Mid-term exam, 4 November 2022

All aids are allowed, except a computer and personal assistance, as well as the use of any external information pertaining to the specific data and questions. Restricted use of some computer-like devices (including tablets and smartphones) is permitted under the rules described at the VHM 801 course homepage. The exam consists of one question with four parts, or subquestions (labeled by letters **a) − d)**, and with weights as indicated), that should all be answered to achieve the maximal 15 points. The subquestions can be answered independently of each other. The mid-term exam accounts for 15% of the course mark; however, every student may choose to waive the result of the mid-term exam. The duration of the mid-term exam is 1 hour.

Generally, **statistical models and methods should be specified**, and all statistical analyses should be summarized in conclusions.

Question 1 (*15 points*)

The first two parts — **a)** and **b)** — are based on measurements of how long aircraft air conditioning units operated after being repaired. The operating times (in hours) for 27 such units are listed in the table below, and Minitab listings are provided at the end of exam for three versions of the data: on original, square-root transformed and natural-log transformed scales.

Operating times	97	51	11	4	141	18	142	68	77
(hours)	80	1	16	106	206	82	54	31	216
	46	111	39	63	18	191	18	163	24

a) (*4 points*) Carry out a brief descriptive analysis for the operating times (on original scale). Focus on the characteristics of the distribution that you consider important for statistical analysis instead of giving a detailed enumeration of the values of different statistics. If you were asked to provide a summary of the distribution, which statistics would you choose to include/present (for example, the mean and standard deviation)?

b) (*3 points*) There was interest in using these data to estimate the probability of the air conditioners to operate for more than 100 hours after the repair. Because this can be done in different ways, you are only required to answer **one of the two parts** *i)* and *ii)* below. It is not recommended to answer both *i)* and *ii)*, but if you do, your score will be for the best answer among the two.

i) Make suitable assumptions about the parametric distribution of operating times on original or one of the transformed scales, and use that distribution (with its estimated parameters) to compute the probability of an operating time to exceed 100 hours.

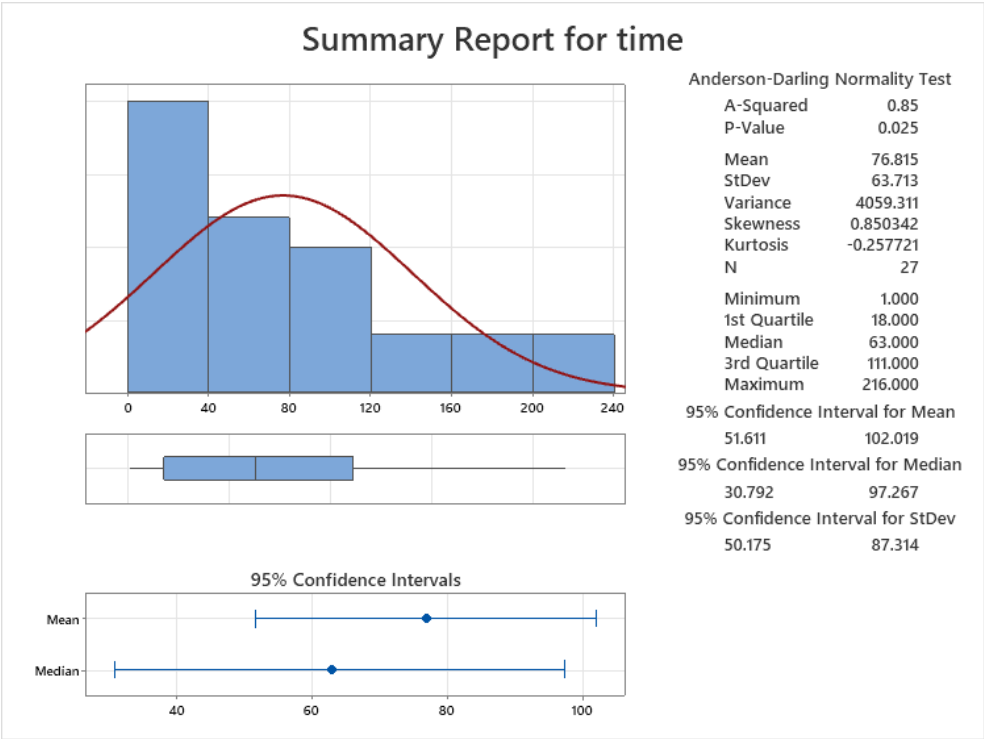
ii) Without making any distributional assumption for the operating times, estimate the probability of air conditioners to operate for more than 100 hours, and supplement the estimate with a 95% confidence interval.

The last two parts — **c)** and **d)** — are completely unrelated to the previous parts.

c) (*4 points*)
Simple choice questions are commonly used for tests and exams, in particular when the exam involves a large number of students. An instructor considers using an exam with 20 questions, each with only two options (correct/incorrect). Assume that the pass mark, or score, is 60%. What chance would a student that is purely guessing have of achieving at least a 60% score and thereby passing the exam? Give your answer also if there are three possible answers for each of the 20 questions (still only one correct answer). If you cannot compute a certain probability exactly, give a narrow range for the probability or explain how you would compute the probability using Minitab or another statistical software (with enough detail to allow the calculation to be carried out).

d) (*4 points*)
Consider now a different scenario where a standardized test has been administered country-wide. The test scores across the country averaged 73%, with a 95% range from 55% to 91%. Assume that a particular class of 25 students had an average score of 70%. Do you think the instructor of that class should be seriously concerned about the performance of the class relative to the national average? If possible from the information provided, make suitable assumptions and use statistical inference to answer the question. If you need additional information to perform statistical inference, explain what you need and how you would do the analysis with this information at hand.

Minitab listings for Question 1, parts **a)** *and* **b)**:



Anderson-Darling Normality Test

A-Squared	0.85
P-Value	0.025
Mean	76.815
StDev	63.713
Variance	4059.311
Skewness	0.850342
Kurtosis	-0.257721
N	27
Minimum	1.000
1st Quartile	18.000
Median	63.000
3rd Quartile	111.000
Maximum	216.000

95% Confidence Interval for Mean

51.611	102.019

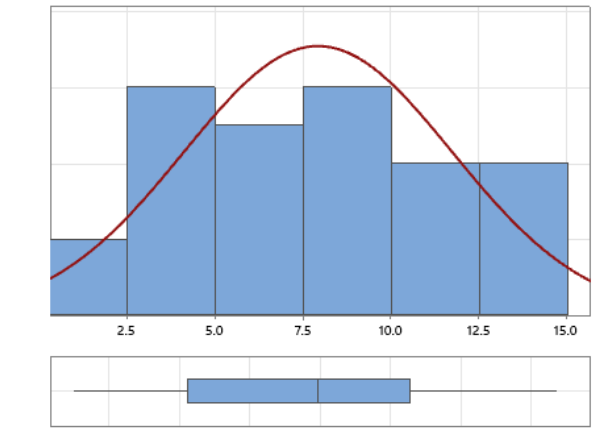
95% Confidence Interval for Median

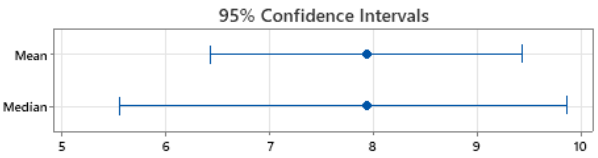
30.792	97.267

95% Confidence Interval for StDev

50.175	87.314

Summary Report for roottime





Anderson-Darling Normality Test

A-Squared	0.20
P-Value	0.858
Mean	7.9294
StDev	3.8046
Variance	14.4751
Skewness	0.120810
Kurtosis	-0.823400
N	27
Minimum	1.0000
1st Quartile	4.2426
Median	7.9373
3rd Quartile	10.5357
Maximum	14.6969

95% Confidence Interval for Mean

6.4244	9.4345

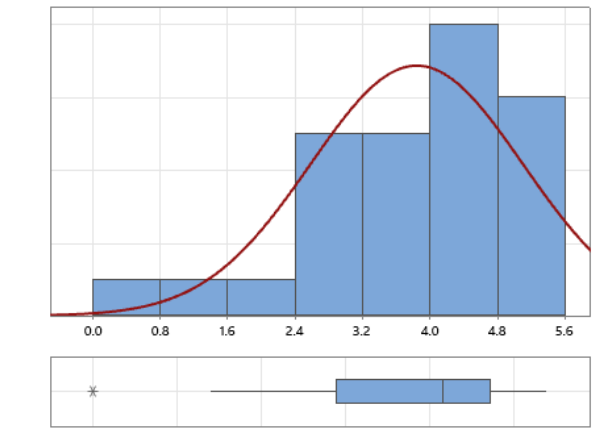
95% Confidence Interval for Median

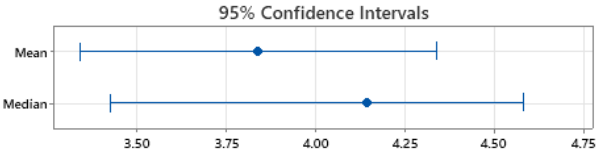
5.5479	9.8621

95% Confidence Interval for StDev

2.9962	5.2140

Summary Report for lntime





Anderson-Darling Normality Test

A-Squared	0.69
P-Value	0.064
Mean	3.8389
StDev	1.2565
Variance	1.5787
Skewness	-1.29384
Kurtosis	2.12793
N	27
Minimum	0.0000
1st Quartile	2.8904
Median	4.1431
3rd Quartile	4.7095
Maximum	5.3753

95% Confidence Interval for Mean

3.3418	4.3359

95% Confidence Interval for Median

3.4264	4.5773

95% Confidence Interval for StDev

0.9895	1.7219