

STAT 151 Final Exam (Version D)

Instructions:

1. This is an open book exam. You have 3 hours to complete the exam.
2. Please download the formula sheet, Table II (z table), Table IV (t table), Table VII (χ^2 table) from Blackboard before the exam begins.
3. This exam consists of 30 multiple-choice questions (each worth 1 mark) and 3 long answer questions. The total mark is 60.
4. For the long answer questions, show all your work. Each step has credits.
5. A template of answer sheet has been sent to you yesterday. If you didn't receive it, please let me know ASAP. Please print the answer sheet out or copy the answer sheet down on paper by hand.
6. Put all answers on the answer sheet in the allotted 3 hours (1 PM-4 PM on April 23rd) and you will have 15 minutes to submit the pictures of the completed answer sheet to Crowdmark after 4 PM. There will be a 100% late penalty if you fail to submit the answer sheet to Crowdmark before 4:15 PM.
7. **IMPORTANT:** After you submit your answer sheet to Crowdmark, you **must** send a copy to me via email. There is no time limit for that but please have it done ASAP. My email address is hur3@macewan.ca.

Multiple choice questions

Use the following information to answer Questions 1-4

The density curve of a continuous random variable X is shown in Figure 1.

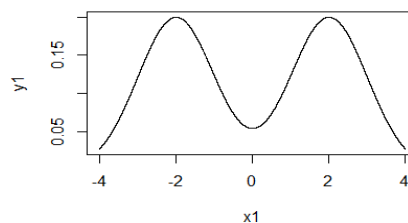


Figure 1: The density curve of a continuous random variable X

1. The shape of this density curve is
 - (a) asymmetric and unimodal.

- (b) asymmetric and bimodal.
 - (c) symmetric at 0 and unimodal.
 - (d) symmetric at 0 and bimodal.
 - (e) two-tailed skewed and unimodal.
2. What is the area under the density curve?
- (a) The area under the density curve is equal to 6.
 - (b) The area under the density curve is equal to 50%.
 - (c) The area under the density curve is equal to 95%.
 - (d) The area under the density curve is equal to 100%.
 - (e) The area under the density curve is unknown.
3. According to the graph, the second quartile (median) of the random variable X is
- (a) -3 or 3 (b) -1 or 1 (c) 0 (d) 0.5 (e) unavailable from the graph.
4. What percentage of data lies between the first quartile (-0.6745) and the second quartile?
- (a) 25% (b) 50% (c) 67.45% (d) 1.349% (e) 100%
-

Use the following information to answer Questions 5-9

Hospital records show that 10% of all patients are admitted for surgical treatment.

5. What is the probability that a patient admitted doesn't need surgical treatment?
- (a) 0.9 (b) 0.81 (c) 0.19 (d) 0.01 (e) 0.1
6. What is the probability that two patients admitted don't need surgical treatment?
- (a) 0.9 (b) 0.81 (c) 0.19 (d) 0.01 (e) 0.1
7. What is the probability that at least one of two patients admitted needs surgical treatment?
- (a) 0.9 (b) 0.81 (c) 0.19 (d) 0.01 (e) 0.1

Let X = the number of patients that need surgical treatment among ten patients admitted in the hospital.

8. What is the distribution of X ? It follows a binomial distribution with the number of trials $n = 10$ and success probability
- (a) $p = 0.81$ (b) $p = 0.9$ (c) $p = 0.01$ (d) $p = 0.1$. (e) $p = 0.19$
9. What is the probability that two patients need surgical treatment among ten patients admitted in the hospital?
- (a) 0.01 (b) 0.0086 (c) 0.45 (d) 0.0043 (e) 0.1937
-

Use the following information to answer Questions 10-15

Students at a university completed the Anxiety Scale questionnaire. Scores on the scale range from 0 (no anxiety) to 20 (extreme anxiety). The mean score was 12.5 and the standard deviation was 4. Assume that the distribution of all scores on the Anxiety Scale is normally distributed.

10. The probability that someone scores less than a 10 on the Anxiety Scale is
(a) 0.2676 (b) 0.1894 (c) 0.7324 (d) 0.8106 (e) 0.5430
 11. The probability that someone scores greater than a 16 on the Anxiety Scale is
(a) 0.2676 (b) 0.1894 (c) 0.7324 (d) 0.8106 (e) 0.5430
 12. The probability that someone scores between a 10 and a 16 on the Anxiety Scale is
(a) 0.2676 (b) 0.1894 (c) 0.7324 (d) 0.8106 (e) 0.5430
 13. The students whose scores are among the top 5% are suggested to find help. The score of a student is 17.9. Does he need help?
(a) He needs help because his score is higher than 16.7575.
(b) He needs help because his score is higher than 5.94.
(c) He doesn't need help because his score is lower than 20 (extreme anxiety).
(d) He needs help because his score is higher than 17.86.
(e) He doesn't need help because his score is lower than 19.08.
 14. Denote the average score of 4 randomly selected students as $\bar{X}$. Then $\bar{X}$ is a random variable. It follows
(a) a normal distribution with mean of 12.5 and standard deviation of 2.
(b) an approximate normal distribution with mean of 12.5 and standard deviation of 2.
(c) a normal distribution with mean of 6.25 and standard deviation of 1.75.
(d) an approximate normal distribution with mean of 6.25 and standard deviation of 1.75.
(e) an approximate normal distribution with mean of 6.25 and standard deviation of 4.
 15. Find the probability that the average score of 4 randomly selected students is less than 12.
(a) 0.5987 (b) 0.3085 (c) 0.7157 (d) 0.4013 (e) 0.6915
-

Use the following information to answer Questions 16-19

In ancient Chinese court system, a defendant is assumed guilty until proven innocent. Suppose that you regard a court trial in ancient China as a hypothesis test with null and alternative hypotheses

$$H_0 : \text{Defendant is guilty} \quad \text{versus} \quad H_a : \text{Defendant is innocent}$$

16. The Type I error occurs
- (a) if the defendant is guilty but the jury concludes that the defendant is innocent.
 - (b) if the defendant is innocent but the jury concludes that the defendant is guilty.
 - (c) if the defendant is guilty and the jury concludes that the defendant is not innocent.
 - (d) if the defendant is innocent and the jury concludes that the defendant is not guilty.
 - (e) if the defendant is guilty but the jury failed to make a decision.
17. The Type II error occurs
- (a) if the defendant is guilty but the jury concludes that the defendant is innocent.
 - (b) if the defendant is innocent but the jury concludes that the defendant is guilty.
 - (c) if the defendant is guilty and the jury concludes that the defendant is not innocent.
 - (d) if the defendant is innocent and the jury concludes that the defendant is not guilty.
 - (e) if the defendant is guilty but the jury failed to make a decision.
18. If you were an innocent defendant,
- (a) you would want β (the probability of making a Type II error) to be small.
 - (b) you would want α (the probability of making a Type I error) to be large.
 - (c) you would want α and β equal.
 - (d) you would want α and β to be 0.
 - (e) you would want α to be 0 (i.e., β to be 1).
19. What are the consequences to the court system if you make $\alpha = 0$?
- (a) If $\alpha = 0$, then a guilty person would always be found innocent.
 - (b) If $\alpha = 0$, then an innocent person would always be found guilty.
 - (c) If $\alpha = 0$, then a guilty person would always be found guilty.
 - (d) If $\alpha = 0$, then an innocent person would always be found innocent.
 - (e) If $\alpha = 0$, then the court always makes a type II error.
-

Use the following information to answer Questions 20-28

Suppose a manufacture of printers for personal computers wishes to manufacturer tests 100 printheads and records the number of characters printed until failure for each. The sample mean $\bar{x}$ is 2.37 and sample standard deviation s is 0.88.

20. A confidence interval for the population mean is in the form of

from $\bar{x} - c$ to $\bar{x} + c$.

The value of c is half length of the confidence interval, i.e., **margin of error**.

For a 90% Confidence Interval for the mean number of characters printed until failure, the **margin of error** (i.e., the half length of the confidence interval) is

- (a) 0.1461 (b) 0.0146 (c) 0.1135 (d) 0.5101 (e) 0.2689
21. Based on the answer to the previous question, a 90% Confidence Interval for the mean number of characters printed until failure is
- (a) from 2.2565 to 2.4835
(b) from 1.8599 to 2.8801
(c) from 2.2239 to 2.5161
(d) from 2.3554 to 2.3846
(e) from 2.1011 to 2.6389
22. A 95% confidence interval for the mean number of characters printed until failure is (1.7405, 2.9995). This interval can be explained as
- (a) The probability that the sample mean of characters printed until failure is between 2.4298 and 2.3102 is 95%.
(b) There is a 95% chance that the mean number of characters printed until failure is smaller than 2.4298 and larger than 2.3102.
(c) The probability that mean number of characters printed until failure is between 2.4298 and 2.3102 is 95%.
(d) The probability that the mean number of characters printed until failure is not between 2.4298 and 2.3102 is only 5%.
(e) It's impossible to determine if the mean number of characters printed until failure is between 2.4298 and 2.3102. But we have 95% confidence that the mean is between 2.4298 and 2.3102.

Let μ = the average number of characters printed until failure
and

$\bar{x}$ = the average number of characters printed until failure of the sample

To **test** if the average number of characters printed until failure is more than 1.5 (in millions of characters) at significance level 0.05, we follow the 6 steps.

23. Step 1. Set up the hypotheses:

- (a) $H_0 : \bar{x} \geq 1.5$ vs. $H_a : \bar{x} < 1.5$
(b) $H_0 : \mu \geq 1.5$ vs. $H_a : \mu < 1.5$

- (c) $H_0 : \bar{x} \leq 1.5$ vs. $H_a : \bar{x} > 1.5$
- (d) $H_0 : \mu \leq 1.5$ vs. $H_a : \mu > 1.5$
- (e) $H_0 : \bar{x} = 1.5$ vs. $H_a : \bar{x} \neq 1.5$

24. Step 2. Decide the significance level α and check the assumptions.

- (a) The significance level $\alpha = 0.05$. The sample mean $\bar{X}$ is normally distributed because the parent distribution is normal.
- (b) The significance level $\alpha = 0.05$. The sample mean $\bar{X}$ is approximately normally distributed because the sample standard deviation is large.
- (c) The significance level $\alpha = 0.05$. The sample mean $\bar{X}$ is approximately normally distributed according to Central Limit Theorem.
- (d) The significance level $\alpha = 0.05$. The distribution of the sample mean $\bar{X}$ is t-distribution with $df = 99$.
- (e) The significance level $\alpha = 0.05$. The distribution of the sample mean $\bar{X}$ can't be determined because the sample size ($n=100$) is too small.

25. Step 3. Calculate the value of the test statistic.

- (a) $\frac{2.37}{0.88/\sqrt{100}}$ (b) $\frac{-2.37}{0.88/\sqrt{100}}$ (c) $\frac{2.37-1.5}{0.88/\sqrt{100}}$ (d) $\frac{1.5-2.37}{0.88/\sqrt{100}}$ (e) $\frac{2.37-1.5}{0.88/\sqrt{100-1}}$

26. Step 4. Find the rejection region (or the P-value).

- (a) The rejection region consists of the left region of -1.984 and the right region of 1.984 (the P-value is between 0.01 and 0.05)
- (b) The rejection region consists of the left region of -1.660 (the P-value is greater than 0.05)
- (c) The rejection region consists of the right region of 1.660 (the P-value is smaller than 0.005)
- (d) The rejection region consists of the left region of -1.984 (the P-value is greater than 0.01)
- (e) The rejection region consists of the right region of 1.984 (the P-value is between 0.005 and 0.01)

27. Steps 5-6. Make a conclusion.

- (a) Since the test statistic value falls in the rejection region (or P-value is less than 0.05), we reject H_0 at $\alpha = 5\%$.
- (b) Since the test statistic value falls in the rejection region (or P-value is less than 0.05), we don't reject H_0 at $\alpha = 5\%$.
- (c) Since the test statistic value doesn't fall in the rejection region (or P-value is greater than 0.05), we don't reject H_0 at $\alpha = 5\%$.
- (d) Since the test statistic value doesn't falls in the rejection region (or P-value is greater than 0.05), we reject H_0 at $\alpha = 5\%$.
- (e) Since the test statistic value falls in both of the rejection region and non-rejection region (or P-value is equal to 0.05), we may or may not reject H_0 at $\alpha = 5\%$.

28. The conclusion that we made in steps 5-6 shows that

- (a) although H_0 is rejected we still can't be 100% sure that the average number of characters printed until failure is more than 1.5 (in millions of characters).
 - (b) H_0 is rejected. Therefore, we are 100% sure that the average number of characters printed until failure is more than 1.5 (in millions of characters).
 - (c) H_0 is not rejected. Therefore, there is a 95% chance that the average number of characters printed until failure is less than or equal to 1.5 (in millions of characters).
 - (d) H_0 is not rejected. Therefore, we are 100% sure that the average number of characters printed until failure is less than 1.5 (in millions of characters).
 - (e) H_0 is rejected. Therefore, the probability that the average number of characters printed until failure is more than 1.5 (in millions of characters) is 95%.
-

Use the following information to answer Questions 29-30

Figure 2 is the scatter plot of PSAT scores and SAT scores of 100 students.

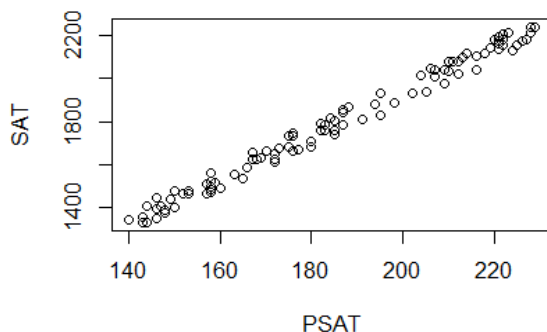


Figure 2: scatter plot of PSAT scores and SAT scores

- 29. According to the scatter plot of PSAT scores and SAT scores, the correlation coefficient (r) of PSAT scores and SAT scores is
 - (a) less than -0.01. (b) equal to 1. (c) equal to 0. (d) greater than 0. (e) greater than 1.5.
- 30. According the result of Minitab, the fitted regression line is

$$SAT = -105.46 + 10.23PSAT$$

The predicted SAT score for a student whose PSAT score is 100 is

- (a) 917.54 (b) 952.3 (c) 1023 (d) 105.46
- (e) unable to be determined because 100 is outside the range of PSAT scores of the data set.

Long-answer Questions

31. (10 marks) An instructor wants to compare the final exam average marks of students who attended lectures regularly and students who did not attend lectures regularly. He selected two random samples. The Table 31 shows the summary of the two samples.

	sample size	sample mean	sample standard deviation
Attended regularly	$n_1 = 49$	$\bar{x}_1 = 75$	$s_1 = 7$
Not attended regularly	$n_2 = 36$	$\bar{x}_2 = 70$	$s_2 = 6$

At 5% significance level, test whether students who attended lectures have higher average.

32. (10 marks) A pediatrician is interested in the 4-5 year-old Canadian child development. She randomly selected 30 children and recorded their heights (unit: cm) as follow:

91.9, 97.8, 111.4, 122.3, 105.4, 95.0, 103.8, 99.6, 96.6, 119.3, 104.8, 101.7, 98.0, 88.2, 91.9, 98, 99.6, 101.2, 112, 103, 111, 102.3, 99, 98.7, 100.2, 89, 88.9, 90.1, 96.7, 93.6.

- (a) (4 marks) In summary, 20 children in this sample are taller than 97 cm and 10 children are shorter than 97 cm. Find an 80% Confidence Interval for the proportion of children aged 4-5 taller than 97 cm.
- (b) (6 marks) Test if more than 50% Canadian children at age 4-5 are taller than 97 cm with significant level 0.05.
33. (10 marks) A researcher performed a study to determine whether the distribution of political party affiliation is different for different income brackets. For each of the income brackets listed, she collected a random sample in Table 1. With significance level 5%, does the data provide sufficient evidence to conclude that the distribution of political party affiliation is different for different income brackets?
(Hint: you need to test if *political party affiliation* is associated with *income bracket*.)

	Income Bracket			
Party	Low	Middle	High	Total
A	101	130	97	328
B	78	113	146	337
C	15	34	9	58
Total	194	277	252	723

Table 1: Data for study of political party affiliation and income brackets