Probability Review

AP Statistics April 28, 2018

Dr. John Holcomb, Cleveland State University

First: Jobs, Jobs, Jobs

- http://247wallst.com/special-report/2017/04/27/the-best-and-worst-jobs-in-america/3/

1. Statistician
> **Annual median wage:** $80,110
> **Projected job growth, 2014-2024:** 34%
> **Total employment:** 33,440

As the increasing digitization of commerce, social interaction, and other aspects of everyday life generates increasing amounts of data, companies will require more people capable of analyzing the large amounts of collected information. The number of statisticians in the United States is projected to grow by 34% in the decade ending in 2024, partially due to the profession's versatile applications. Such workers will likely be needed across a variety of fields such as academia, banking, health care, marketing, government, and sports. Most statisticians have a master's degree and a strong quantitative background — and they are well compensated. The typical statistician earns $80,110 annually, more than twice the $37,040 median for all occupations.

PROBLEM 1

The data below comes from a nutritional study conducted at Ohio University. Eighty-three subjects completed the study in which subjects evaluated variables describing beef, turkey, and emu meat. Each of the three meats was prepared with taco flavoring and subjects were asked to rate the flavor. Participants were categorized by level of education in 3 categories: some college, earned a bachelors degree, and post graduate education and asked if they liked emu meat.

	Liked Emu Meat		
Education Level	No	Yes	Total
Some College	31	18	49
Bachelor’s Degree	6	9	15
Post Graduate	10	9	19
Total	47	36	83

a) Given that a randomly selected study participant had “some college,” what is the probability he or she “liked emu meat”?

Solution

The word "Given" indicates that it is a conditional probability

A = "Liked emu meat"   B = "Some college"

P(Liked emu meat|some college) = P(A|B) =

		Liked Emu Meat	Total
Some College	31	18	49

$$\frac{\#liked\ emu\ \&\ some\ college}{\#some\ college} = \frac{18}{49} = .367$$

b) Among study participants, are the events "liked emu meat" and "some college" mutually exclusive?

Solution

Two **events** are **mutually exclusive** if they cannot occur at the same time.

Some participants (18 of them) did "like emu meat" and had "some college", so it is possible that both events occurred. Thus, these two events are *not* mutually exclusive.

c) Among study participants, are the events "liked emu meat" and "some college" independent? Justify your answer based on probabilities calculated from the table.

Solution

Two **events**, *A* and *B*, are **independent** if the fact that B occurred does not affect the probability of A occurring.

A = "Liked emu meat" *B* = "Some college"

Does $P(A|B) = P(A)$? From part a) = $P(A|B) = .367$

Find $P(A)$

	Liked Emu Meat		
Education Level	No	Yes	Total
Some College	31	18	49
Bachelor's Degree	6	9	15
Post Graduate	10	9	19
Total	47	36	83

$$P(A) = \frac{36}{83} = .434$$

No, the events are *not* independent because the two calculated probabilities (.367 & .434) are not the same.

Can also determine independence by determining if $P(A)*P(B)=P(A\cap B)$

$$P(A)=\frac{\#\,liked\ emu}{total\,\#}=\frac{36}{83}=.434$$

$$P(B)=\frac{\#\,some\ college}{total\,\#}=\frac{49}{83}=.590$$

$$P(A)*P(B)=(.434)*(.590)=.256$$

$$P(A\cap B)=\frac{\#\,liked\ emu\ \&\ some\ college}{total\,\#}=\frac{18}{83}=.217$$

No, the events are *not* independent because .256 does not equal .217.

Mutually Exclusive (Disjoint)

$$\neq$$

Independence

Mutually Exclusive: Can both events occur at the same time?

Independent: Have to calculate and see does

$$P(A|B) = P(A) \text{ or } P(A \cap B) = P(A) * P(B)$$

d) Regardless of your answer to part c), assume that "liked emu meat" is independent of "some college." Also assume the probability of "liked emu meat" is .434 and the probability of "some college" is .590. If these assumptions holds true, how many participants would we expect to have "liked emu meat" and have "some college?"

Solution

If *A* = "liked emu meat" and *B* = "some college" and *A* and *B* are independent, then $P(A) * P(B) = P(A \cap B)$

$$P(A) = .434 \quad P(B) = .590$$

$$P(A) * P(B) = (.434) * (.590) = .256 = P(A \cap B)$$

.256 then is the probability of "liked emu meat" and "some college," so the number of participants we expect is

$$P(A \cap B) * n = .256(83) = 21.2$$

DO NOT ROUND AN EXPECTED VALUE

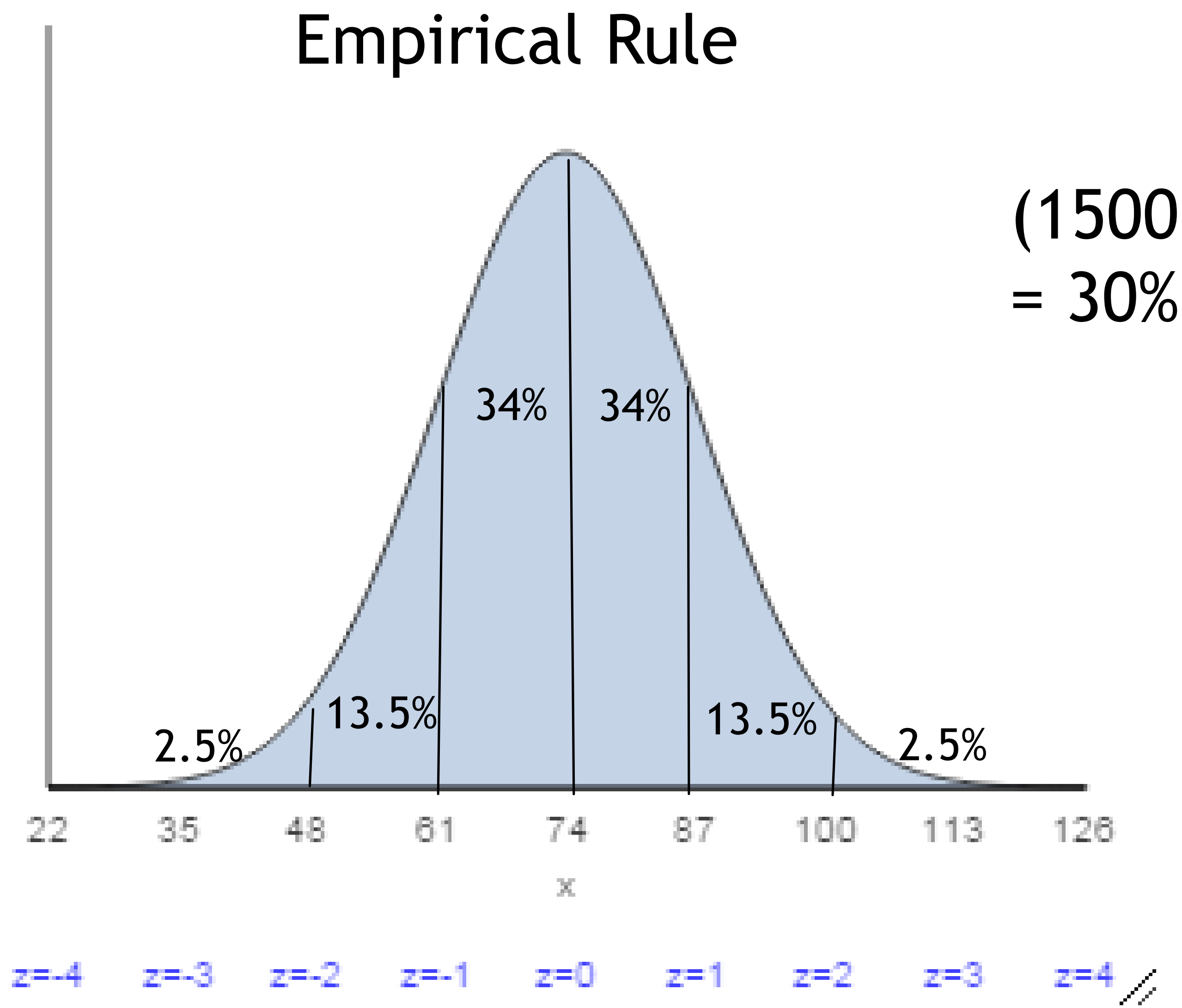
Problem 2

Resting pulse (in beats per minute - bpm) for a population of 5000 U.S. adults is normally distributed with a mean of 74 beats per minute and a standard deviation of 13 beats per minute.

Which of the following is the widest interval of beats per minute that contains approximately 1500 adults?

a) 90 bpm to 115 bpm
b) 61 bpm to 75 bpm
c) 60 bpm to 72 bpm
d) 69 bpm to 79 bpm

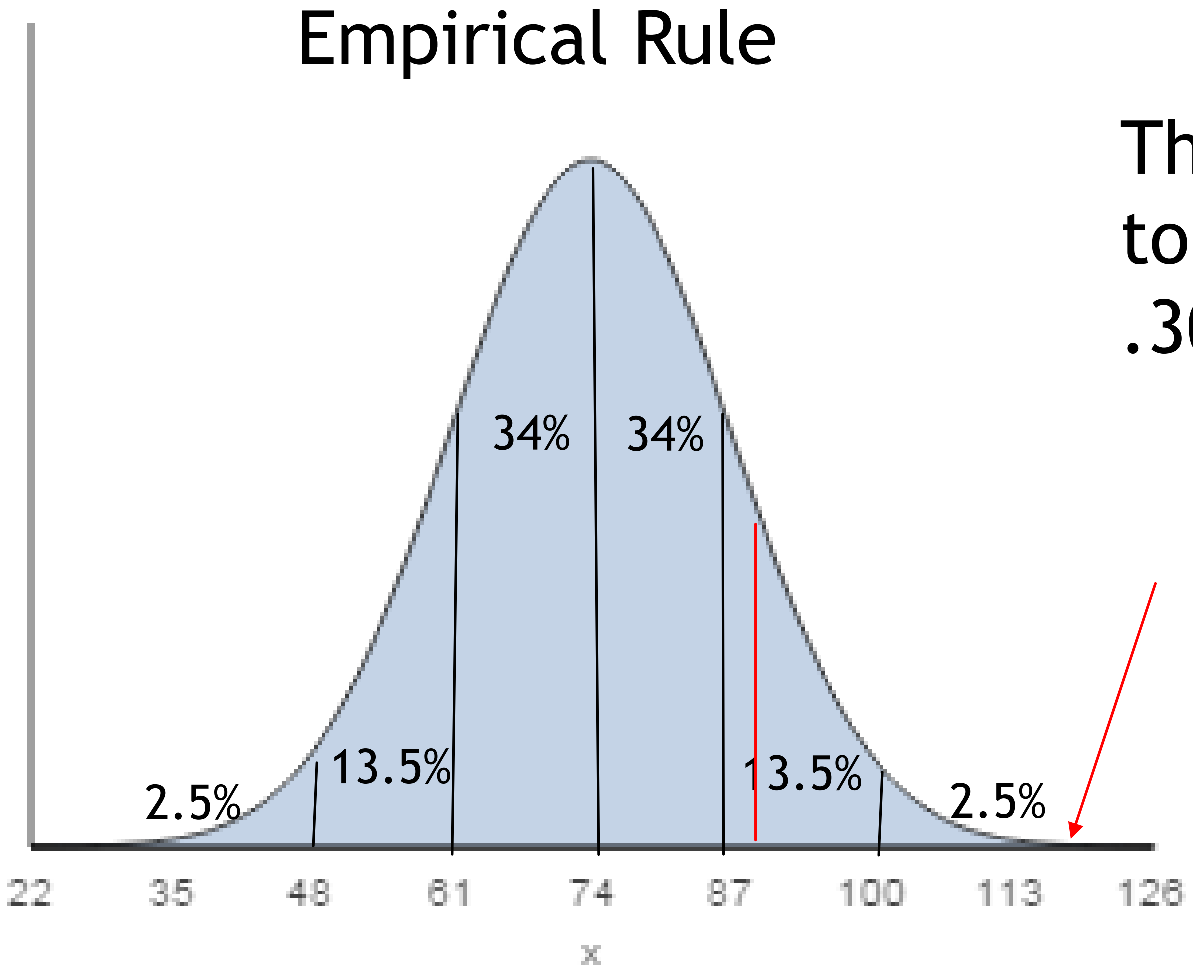
Population is Normal $\mu = 74, \sigma = 13$



(1500/5000)*100%
= 30%

Possible Answers:

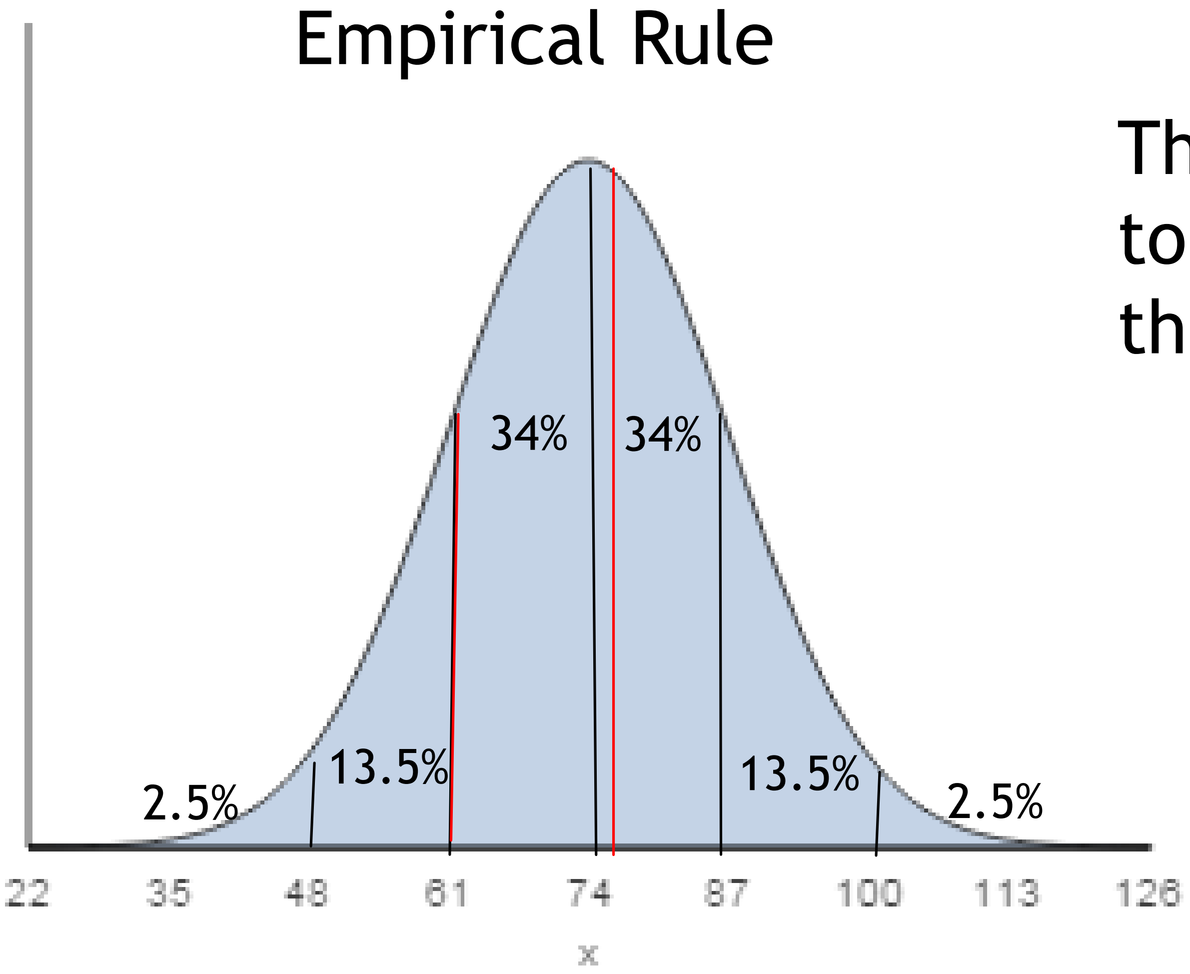
~~a)~~ 90 bpm to 115 bpm



The area has to be less than .30

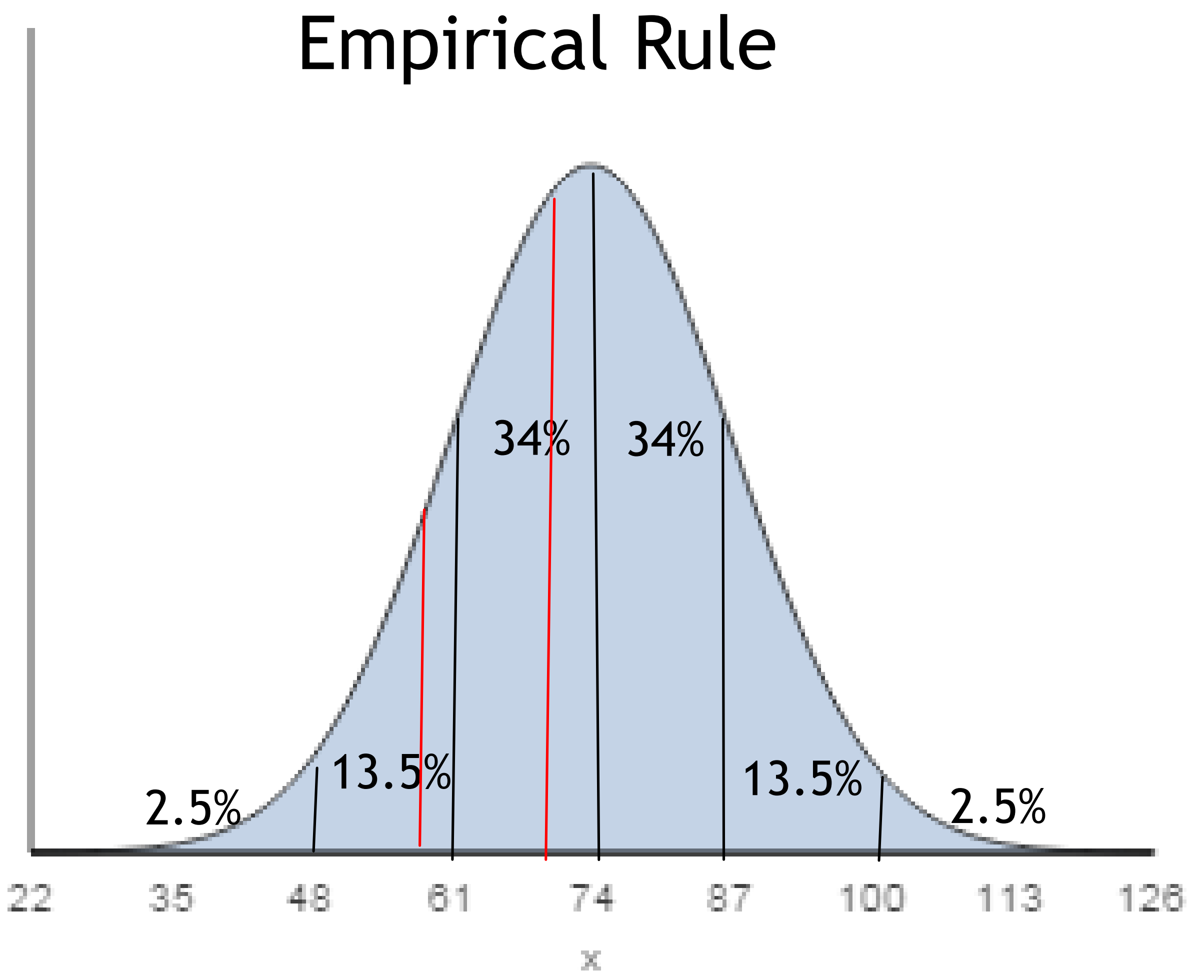
Possible Answers:

~~b)~~ 61 bpm to 75 bpm



The area has to be more than .30

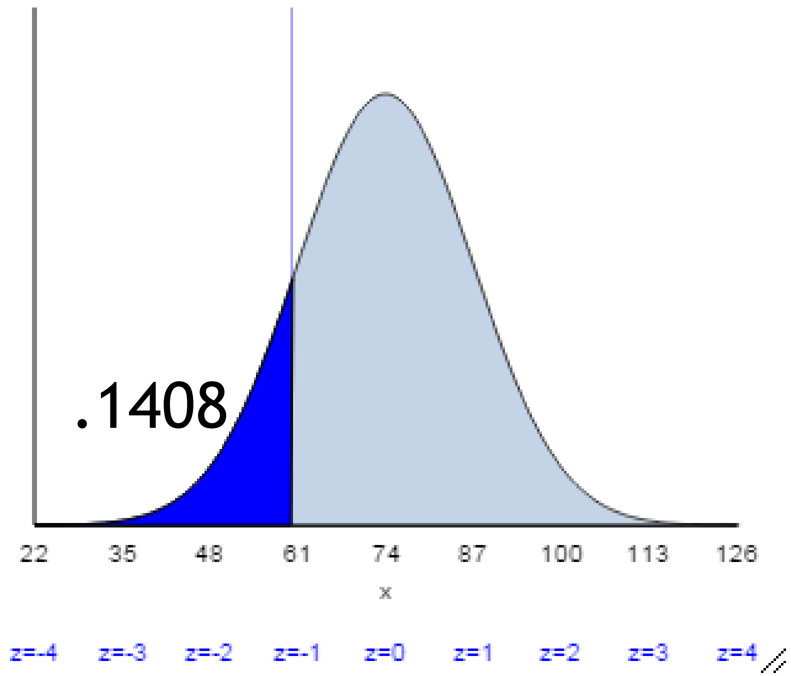
Possible Answers:
c) 60 bpm to 72 bpm

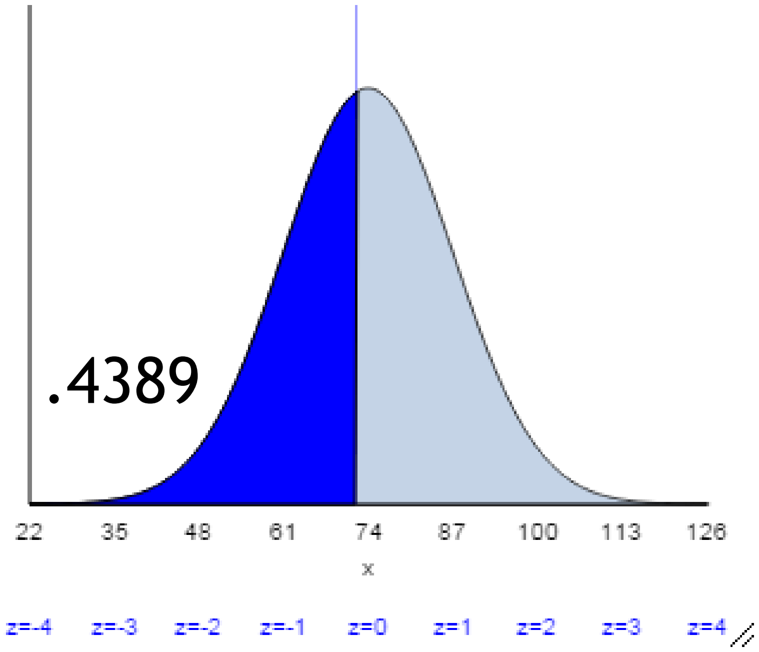


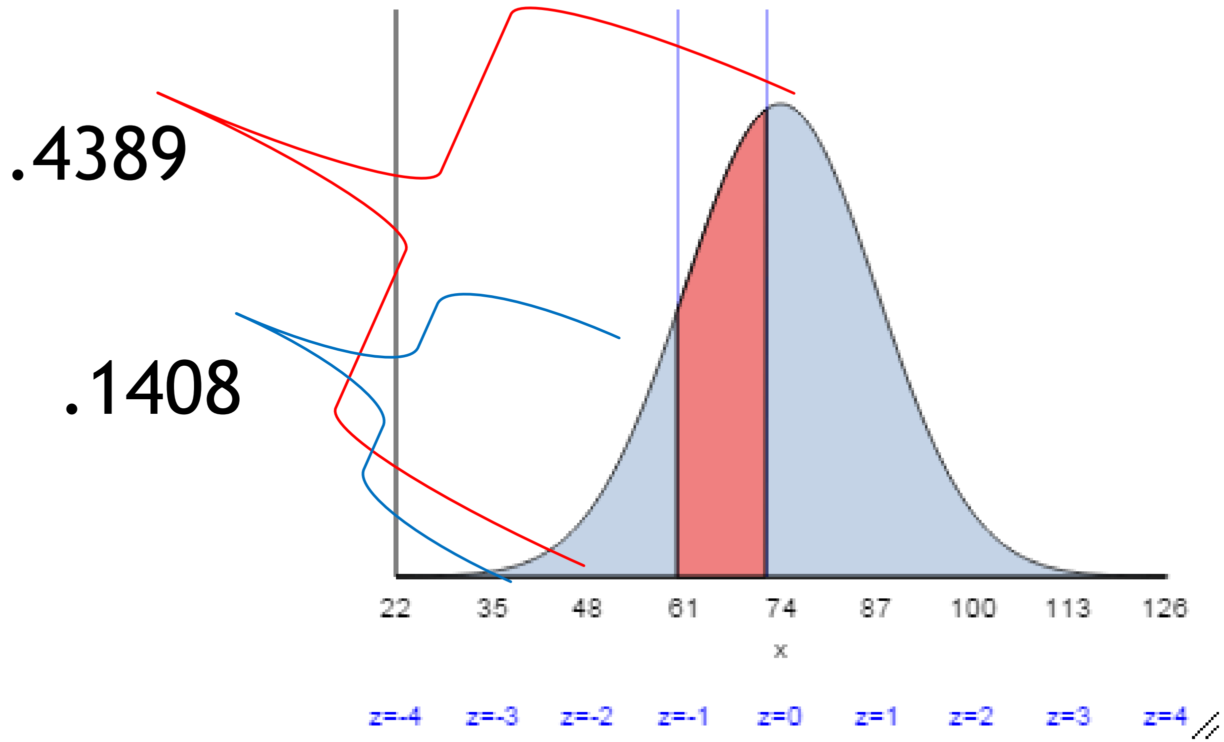
Possible Answers:
c) 60 bpm to 72 bpm

$$Z_1 = \frac{X-\mu}{\sigma} = \frac{60-74}{13} = -1.076$$

$$Z_2 = \frac{X-\mu}{\sigma} = \frac{72-74}{13} = -.154$$





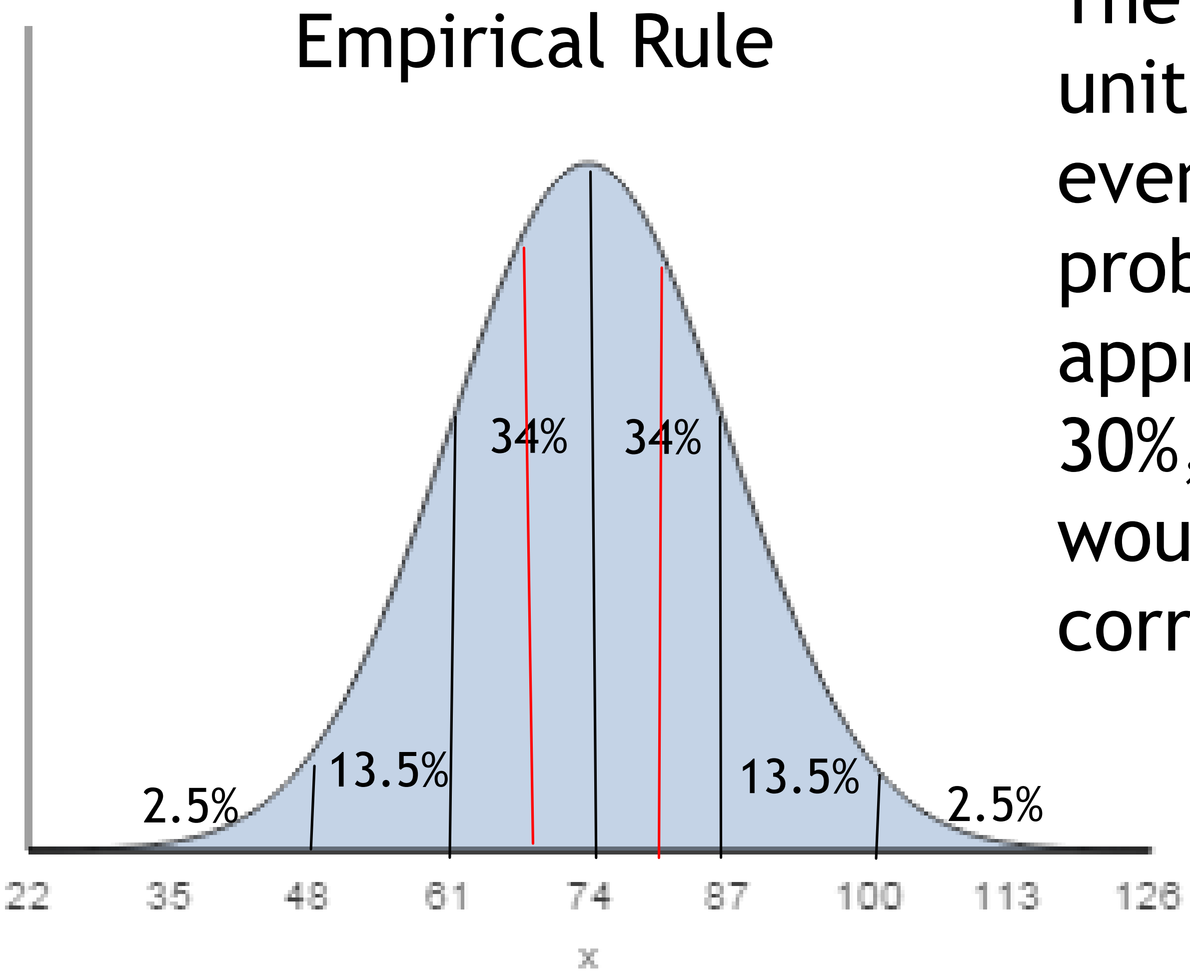


$$.4389 - .1408 = .2981 \approx .30$$

```
normalcdf(60,72,74,13)
                .2981081547
```

Possible Answers:

~~d)~~ 69 bpm to 79 bpm



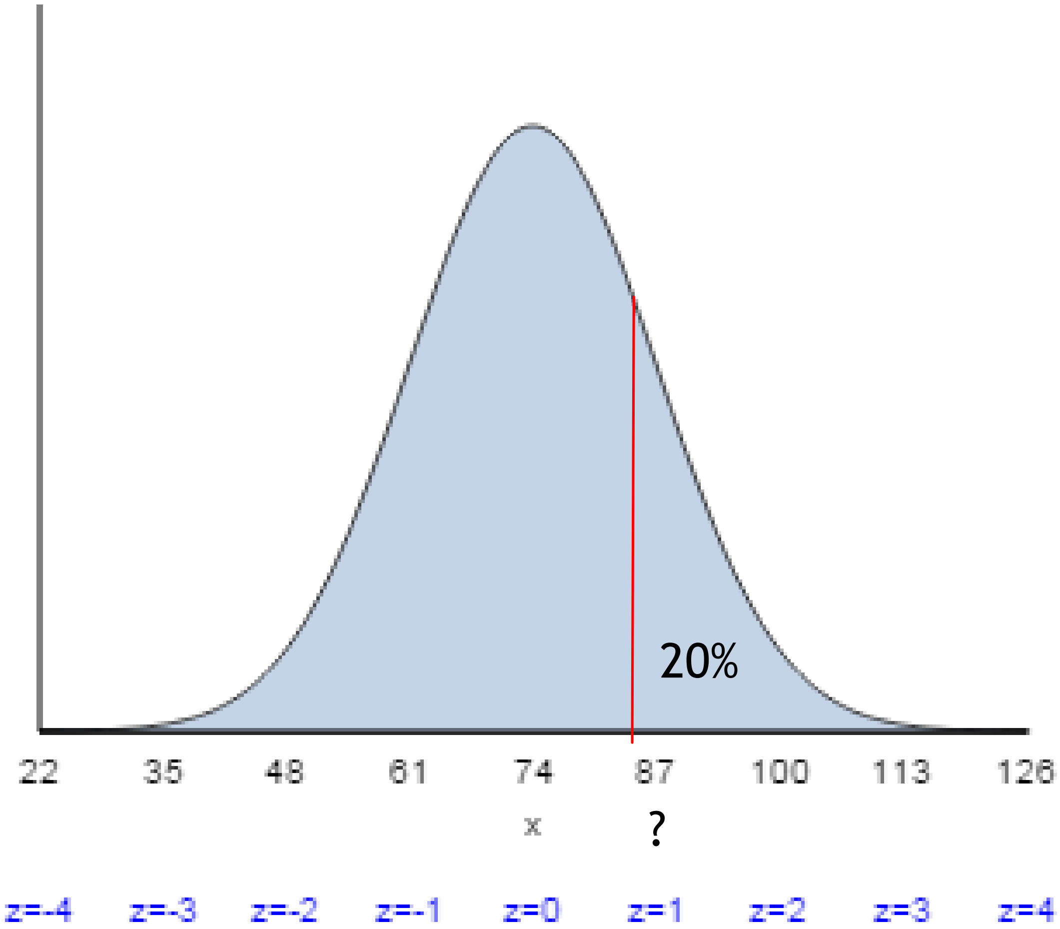
The interval is 10 units wide, so even if the probability is approximately 30%, answer c) would be the correct answer.

Resting pulse (in beats per minute - bpm) for a population of 5000 U.S. Adults is normally distributed with a mean of 74 beats per minute and a standard deviation of 13 beats per minute.

Which of the following is the widest interval of beats per minute that contains approximately 1500 adults?

a) 90 bpm to 115 bpm
b) 61 bpm to 75 bpm
c) 60 bpm to 72 bpm
d) 69 bpm to 79 bpm

b) What resting pulse has 20% of the population mentioned previously with higher beats per minute?



At what point is 20% above (or 80% below)? Use the body of the Normal Table. Z=.84 with .7995 closest.

$$Z = \frac{X-\mu}{\sigma} \Rightarrow .84 = \frac{X-74}{13}$$

$$\Rightarrow .84(13) = X - 74$$

$$\Rightarrow X = 84.92 \approx 85$$

```
invNorm(.8,74,13)
                84.94107604
```

c) What is the probability that at most 2 of 6 adults selected at random have a resting pulse greater than the one calculated in part (b)?

From part (b), it was found that 20% of adults have a resting pulse greater than 85 bpm.

What kind of situation is this?

Drawing a finite number of participants (6)

Only two possible outcomes (>85 or not)

A constant probability of success (.2)

Independent trials (at random)

Let Y denote the number of adults out of 6 with bpm greater than 85. Y is a binomial random variable with n = 6 and p = .20.

What does at most 2 mean?

$a)P(Y \geq 2)$

$b)\ P(Y < 2)$

$c)\ P(Y > 2)$

$d)\ P(Y \leq 2)$

Binomial Formula:

$$P(Y = y) = \binom{n}{y} p^y (1 - p)^{n-y}$$

$$P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2)$$

$$= \binom{6}{0} .2^0 (.8)^6 + \binom{6}{1} .2^1 (.8)^5 + \binom{6}{2} .2^2 (.8)^4$$

$$= .262 + .393 + .246 = .901$$

d) If 6 adults are selected at random, what is the expected number of adults who will have a resting pulse rate of 85 bpm or higher?

Solution

We know from part (b) that 20% of adults have a resting pulse of 85 bpm or higher. We know from part (c) that if *Y* is the number selected from 6 with a resting pulse of 85 bpm or higher than *Y* is binomial with $n = 6$ and $p = .2$.

$$E(Y) = np = 6(.20) = 1.2$$

DO NOT ROUND AN EXPECTED VALUE

e) What is the probability that a randomly selected sample of 6 adults will have a mean resting pulse greater than 86 beats per minute?

Let $\bar{X}$ denote the mean resting pulse for 6 randomly selected adults. Because resting pulses have a normal distribution, the distribution of $\bar{X}$ is normal with a mean of 74 bpm and a standard deviation of

$$\frac{\sigma}{\sqrt{n}} = \frac{13}{\sqrt{6}} = 5.31\text{bpm}$$

$$P(\bar{X} > 86) = P\left(Z > \frac{86-74}{(13/\sqrt{6})}\right) = P(Z > 2.26) = 1 - .9881 = .0119$$

Common Mistakes

1. Not showing enough work

Correct $$E(Y) = np = 6(.20) = 1.2$$

Dinged $$n = 6 \quad p = .20 \qquad E(Y) = 1.2$$

2. Rounding Expected Values

Expected Values are averages, so it is okay to have a fractional person.

3. Incorrectly using the distribution of X instead of $\bar{X}$ with normality (or vice versa);

What is the probability that a randomly selected sample of 6 adults will have a mean resting pulse greater than 86 beats per minute?

Correct:

$$P(\bar{X} > 86) = P\left(Z > \frac{86-74}{(13/\sqrt{6})} \right) = P(Z > 2.26) = 1 - .9881 = .0119$$

Incorrect:

$$P(\bar{X} > 86) = P\left(Z > \frac{86-74}{13} \right) = P(Z > .92) = 1 - .8220 = .1780$$

4. Confusing notation for parameters and statistics

Population parameters

$$\mu \qquad \sigma \qquad p$$

Sample statistics (estimators)

$$\bar{X} \qquad s \qquad p$$

5. Forgetting the difference between variance and standard deviation

$$\mu = E(X) = \sum_{i=1}^{k} x_i p_i$$

$$\sigma^2 = Var(X) = E(X^2) - [E(\mathrm{X})]^2 = \sum_{i=1}^{k} x_i^2 p_i - [E(x)]^2$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{E(X^2) - [E(\mathrm{X})]^2} = \sqrt{\sum_{i=1}^{k} x_i^2 p_i - [E(x)]^2}$$

6. Non-labelled calculator speak

```
normalcdf(60,72,74,13)
                .2981081547
```

Correct: normalcdf(lower=60, upper=72, μ =74, σ =13)

Dinged: normalcdf(60, 72, 74, 13)

GOOD LUCK!!!!