Math 10 - Compilation of Sample Exam Questions + Answers

Sample Exam Question 1

We have a population of size N. Let p be the independent probability of a person in the population developing a disease. Answer the following questions in terms of N and p.

a) What is the probability that a person in the population NOT developing the disease? (1 pt)

b) If $N = 2$, what is the probability that no one develops the disease? If your answer involves N, please replace it by the number 2. (1 pt)

c) For a general size N, what is the probability that no one develops the disease? (1 pt)

d) What is the probability that, for a general size N, at least one person develops the disease? (1 pt)

Answer: a) $1 - p$, b) $(1 - p)^2$, c) $(1 - p)^N$, d) $1 - (1 - p)^N$.

Sample Exam Question 2

Suppose 3 fair dice are rolled independently. Let their outcomes be D_1, D_2, D_3. Please simplify your answer as much as possible.

a) Suppose that $D_1 = 5$ and $D_2 = 6$, what is the probability that D_3 is a different outcome? (1 pt)

b) In general, what is the probability that D_3 is different from D_1 and D_2 GIVEN that $D_1 \neq D_2$? (1 pt)

c) In general, what is the probability that none of the rolls had the same outcome? (2 pts)

d) What is the probability that at least two of the rolls had the same outcome? (1 pt)

Answer: a) $\frac{2}{3}$, b) $\frac{2}{3}$, c) $\frac{5}{6}\frac{4}{6} = \frac{20}{36} = \frac{5}{9}$, d) $1 - \frac{5}{9} = \frac{4}{9}$.

Sample Exam Question 3

Police officers in a town are stopping drivers at random and subjecting them to a breathalyzer test that detects alcohol intoxication. During the operation, there were 5000 drivers on the road.

One of those drivers were chosen at random, say Tommy, and he tested positive. The police were using breathalyzers that detects drunkenness with 100 % probability of IF the driver is drunk. (positive result) If the driver is sober, it has a 99 % probability of reporting that the driver is NOT drunk. (negative result)

a) The police wants to deport Tommy because they think the probability that he is driving drunk is 99 %. Is their claim true or false based on the information so far? Explain. (2 pts)

Now you are told that out of the 5000 drivers on the road, 5 of them are drunk, while the rest are sober.

b) If all 4995 sober drivers are given the breathalyzer test, approximately how many would get a positive result on average? (1 pt)

c) The police was picked a driver at random from the 5000. What is the probability that he or she is sober? You do not have to simplify your answer. (1 pt)

d) Using Bayes Theorem, write down a numerical fraction for the approximate probability that a randomly selected driver is not drunk, given that he or she got a positive result. You are given that the probability that a randomly selected driver will test positive is $\frac{55}{5000}$. (4 pts)

Answer: a) False, you need more information, e.g. the base rate, b) $4995 \cdot 0.01 = 49.95 \approx 50$, c) $\frac{4995}{5000}$, d) $D =$ drunk, $Pos =$ tested positive. By Bayes Theorem: $P(\text{ not } D|Pos) = \frac{P(Pos|\text{ not } D)P(\text{ not } D)}{P(Pos)} = \frac{0.001\frac{4995}{5000}}{\frac{55}{5000}} \approx \frac{50}{55} = \frac{10}{11}$.

Sample Exam Question 4

The proportion of scores for 1,000 students in a class are well approximated by a normal distribution, with mean 50 and standard deviation 10. Approximately how many students scored 70 and above? Show your work. Illustrations are allowed. (2 pts)

Answer: using the 68-95 heuristic/rule from the textbook, approximately 95 % of students scored between 30 and 70. So 2.5 % scored 70 and above. Then, approximately $1000 \cdot 0.025 = 25$ is the answer.

Sample Exam Question 5

The weights for 1000 bars of gold are well approximated by a normal distribution, with mean 100 grams and standard deviation 20 grams. You may round your final answers to the nearest integer.

a) Approximately how many bars of gold weigh between 100 grams and 130 grams inclusive? (2 pts)

b) Approximately how many bars of gold weigh less than 80 grams?

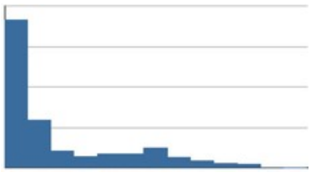
Answer: Let X be the weight of a randomly selected bar of gold. $P(X \leq k) = P(Z \leq \frac{k-100}{20})$, where Z is standard normal (z-tables).

a) So, $P(X \leq 100) = P(Z \leq \frac{100-100}{20} = 0) = 0.50$, and $P(X \leq 130) = P(Z \leq \frac{130-100}{20}) = P(Z \leq 1.5) = 0.9332$. So, $P(X \leq 130) - P(X \leq 100) = 0.9332 - 0.50 = 0.4332$. Ans: approximately $0.4332 \cdot 1000 \approx 433$ bars.

b) Also, $P(X \leq 80) = P(Z \leq \frac{80-100}{20}) = P(Z \leq -1) = 0.1587$. Ans: $0.1587 \cdot 1000 \approx 159$.

Sample Exam Question 6

The incomes of people in country X has a distribution that looks like the one below, with population mean μ and variance σ^2.



a) If you take a large simple random sample of n incomes from country X, what is a good approximation of the sampling distribution of the sample mean M? What are the mean and variance of this approximation? (2 pt)

b) Which theorem made the approximation in the previous question possible? (1 pt)

c) If another researcher independently took another large simple random sample of n incomes from country X, what is the probability that his sample mean would be in the interval $[\mu - \frac{\sigma}{\sqrt{n}}, \mu + \frac{\sigma}{\sqrt{n}}]$? (2 pts)

Answer: a) Normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. b) Central Limit Theorem. c) Using the 68-95 heuristic/rule from the textbook, the probability is 0.68.

Sample Exam Question 7

You are a small employee in a big chain of supermarkets. Your boss wants to know if the companys image is attracting more men than women, or is it roughly 50/50. Population has two types: Men and Women. You take a sample of size $n = 10$. You find that your sample proportion of men is 60 % or 0.60 or $\frac{3}{5}$.

You want to construct a numerical 90 % confidence interval for the population proportion of men, because you took Math 10 before. How would you do it?

Your answer must be in the form $[a, b]$, where a and b are numbers that may include square roots. You do not have to evaluate square roots. You do not have to adjust the limits of the interval by $\frac{0.5}{n}$. *(6 pts).*

Answer: formula for the confidence interval is $[p - Z_\alpha S_p, p + Z_\alpha S_p]$, where $p = \frac{3}{5}$, $Z_\alpha = 1.64$, and the standard error $S_p = \sqrt{\frac{\frac{3}{5} \cdot \frac{2}{5}}{10}} = \sqrt{\frac{6}{250}}$.

Sample Exam Question 8

Suppose you have a block of metal that is exactly 500 grams in weight. Sorry, no imperial units allowed in my class.

You have an electronic weighing scale that may or may not be faulty. You put this block of metal on the weighing scale $n = 25$ times, and record the results.

Each result varies a little due to various reasons (positioning, random mechnical errors etc), but you are hoping to find out if the scale is correct on average, or systematically giving you lower/higher than the actual weight.

If the scale is correct on average, it would produce results drawn from a normal distribution with mean 500 grams and unknown variance σ^2. (E.g. real life scales will tell you they are intended to be accurate within X grams)

The sample mean of your $n = 25$ data points is $M = 490$ grams.

1. Can you conclude that the electronic weighing scale is faulty and is systematically giving you results that are lower than the actual weight, on average? Since $M = 490 < 500$ true weight? Explain. *(2 pts)*

2. Using this set of sample data, you calculated an estimate of the standard deviation $s = 20$ grams. What sampling distribution of the mean would you use? State all the parameters of this sampling distribution (using μ for the real mean). Why do you use this sampling distribution? *(3 pts)*

3. Using the statistics produced by your sample, construct a 95 % confidence interval for the mean. *(4 pts)*

Answers

1) No. Even if the scale is correct on average, you could have gotten 490 by chance.

2) No population variance, have to use t distribution. Mean μ, standard error $\frac{20}{\sqrt{25}} = 4$, degrees of freedom 24.

3) Degrees of freedom $= 25 - 1 = 24$. Find the t-value, $t = 2.06$. Formula: $[M - t \cdot SE, M + t \cdot SE]$, where SE is the standard error given.

Lecture 1 - Class Exercises

Summation Notation Practice

$X_1 = 1, \quad X_2 = 2, \quad X_3 = 3, \quad X_4 = 4, \quad X_5 = 5$

1. Calculate $\sum_{i=1}^{5} X_i$.

2. Calculate $\sum_{i=1}^{3} X_i$.

3. Calculate $\sum_{i=1}^{3} X_i^2$.

4. Calculate $\left(\sum_{i=1}^{2} X_i\right)^2$.

5. Calculate $\left(\sum_{i=1}^{2} X_i^2\right)^2$.

$Y_1 = 1, \quad Y_2 = 2$

6. Calculate $\sum_{i=1}^{2} X_i Y_i$.

7. Calculate $\sum_{i=1}^{2} X_i^2 Y_i$.

Linear Transformation Practice

$X_1 = 1, \quad X_2 = 2, \quad X_3 = 3, \quad X_4 = 4, \quad X_5 = 5$

8. Calculate Z_1 and Z_2, where $Z_i = 2X_i + 1$.

9. Is $Z_i = 5X_i^2 - 7$ a linear transformation of the X_is?

10. If you plotted $Z_i = 10X_i - 2$ with Z_i on the vertical axis and X_i on the horizontal axis, then connect all the points with an infinitely long line, where would the vertical intercept be?

Logarithm Practice

11. What is the value of $\log_{10}(10000)$?

12. What is the value of $\log_2(16)$?

13. What is the value of $\log_3(27)$?

$W_1 = 4, \quad W_2 = 16, \quad W_3 = 64, \quad W_4 = 256$

14. If you plotted the value of W_i on the vertical axis and corresponding i on the horizontal axis, would you be able to connect the resulting (i, W_i) points with a line? *(note: "lines" in mathematics are always straight, unless specified otherwise)*

15. Calculate $U_i = \log_4(W_i)$ for $i = 1, 2, 3, 4$.

16. If you plotted the value of U_i on the vertical axis and corresponding i on the horizontal axis, would you be able to connect the resulting (i, U_i) points with a line?

17. Is $U_i = \log_4(W_i)$ a linear transformation of the W_is?

Answers

1) 15
2) 6
3) 14
4) 9
5) 25
6) 5
7) 9
8) 3, 5
9) No
10) -2
11) 4
12) 4
13) 3
14) No
15) 1, 2, 3, 4
16) Yes
17) No

Lecture 2 - Class Exercises

Central Tendency and Variability Practice

Our favourite dataset : $X_1 = 1, \quad X_2 = 2, \quad X_3 = 3, \quad X_4 = 4, \quad X_5 = 5$

1. Calculate the mean of this set of data.

2. Calculate $\frac{1}{5}\sum_{i=1}^{5} X_i$.

3. Calculate the median of this set of data.

4. What is the mode of this set of data? Trick question: it is possible for there to be no mode.

5. Compare the median and the mean for this set of data. Are they the same? Why or why not?

New dataset : $Y_1 = 1, \quad Y_2 = 3, \quad Y_3 = 7, \quad Y_4 = 1$

6. Calculate the mean of this new set of data.

7. Calculate the median of this new set of data.

8. Compare the median and the mean for this set of data. Are they the same? Why or why not?

9. What is the mode of this set of data? Not a trick question this time.

10. Which one "balances the scale"? The mean or the median?

11. Which one minimizes the sum of squared deviations? The mean or the median?

12. Which one minimizes the sum of absolute deviations ? The mean or the median?

13. Consider our favourite dataset : $X_1 = 1, \quad X_2 = 2, \quad X_3 = 3, \quad X_4 = 4, \quad X_5 = 5$. Let us apply a linear transformation $Z_i = 5X_i - 2$ to each one of them. Without calculating each Z_i, can you deduce what their mean would be?

14. If I tell you the variance of our favourite dataset in question 13 is 2. Can you deduce what the variance of the Z_is will be?

Answers

1) 3
2) 3

3) 3
4) N.A.
5) The same. Symmetric/uniform distribution.
6) 3
7) 2
8) Not the same. Skewed.
9) 1
10) Mean
11) Mean
12) Median
13) Yes. Mean of Z is mean of X times 5 minus 2.
14) Yes. $5^2 = 25$ times 2.

Lecture 3 - Class Exercises

Sample Exam Question

Suppose you are given a set of data: $X_1 = 10$, $X_2 = 20$, $X_3 = 30$, $X_4 = 110$, $X_5 = 130$, $X_6 = 300$.

1. Calculate the mean and median of this set of data. Show your work. (2 points)

2. Give a short reason why the mean and the median are the same, or why they are not the same. (1 points)

3. The population variance of this set of data is around 10067. You are told that after applying a linear transformation $Y_i = aX_i - b$ to the data, the population variance of the set of Y_is is around 40268. What is the value of $a > 0$ in the linear transformation? (1 points)

4. If we convert our set of data and its transformation in question 3 into a set of bivariate data (X_i, Y_i), what would the correlation coefficient r be? Explain your answer. (2 points)

5. Suppose we convert our set of data and its transformation in question 3 into a set of bivariate data (X_i, Y_j) by pairing each X_i and Y_j at random. What is a good guess for how the resulting correlation coefficient r might differ from the r calculated in question 4? Why would there be or not be a difference? (2 points)

6. Suppose we apply the transformation $Z_i = \frac{1}{1000}X_i^3 + 50$ to our data. We then covert these data into a set of bivariate data (X_i, Z_i). Would the correlation coefficient be $r = 1$? Why or why not? (2 points)

7. Suppose you were given a set of bivariate data (R_i, S_i) with correlation coefficient $r = 0.65$. Suppose we change the unit of measure of R_i from inches to cm (1 inch is 2.54 cm). What would the new r be? Explain your answer. (2 points)

Answers

1) mean = 100, median = 70.
2) positively skewed. mean is affected more by large values and outliers.
can also give the balance scale argument.
3) $a = 2$.
4) $r = 1$. Perfect linear relationship.
5) r will probably be lower. Less/no correlation if randomized.
6) Will not be 1. Points will not have a perfect linear relationship.
7) New $r = 0.65$. r is not affected by change in units of measurement of one of the variables.

Lecture 5 - Class Exercises

1. Given 4 dice rolls, write down an expression for the probability that at least two rolls have the same outcome. You do not have to simplify your answer.

2. Given 7 dice rolls, what is the probability that **none** of the rolls had the same outcome? (in mathematics, this is known as the pigeonhole principle)

Recall that for the Binomial Distribution, the probability of getting k successes in n trials is ${}^nC_k\, p^k(1-p)^{n-k}$, where ${}^nC_k = \frac{n!}{k!(n-k)!}$.

3. Suppose the probability of being admitted to a college is 0.50 for every student who applies. If 3 students applied, what is the probability that 0 or 2 students got admitted? Give a simplified numerical answer and show your work.

4. If instead, the probability of being admitted is 0.30 and 10 students applied, write down a numerical expression for the probability that between 1 and 3 students got admitted. You do not have to simplify your answer.

5. Suppose there are two candidates running for president, Mr T and Mrs H. There are only 1000 voters. 400 of them voted for Mrs H, while 600 of them voted for Mr T.

Suppose you took a simple random sample of size 100 from these 1000 voters. Write down a numerical expression of the probability that 35 in your sample voted for Mr T. You do not have to simplify your answer.

Answers

1) $1 - \frac{5}{6} \cdot \frac{4}{6} \cdot \frac{3}{6}$.

2) 6 possible outcomes, 7 dice. So, the probability is zero.

3) $P($ 0 success in 3 trials $) + P($ 2 success in 3 trials $) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}$.

4) ${}^{10}C_1\ (0.3)^1(0.7)^9 +^{10} C_2\ (0.3)^2(0.7)^8 +^{10} C_3\ (0.3)^3(0.7)^7$

5) ${}^{100}C_{35}\ (0.6)^{35}(0.4)^{65}$

Lecture 6 - Class Exercise

The scores of a class of $10,000$ students are approximately normally distributed with mean 50 and standard deviation 10.

Hint: you can do this entire question using just the 68-95-99.7 rule, and I recommend you do.

1. Approximately how many students scored between 40 and 60 inclusive? (1 pt)

2. Approximately how many students scored between 50 and 70 inclusive? (2 pts)

3. Approximately how many students, in total, scored ≥ 70 or scored ≤ 20? (4 pts)

Answers
1) 6800
2) 4750
3) 250 + 15 = 265

Lecture 7 - Class Exercise

Suppose that the mean annual salary for a particular job is $\$60,000$ with a standard deviation of $\$10,000$. The distribution of salaries for this job is heavily positively skewed (long tail to the right).

1. What is the approximate probability that the mean of a simple random sample of 100 salaries lies in the interval $[58360, 61640]$? *(4 pts)*

2. Suppose you obtained a new independent simple random sample of 100 salaries, and found that the sample mean for this new sample lies in the interval $[60000 - 1.64 \cdot (10000), 60000 + 1.64 \cdot (10000)] = [58360, 61640]$. We say that the sample mean is within 1.64 standard deviations of the population mean.

Question: would the population mean be also within 1.64 standard deviation of this new sample mean? Why or why not? *(1 pt)*

There are currently 10,000 drivers on the road in a town. The police takes a simple random sample of $n = 100$ drivers and found that 3 of them were drunk. Let's just say whatever test they are using for drunkenness has perfect 100 % accuracy.

3. Based on this sample, should the police conclude that close to 3 percent of the 10,000 drivers were driving drunk? Why or why not? *(2 pts)*

4. Suppose out of those 10,000 drivers, the proportion that were drunk is 0.01 or 1 percent. Suppose a new simple random sample of size 100 was to be taken in the future. Let p be the proportion of drunk drivers in this new sample.

What is the probability that p would be greater than or equal to 0.03? For this question, assume that $\sqrt{\frac{0.99 \cdot 0.01}{100}} = 0.01$. *(4 pts)*

Answers

1) Approximately 90% is a perfectly fine answer.

2) Yes. Distance here is symmetric: if a number A is X away from another number B, then B is also X away from A. i.e. $|x - y| = |y - x|$.

3) No. If we pick another random sample, we will get a different sample proportion. The sample mean is an estimator for the population mean. However, before we can draw conclusions, we need to quantify how good of an estimate this sample mean is. i.e. the variability of the estimator.

4) Using the 68-95 heuristic, you will get 0.025.