

MAT 100: Exam 3
Fall – 2022
12/14/2022
85 Minutes

Name: Caleb McWhorter — Solutions

Write your name on the appropriate line on the exam cover sheet. This exam contains 15 pages (including this cover page) and 14 questions. Check that you have every page of the exam. Answer the questions in the spaces provided on the question sheets. Be sure to answer every part of each question and show all your work.

Question	Points	Score
1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
9	10	
10	10	
11	10	
12	10	
13	10	
14	10	
Total:	140	

1. (10 points) Mark the following statements as True (T) or False (F):

- (a) T : A notion of center is robust if it is resistant to outliers.
- (b) F : The mean of a dataset is always larger than its median.
- (c) T : The standard deviation of a dataset measures its ‘spread.’
- (d) F : The IQR measures the spread about the mean.
- (e) F : The z -score of a value measures how far it is from the mean.
- (f) T : The larger the magnitude of a z -score, the more ‘unusual’ the corresponding value.
- (g) F : For any dataset, P_{50} will always be the median.
- (h) T : The larger the value of z_{x_0} , the larger $P(X < x_0)$ will be.
- (i) T : By the Central Limit Theorem, if one takes a simple random sample from a distribution with mean μ and standard deviation σ , then the distribution of means of groups, \bar{X} , of size n is given by $N(\mu, \sigma/\sqrt{n})$.
- (j) T : For fixed p , if n is ‘large’, $B(n, p) \approx N(np, \sqrt{np(1-p)})$.

2. (10 points) Being as detailed and accurate as possible, answer the following:

(a) Define what it means for a measurement of center to be robust.

A measurement of center is robust if it is resistant to outliers.

(b) Given a value x from a normal distribution, what does z_x measure?

Because $z_x = \frac{x-\mu}{\sigma}$, z_x measures the number of standard deviations x is below or above the mean.

(c) For a finite collection of numbers, what is the IQR of the dataset and what does it measure?

The IQR is $IQR = Q_3 - Q_1$, where Q_1 and Q_3 are the first and third quartiles. The IQR measures the interquartile range, i.e. the spread about the median.

(d) For a finite collection of numbers, what does the standard deviation measure and how is it defined?

The standard deviation is $\sigma = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$, where n is the number of data values, the x_i are the data values, and \bar{x} is the mean data value. The standard deviation measures the 'spread' about the mean.

(e) What assumption is required for events in a binomial distribution?

A count of the number of times an event occurs in a sequence of trials follows a binomial distribution if the following conditions are met:

1. The event either occurs or not.
2. The event has a fixed probability of occurring.
3. The trials are independent from each other.
4. There are a fixed number of trials.

3. (10 points) Suppose that you have the following dataset:

-2 0 1 3 8

Showing all your work, compute the mean and standard deviation for this dataset.

Solution. The mean is...

$$\bar{x} = \frac{\sum x_i}{n} = \frac{-2 + 0 + 1 + 3 + 8}{5} = \frac{10}{5} = 2$$

We then compute the terms required to find the standard deviation:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
-2	-4	16
0	-2	4
1	-1	1
3	1	1
8	6	36
		Total: 58

Therefore, the standard deviation is...

$$\sigma = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{1}{4} \cdot 58} = \sqrt{14.5} \approx 3.80789$$

4. (10 points) Suppose you have the following dataset:

4 13 24 30 6 10 14 1 8 3 28

Showing all your work, compute the 5-number summary for this dataset.

Solution. First, we place the data in order:

1 3 4 6 8 10 13 14 24 28 30

There are $n = 11$ numbers in the dataset. We know the 5-number summary consists of the min, Q_1 , median, Q_3 , and the max. We know Q_1 is the 25th percentile, the median is the 50th percentile, and Q_3 is the 75th percentile.

The p th percentile of a collection of n numbers is the (np_d) th number in the dataset—rounding upwards and if this is an exact integer, taking the average of that number and the next in the dataset, where p_d is the percentile written as a decimal. We have...

$$P_{25} = 0.25 \cdot 11 = 2.75 \rightsquigarrow \text{3rd number}$$

$$P_{50} = 0.50 \cdot 11 = 5.5 \rightsquigarrow \text{6th number}$$

$$P_{75} = 0.75 \cdot 11 = 8.25 \rightsquigarrow \text{9th number}$$

Therefore, $P_{25} = 4$, $P_{50} = 10$, and $P_{75} = 24$. Therefore, the 5-number summary is...

Min	Q_1	Median	Q_3	Max
1	4	10	24	30

This aligns with the other way we saw of computing the 5-number summary. We find the median of the dataset. There are 11 numbers, so the median is the $\frac{11}{2} = 5.5 \rightsquigarrow$ 6th value in the dataset, which is 10. To find Q_1 , we find the median of the values below this median. There are 5 such values. Therefore, this median is the $\frac{5}{2} = 2.5 \rightsquigarrow$ 3rd value in the dataset, which is 4. To find Q_3 , we find the median of the values above the median. There are 5 such values. Therefore, this median is the $\frac{5}{2} = 2.5 \rightsquigarrow$ 3rd value above the median, which is 24. This gives the same 5-number summary as the one given above.

5. (10 points) Suppose you have the following dataset:

4 13 24 30 6 10 14 1 8 3 28

Showing all your work, compute P_{20} and P_{60} .

Solution. There are $n = 11$ numbers. First, we place the dataset in order:

1 3 4 6 8 10 13 14 24 28 30

The p th percentile of a collection of n numbers is the (np_d) th number in the dataset—rounding upwards and if this is an exact integer, taking the average of that number and the next in the dataset, where p_d is the percentile written as a decimal. We have...

$$P_{20} = 0.20 \cdot 11 = 2.2 \rightsquigarrow \text{3rd number}$$

$$P_{60} = 0.60 \cdot 11 = 6.6 \rightsquigarrow \text{7th number}$$

Therefore, we have...

$$P_{20} = 4$$

$$P_{60} = 13$$

6. (10 points) Suppose you have the following dataset:

4 13 24 30 6 10 14 1 8 3 28

Give a box plot for this dataset.

Solution. A box plot is a visual representation of the 5-number summary. We then first need to find the 5-number summary—consisting of the min, Q_1 , median, Q_3 , and max for the data. Putting the data in order, we have...

1 3 4 6 8 10 13 14 24 28 30

There are $n = 11$ numbers in the dataset. We know the 5-number summary consists of the min, Q_1 , median, Q_3 , and the max. We know Q_1 is the 25th percentile, the median is the 50th percentile, and Q_3 is the 75th percentile.

The p th percentile of a collection of n numbers is the (np_d) th number in the dataset—rounding upwards and if this is an exact integer, taking the average of that number and the next in the dataset, where p_d is the percentile written as a decimal. We have...

$$P_{25} = 0.25 \cdot 11 = 2.75 \rightsquigarrow \text{3rd number}$$

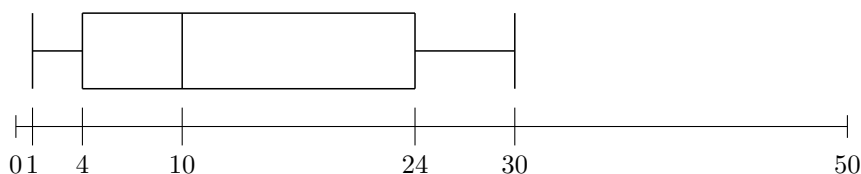
$$P_{50} = 0.50 \cdot 11 = 5.5 \rightsquigarrow \text{6th number}$$

$$P_{75} = 0.75 \cdot 11 = 8.25 \rightsquigarrow \text{9th number}$$

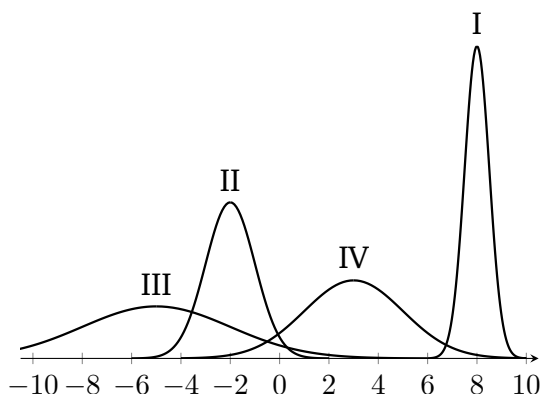
Therefore, $P_{25} = 4$, $P_{50} = 10$, and $P_{75} = 24$. Therefore, the 5-number summary is...

Min	Q_1	Median	Q_3	Max
1	4	10	24	30

This aligns with the other way we saw of computing the 5-number summary. We find the median of the dataset. There are 11 numbers, so the median is the $\frac{11}{2} = 5.5 \rightsquigarrow$ 6th value in the dataset, which is 10. To find Q_1 , we find the median of the values below this median. There are 5 such values. Therefore, this median is the $\frac{5}{2} = 2.5 \rightsquigarrow$ 3rd value in the dataset, which is 4. To find Q_3 , we find the median of the values above the median. There are 5 such values. Therefore, this median is the $\frac{5}{2} = 2.5 \rightsquigarrow$ 3rd value above the median, which is 24. This gives the same 5-number summary as the one given above. We can then construct a box plot of the data:



7. (10 points) Below are four different normal distributions, labeled I – IV:



- Arrange these normal distributions by the size of their means.
- Arrange these normal distributions by the size of their standard deviations.

Solution.

- The mean of a normal distribution, μ , is its ‘center’ and is located at the vertical line of symmetry for the distribution. Examining the plot and finding the mean of each normal distribution, we have...

I	II	III	IV
8	-2	-5	3

Placing these means in order yields...

$$\mu_{\text{III}} < \mu_{\text{II}} < \mu_{\text{IV}} < \mu_{\text{I}}$$

- The standard deviation of a normal distribution, σ , is the amount of ‘spread’ of the dataset. The standard deviation is the distance from the mean of a normal distribution, i.e. its center, to where the curve switches from ‘bending downward’ to ‘bending upward’ (either on its left or its right). Examining the plot, we can estimate the standard deviation for each of the normal distributions:

I	II	III	IV
0.5	1	3	2

Placing these standard deviations in order yields...

$$\sigma_{\text{I}} < \sigma_{\text{II}} < \sigma_{\text{IV}} < \sigma_{\text{III}}$$

8. (10 points) Dwight and Danny are salesmen at two different paper companies. Dwight earns \$20,000 a year in commissions while Danny earns \$15,200. Commissions at Dwight's company are normally distributed with mean \$15,620 and standard deviation \$2,100. Danny works at a company whose sales commissions have distribution $N(\$12350, \$1300)$. Relative to other workers at their companies, who is the better salesperson? Be sure to fully justify your answer.

Solution. Recall that the z -score can be used as a measure of 'unusualness' for a normal distribution. If $z_x < 0$ is below the mean. If $z_x = 0$, then x is equal to the mean. If $z_x > 0$, then x is greater than the mean. The larger the magnitude of z_x , i.e. the larger the value of $|z_x|$, the more 'unusual' the score. We have...

$$z_{\text{Dwight}} = \frac{x - \mu}{\sigma} = \frac{\$20000 - \$15620}{\$2100} = \frac{\$4380}{\$2100} \approx 2.09$$
$$z_{\text{Danny}} = \frac{x - \mu}{\sigma} = \frac{\$15200 - \$12350}{\$1300} = \frac{\$2850}{\$1300} \approx 2.19$$

Because both z -scores are positive, both men have above average pay at their companies—but of course this was easily observed. Because $z_{\text{Danny}} > z_{\text{Dwight}}$, Danny makes more relative to others at his company than Dwight does at his company. Using this as our basis for comparison, we declare Danny to be the better salesperson.

9. (10 points) Suppose that professional YouTube channels earn a yearly salary that is approximately normally distributed with mean \$46,293 with standard deviation \$8,527. What is the minimum amount your 'professional' YouTube channel have to earn to be in the top 6% of YouTube channels?

Solution. Let X denote this minimum necessary salary to be in the top 6% of YouTube channels. This would mean X is greater than 94% of incomes for 'professional' YouTube channels. But then we know that $z_X \rightsquigarrow 0.94$. We can examine a z -score table to find the closest z -value such that $z_X \rightsquigarrow 0.94$, which is $z_X \approx 1.555$. But then...

$$z_X = \frac{X - \mu}{\sigma}$$

$$1.555 = \frac{X - \$46,293}{\$8,527}$$

$$X - \$46,293 = \$13,259.49$$

$$X = \$59,552.49$$

Therefore, the least you could make and still be in the top 6% of 'professional' YouTube channel earnings is \$59,552.49.

10. (10 points) Suppose that you have a normal distribution with mean 140.3 and standard deviation 16.4. Showing all your work, find the following:

- (a) $P(X = 140.3)$
- (b) $P(X \leq 130)$
- (c) $P(X \geq 130)$
- (d) $P(X \leq 180)$
- (e) $P(130 \leq X \leq 180)$

Solution.

(a) Because a normal distribution is a continuous distribution, we have...

$$P(X = 140.3) = 0.$$

(b)

$$z_{130} = \frac{130 - 140.3}{16.4} = \frac{-10.3}{16.4} = -0.63 \rightsquigarrow 0.2643$$

Therefore, $P(X \leq 130) = 0.2643$.

(c) Using (b), we have...

$$P(X \geq 130) = 1 - P(X \leq 130) = 1 - 0.2643 = 0.7357$$

(d)

$$z_{180} = \frac{180 - 140.3}{16.4} = \frac{39.7}{16.4} = 2.42 \rightsquigarrow 0.9922$$

Therefore, $P(X \leq 180) = 0.9922$.

(e) Using (b) and (d), we have...

$$P(130 \leq X \leq 180) = P(X \leq 180) - P(X \leq 130) = 0.9922 - 0.2643 = 0.7279$$

11. (10 points) Suppose you have a normal distribution with mean 140.3 and standard deviation 16.4. You take a simple random sample from this distribution of size 8 and take the sample mean, \bar{X} . Showing all your work, find the following:

- (a) $P(\bar{X} = 140.3)$
- (b) $P(\bar{X} \leq 130)$
- (c) $P(\bar{X} \geq 130)$
- (d) $P(\bar{X} \leq 180)$
- (e) $P(130 \leq \bar{X} \leq 180)$

Solution. Because you have a simple random sample and are drawing from a normal distribution (so the fact that the sample size of $n = 8$ not being ‘sufficiently large’ does not matter), the Central Limit Theorem applies. Therefore, using the Central Limit Theorem, we know the distribution of sample means of sample size 8 is $N(\mu, \sigma/\sqrt{n}) = N(140.3, 16.4/\sqrt{8}) = N(140.3, 5.798)$.

- (a) Because a normal distribution is a continuous distribution, we have...

$$P(\bar{X} = 140.3) = 0.$$

- (b)

$$z_{130} = \frac{130 - 140.3}{5.798} = \frac{-10.3}{5.798} = -1.78 \rightsquigarrow 0.0375$$

Therefore, $P(\bar{X} \leq 130) = 0.0375$.

- (c) Using (b), we have...

$$P(\bar{X} \geq 130) = 1 - 0.0375 = 0.9625$$

- (d)

$$z_{180} = \frac{180 - 140.3}{5.798} = \frac{39.7}{5.798} = 6.85 \rightsquigarrow 1.$$

Therefore, $P(\bar{X} \leq 180) = 1.$

- (e) Using (b) and (d), we have...

$$P(130 \leq \bar{X} \leq 180) = P(\bar{X} \leq 180) - P(\bar{X} \leq 130) = 1. - 0.0375 = 0.9625$$

12. (10 points) Suppose that only 20% of professors at a college use TokTik. You take a simple random sample of 11 professors from this college.

- (a) Find the probability that exactly 3 of the professors use TokTik.
- (b) Find the probability that at most 3 of the professors use TokTik.
- (c) Find the probability that less than 3 of the professors use TokTik.
- (d) Find the probability that at least 8 of the professors use TokTik.
- (e) Find the probability that at least 1 of the professors use TokTik.

Solution. Each professor surveyed either uses TokTik or not. We assume that the probability that a professor uses TokTik is constant. We assume that the professors surveyed answer independently. Finally, there are a fixed number of professors surveyed. Therefore, the count X of professors surveyed that use TokTik is given by a binomial distribution $B(n, p) = B(11, 0.20)$.

(a)

$$P(X = 3) = 0.2215$$

(b)

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.0859 + 0.2362 + 0.2953 + 0.2215 \\ &= 0.8389 \end{aligned}$$

(c)

$$\begin{aligned} P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.0859 + 0.2362 + 0.2953 \\ &= 0.6174 \end{aligned}$$

(d)

$$\begin{aligned} P(X \geq 8) &= P(X = 8) + P(X = 9) + P(X = 10) + P(X = 11) \\ &= 0.0002 + 0. + 0. + 0. \\ &= 0.0002 \end{aligned}$$

(e)

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.0859 = 0.9141$$

13. (10 points) In a large lecture class, there are 263 students. From historical academic records, it is estimated that 16% of the students will fail the course.

- (a) Approximately how many students do you estimate will likely fail this course?
- (b) What is the probability that less than 30 students will fail this course?

Solution. We have a fixed sample size of $n = 263$ students. Each sampled student will either pass or fail the course. We assume the probability that they fail the course is fixed at $p = 0.16$. Furthermore, we assume the sampled students are independent. Then the number of students that will fail the exam is given by a binomial distribution, $B(n, p) = B(263, 0.16)$, which has mean $np = 263(0.16) = 42.08$ and standard deviation $\sqrt{np(1-p)} = \sqrt{263(0.16)(1-0.16)} = \sqrt{263(0.16)(0.84)} = \sqrt{35.3472} = 5.945$. However, the given binomial table does not have $n = 263$ nor $p = 0.16$.

However, we shall assume the sample was a simple random sample and $np = 263(0.16) =$ and $n(1-p) = 263(0.84) =$. Therefore, we can approximate this binomial distribution with the normal distribution $N(np, \sqrt{np(1-p)}) = N(42.08, 5.945)$.

- (a) We would estimate the ‘average’ number of students will fail the course, i.e. the mean. Either considering this as the binomial distribution $B(263, 0.16)$ or using the normal approximation $N(42.08, 5.945)$, we then estimate that approximately 42 students will fail the course.

- (b) Using the normal approximation, we have...

$$z_{30} = \frac{30 - 42.08}{5.945} = \frac{-12.0835}{5.945} = -2.03 \rightsquigarrow 0.0212$$

But then $P(X < 30) \approx 0.0212$.

14. (10 points) You are taking a history class. The professor states that the class average is approximately normally distributed with mean 72.6 with standard deviation 8.3. Suppose you take a simple random sample of 5 students from the course. What is the probability that the average of these five student's averages is less than 65?

Solution. The sample is a simple random sample of size $n = 5$. Because the underlying distribution is normal, the Central Limit Theorem applies. Therefore, the distribution of sample means of size 5 is $N(\mu, \sigma/\sqrt{n}) = N(72.6, 8.3/\sqrt{5}) = N(72.6, 3.712)$. But then...

$$z_{65} = \frac{65 - 72.6}{3.712} = \frac{-7.6}{3.712} = -2.05 \rightsquigarrow 0.0202$$

Therefore, $P(\bar{X} \leq 65) = 0.0202$.