

MATH 203: Principles of Statistics 1

Practice Final Examination: Solutions

Fall 2024

This practice exam is a reformatted version of a real past exam.

Note: Numerical solutions follow full precision. In practice, rounding to 3 or 4 decimal places is sufficient.

MATH 203: Specific Instructions.

Please write your answers on the exam in the boxes provided for each question; only answers written inside the correct boxes will be marked. Extra space for answers can be found at the end of the exam paper. Please write clearly.

This paper contains ten questions. Each question carries 10 marks. Credit will be given for all questions attempted. The total mark available is 100 but rescaling of the final mark may occur.

Only non-programmable and non text-storing calculators may be used. Dictionaries and Translation dictionaries are permitted.

A formula sheet is provided.

1. The Montreal police department has compiled data relating the age of drivers to the probability of being involved in an accident in a one year period:

Age group	Proportion of drivers	Conditional Probability of accident
Under 25	0.16	0.055
25-64	0.70	0.023
65 and over	0.14	0.04

For example, the conditional probability of having an accident given that the driver falls in the Under 25 age group is 0.055. What is the probability that during one year

- (a) there is an accident and the driver is over 65? 2 MARKS
- (b) there is an accident? 4 MARKS
- (c) the driver is over 65, given that there is an accident? 4 MARKS

Show your working.

Q1

Let A_1 , A_2 and A_3 denote the partition by the three driver age groups, and A denote the event that an accident occurs.

- (a) We want

$$P(A \cap A_3) = P(A_3) \times P(A|A_3) = 0.14 \times 0.04 = 0.0056.$$

- (b) We want

$$\begin{aligned}
 P(A) &= P(A \cap A_1) + P(A \cap A_2) + P(A \cap A_3) \\
 &= P(A_1) \times P(A|A_1) + P(A_2) \times P(A|A_2) + P(A_3) \times P(A|A_3) \\
 &= (0.16 \times 0.055) + (0.7 \times 0.023) + (0.14 \times 0.04) \\
 &= 0.0305
 \end{aligned}$$

- (c) We want

$$\begin{aligned}
 P(A_3|A) &= \frac{P(A|A_3)P(A_3)}{P(A)} \\
 &= \frac{0.14 \times 0.04}{0.0305} \\
 &= 0.1836066
 \end{aligned}$$

2. Seat pitch on a passenger airline is the distance from the back of one seat to the front of the one directly behind it. It is known that the greater the seat pitch the more comfortable the seat. The seat pitch for an economy seat is considered comfortable if it is between 33.25 and 34.75 inches. The seat pitch for all economy seats is Normally distributed with mean 34 inches and standard deviation 0.5 inch.

(a) For a randomly selected economy seat, what is the probability that it will be comfortable? 3 MARKS

(b) If six economy seats are selected at random and independently, what is the probability that **fewer than three** seats will not be considered comfortable? 5 MARKS

(c) If ten economy seats are to be selected at random and independently, what is the **expected number** of seats that will be considered comfortable? 2 MARKS

Justify your answer in each case.

Q2

Let Y denote the pitch rv, so that $Y \sim Normal(34, (0.5)^2)$.

(a) We want

$$P(33.25 \leq Y \leq 34.75) = P\left(\frac{-0.75}{0.5} \leq \frac{Y - 34}{0.5} \leq \frac{0.75}{0.5}\right)$$

which equals

$$P(-1.5 \leq Z \leq 1.5)$$

where $Z \sim Normal(0, 1)$. We compute this from the Normal tables as

$$\Phi(1.5) - \Phi(-1.5) = 0.9332 - 0.0668 = 0.8664$$

as by symmetry for $z < 0$, $\Phi(z) = 1 - \Phi(-z)$, so $\Phi(-1.5) = 1 - \Phi(1.5)$

(b) The probability that a seat is not considered comfortable is $q = 1 - 0.8664 = 0.1336$, and therefore the number, X , of seats (out of six) that will not be comfortable has a $Binomial(6, q)$ distribution. Therefore the probability that fewer than three will not be considered comfortable is

$$P(X = 0) + P(X = 1) + P(X = 2) = \binom{6}{0} q^0 p^6 + \binom{6}{1} q^1 p^5 + \binom{6}{2} q^2 p^4 = 0.9651685.$$

(c) Here $X \sim Binomial(10, p)$ with $p = 1 - q = 0.8664$, so the expected number is

$$E(X) = np = 10 \times 0.8664 = 8.664.$$

3. The area (in square cm) of the bacteria-free zone on a Petri dish is known to be a continuous random variable with a mean of 3.6 and a standard deviation of 0.5.
- (a) If one Petri dish is examined, what is the probability that its area will be exactly equal to 3.6 square cm? 2 MARKS
 - (b) If 100 such dishes are examined, what is the approximate probability that the average area will be less than or equal to 3.7 square cm? 5 MARKS
 - (c) In a data sample, approximately what proportion of the sampled areas will lie between 2.6 and 4.6 square cm ? 3 MARKS

Justify your answers.

Q3

- (a) This probability is zero, as the distribution is continuous
- (b) Here we use the Central Limit Theorem to write for the sample average area, $\bar{X}$,

$$\bar{X} \rightsquigarrow Normal(\mu, (\sigma/\sqrt{n})^2) = Normal(3.6, (0.5/\sqrt{100})^2) = Normal(3.6, (0.05)^2)$$

so therefore

$$P(\bar{X} \leq 3.7) = P\left(\frac{\bar{X} - 3.6}{0.05} \leq \frac{3.7 - 3.6}{0.05}\right) = P(Z \leq 2) = 0.9772.$$

from standard Normal tables.

- (c) There are two possible solutions
 - If we are willing to assume that the distribution is mound-shaped and symmetric, we can use the Empirical Rule to say that the proportion is approximately 95%, as (2.6, 4.6) is precisely $(\mu - 2\sigma, \mu + 2\sigma)$
 - We are not told anything about the distribution, so it is not possible to say what the proportion is.

We may use Chebychev's Rule with $k = 2$ to bound (but not approximate) the proportion.

4. A car manufacturer advertises that its new subcompact models get an average of 51 miles per gallon. A consumer research group believes the average is lower than 51 miles per gallon. To justify their claim, they randomly sample 35 cars and the results show a sample mean of 48.2 miles per gallon and a sample standard deviation of 4.2 miles per gallon. A statistical hypothesis test is to be carried out using $\alpha = 0.05$.
- (a) State the null hypothesis and the alternative hypotheses to be tested. Use mathematical notation. 2 MARKS
 - (b) Conduct a statistical test of hypothesis and state your conclusion. 6 MARKS
 - (c) Is it necessary to make a Normal assumption for the original data for the hypothesis test to be statistically valid ? Justify your answer. 2 MARKS

Q4

We have a one sample problem for parameter μ , with $n = \text{large}$.

(a)

$$H_0 : \mu = 51$$

Null hypothesis

$$H_a : \mu < 51$$

Alternative hypothesis

(b) We use test statistic

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - 51}{S/\sqrt{35}}$$

with observed value

$$T_{obs} = \frac{\bar{x} - 51}{4.2/\sqrt{35}} = \frac{48.2 - 51}{4.2/\sqrt{35}} = -3.9440532.$$

We have that the Rejection Region is $\{T \leq -z_{0.05}\} = \{T \leq -1.6449\}$ from the t-table, final row. The test statistic lies in the Rejection Region, so we reject H_0 in favour of H_a .

(c) It is not necessary to make the Normal assumption as the sample size is large ($n \geq 30$).

5. A cream with alpha-hydroxy acid, advertises that it can reduce wrinkles and improve the skin. In a study published in Archives of Dermatology (June 1996), 49 participants used a cream with alpha-hydroxy acid for 22 weeks. At the end of the study period, a dermatologist judged whether each participant exhibited any improvement in the condition of their skin: 30 showed improvement and 19 did not show improvement.
- (a) Find a 95% confidence interval for the true population proportion of participants with improved skin after using this cream. Interpret the interval. 5 MARKS
- (b) Do the data provide sufficient evidence to conclude that the cream will improve the skin of **more than** 60% of subjects? Carry out a hypothesis test with $\alpha = 0.01$ and report the result. 5 MARKS

Q5

We have a one sample problem for parameter p , with n large.

- (a) From the formula sheet, the 95% C. I. uses $\alpha = 0.05$ and the interval

$$\left(\hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

with $\hat{p} = 30/49 = 0.6122449$. Thus the interval is

$$(0.4758182, 0.7486716)$$

We can say that the true proportion with improved skin would lie within the interval formed by using the C. I. formula at a rate of $1 - \alpha = 0.95$ over an infinite number of independent replications of this experiment with the same sample size

- (b) We have a one sample problem for parameter p , with $n\hat{p}, n(1 - \hat{p}) \geq 15$; we test

$$H_0 : p = 0.6$$

$$H_a : p > 0.6$$

We use test statistic

$$T = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{\hat{p} - 0.6}{\sqrt{\frac{0.6 \times 0.4}{49}}}$$

with observed value

$$T_{obs} = \frac{0.6122449 - 0.6}{\sqrt{\frac{0.6 \times 0.4}{49}}} = 0.1749636$$

We have that the Rejection Region is $\{T \geq z_{0.01}\} = \{T \geq 2.3263\}$ from the t-table, final row. The test statistic does not lie in the Rejection Region, so we do not reject H_0 in favour of H_a .

6. Researchers compared the hatching rates of two groups of Pacific tree frog eggs. One group of eggs was shielded with ultraviolet blocking sun shades, while the second group was not. The number of eggs successfully hatched in each group is provided in the following table:

	Sun-shaded eggs	Unshaded eggs
Number hatched	34	31
Total number	70	80

- (a) Find a 95% confidence interval for the difference between the population proportion of hatched eggs for sun-shaded and the population proportion of hatched eggs for unshaded. 5 MARKS
- (b) Test the null hypothesis that the population proportions of these two groups are the same against the alternative that they are different using $\alpha = 0.05$. 5 MARKS

Q6

We have a two sample problem for parameter $p_1 - p_2$, with sample size large. From the data, we have

- $\hat{p}_1 = 34/70 = 0.4857143$,
- $\hat{p}_2 = 31/80 = 0.3875$,
- $\hat{p} = 65/150 = 0.4333333$.

- (a) From the formula sheet, the 95% C. I. uses $\alpha = 0.05$ and the interval

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \times \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{m} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n}},$$

Thus the interval is $(-0.0602346, 0.2566632)$.

- (b) We test

$$H_0 : p_1 = p_2$$

$$H_a : p_1 \neq p_2$$

We use test statistic

$$T = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{m} + \frac{1}{n} \right)}} \quad T_{obs} = \frac{0.4857 - 0.3875}{\sqrt{0.4333 \times (1 - 0.4333) \left(\frac{1}{70} + \frac{1}{80} \right)}} = 1.2110112$$

We have that the Rejection Region is $\{|T| \geq z_{0.025}\} = \{|T| \geq 1.9600\}$ from the t-table, final row. The test statistic does not lie in the Rejection Region, so we do not reject H_0 in favour of H_a .

7. An experiment was conducted to compare the mean reaction times to two types of traffic signs: prohibitive (No Left Turn) and permissive (Left Turn Only). Ten drivers were included in the experiment. Each driver was presented 40 traffic signs, 20 prohibitive and 20 permissive, in random order. The mean time to reaction and the number of correct actions were recorded for each driver. The mean reaction times (in milliseconds) to the 20 prohibitive and 20 permissive traffic signs are shown in the following table for each of the ten drivers.

Driver	Prohibitive	Permissive	Difference
1	824.00	702.00	122.00
2	866.00	725.00	141.00
3	841.00	744.00	97.00
4	770.00	663.00	107.00
5	829.00	792.00	37.00
6	764.00	708.00	56.00
7	857.00	747.00	110.00
8	831.00	685.00	146.00
9	846.00	742.00	104.00
10	759.00	610.00	149.00
Mean:	818.7	711.8	106.9
Standard deviation:	39.67	50.95	36.95

- (a) Do the data provide sufficient evidence to indicate a difference in mean reaction times to prohibitive and permissive traffic signs? Carry out a hypothesis test using $\alpha = 0.05$. State the main assumption needed to make the test statistically valid.

6 MARKS

- (b) Construct a 95% confidence interval for the mean reaction time to permissive traffic signs. State the main assumption needed to make the interval statistically valid.

4 MARKS

Here we have a paired comparison, with $n = 10$ small. We have $\bar{d} = 106.9$ and $s_d = 36.95$, and must compare against the t_9 distribution.

(a) We test

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

We use test statistic

$$T = \frac{\bar{D}}{S_d/\sqrt{n}} \quad T_{obs} = \frac{106.9}{36.95/\sqrt{10}} = 9.1487816$$

We have that the Rejection Region is $\{|T| \geq t_{0.025,9}\} = \{|T| \geq 2.2622\}$ from the t-table row for $\nu = 9$. The test statistic lies in the Rejection Region, so we do reject H_0 in favour of H_a .

Here we must assume that the difference data are Normally distributed.

(b) Here we have $\bar{y} = 711.8$ and $s_2 = 50.95$. From the formula sheet, the 95% C. I. uses $\alpha = 0.05$ and the interval is

$$\bar{y} \pm t_{\alpha/2, n-1} \times \frac{s_2}{\sqrt{n}}$$

that is

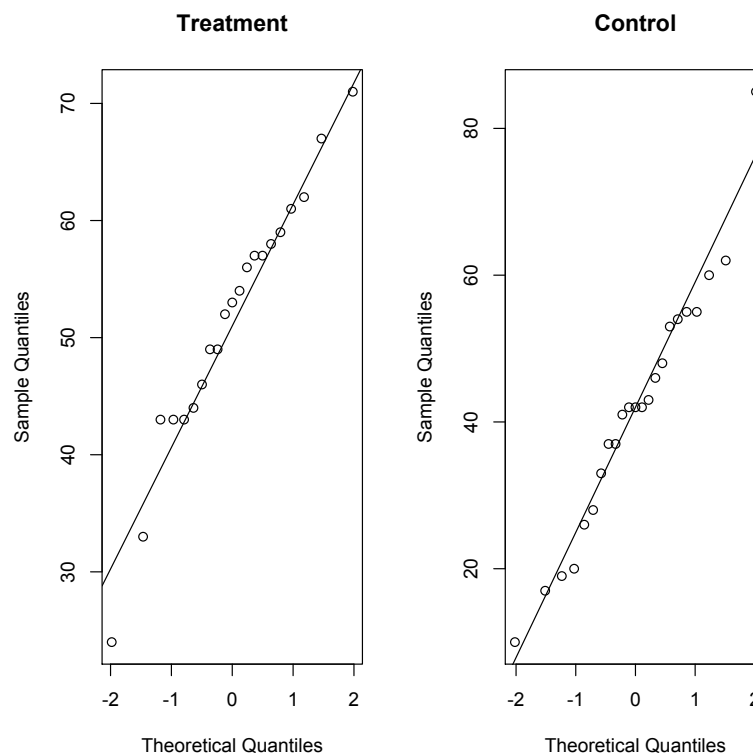
$$(675.3518755, 748.2481245).$$

Here we must assume that the permissive sign data are Normally distributed.

8. An educator believes that new directed reading activities in the classroom will help elementary school pupils improve some aspects of their reading ability. She arranges for a third-grade class of 21 students to take part in these activities for an eight-week period. A control classroom of 23 third-graders follows the same curriculum without the activities. At the end of the eight weeks, all students are given a Degree of Reading Power (DRP) test, which measures the aspects of reading ability that the treatment is designed to improve. A summary of the results is shown in the table.

Group	n	$\bar{x}$	s
Treatment	21	51.48	12.11
Control	23	41.52	15.45

Below are Normal probability (Q-Q) plots for the data (here the sample data have not been standardized):



- Test the null hypothesis that the population mean DRP scores of the two groups are the same against the alternative hypothesis that they are different, using $\alpha = 0.05$.
6 MARKS
- State the necessary assumptions for your test in (a) to be valid.
2 MARKS
- Are the assumptions met? Justify your answer.
2 MARKS

- (a) Here we have a two sample problem, with m and n small, so we must assume that the population variances are equal, with pooled estimate

$$s_p^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2} = 194.8689762$$

so therefore $s_p = 13.9595479$. The test statistic is

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad \text{with} \quad T_{obs} = \frac{51.48 - 41.52}{13.9595479 \sqrt{\frac{1}{21} + \frac{1}{23}}} = 2.3639333$$

We have that the Rejection Region is $\{|T| \geq t_{0.025,42}\} \approx \{|T| \geq 1.9600\}$ from the t-table row for $\nu = \infty$, which we can use as an approximate upper point. Note also that the upper 0.025 point of the t_{30} distribution is 2.0423, and the test statistic exceeds that. We know that $t_{0.025,42} < t_{0.025,30}$, so the information from the table is sufficient to complete the test.

The test statistic lies in the Rejection Region, so we do reject H_0 in favour of H_a .

- (b) Here we must assume that the original data are Normally distributed, and that the population variances are equal. We also assume independent samples.
- (c) The QQ plots indicate that Normality is a reasonable assumption, and the two sample variances do seem fairly similar; it is reasonable to conclude that the assumptions are met.

Note, we could also compute the large sample version of the test statistic for comparison

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \quad \text{with} \quad T_{obs} = \frac{51.48 - 41.52}{\sqrt{\frac{12.11^2}{21} + \frac{12.11^2}{23}}} = 2.3903524$$

and therefore we can still conclude that the observed version of the test statistic is large enough to reject the null hypothesis.

9. A personality questionnaire was administered to independent samples of Cocaine abusers and College students to measure their degree of sociability. The results are summarized in the table.

Group	n	$\bar{x}$	s
Cocaine abusers	450	10.4	4.3
College students	589	12.5	4.0

- (a) Construct a 95% confidence interval for the difference in population mean scores between Cocaine abusers and College students. 4 MARKS
- (b) Test the null hypothesis that the population mean scores of these two groups are the same against the alternative that they are different using $\alpha = 0.05$. 6 MARKS

Q9

Here we have a two sample problem with n and m large.

- (a) From the formula sheet, the 95% C. I. uses $\alpha = 0.05$ and the interval

$$(\bar{X} - \bar{Y}) \pm z_{\alpha/2} \times \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}},$$

Thus the interval is $(-2.6120575, -1.5879425)$.

- (b) We test

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

We use test statistic

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)}} \quad T_{obs} = \frac{10.4 - 12.5}{\sqrt{\left(\frac{18.49}{450} + \frac{16}{589}\right)}} = -8.038159$$

We have that the Rejection Region is $\{|T| \geq z_{0.025}\} = \{|T| \geq 1.9600\}$ from the t-table, final row. The test statistic lies in the Rejection Region, so we do reject H_0 in favour of H_a .

10. State whether each of these statements is True or False. No justification is necessary.

- (a) It is always true that $P(A|B) + P(A|B^c) = 1$. 1 MARK
- (b) If $P(A \cup B) = P(A) + P(B)$ then A and B are mutually exclusive. 1 MARK
- (c) If A and B are mutually exclusive then they are also independent. 1 MARK
- (d) You can use a statistical test to evaluate the hypotheses 1 MARK

$$H_0 : \bar{x} = 0$$

$$H_a : \bar{x} > 0.$$

- (e) If $P(A \cup B) = P(A) + P(B)$ then A and B are independent. 1 MARK
- (f) The confidence interval for μ defined by

$$\left(\bar{X} - z_{0.025} \frac{S}{\sqrt{n}}, \bar{X} + z_{0.025} \frac{S}{\sqrt{n}} \right)$$

will, with probability 0.95 contain $\bar{X}$. 1 MARK

- (g) As the sample size increases, the standard deviation of the sampling distribution of $\bar{X}$ decreases. 1 MARK
- (h) The width of a confidence interval depends on the sample size. 1 MARK
- (i) In a hypothesis test, a p-value is the probability that the null hypothesis is true. 1 MARK
- (j) If the sample size is large enough, the Central Limit Theorem says that observations in the sample can be regarded as being approximately Normally distributed. 1 MARK

Q10

- (a) FALSE: There is no mathematical basis for this formula
- (b) TRUE: By the general addition rule, it is always true that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

and so we must have $P(A \cap B) = 0$ and A and B are mutually exclusive

- (c) FALSE: $P(A \cap B) = 0$ does not imply $P(A \cap B) = P(A)P(B)$
- (d) FALSE: $\bar{x}$ is a fixed number, so not amenable to statistical hypothesis testing.
- (e) FALSE: as in (b) and (c), $P(A \cap B) = 0$ does not imply $P(A \cap B) = P(A)P(B)$
- (f) FALSE: $\bar{X}$ is in the C. I. with probability 1.
- (g) TRUE: the variance is σ^2/n where σ is fixed.
- (h) TRUE: the width of the C. I. is proportional to $1/\sqrt{n}$
- (i) FALSE: the p-value is the probability of observing a test statistic more extreme than the one was actually observed assuming H_0 is true.
- (j) FALSE: the Central Limit Theorem says that sample mean random variable can be regarded as being approximately Normally distributed, it says nothing about the distribution of the original data.

Extra space

Extra space

Extra space