AP* Statistics Review

Normal Models and Sampling Distributions

Teacher Packet

Characteristics of a normal model

- The shape is unimodal, symmetric, and mound-shaped.
- The mean is equal to the median.
- A normal model is continuous, although it is often used to approximate a discrete distribution like a histogram.
- The shape, center, and spread of a normal model can be quickly given by writing $N(\mu, \sigma)$.

The 68%-95%-99.7% (Empirical) Rule

- About 68% of the area is in the interval $(\mu - \sigma, \mu + \sigma)$, or within 1 standard deviation of the mean
- About 95% of the area is in the interval $(\mu - 2\sigma, \mu + 2\sigma)$, or within 2 standard deviations of the mean
- About 99.7% of the area is in the interval $(\mu - 3\sigma, \mu + 3\sigma)$, or within 3 standard deviations of the mean

What's a z score?

- A z score tells how many standard deviations above or below the mean a data point is. A z score of +1 means the point is one standard deviation above the mean. A z score of -2 means the point is two standard deviations below the mean.
- Calculate a z score using this formula: $z = \dfrac{x - \mu}{\sigma}$, where x is the specific value of interest in the distribution.
- Z scores make it possible to compare quantities measured on different scales, such as SAT scores and ACT scores.

Finding probabilities using a normal model (forwards!)

- Strategy: $x \rightarrow z \rightarrow P$.
- Sketch a normal model, draw a vertical line at the x value, and shade the area of interest.
- Take the x value and find the z score using the formula.
- Then look up the z score on the normal table to find the probability below (to the left of) the line. (See the last page for calculator usage)
- If you want to find the area to the right of the line, subtract the P value off the table from 1.
- If you want to find the area between two lines, find the z scores for each x value, look up P values on the table for each z score, and subtract the P values.

Finding an x value given a normal probability (backwards!)

- Strategy is $P \rightarrow z \rightarrow x$.
- Sketch a normal model, draw a vertical line where you think the x value will be, and label the area/probability you have been given.
- Remember, the table lists the area to the left of the line. If you have been given an area to the right, subtract from 1 to get the area on the left.
- Find the P value inside the chart and move outwards to read the z score.
- Plug the z-score as well as the mean and standard deviation into the formula and solve for x.

Sampling Distributions and The Central Limit Theorem

Consider taking many (theoretically, all possible) samples of size n from a population. Take the average $\bar{x}$ of each sample. All of these sample means make up the sampling distribution, which can be graphed as a histogram.

- The mean of the sampling distribution is the same as the mean of the population: $\mu_{\bar{x}} = \mu$

- The standard deviation of the sampling distribution gets smaller according to this equation: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

- The Central Limit Theorem states that as the sample size n increases, the sampling distribution becomes more normal (regardless of the shape of the population). In practice, if $n \geq 30$, we assume the distribution is approximately normal.

When do you use the Central Limit Theorem?

Use the Central Limit Theorem when a question asks you to calculate a probability about an average or mean.

Example:
The amount of dirt loaded into a dump truck varies normally with a mean of 750 pounds and a standard deviation of 40 pounds. A) Find the probability that a randomly chosen truck would hold over 800 pounds of dirt. B) Find the probability that the average weight in a random sample of 5 trucks would be over 800 pounds.

Solution:

A) x = weight of dirt in a dump truck

x follows a $N(750,40)$ model

$z = \dfrac{800-750}{40} = 1.25$, $P(x > 800)) = P(z > 1.25) = 1 - .8944 = .1056$

Nearly 11% of the trucks (10.56%) would hold over 800 pounds.

B) $\bar{x}$ = average weight of dirt in a sample of 5 dump trucks

$\bar{x}$ follows a $N\left(750, \dfrac{40}{\sqrt{5}} = 17.89\right)$ model

$z = \dfrac{800-750}{17.89} = 2.80$

$P(\bar{x} > 800)) = P(z > 2.80) = 1 - .9974 = .0026$

Less than one percent (0.26%) of groups of 5 dump trucks would have an average weight over 800 pounds.

Multiple Choice Questions on Normal Models and Sampling Distributions

1. If heights of 3rd graders follow a normal distribution with a mean of 52 inches and a standard deviation of 2.5 inches, what is the z score of a 3rd grader who is 47 inches tall?

    (A)   -5
    (B)   -2
    (C)   2
    (D)   5
    (E)   26.2

2.   Suppose that a normal model describes the acidity (pH) of rainwater, and that water tested after last week's storm had a z-score of 1.8. This means that the acidity of the rain

    (A)   had a pH 1.8 higher than average rainfall.
    (B)   had a pH of 1.8.
    (C)   varied with standard deviation 1.8
    (D)   had a pH 1.8 standard deviations higher than that of average rainwater.
    (E)   had a pH 1.8 times that of average rainwater.

3. In a factory, the weight of the concrete poured into a mold by a machine follows a normal distribution with a mean of 1150 pounds and a standard deviation of 22 pounds. Approximately 95% of molds filled by this machine will hold weights in what interval?

    (A)   1084 to 1216 pounds
    (B)   1106 to 1150 pounds
    (C)   1106 to 1194 pounds
    (D)   1128 to 1172 pounds
    (E)   1150 to 1194 pounds

4. Which of the following are true?

    I.   In a normal distribution, the mean is always equal to the median.
    II.   All unimodal and symmetric distributions are normal for some value of μ and σ.
    III.   In a normal distribution, nearly all of the data is within 3 standard deviations of the mean, no matter the mean and standard deviation.

    (A)   I only
    (B)   II only
    (C)   III only
    (D)   I and III only
    (E)   I, II, and III

5.   The height of male Labrador Retrievers is normally distributed with a mean of 23.5 inches and a standard deviation of 0.8 inches. (The height of a dog is measured from his shoulder.)   Labradors must fall under a height limit in order to participate in certain dog shows. If the maximum height is 24.5 inches for male labs, what percentage of male labs are <u>not</u> eligible?

    (A)   0.1056
    (B)   0.1250
    (C)   0.8750
    (D)   0.8944
    (E)   0.9750

6.   The heights of mature pecan trees are approximately normally distributed with a mean of 42 feet and a standard deviation of 7.5 feet. What proportion of pecan trees are between 43 and 46 feet tall?

    (A)   0.1501
    (B)   0.2969
    (C)   0.4470
    (D)   0.5530
    (E)   0.7031

7.   Heights of fourth graders are normally distributed with a mean of 52 inches and a standard deviation of 3.5 inches. Ten percent of fourth graders should have a height below what number?

    (A)   -1.28 inches
    (B)   45.0 inches
    (C)   47.5 inches
    (D)   48.9 inches
    (E)   56.5 inches

8. A large college class is graded on a total points system. The total points earned in a semester by the students in the class vary normally with a mean of 675 and a standard deviation of 50. Another large class in a different department is graded on a 0 to 100 scale. The final grades in that class follow a normal model with a mean of 82 and a standard deviation of 6. Jessica earns 729 points in the first class, while Ana scores 90 in the second class. Which student did better and why?

  (A) Jessica did better because her score is 54 points above the mean while Ana's is only 8 points above the mean

  (B) The students did equally well because both scored above the mean.

  (C) Ana did better because her score is 1.33 standard deviations above the mean while Jessica's is only 1.08 standard deviations above the mean.

  (D) Neither student did better; they cannot be compared because their classes have different scoring systems.

9. The distance Jonathan can throw a shot put is skewed to the right with a mean of 14.2 meters and a standard deviation of 3.5 meters. Over the course of a month, Jonathan makes 75 throws during practice. Assume these throws can be considered a random sample of Jonathan's shot put throws. What is the probability that Jonathan's average shot put distance for the month will be over 15.0 meters?

  (A) 0.0239

  (B) 0.4096

  (C) 0.5224

  (D) 0.5904

  (E) 0.9761

10. Heights of fourth graders are normally distributed with a mean of 52 inches and a standard deviation of 3.5 inches. For a research project, you plan to measure a simple random sample of 30 fourth graders. For samples such as yours, 10% of the samples should have an average height below what number?

  (A) 47.52 inches

  (B) 51.18 inches

  (C) 51.85 inches

  (D) 52.82 inches

  (E) 56.48 inches

Free Response Questions on Normal Models and Sampling Distributions

Free Response 1.

A machine is used to fill soda bottles in a factory. The bottles are labeled as containing 2.0 liters, but extra room at the top of the bottle allows for a maximum of 2.25 liters of soda before the bottle overflows.   The standard deviation of the amount of soda put into the bottles by the machine is known to be 0.15 liter.

(a)   Overfilling the bottles causes a mess on the assembly line, but consumers will complain if bottles contain less than 2 liters.   If the machine is set to fill the bottles with an average of 2.08 liters, what proportion of bottles will be overfilled?

(b)   If management requires that no more than 3% of bottles should be overfilled, the machine should be set to fill the bottles with what mean amount?

(c)   Complaints from consumers about underfilled bottles leads the company to set the mean amount to 2.15 liters. In this situation, what standard deviation would allow for no more than 3% of the bottles to be overfilled?

Free Response 2.
The distribution of scores for persons over 16 years of age on a common IQ test is approximately normal with mean 100 and standard deviation 15.

a) What is the probability that a randomly chosen adult has an IQ score on this test over 105?

b) What are the mean and standard deviation of the average IQ score on this test for an SRS of 60 people?

c) What is the probability that the average IQ score on this test of an SRS of 60 people is 105 or higher?

d) Would your *method* of answering (a) or (c) be affected if the distribution of IQ scores on this test in the adult population were distinctly nonnormal (for example, if they were skewed)? Explain which parts could be answered the same way, which could not, and how you know.

Key to Normal Models and Sampling Distributions Multiple Choice

1. B   Distractors include switching the order of x and μ, and forgetting to divide by the standard deviation.

2. D   z score is number of standard deviations from mean

3. C   Distractors include adding/subtracting 1 or 3 standard deviations instead of 2.

4. D   Unimodal and symmetric doesn't necessarily mean normal.

5. A   Distractor is finding eligible labs *under* the height limit instead of those who are not eligible.

6. A   Distractors are finding proportion of trees under or over 43 or 46 feet instead of those between 43 and 46 feet.

7. C   z = -1.28; distractor is solving z score formula incorrectly

8. C   z score (or number of standard deviations above the mean) is relevant; number of points above the mean is irrelevant

9. A $\bar{x}$ follows $N\left(14.2, \dfrac{3.5}{\sqrt{75}} = .404\right)$ model with z score of 1.98; distractors include not modifying the standard deviation, and finding the probability below 15 m instead of above 15 m

10. B   z = -1.28; $\bar{x}$ follows $N\left(52, \dfrac{3.5}{\sqrt{30}} = .639\right)$ model; distractors include not modifying standard deviation, incorrectly solving z score formula for x

Rubrics for Normal Models and Sampling Distributions Free Response

1. **Solution**

Part (a):

Let x = amount of soda put into the bottle

x follows $N(2.08, 0.15)$ model

$$z = \frac{2.25 - 2.08}{0.15} = 1.13$$

$$P(x > 2.25) = P(z > 1.13) = 1 - 0.871 = 0.129$$

12.9% of bottles would be overfilled.

Part (b):

x follows $N(\mu, 0.15)$.

A z score of 1.88 separates the normal model into 97% below the line and 3% above the line: $P(z > 1.88) = 0.0300$

We require that only 3% have over 2.25 liters of soda.

$$z = \frac{x - \mu}{\sigma} \qquad 1.88 = \frac{2.25 - \mu}{0.15} \qquad \mu = 2.25 - (1.88)(0.15) = 1.97 \text{ liters}$$

Part (c):

x follows $N(2.15, \sigma)$.

A z score of 1.88 separates the normal model into 97% below the line and 3% above the line: $P(z > 1.88) = 0.0300$

We require that only 3% have over 2.25 liters of soda.

$$z = \frac{x - \mu}{\sigma} \qquad 1.88 = \frac{2.25 - 2.15}{\sigma} \qquad \sigma = \frac{2.25 - 2.15}{1.88} = 0.053 \text{ liters}$$

Scoring

Part (a) can be essentially correct (E) or incorrect (I). Parts (b) and (c) can be essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct if the proportion is correct (except for minor arithmetic errors) and supporting work (not including calculator input) is shown.

Part (a) is incorrect if the proportion is calculated incorrectly or the correct proportion is given with no supporting work.

Part (b) is essentially correct if the mean is correct (except for minor arithmetic errors) and supporting work (not including calculator input) is shown.

Part (b) is partially correct if the z score is incorrect but the method for calculating the mean was correct based on that z score and work is shown
OR
the z score is correct but the mean is incorrectly calculated.
Part (b) is incorrect if the z score is wrong and the mean is wrong
OR
The z score is wrong and the mean is correct but no supporting work is given.

Part (c) is essentially correct if the standard deviation is correct (except for minor arithmetic errors) and supporting work (not including calculator input) is shown.
OR
if the same incorrect z score is used from part (b) and the standard deviation is calculated correctly based on that z score with work shown (a student will not be penalized twice for having the wrong z score).
Part (c) is partially correct if the z score is correct but the standard deviation is incorrectly calculated.
Part (c) is incorrect if the standard deviation is correct but no supporting work is given.

4   Complete Response

All parts essentially correct.

3   Substantial Response

Two parts essentially correct and one part partially correct

2   Developing Response

Two parts essentially correct and no parts partially correct
OR
One part essentially correct and two parts partially correct

1   Minimal Response

One part essentially correct and either zero or one part partially correct
OR
No parts essentially correct and two parts partially correct

2. **Solution**   **Part (a):**
   Let x = score on this IQ test
   x follows $N(100, 15)$ model
   $$z = \frac{105 - 100}{15} = 0.33$$
   $$P(x > 105) = P(z > 0.33) = 1 - .6293 = 0.3707$$

Part (b):
   The average of the sample means is equal to the average of the population, so
$\mu_{\bar{x}} = 100$.

   The standard deviation of the groups of size 60 is $\sigma_{\bar{x}} = \dfrac{15}{\sqrt{60}} = 1.94$.

Part (c):
   $\bar{x}$ follows $N(100, 1.94)$ model
   $$z = \frac{105 - 100}{1.94} = 2.58$$
   $$P(\bar{x} > 105) = P(z > 2.58) = 1 - .9951 = 0.0049$$

Part (d):
   If the population was not normal, then we could not use z scores and a normal model to calculate the probability in part (a). However, since the sample size is large ($n = 60$), the Central Limit Theorem states that the sampling distribution will be approximately normal, so we could do part (c) the same way.

Note: To receive complete credit for part (a) or part (c), a student must show how the probability is computed. Since part (a) and part (c) involve different normal distributions, it is important to identify which normal distribution is used in which part. This can be done by showing the z score calculation, displaying the mean and standard deviation within the probability statements, or by listing the mean and standard deviation and displaying an appropriate graph.

Scoring
   Parts (a) and (c) can be either essentially correct (E) or incorrect (I). Parts (b) and (d) can be essentially correct (E), partially correct (P), or incorrect (I).

   Part (a) is essentially correct if the probability is calculated correctly (except for minor arithmetic errors) and work is shown. Calculator input does not count as work.
   Part (a) is incorrect if no work is shown.

   Part (b) is correct if both the mean and standard deviation are computed correctly and work is shown for the standard deviation calculation.

        Part (b) is partially correct if only one of the mean or standard deviation is computed correctly (with work shown for the standard deviation).

     Part (c) is essentially correct if the probability is calculated correctly using the mean and standard deviation from part (b) with work shown.

     Part (c) is incorrect if no work is shown.

     Part (d) is essentially correct if the student states that part (a) could NOT be done but part (c) could be done AND justification is given. Justification can be either stating that the sample size is large enough or referring to the Central Limit Theorem.

     Part (d) is partially correct if the student correctly answers for either part (a) or part (c) with justification

OR

     the student answers both part (a) and part (c) correctly with weak justification.

     Part (d) is incorrect if the answers are incorrect

OR

     the answers are correct but no justification is given.

Each essentially correct response is worth 1 point; each partially correct response is worth half a point.

4   Complete Response

3   Substantial Response

2   Developing Response

1   Minimal Response

If a response is between two scores (for example, 2.5 points), use a holistic approach to determine whether to score up or down depending on the strength of the response and communication.

(Optional) Using a TI-84 to find normal probabilities

Note: cdf stands for cumulative distribution function. This adds up probabilities for a range of outcomes. pdf stands for probability distribution function. Students do not generally use Normalpdf.

Finding the area under a normal curve

- First, draw and label a graph on paper, shading the area you are looking for and labeling the graph with the information you know.
- Go to DISTR/normalcdf.
- If you know the z-scores for the area you want, in the parentheses put the lower z limit and then the upper z limit separated by commas. Use –1E99 for negative infinity and 1E99 for positive infinity. After you hit ENTER, the calculator will tell you the area under the curve (probability) between your 2 z values.

$$P(z_{low} < z < z_{high}) = \text{normalcdf}(z_{low}, z_{high})$$

- If you do not know the z scores, you may let the calculator find them for you. In the parentheses, put the lower x limit, the upper x limit, the mean of the normal distribution, and the standard deviation of the normal distribution. The latter two values are used to convert your x values into z values.

$$P(x_{low} < x < x_{high}) = \text{normalcdf}(x_{low}, x_{high}, \mu, \sigma)$$

Finding a z-score corresponding to a given area under the normal curve

- First, draw and label a graph on paper, shading the area you are looking for and labeling the graph with the information you know.
- Go to DISTR/invNorm.
- You must input the proportion (area) to the left of the z-score you are looking for. Let's call this A. If you are given the area to the right, subtract that from 1 to get A. Put this number in the parentheses, and the calculator will give you the z-score that separates the distribution into two parts: one with area A to the left of that z-score and one with area 1-A to the right of that z-score.

$$z = \text{InvNorm}(A)$$

- If you want to find the x value instead of the z score, you need to tell the calculator the mean and standard deviation, too.

$$x = \text{InvNorm}(A, \mu, \sigma)$$

AP Statistics Exam Connections

The list below identifies free response questions that have been previously asked on the topic of Normal Models and Sampling Distributions. These questions are available from the CollegeBoard and can be downloaded free of charge from AP Central.

http://apcentral.collegeboard.com.

Free Response Questions	
2003	Question 3
2004 Form B	Question 3
2006 Form B	Question 3