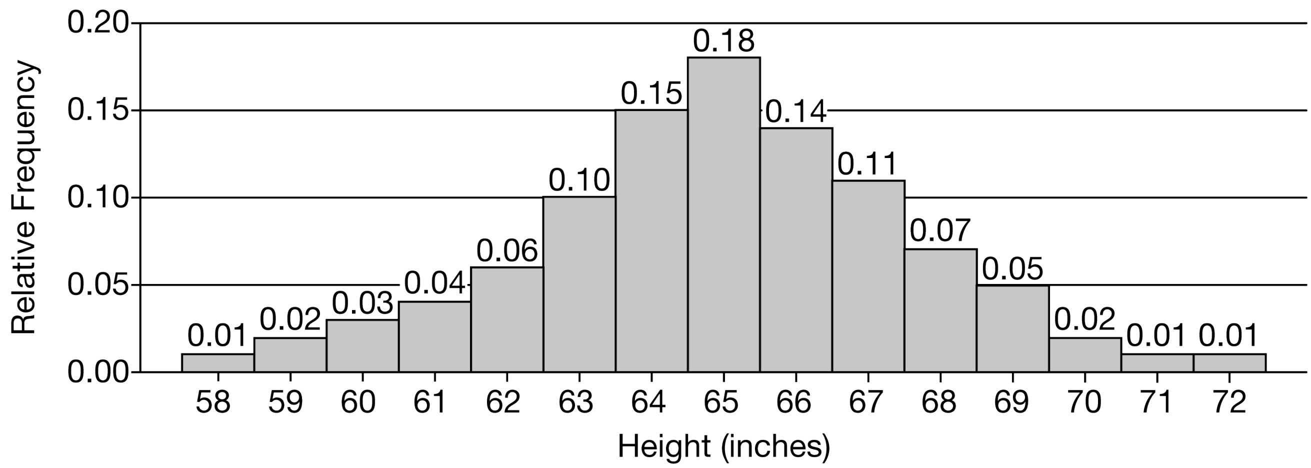
The Normal Distribution, Revisited Quiz

1. Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

The following histogram shows the relative frequencies of the heights, recorded to the nearest inch, of a population of women. The mean of the population is 64.97 inches, and the standard deviation is 2.66 inches.



One woman from the population will be selected at random.

(a) Based on the histogram, what is the probability that the selected woman will have a height of at least 67 inches? Show your work.

(b) What is the area of the bar corresponding to a height of 67 inches in the graph, and what does the area represent in terms of probability?

(c) The histogram displays a discrete probability model for height. However, height is often considered a continuous variable that follows a normal model. Consider a normal model that uses the mean and standard deviation of the population of women as its parameters.

(i) Use the normal model and the relationship between area and relative frequency to find the probability that the randomly selected woman will have a height of at least 67 inches. Show your work.

(ii) Does your answer in part (c-i) match your answer in part (a) ? If not, give a reason for why the answers might be different.

(d) Let the random variable H represent the height of a woman in the population. $P(H < 60)$ represents the probability of randomly selecting a woman with height less than 60 inches. Based on the information given, the probability can be found using either the discrete model or the normal model.

(i) Give an example of a probability of H that can be found using the discrete model but not the normal model. Explain why.

(ii) Give an example of a probability of H that can be found using the normal model but not the discrete model.

The Normal Distribution, Revisited Quiz

Explain why.

Part A, B, C, and D

The primary goals of this question are to assess a student's ability to (1) calculate a probability from a discrete model; (2) recognize that area is related to probability; (3) extend the discrete model to a continuous model using the normal curve; and (4) determine when a discrete model and continuous model are appropriate for finding a certain probability.

Scoring

Parts (a), (b), (c), and (d) are scored as essentially correct (E), partially correct (P), or incorrect (I).

Each essentially correct (E) part counts as 1 point.

Each partially correct (P) part counts as 1/2 point.

If a response is between two scores (for example, 2 1/2 points), use a holistic approach to decide whether to score up or down, depending on the overall strength of the response and communication.

Reasons to score up:

· All notation is correct and clearly marked

· All explanations are clear

· No wrong information is included that was not part of the scoring (for example, saying sample size must be greater than 30 when that has nothing to do with the problem)

· No minor calculation errors are made, if they are not part of the scoring

· Interpretation parts are especially strong

Reasons to score down:

· Notation is not wrong, but is spotty and not clearly marked

· Explanations are not wrong, but are hard to follow

· Wrong or extraneous information is included but not part of scoring

· Minor calculation errors that are not part of the scoring are made

Interpretation parts are scored an E but are considered a weak E

0	1	2	3	4 ✓

Parts (a) through (d) sum to 4 points

The Normal Distribution, Revisited Quiz

OR

Parts (a) through (d) sum to $3^1/_2$ points AND a holistic approach is used to decide to score up

- ☐ Part (a) essentially correct
- ☐ Part (a) partially correct
- ☐ Part (a) incorrect
- ☐ Part (b) essentially correct
- ☐ Part (b) partially correct
- ☐ Part (b) incorrect
- ☐ Part (c) essentially correct
- ☐ Part (c) partially correct
- ☐ Part (c) incorrect
- ☐ Part (d) essentially correct
- ☐ Part (d) partially correct
- ☐ Part (d) incorrect

Solution

Part (a):

The probability is equal to the sum of the relative frequencies of the heights from 67 to 72.

$0.11 + 0.07 + 0.05 + 0.02 + 0.01 + 0.01 = 0.27.$

Scoring

Part (a) is scored as follows:

Essentially correct (E) if the response correctly sums the relative frequencies of the values from 67 to 72 and shows the work.

Partially correct (P) if the response provides the correct probability with no work;

OR

if the response provides the answer of 0.11 (equals 67) or 0.16 (greater than 67) with work.

Incorrect (I) if the response does not satisfy the criteria for E or P.

Solution

Part (b):

The Normal Distribution, Revisited Quiz

The bar corresponding to a height of 67 inches is a rectangle with base 1 (67.5 minus 66.5) and height 0.11. So the area is 0.11, which is also equal to the probability that a randomly selected woman has a height of 67 inches.

Scoring

Part (b) is scored as follows:

Essentially correct (E) if the response satisfies the following two components.

Indicates the bar is a rectangle with base 1, height 0.11, and area 0.11.

Indicates the area of 0.11 is equal to the probability that a randomly selected woman has a height of 63 inches.

Partially correct (P) if the response satisfies only one of the two components.

Incorrect (I) if the response does not satisfy the criteria for E or P.

Solution

Part (c):

(i) For a normal model with mean 64.97 inches and standard deviation 2.66 inches, the *z*-score for 67 inches is $z = \frac{67-64.97}{2.66} \approx 0.76.$ Using area as probability, the area under the standard normal curve to the right of 0.76 is approximately 0.2236.

(ii) The probability of 0.2236 is a little less than the answer in part (a) of 0.27 because there is less area in the tail of the normal curve than in the bars of the histogram.

Scoring

Part (c) is scored as follows:

Essentially correct (E) if the response satisfies the following four components.

· Finds the correct *z* Alt text: z-score in part (c-i).

· Finds the correct area in part (c-i).

· Recognizes the two probabilities are slightly different in part (c-ii).

· Provides a reasonable explanation of why they might be different.

Partially correct (P) if the response satisfies only two or three of the four components.

Incorrect (I) if the response does not satisfy the criteria for E or P.

Solution

Part (d):

(i) The probability that the selected woman has a height equal to 63 inches, $P(H = 63)$, is an example of a probability that can be found with the discrete model as the area (or height) of the bar that corresponds to 63 inches. However, there is no area above a single number on the normal model. (ii) The probability that the selected woman has a height greater than

The Normal Distribution, Revisited Quiz

63.5 inches, $P(H > 63.5)$, is an example of a probability that can be found with the normal model by standardizing 63.5. However, it is not clear on the discrete model what the relative frequency of fractional heights might be.

Scoring

Part (d) is scored as follows:

Essentially correct (E) if the response satisfies the following four components.

· Provides an example of a probability that can be found using the discrete model in part (d-i).

· Provides a reasonable explanation as to why the discrete model is appropriate AND why the normal model is not appropriate in part (d-i).

· Provides an example of a probability that can be found using the normal model in part (d-ii).

· Provides a reasonable explanation as to why the normal model is appropriate AND why the discrete model is not appropriate in part (d-ii).

Partially correct (P) if the response satisfies two or three of the four components.

Incorrect (I) if the response does not satisfy the criteria for E or P.

2. The mean and standard deviation of the sample data collected on continuous variable x are -0.25 and 0.03, respectively. The following table shows the relative frequencies of the data in the given intervals.

Interval	**Relative Frequency**
$-0.34 \leq x < -0.31$	0.02
$-0.31 \leq x < -0.28$	0.15
$-0.28 \leq x < -0.25$	0.33
$-0.25 \leq x < -0.22$	0.36
$-0.22 \leq x < -0.19$	0.11
$-0.19 \leq x < -0.16$	0.03

Based on the table, do the data support the use of a normal model to approximate population characteristics?

(A) Yes, because the sum of the relative frequencies is 1.00.

(B) Yes, because the distribution of relative frequencies is very close to the empirical rule for normal models.

(C) No, because the values are negative and normal models are used for positive values.

(D) No, because the distribution of relative frequencies is very far from the empirical rule for normal models.

(E) No, because the sample size and the population parameters are not known.

The Normal Distribution, Revisited Quiz

Answer B

Correct. Based on the table, 69% of the data falls within 1 standard deviation of the mean, 95% falls within 2 standard deviations of the mean, and 100% falls within 3 standard deviations of the mean. This is very close to the empirical rule of 68% of the data within 1 standard deviation of the mean, 95% within 2 standard deviations of the mean, and 99.7% within 3 standard deviations of the mean.

3. The continuous random variable N has a normal distribution with mean 7.5 and standard deviation 2.5. For which of the following is the probability equal to 0 ?

(A) $P(N = 8)$ ✓

(B) $P(N > 8)$

(C) $P(N < 8)$

(D) $P(7 < N < 8)$

(E) $P(N < 7)$ or $P(N > 8)$

Answer A

Correct. The normal model requires area to find probability. At the point 8, there is no interval and no area under the curve. No area technically means a probability of 0.

4. Data will be collected on the following variables. Which variable is most likely to be approximated by a normal model?

(A) The distribution of the number of books read last week by middle school students, where the right tail of the distribution is longer than the left

(B) The distribution of life span, in minutes, for batteries of a certain size, where most life spans cluster around the center of the distribution but with some very low and some very high life spans ✓

(C) The distribution of ages, in years, of the students at a certain college, where most students are between 18 and 22 years old, but ages greater than 22 will probably be more spread out than ages less than 18

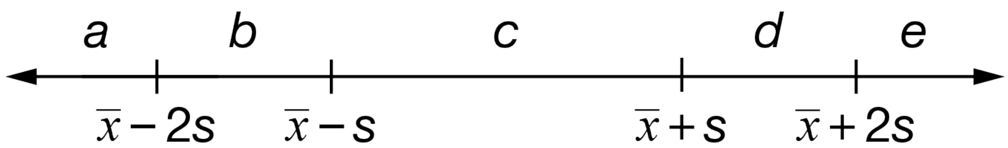
(D) The distribution of the number of birthdays per month for the employees at a certain company, where the number of birthdays in each month is approximately equal

(E) The distribution of the length of a stay, in days, in a hospital after surgery, where many patients have very short hospital stays, but some stays are quite lengthy and considered high outliers

Answer B

Correct. Battery life span is probably approximately normal, with most life spans clustering around the mean at the center but with some very low and some very high life spans.

5. Clara recorded 50 numerical observations on a certain variable and then calculated the mean $\bar{x}$ and the standard deviation s for the observations. To help decide whether a normal model is appropriate, she created the following chart.



In Clara's chart, the letters $a, b, c, d,$ and e represent the number of observations falling in each interval. Which of the following list of counts for $a, b, c, d,$ and e, respectively, is the best indicator that the variable can be modeled with a normal approximation?

(A) 1, 7, 34, 7, 1 ✓

(B) 1, 10, 28, 10, 1

(C) 2, 8, 30, 8, 2

(D) 2, 4, 38, 4, 2

(E) 5, 5, 30, 5, 5

Answer A

Correct. For a normal approximation, about 68% of the 50 observations, or 34 observations, should fall within 1 standard deviation of the mean, as shown in this list. This list has 48 of the 50 observations (96%) falling within 2 standard deviations of the mean, which is close to what would be observed in a normal model (95%).

6. The distribution of time needed to complete a certain programming task is approximately normal, with mean 47 minutes and standard deviation 6 minutes. Which of the following is closest to the probability that a randomly chosen task will take less than 34 minutes or more than 60 minutes to complete?

(A) 0.0151

(B) 0.0303 ✓

(C) 0.4849

(D) 0.9697

(E) 0.9849

The Normal Distribution, Revisited Quiz

Answer B

Correct. If x represents "the time needed to complete the task," then $P(x < 34) + P(x > 60) = P\left(\frac{x-\mu}{\sigma} < \frac{34-47}{6}\right) + P\left(\frac{x-\mu}{\sigma} > \frac{60-47}{6}\right) = P\left(z < -\frac{13}{6}\right) + P\left(z > \frac{13}{6}\right) \approx$
.

7. A machine is used to fill bags with a popular brand of trail mix. The machine is calibrated so the distribution of the weights of the bags of trail mix is normal, with mean 240 grams and standard deviation 3 grams. Of the following, which is the least weight of a bag in the top 5 percent of the distribution?

(A) 234 grams

(B) 240 grams

(C) 243 grams

(D) 246 grams

(E) 248 grams

Answer D

Correct. The upper 5% of the area under the standard normal curve is to the right of $z \approx 1.65$ (because $P(z < 1.65) \approx 0.95$). The weight of a bag with that z-score is $3(1.65) + 240 = 244.95$. Of the choices shown, 246 is the least value that is greater than 244.95.