Exam information	
Course code and title	STAT7203 Probability Models & Data Analysis
Semester	Semester 2, 2022
Exam type	Online, non-invigilated, end-of-semester examination
Exam technology	File upload to Blackboard Assignment
Exam date and time	Your examination will begin at the time specified in your personal examination timetable. If you commence your examination after this time, the end for your examination does NOT change. The total time for your examination from the scheduled starting time will be: 2 hours 10 minutes (including 10 minutes reading time during which you should read the exam paper and plan your responses to the questions). A 15-minute submission period is available for submitting your examination after the allowed time shown above. If your examination is submitted after this period, late penalties will be applied unless you can demonstrate that there were problems with the system and/or process that were beyond your control.
Exam window	You must commence your exam at the time listed in your personalised timetable. You have from the start date/time to the end date/time listed in which you must complete your exam.
Permitted materials	This is an open book exam.
Recommended materials	Printer/Scanner
Instructions	You will need to download the question paper included within the Blackboard Test. Once you have completed the exam, upload the completed exam answers file to the Blackboard assignment submission link. You may submit multiple times, but only the last uploaded file will be graded. Write your answers on blank paper (clearly label your solutions so that it is clear which problem it is a solution to) or annotate an electronic file on a suitable device. You must submit your answers as a **single PDF file**.
Who to contact	Given the nature of this examination, responding to student queries and/or relaying corrections to exam content during the exam may not be feasible. If you have any concerns or queries about a particular question or need to make any assumptions to answer the question, state these at the start of your solution to that question. You may also include queries you may have made with respect to a particular question, should you have been able to 'raise your hand' in an examination-type setting. If you experience any interruptions to your examination, please collect evidence of the interruption (e.g. photographs, screenshots or emails). If you experience any issues during the examination, contact ONLY the Library AskUs service for advice as soon as practicable:

	Chat: support.my.uq.edu.au/app/chat/chat_launch_lib
	Phone: +61 7 3506 2615
	Email: examsupport@library.uq.edu.au
	You should also ask for an email documenting the advice provided so you can provide this as evidence for a late submission.
Late or incomplete submissions	In the event of a **late submission**, you will be required to submit evidence that you completed the assessment in the time allowed. This will also apply if there is an **error in your submission** (e.g. corrupt file, missing pages, poor quality scan). We **strongly recommend** you use a phone camera to take time-stamped photos (or a video) of every page of your paper during the time allowed (even if you submit on time).
	If you submit your paper after the due time, then you should send details to SMP Exams (exams.smp@uq.edu.au) as soon as possible after the end of the time allowed. Include an explanation of why you submitted late (with any evidence of technical issues) AND time-stamped images of every page of your paper (eg screen shot from your phone showing both the image and the time at which it was taken).
Important exam condition information	Academic integrity is a core value of the UQ community and as such the highest standards of academic integrity apply to all examinations, whether undertaken in-person or online.
	This means:
	• You are permitted to refer to the allowed resources for this exam, but you cannot cut-and-paste material other than your own work as answers.
	• You are not permitted to consult any other person – whether directly, online, or through any other means – about any aspect of this examination during the period that it is available.
	• If it is found that you have given or sought outside assistance with this examination, then that will be deemed to be cheating.
	If you submit your online exam after the end of your specified reading time, duration, and 15 minutes submission time, the following penalties will be applied to your final examination score for late submission:
	• Less than 5 minutes – 5% penalty
	• From 5 minutes to less than 15 minutes – 20% penalty
	• More than 15 minutes – 100% penalty
	These penalties will be applied to all online exams unless there is sufficient evidence of problems with the system and/or process that were beyond your control.
	Undertaking this online exam deems your commitment to UQ's academic integrity pledge as summarised in the following declaration:
	"I certify that I have completed this examination in an honest, fair and trustworthy manner, that my submitted answers are entirely my own work, and that I have neither given nor received any unauthorised assistance on this examination".

[2 marks]

1. Suppose that a fair coin is tossed twice, and consider the following three events:

- $A = \{\text{head on first toss}\}$,

- $B = \{\text{head on second toss}\}$, and

- $C = \{\text{Both tosses the same}\}$.

Are these events independent? Mathematically justify your answer.

[2 marks each]

2. A box contains 100 balls, of which r are red. Suppose that the balls are drawn from the box one at a time, at random, without replacement.

(a) Determine the probability that the first ball drawn will be red.

(b) Determine the probability that the last ball drawn will be red.

[2 marks each]

3. Consider a system comprising two components (A and B) connected in parallel.
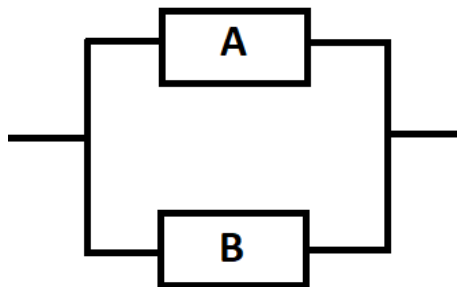
The system is working if there is a path from left to right through working components. The cumulative distribution function of the time to failure in years for each component is

$$F(t) = \begin{cases} 1 - t^{-2}, & t \geq 1, \\ 0, & \text{else.} \end{cases}$$

(a) Assuming the time to failure for components A and B are independent, determine the probability that the system is still working after 2 years.

(b) What is the probability that component A has failed in its first 2 years of operation, given the system is still working after 2 years.

[2 marks each]

4. A pair of random variables (X, Y) has a joint probability distribution such that $X \sim \text{Exp}(2)$, and the conditional distribution of Y given $\{X = x\}$ is $\mathcal{N}(x, 6x)$.

(a) Are X and Y independent? Justify your answer.

(b) Write down the joint probability density function of (X, Y), clearly specifying the support of the distribution.

(c) Using the formula $\mathbb{E}Y = \mathbb{E}\left[\mathbb{E}\left[Y \mid X\right]\right]$, find the expectation of Y.

(d) The moment generating function of Y is

$$M_Y(s) = \frac{2}{2 - s - 3s^2}, \quad -1 < s < \frac{2}{3}.$$

From this, or otherwise, find the variance of Y.

[2 marks each]

5. ***Note:*** *For The following questions, work out your answers 'by hand'. You may still use **R** (or any other programming language) to obtain probabilities and quantiles from the appropriate distributions and calculate your final answers.*

One hundred and eighteen fourth-grade children from four American public schools completed a survey about their video game playing habits. Students were asked about their preferred genre (Action, Adventure, Simulation), time spent playing video games, and the strategies they used to improve at the games they play most often.

(a) The 19 students that preferred Adventure video games spent an average of 4.42 hours per week playing video games, with a sample standard deviation of 4.00 hours. The 22 students that preferred Simulation video games spent an average of 2.57 hours per week playing video games, with a sample standard deviation of 1.93 hours.

Assuming the underlying distributions are normal, is there any evidence of a difference in the mean time spent playing video games between students that prefer Adventure video games and students that prefer Simulation video games? State the null and alternative hypotheses, and use an appropriate test statistic to determine the p-value. What do you conclude?

(b) Out of 118 students surveyed, 22 students preferred the Simulation genre of video games. Construct a 90% confidence interval for the proportion of all students who prefer the Simulation genre. State why your assumptions or approximations are reasonable.

(c) Out of 87 male students surveyed 21 said they used 'cheat codes' to improve their game play whereas 11 out of the 79 female students surveyed said they use 'cheat codes'. Construct a 95% confidence interval for the difference between males and females in the proportion that use 'cheat codes' to improve their game play. State why your assumptions or approximations are reasonable.

[2 marks each]

6. Researchers surveyed older adults between 45 and 85 years of age who regularly played digital games. Participants in the study completed a questionnaire covering a variety of demographic characteristics, genre and platform preferences and average time spent playing digital games per day.
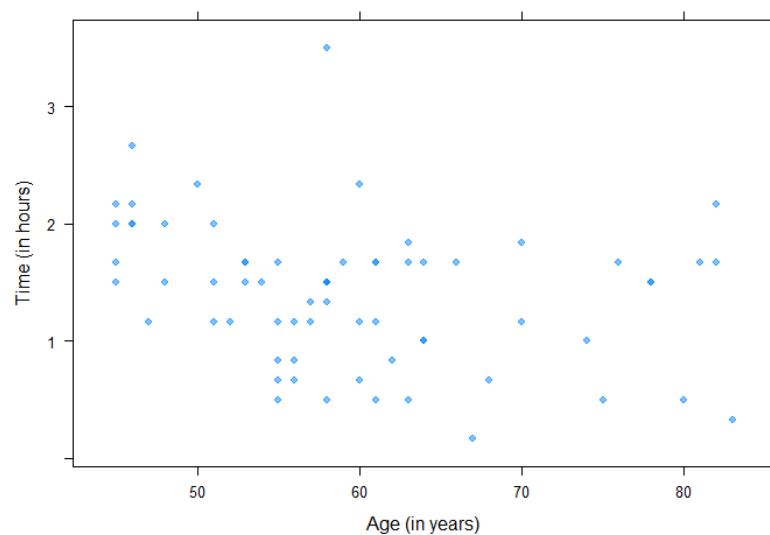
A linear regression model is constructed for the average time spent playing games per day in terms of the person's age.

The edited output from `R` after fitting is given below.

```
summary(lm(formula = PlayingTime ~ Age, data = Games))

Coefficients:
             Estimate   Std. Error
(Intercept)  2.41362    0.42606
Age         -0.01697    0.00704
---

Residual standard error: 0.5904 on 64 degrees of freedom
Multiple R-squared:  0.08325, Adjusted R-squared:  0.06893
```

(a) The following figures were generated to check the assumptions underlying the linear regression. State the assumptions of the linear regression model and comment on their validity for this data with reference to the figures below.
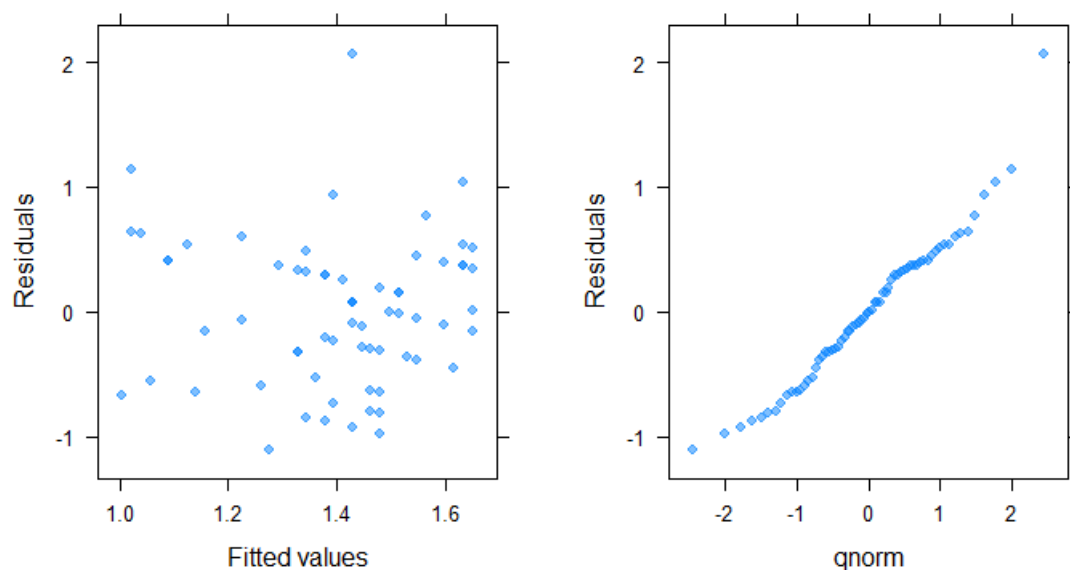

Figure 1: Left: Plot of residuals against fitted values from the linear regression. Right: Plot of residuals against quantiles of the standard normal distribution.

(b) How many participants were in the study?

(c) What is the estimated mean time spent playing digital games for an adult aged 60?

(d) Assume the model assumptions hold. Give a 95% confidence interval for the coefficient of age in this linear regression.

(e) Assume the model assumptions hold. Does the regression analysis provide evidence of an association between time spent playing digital games and age? State the null and alternative hypotheses, and use an appropriate test statistic to determine the p-value. What do you conclude?

[2 marks each]

7. A study examined the efficacy of an Internet-based telepsychology program for the treatment of fear of public speaking. The study recruited 127 participants who were randomly assigned to one of the following experimental conditions: an Internet-based self-administered treatment program (SA), the same program applied by a therapist face to face (TA), and a waiting-list control group (WL). A number of participants voluntarily withdrew from the study before completion. At the end of the treatment period, all participants who completed the study, filled out a questionnaire which gave a score on the Social Avoidance and Distress Scale (SAD). Researchers wanted to analyse data from the study using ANOVA to test whether there are systematic differences between the treatments in the mean score on the Social Avoidance and Distress Scale. The sample means and sample standard deviations of the SAD scores for the three treatment groups are given in the table below.

	n	$\bar{x}$	s
SA	30	7.67	5.81
TA	22	8.29	5.14
WL	25	11.28	7.45

Below is a boxplot of the data.
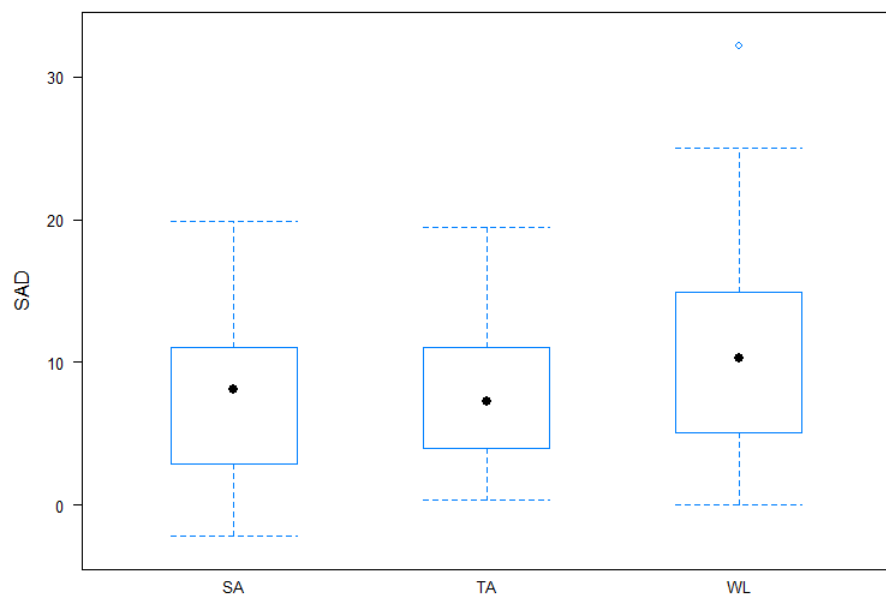
(a) Do the boxplots appear consistent with the assumptions of ANOVA? Justify you answer.

(b) Complete the following ANOVA table.

Source	DF	SS	MS	F
Groups		194.1		
Total		3059.9		

(c) Determine the p-value for the ANOVA hypothesis test. What do you conclude?

40 marks in total

END OF EXAMINATION