2.2C Activity "Solving Problems with the Normal Distribution" - updated aug2020

1. **Normal vs Not Stated**
   - If the **problem states the data is normal or approximately normal**, it is acceptable to calculate z-scores and probabilities.
   - If the data is given and the **problem does not state if the data is normal,** then the data should be graphed (dot plot, stem plot, histogram, boxplot, normal probability plot) and the graphs should be analyzed for the overall pattern (shape, center, spread).

2. **Required work**
   - **State** the distribution, the mean, and standard deviations
   - **Sketch** the normal curve for the data given (mark the mean)
   - **Shade** appropriate area and label data values on the normal curve.
   - **Calculate z-scores.** This is the ground work to help you assess the reasonableness of your answer. ***Mentally, estimate answer using "68-95-99.7 Rule."***

   $$\textbf{Formula for calculating z} - \textbf{scores} \quad z = \frac{x-\mu}{\sigma}$$

   - **Provide** the appropriate probability statement using the z-scores
   - **Answer in CONTEXT.** Your final answer should state the calculated probability in the context of the problem.

3. **TI-84 Calculator commands:**

   a. To calculate the area under the normal curve:
   - Select 2nd Distr, **2:Normalcdf**
   - then enter: (start value, end value, mean, std dev).

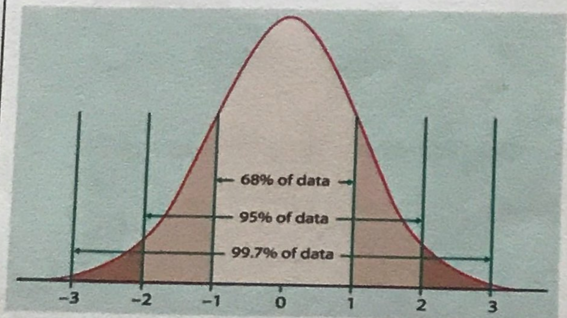
   b. To calculate the zscore (standardized value) from the percentile:
   - Select 2nd Distr, **3:invNorm**
   - then enter the (percentile, mean, std dev).

Z-scores (Normal Probability) Mechanics

Example: In the United States the average number of calories consumed per day by adolescent boys is approximately normal with a mean of 3,000 and a standard deviation of 500. I fabricated this data for this problem set.

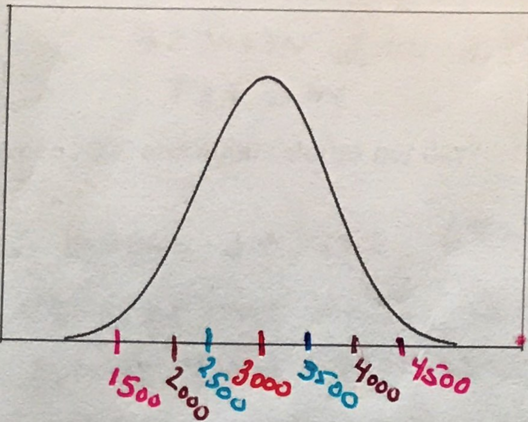
Draw and label a normal curve with this information:

State distribution: N (0, 1)



EMPIRICAL RULE

N (3,000, 500)



See page #1 for required work:

1. What percentage of these adolescent boys consumes less than 4200 calories per day?
2. skip

N(3,000, 500)

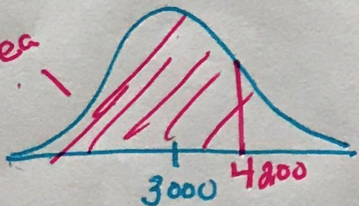
$$Z = \frac{4200 - 3000}{500}$$

$$Z = 2.4$$

$$P(Z \leq 2.4) = .9918$$

Think? What is a reasonable answer? $.5 + \frac{.95}{2} = .975$ ✓

area



∴ ABOUT 99% OF THE BOYS CONSUME LESS THAN 4,200 CALORIES/DAY

3. skip
4. What percentage of these adolescent boys consumes more than 2000 calories per day?

N(3,000, 500)

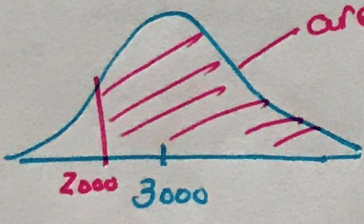
$$Z = \frac{2000 - 3000}{500}$$

$$Z = -2$$

$$P(Z \geq -2) = .9772$$

Think! what is a reasonable answer? +/- 2SD → 95% → .975

area



∴ About 98% OF THE BOYS CONSUME MORE THAN 2,000 CAL/DAY.

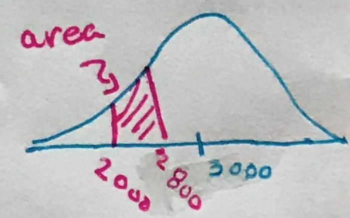
Source: Karen Abbey

PG3

5. skip

6. What percentage of these boys consumes between 2000 and 2800 calories per day?

N(3,000,500)



$$Z=\frac{2000-3000}{500} \qquad Z=\frac{2800-3000}{500}$$

$$Z=-2 \qquad Z=-0.4$$

$$P(-2 \le Z \le -0.4) = 0.3218$$

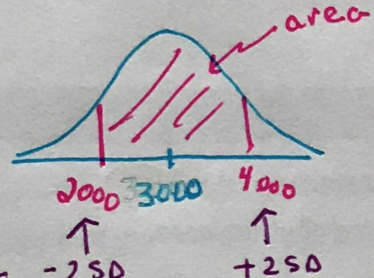
Check with
normalcdf (2,000, 2800, 3000, 500)
0.3218 ✓

∴ ABOUT 32% OF BOYS CONSUME BETWEEN 2,000–2,800 CALORIES PER DAY.

7. What percentage of these boys consumes between 2000 and 4000 calories per day?

8. skip

N(3,000,500)

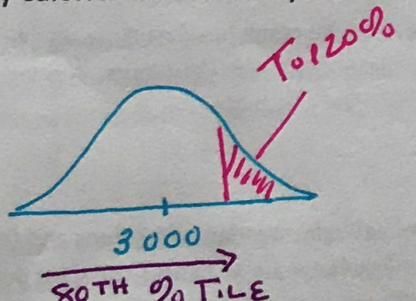


$$P(-2 \le Z \le 2) = .95$$

∴ BASED ON THE 68-95-99.7 RULE, 95% OF THE BOYS CONSUME 2,000–4,000 CALORIES PER DAY BECAUSE THAT IS +/- 2 SD'S FROM THE MEAN

9. How many calories do these boys consume if they are in the top 20% of all adolescent boys?

N(3,000,500)



80% %TILE $\xrightarrow{\text{INVNORM}}$ $Z=.8416$

$$Z = 0.8416 = \frac{X-3000}{500}$$

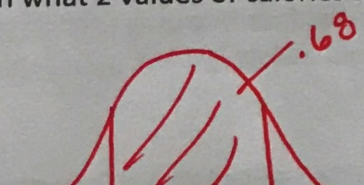
SOLVE FOR X → $\boxed{X = 3420.8}$

Check w/
INVNORM (0.8, 3,000, 500) = 3420.8

∴ THE TOP 20% OF BOYS CONSUME ABOUT 3,420 CALORIES OR MORE PER DAY.

10. Between what 2 values of calories do 68% of these boys consume?

N(3,000,500)



68% OF THESE BOYS CONSUME BETWEEN 2,500 AND 3,000 CALORIES.

Source: Karen Abbey

11. Adolescent girls eat 2200 calories on average per day with a standard deviation of 350 calories. Jane eats 2500 calories on her birthday and John eats 3600. Calculate and interpret their z-scores.

Girls
$N(2{,}200, 350)$

Jane → $Z = \frac{2500-2200}{350}$

$\boxed{Z = 0.86}$

Boys
$N(3{,}000, 500)$

John → $Z = \frac{3{,}600-3{,}000}{500} =$

$\boxed{Z = 1.2}$

∴ COMPARED TO PEERS John eats more CALORIES THAN JANE. JANE EATS ALMOST 1 MORE S.D THAN GIRLS. WHILE JOHN EATS 1.2 MORE S.D. THAN BOYS

Summary

There are 2 ways to measure a quantitative variable: by percentiles and z-scores.

- **Z-scores and percentiles help you interpret the meaning of your score.**

Percentiles tell you how your score compares to other student's scores. For instance, if 50 students take a test in Statistics and your score is a 90, if 40 students have scores below yours and everyone else scored higher, then you are in the 80^{th} percentile. (40/50 = 80%)

Z-scores are a way to measure how many standard deviations you are from the mean in a particular distribution. The z-scores enable you to compare scores from different distributions such as your SAT and ACT scores. A standardized score would be calculated for each observation in each distribution.

The Empirical Rule is the **68-95-99.7 rule** which shows the relationship between standard deviations and the percentiles. This rule can be used to determine what proportion of observations will lie between 1, 2 and 3 standard deviations from the mean. For instance, 68% of the observations fall within one standard deviation from the mean.

The normal curve is a symmetric, single peak, bell-shaped curve with a mean that is the center of the distribution. The normal model is a mathematical model for the pattern of the data. We use the normal probability tables to convert standardized scores to probabilities. The probabilities can only be used when the underlying data is normal or approximately normal. The area under a Normal Curve (a density curve) is equal to 1, or 100%.

The Standard Normal Distribution Tables are used to calculate probabilities. The Standard Normal Distribution has a mean of zero and a standard deviation of 1, or N(0,1). The table shows the estimates for the area under the normal curve, this represents the probability. ***Table A shows the probability or percentage of area to the left of the z-score value.***