use the R command `pnorm(45, 40, 4, lower.tail = F)` to find the probability. Let Y denote the number of years until a year occurs having a rainfall of over 45 inches. Therefore, Y follows the geometric distribution

$$Y \sim \text{geom}(0.1056).$$

The probability that it takes at least 5 years until a year occurs having a rainfall of over 45 inches is

$$P(Y \geq 5) = 1 - P(Y \leq 4) = 1 - (1 - (1 - 0.1056)^4) = 0.6399.$$

Alternatively, we can use the R command `pgeom(3, 0.1056, lower.tail = F)` to find the probability.

5.2   Question 2

Assume that $X_1, ..., X_{10}$ are independent normally distributed random variables, each with mean 5 and variance 1.

Question 2.1 Find the probability that $X_i > 5$ for all $i = 1, ..., 10$.

Solution. For all $X_1, ..., X_{10}$, the probability of $X_i > 5$ is

$$P(X_i > 5) = 1 - P(X_i \le 5) = 1 - \Phi(0) = 0.5.$$

Question 2.2 Find the probability that at least two of the X_is are greater than 5.

Solution. Let A denote the number of X_is that are greater than 5. Since then, A follows the binomial distribution $A \sim \text{binom}(10, 0.5)$.

$$P(A \ge 2) = 1 - P(A \le 1) = 1 - \binom{10}{0} \cdot 0.5^{10} - \binom{10}{1} \cdot 0.5^{10} = 0.9893.$$

Question 2.3 Find the probability that the sample mean of the X_is is greater than 5.1.

Solution. For $X_1, ..., X_{10}$ that are independent normally distributed random variables with mean 5 and variance 1, the sample mean of X_is is normally distributed with mean 5, and variance $\frac{1}{10}$.

$$P(\bar{X} > 5.1) = 1 - P(\bar{X} \le 5.1) = 1 - \Phi\left(\frac{5.1 - 5}{\sqrt{1/10}}\right) = 1 - 0.6255 = 0.3745,$$

R command `pnorm(5.1, 5, sqrt(1/10), lower.tail = F)` returns $P(\bar{X} > 5.1) = 0.3759$.

Question 2.4 Without doing any further calculations, write down the probability that the sum of all the X_is is greater than 51. Why were no further calculations necessary?

Solution. The probability that the sum of all the X_is is greater than 51 is the same as the probability that the sample mean of the X_is is greater than 5.1 since if the sample mean is greater than 5.1, the sum would definitely be greater than 51.

$$\therefore P\left(\sum_{i=1}^{10} X_i > 51\right) = P(\bar{X} > 5.1) = 0.3745 \text{ or } 0.3759.$$

Week 6

Exercise Sheet 6

6.1 Question 1

Here are the weights of some rather greedy dogs (in kgs): 39.6 40.2 40.9 41.4 39.8 39.4 43.6 41.8 40.9.
Question 1.1 Assuming a Normal model for the population of such dogs, calculate the 92% confidence interval for μ and interpret this in words.
Solution. Let X denote the weights of the dogs. We first compute the sample mean and sample variance of the data.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{1}{9} \cdot (39.6 + 40.2 + 40.9 + 41.4 + 39.8 + 39.4 + 43.6 + 41.8 + 40.9) = 40.84$$

$$\because S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{\sum X_i^2 - 2\bar{X}\sum X_i + n\bar{X}^2}{n-1} = \frac{\sum X_i^2 - n\bar{X}^2}{n-1}$$

$$\therefore S^2 = \frac{39.6^2 + 40.2^2 + 40.9^2 + 41.4^2 + 39.8^2 + 39.4^2 + 43.6^2 + 41.8^2 + 40.9^2 - 9 \cdot \bar{X}^2}{9-1} = 1.75$$

Consider the pivotal quantity

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \,\dot{\sim}\, \mathrm{t}(\mathrm{df} = n-1).$$

$$\therefore P\left(t_{\alpha/2} < \frac{\sqrt{n}(\bar{X} - \mu)}{S} < t_{1-\alpha/2}\right) = 1 - \alpha$$

$$\therefore P\left(\bar{X} - t_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

As a result, the $(1-\alpha) \cdot 100\%$ confidence interval for μ is

$$\left(\bar{X} - t_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \ \bar{X} + t_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right)$$

In this case, since we are required to find the 92% confidence interval for μ, we take $\alpha = 0.08$ as the confidence coefficient. We compute $t_{8,0.96}$ using the R command `qt(0.96, df = 8)` which returns $t_{8,0.96} = 2.00$. Therefore, the 92% confidence interval for μ is

$$\left(40.84 - 2 \cdot \frac{\sqrt{1.75}}{3}, \ 40.84 + 2 \cdot \frac{\sqrt{1.75}}{3} \right) = (39.96, 41.72).$$

This represents that we are 92% confident that the population mean of the weights of the dogs is between the range 39.96 and 41.72.

Question 1.2 What caused the difference in confidence interval for μ in 2 exercises?
Solution. The difference in confidence interval for μ is caused by different distributions we used to approximate the random variables. In last exercise, we used Standard Normal Distribution instead of Student T Distribution since the population variance were known and given to us ($\sigma^2 = 2$). Note that the reason of one of the confidence intervals being wider than the another is that T Distribution has fatter tails and therefore has larger quantiles, which evantually lead to a wider confidence interval.

6.2   Question 2

We decide to model cars entering a motorway service station with a Poisson Process of rate 1500 / day.
Question 2.1 Which exact distribution would we use to model the total number of cars that enter during June?
Solution. We should use Poisson Distribution. Let Y denote the number of cars entering the motorway service station every day. According to what we are given, $Y \sim \text{Pois}(1500)$. Then, let X denote the number of cars entering the motorway service station in June, which is a 30-day period, X follows the Poisson Distribution $X \sim \text{Pois}(45000)$.

Question 2.2 Use the CLT to calculate the probability that greater than 44800 but less than 45200 cars entered during June, defining the appropriate random variables formally. Are the conditions for the CLT satisfied?
Solution. Let X denote the total number of cars entering the service station in June. We can consider X as the sum of 30 I.I.D random variables with mean and variance $\mu = \sigma^2 = 1500$. Therefore, we can approximate

$$X \dot\sim N(30 \cdot 1500, 30 \cdot 1500) = N(45000, 45000).$$

As a result, the probability that greater than 44800 but less than 45200 cars entered during June is

$$P(44800 - 0.5 < X < 45200 + 0.5) = \Phi\left(\frac{45200.5 - 45000}{\sqrt{45000}} \right) - \Phi\left(\frac{44799.5 - 45000}{\sqrt{45000}} \right) = 0.6578$$

Alternatively, the R command `pnorm(45200.5, 45000, sqrt(45000)) - pnorm(44799.5, 45000, sqrt(45000))` returns 0.6554.

Week 7

Exercise Sheet 7

7.1   Question 1

Solution. Let X denote the weights of the dogs. According to Question 1 from Exercise Sheet 6, we obtain our sample mean and sample standard deviation:

$$\bar{X} = 40.84, \ \ S^2 = 1.75.$$

With the confidence coefficient α, consider the pivotal quantity

$$Q = \frac{(n-1)S^2}{\sigma^2} \ \dot{\sim} \ \chi^2_{n-1}$$

$$\therefore P\left(\chi^2_{\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2}\right) = 1 - \alpha$$

$$\therefore P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right) = 1 - \alpha$$

As a result, the $(1-\alpha) \cdot 100\%$ confidence interval for σ^2 is

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \ \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right)$$

While the $(1-\alpha) \cdot 100\%$ confidence interval for σ is therefore

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}}, \ \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}}}\right)$$

In this case, since we are required to to find the 92% confidence interval for σ, we take $\alpha = 0.08$ as the confidence coefficient. We compute $\chi^2_{\alpha/2} = 2.54$ using the R command `qchisq(0.04, df = 8)` and $\chi^2_{1-\alpha/2} = 16.17$ using the R

27

command `qchisq(0.96, df = 8)`. Therefore, the 92% confidence interval for σ is

$$\left(\sqrt{\frac{8 \cdot 1.75}{16.17}}, \ \sqrt{\frac{8 \cdot 1.75}{2.54}} \right) = (0.93, 2.35)$$

This represents that we are 92% confident that the population standard deviation of the weights of the dogs is between the range 0.93 and 2.35.

7.2 Question 2

Question 2.1
Solution

$$H_0 : \sigma^2 > 25$$

$$H_1 : \sigma^2 \leq 25$$

$$\text{Test statistics: } \chi^2 = \frac{(n-1)s^2}{\sigma^2} \ \dot{\sim} \ \chi_9^2$$

$$\text{Observed test statistics: } \chi^2 = \frac{(10-1) \cdot 20.2}{25} = 7.272$$

$$\text{P-Value: } P = P(\chi_9^2 \leq 7.272) = 0.3912$$

Question 2.2
Solution

$$H_0 : \sigma^2 < 25$$

$$H_1 : \sigma^2 \geq 25$$

$$\text{Test statistics: } \chi^2 = \frac{(n-1)s^2}{\sigma^2} \ \dot{\sim} \ \chi_9^2$$

$$\text{Observed test statistics: } \chi^2 = \frac{(10-1) \cdot 20.2}{25} = 7.272$$

$$\text{P-Value: } P = P(\chi_9^2 \geq 7.272) = 0.6088$$

Week 8

Exercise Sheet 8

We want to test whether the mean lifetime of a particular type of bulb is equal to 20 days. A random sample of 25 such bulbs from brand A was selected, and their lifetime was recorded. The sample mean was 18.3 days and you may assume that the population standard deviation $\sigma_A = 4$ days. For the purpose of this question you can assume that the lifetime of these bulbs in days follows a normal distribution.

Question 1 Perform a suitable statistical test to decide whether the lifetime of these types of bulbs is indeed 20 days. What do you conclude?
Solution. We first set up our null hypothesis and alternative hypothesis.

$$H_0 : \mu = 20, \ H_1 : \mu \neq 20.$$

Test-Statistic: $\quad T_0 = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1)$

Observed Test-Statistic: $\quad t = \dfrac{18.3 - 20}{4/\sqrt{25}} = -2.125$

P-Value: $\quad P = P(T_0 \leq -2.125) + P(T_0 \geq -2.125) = 2 \cdot P(T_0 \leq -2.125) = 0.0336$

We would fail to reject the null hypothesis if we use a significance level less than the p-value 0.0336.

Question 2 If the mean lifetime is actually $\mu_x = 19.5$ days, and you assume a 1% significance level, calculate the power of the test and the probability of a type II error.
Solution.

$$H_0 : \mu = 20, \ H_1 : \mu \neq 20.$$

Original Test-Statistic: $T_0 = \dfrac{\bar{x} - 20}{4/\sqrt{25}} \sim \mathrm{N}(0,1)$

Rejection Region: $\bar{x} < 20 - Z_{1-\alpha/2} \cdot \dfrac{4}{5}$ or $\bar{x} > 20 + Z_{1-\alpha/2} \cdot \dfrac{4}{5}$

$\qquad\qquad \bar{x} < 17.9393$ or $\bar{x} > 22.0607$

New Test Statistic: $\dfrac{\bar{x} - 19.5}{4/\sqrt{25}} \sim \mathrm{N}(0,1)$ under H_1

Power: $P(\text{Reject } H_0 \mid \mu = 19.5) = P\!\left(Z < \dfrac{17.9393 - 19.5}{0.8} \right) + P\!\left(Z > \dfrac{22.0607 - 19.5}{0.8} \right) = 0.0262$

Probability of Type II Error: $P = 1 - 0.0262 = 0.9738$

Question 3 Construct a 95% confidence interval for σ_B^2 and draw a conclusion.
Solution.
$$\alpha = 1 - 0.95 = 0.05$$

Confidence Interval: $\left(\dfrac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}, \ \dfrac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} \right) = \left(\dfrac{(30-1)\cdot 3^2}{45.7223}, \ \dfrac{(30-1)\cdot 3^2}{16.0471} \right) = (5.7084, \ 16.2647).$

Generally, although $\sigma_A^2 = 16$ falls within the 95% confidence interval for σ_B^2, it is at a relatively high position in the interval. As a result, the light bulbs produced by brand B may have less variability in lifetime.

Question 4
Solution. We can compute the test-statistic of σ_A^2/σ_B^2 and decide whether to reject the null hypothesis using F-distribution.

Assumptions. First, the lifetime of all the light bulbs are independent and identically distributed. Second, the lifetime of these light bulbs follows a normal distribution.
Relevance. Unbiased estimators are required for both tests.