

1. A Baylor Heart and Vascular Hospital study (<http://www.dallasnews.com/lifestyles/health-and-fitness/health/20140821-broke-college-kids-beware-baylor-study-links-instant-noodles-to-health-hazards.ece>) in the United States that linked instant noodles consumption by South Koreans to some risks for heart disease. If they made this realization by looking at someone's heart health score and the number of cups of noodles they eat per week what would the variable(s) be in this study and what type(s) would they be?

- i) Heart health score is one variable but we don't have enough info to say which type.
 ii) Number of cups of noodles per week is a discrete, quantitative variable.

2. When looking for a mate men and females have different tastes. On average, from student surveys, I've found that men prefer females to be about 5' 6" and females prefer males to be about 5' 11". What would the variable(s) be in this study and what type(s) would they be?

- i) Gender is a categorical variable
 ii) Height is a continuous quantitative variable that we measure discretely

3. Below is a table of the number of goals seen in a random sample of 62 English Premier League games in 2013. Fill in the relative frequency and percentage column and write a sentence about the frequency data.

Number of Goals	Frequency	Relative Frequency	Percentage
0	3	$\frac{3}{62} = .0483871$	4.84%
1	21	$\frac{21}{62} = .33870968$	33.87%
2	11	$\frac{11}{62} = .17741935$	17.74%
3	10	$\frac{10}{62} = .16129032$	16.13%
4	11	$\frac{11}{62} = .17741935$	17.74%
5	5	$\frac{5}{62} = .08064516$	8.06%
6	0	$\frac{0}{62} = 0$	0%
7	1	$\frac{1}{62} = .01612903$	1.61%
Totals	62	≈ 1	$\approx 100\%$

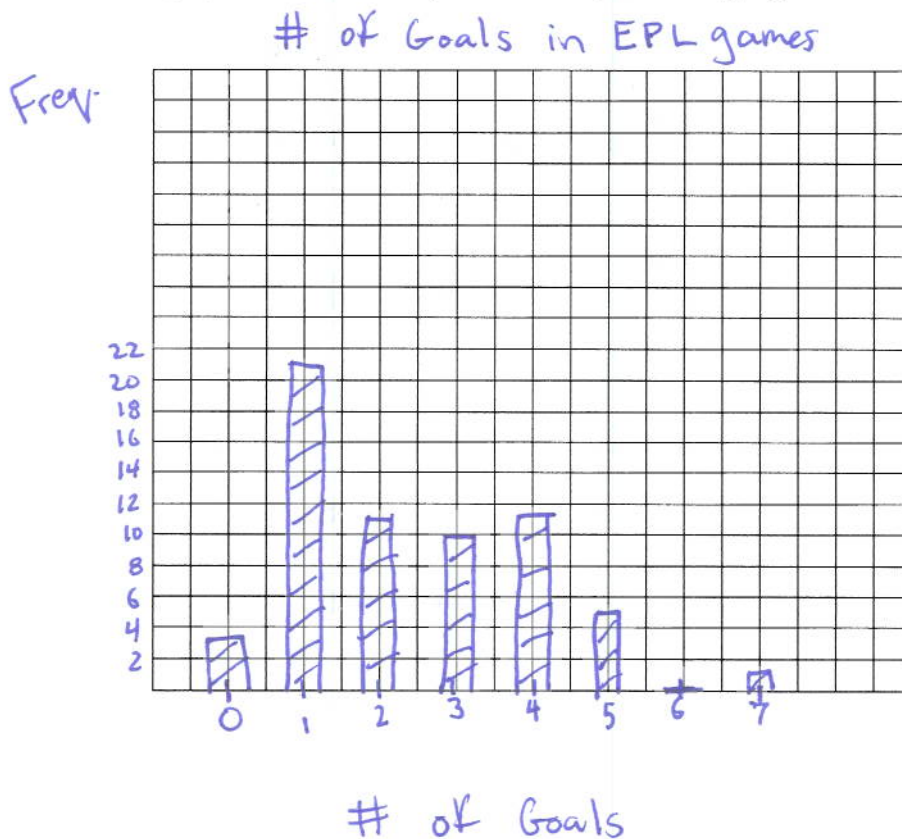
1 goal is the most occurring # of goals per game.

4. I ate a giant bag of M&M's and I kept track of how many of each color I saw and recorded it in the table below. Fill in the relative frequency and percentage column and write a sentence about the frequency data.

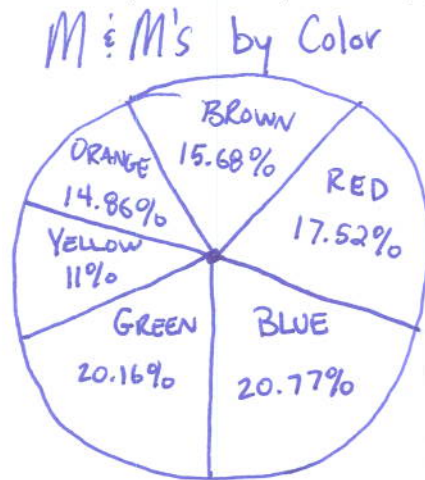
Color	Frequency	Relative Frequency	Percentage
Blue	102	$\frac{102}{491} = .20773931$	20.77%
Green	99	$\frac{99}{491} = .20162933$	20.16%
Red	86	$\frac{86}{491} = .17515275$	17.52%
Brown	77	$\frac{77}{491} = .15682281$	15.68%
Orange	73	$\frac{73}{491} = .14867617$	14.87%
Yellow	54	$\frac{54}{491} = .10997963$	11.00%
Totals	491	≈ 1	$\approx 100\%$

Blue and Green M&M's are most common.

5. Draw a bar graph for the data in question three; include appropriate labels.



6. Draw a pie chart for the data in question four; include appropriate labels.



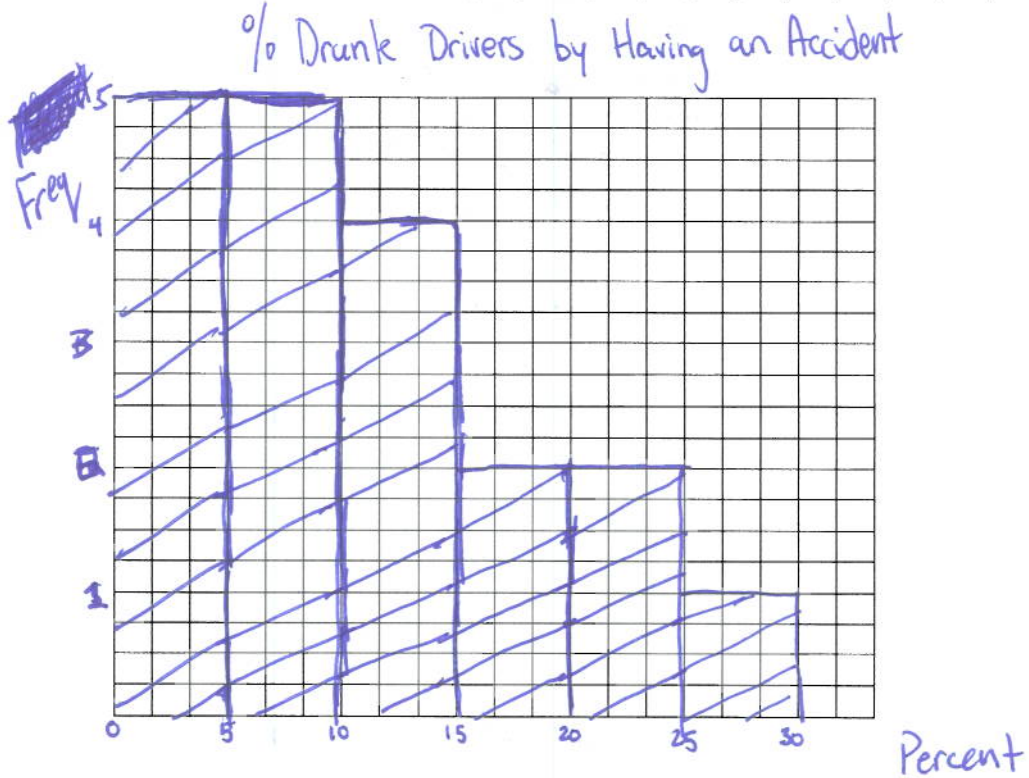
7. Draw a side by side bar chart for the following data; include appropriate labels and write a sentence about your results

Gender	Sex within the last 24 hours	No sex within the last 24 hours	Total
Male	5	6	11
Female	7	30	37
Total	12	36	48



8. Draw a histogram for the following data about the percent of drunk drivers caught by having an accident vs other from , using bars with 5% width; include appropriate labels and write a sentence about your results and make a statement about how the graph is skewed.

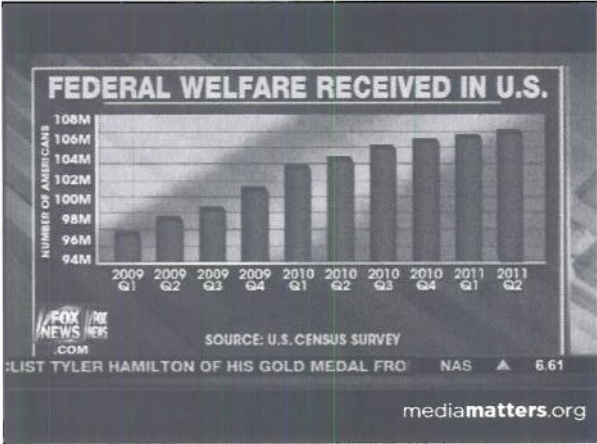
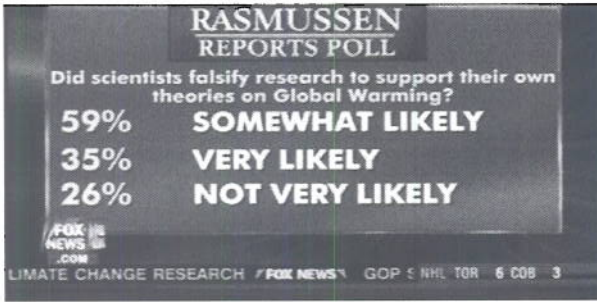
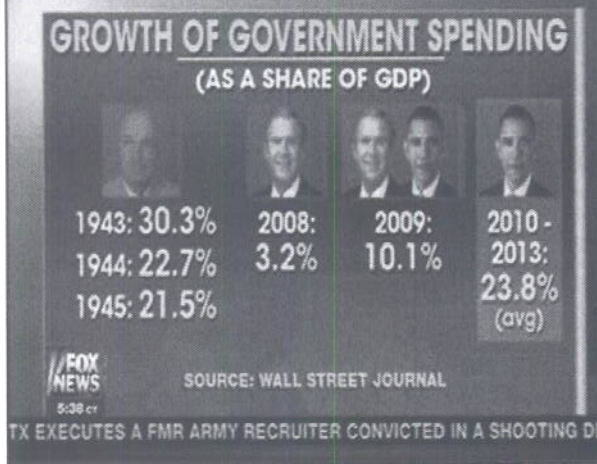
26%, 20%, 22%, 17%, 18%, 13%, 10%, 14%, 8%, 11%, 7%, 7%, 4%, 5%, 3%, 2%, 4%, 5%, 1%



	#
0-5	5
5-10	5
10-15	4
15-20	2
20-25	2
25-30	1

This graph is skewed right.


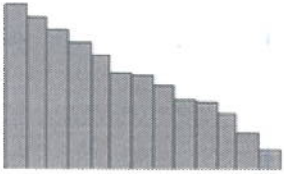
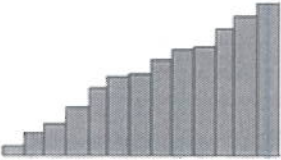
9. Fox News has been known, along with other networks, to use misleading graphs to tell a version of news that fits their agenda. (<http://simplystatistics.org/2012/11/26/the-statisticians-at-fox-news-use-classic-and-novel-graphical-techniques-to-lead-with-data/>) What is misleading about these graphs?

Graph	Why is this misleading?
 <p>FEDERAL WELFARE RECEIVED IN U.S.</p> <p>NUMBER OF AMERICANS</p> <p>108M 106M 104M 102M 100M 98M 96M 94M</p> <p>2009 Q1 2009 Q2 2009 Q3 2009 Q4 2010 Q1 2010 Q2 2010 Q3 2010 Q4 2011 Q1 2011 Q2</p> <p>FOX NEWS .COM SOURCE: U.S. CENSUS SURVEY</p> <p>LIST TYLER HAMILTON OF HIS GOLD MEDAL FRO NAS ▲ 6.61</p> <p>mediamatters.org</p>	<p>The graph doesn't start at zero. 2009 Q4 appears to be twice as much as 2009 Q1 but it is not.</p>
 <p>RASMUSSEN REPORTS POLL</p> <p>Did scientists falsify research to support their own theories on Global Warming?</p> <p>59% SOMEWHAT LIKELY</p> <p>35% VERY LIKELY</p> <p>26% NOT VERY LIKELY</p> <p>FOX NEWS .COM</p> <p>CLIMATE CHANGE RESEARCH / FOX NEWS \ GOP < NHL TOR 6 COB 3</p>	<p>The percentages don't add up to 1.</p>
 <p>GROWTH OF GOVERNMENT SPENDING (AS A SHARE OF GDP)</p> <p>1943: 30.3% 2008: 3.2% 2009: 10.1% 2010 - 2013: 23.8% (avg)</p> <p>1944: 22.7%</p> <p>1945: 21.5%</p> <p>FOX NEWS</p> <p>SOURCE: WALL STREET JOURNAL</p> <p>TX EXECUTES A FMR ARMY RECRUITER CONVICTED IN A SHOOTING DEA</p>	<p>The numerical summaries have different units of measure → Total Expenditures → Deficits</p>

NOTE: 2010-2013 is total expenditures and 2008-2009 are deficits

FREEBIE: The picture to the left is Truman and he wasn't president until 04/1945 when FDR died

10. Describe the shape of the following histograms and circle the correct inequality or equality

Graph	Shape (Name the Shape)	Mean and Median (circle one)
	Bell shaped	a. mean = median b. <i>mean < median</i> c. <i>mean > median</i>
	Skewed Right	a. <i>mean = median</i> b. <i>mean < median</i> c. mean > median
	Skewed Left	a. <i>mean = median</i> b. mean < median c. <i>mean > median</i>

11. Below is the data for the number of absences per student in a boring, non-Statistics 201 class. Please calculate the following numerical summaries:

0, 0, 1, 1, 1, 1, 2, 3, 3, 4, 4, 4, 4, 6, 6, 6, 7, 8, 8, 9, 9, 9

Note: $\sum x = 96$

$$\sum(x - \bar{x})^2 = 199.09$$

$$\text{variance} = \frac{\sum(x - \bar{x})^2}{n-1}$$

$$\text{Position of the } p\text{th percentile} = \left(\frac{p}{100}\right) * (n + 1)$$

Mean: $\bar{x} = \frac{\sum x}{n} = \frac{96}{22} = 4.\overline{36}$

First Quartile: $.25(22+1) = 5.75^{\text{th}}$ position $\rightarrow \frac{1+1}{2} = 1$

Median: $.5(22+1) = 11.5^{\text{th}}$ position $\rightarrow \frac{4+4}{2} = 4$

Third Quartile: $.75(22+1) = 17.25^{\text{th}}$ position $\rightarrow \frac{7+8}{2} = 7.5$

Mode: 1 and 4

Variance: $s^2 = \frac{\sum(x - \bar{x})^2}{n-1} = \frac{199.09}{22-1} = 9.48047619$

Standard Deviation: $s = \sqrt{s^2} = \sqrt{9.48047619} = 3.07903819$

12. Below is the data for a sample of the attractiveness ratings, on a scale from 0 to 10, for you by your peers in Stat 201. Please calculate the following numerical summaries:

6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10

Note: $\sum x = 276$

$$\sum(x - \bar{x})^2 = 61.54$$

$$\text{variance} = \frac{\sum(x - \bar{x})^2}{n - 1}$$

$$\text{Position of the } p\text{th percentile} = \left(\frac{p}{100}\right) * (n + 1)$$

Mean: $\bar{x} = \frac{\sum x}{n} = \frac{276}{35} = 7.88571429$

First Quartile: ~~0.25(35)~~
 $.25(35 + 1) = 9^{\text{th}}$ position $\rightarrow 7$

Median: $.5(35 + 1) = 18^{\text{th}}$ position $\rightarrow 8$

Third Quartile: $.75(35 + 1) = 27^{\text{th}}$ position $\rightarrow 9$

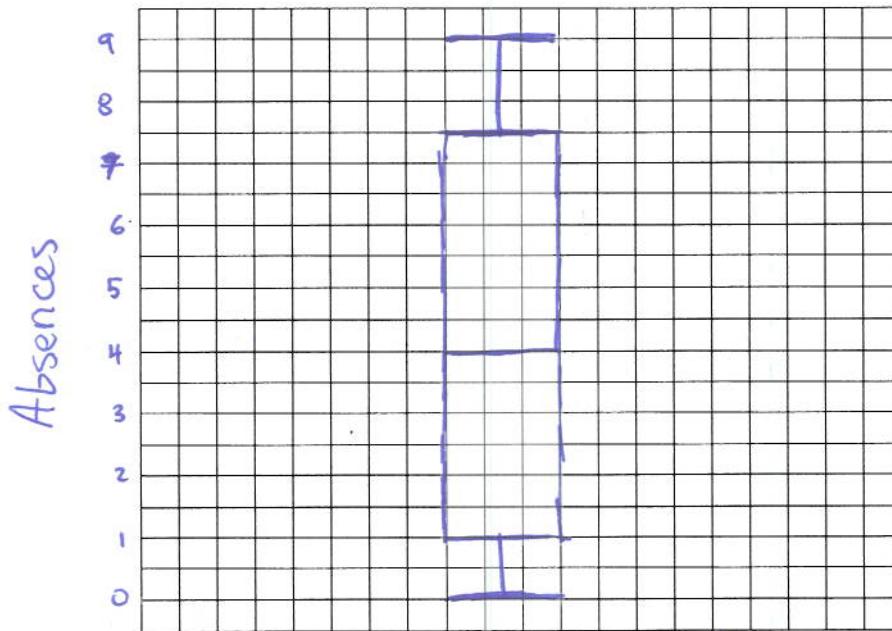
Mode: 7 and 8

Variance: $s^2 = \frac{\sum(x - \bar{x})^2}{n - 1} = \frac{61.54}{35 - 1} = 1.81$

Standard Deviation: $s = \sqrt{s^2} = \sqrt{1.81} = 1.3453624$

13. Draw a box plot for the data in question eleven; include appropriate labels and comment on how the graph is skewed.

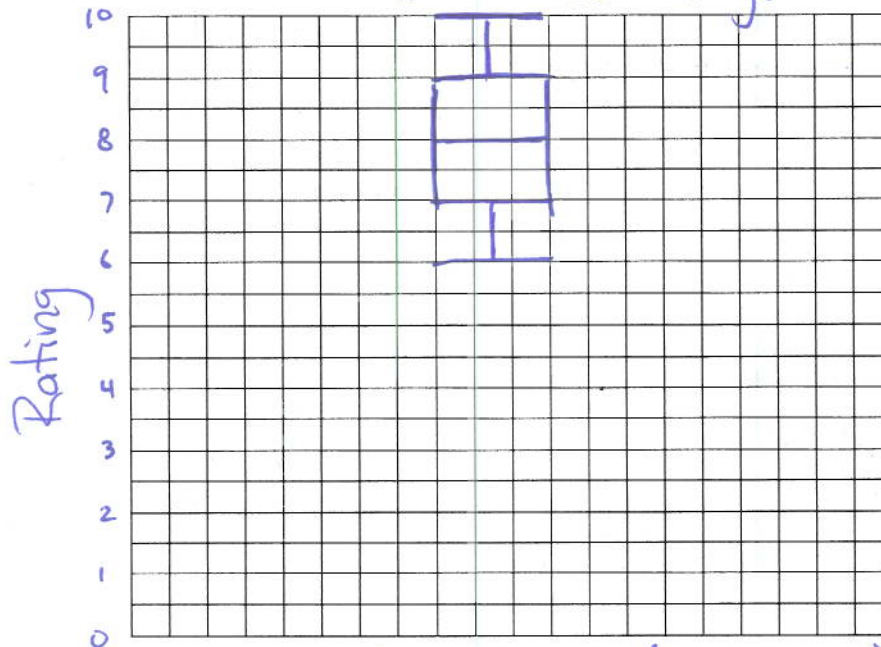
Absences by Students



- i) This graph is skewed very little ($4.36 \approx 4$)
 ii) This graph is slightly skewed right ($4.36 > 4$)

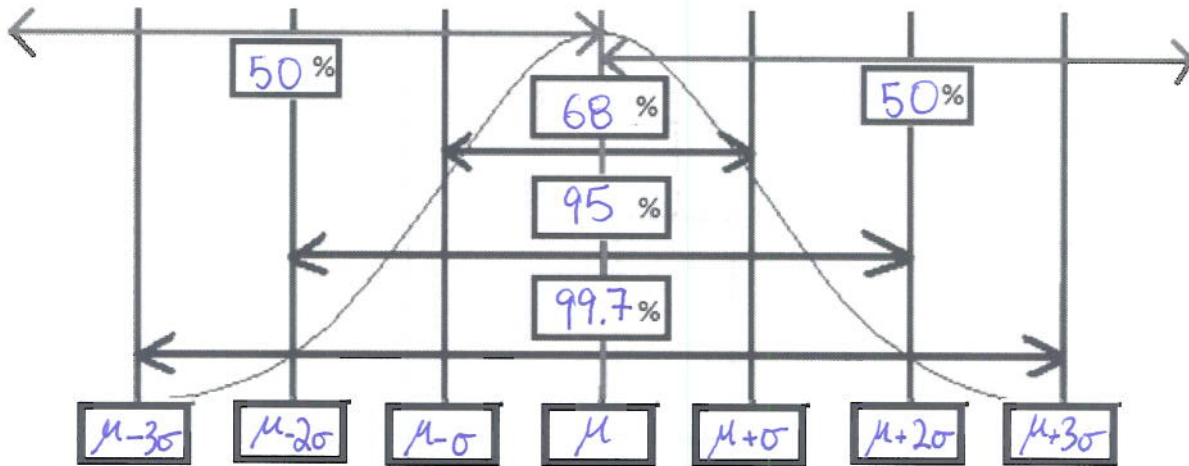
14. Draw a box plot for the data in question twelve; include appropriate labels and comment on how the graph is skewed.

Attractiveness Ratings



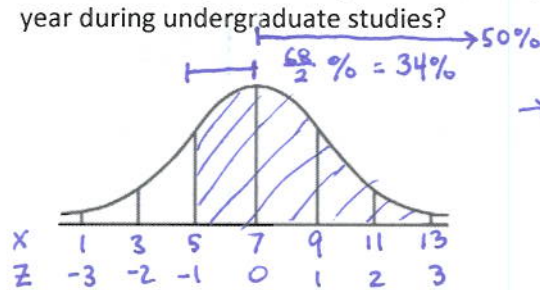
- i) This graph is skewed very little ($7.8857 \approx 8$)
 ii) This graph is slightly skewed left ($7.8857 < 8$)

15. Fill in the values according to the Empirical Rule.



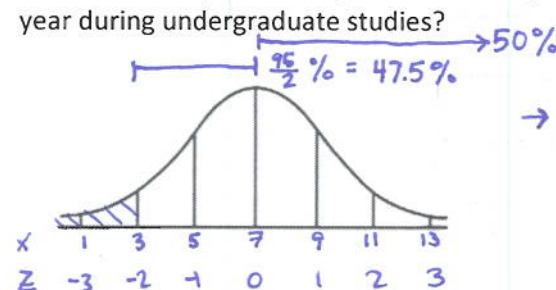
16. Suppose that the distribution of the average number of 'hook-up' partners per year during undergraduate studies is known to be bell-shaped and symmetric with a mean of 7 and a standard deviation of 2.

a. What percentage of undergraduates averaged more than five 'hook-up' partners per year during undergraduate studies?



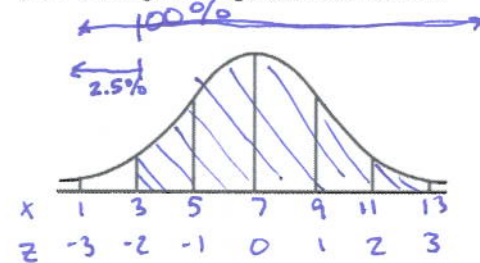
$50 + 34 = 84$
 → 84% of undergraduates averaged more than five 'hook-up' partners per year.

b. What percentage of undergraduates averaged less than three 'hook-up' partners per year during undergraduate studies?



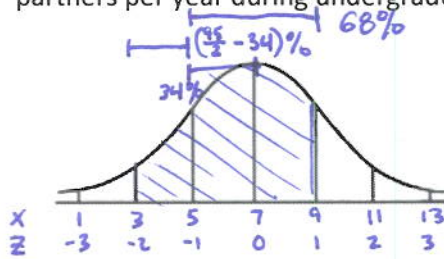
$100 - 50 - 47.5 = 2.5$
 → 2.5% of undergraduates averaged less than three 'hook-up' partners per year.

c. What percentage of undergraduates averaged more than three 'hook-up' partners per year during undergraduate studies?



$100 - 2.5 = 97.5$ [Complement Rule]
 → 97.5% of undergraduates averaged more than three 'hook-up' partners per year.

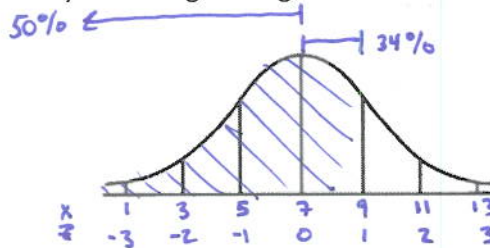
- d. What percentage of undergraduates averaged between three and nine 'hook-up' partners per year during undergraduate studies?



$$68 + \left(\frac{95}{2} - 34\right) = 81.5$$

→ 81.5% of undergraduates averaged between three and nine 'hook-up' partners per year.

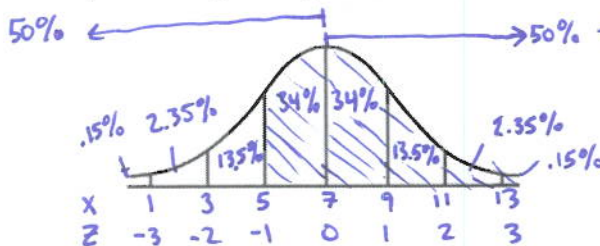
- e. What percentage of undergraduates averaged less than nine 'hook-up' partners per year during undergraduate studies?



$$50 + 34 = 84$$

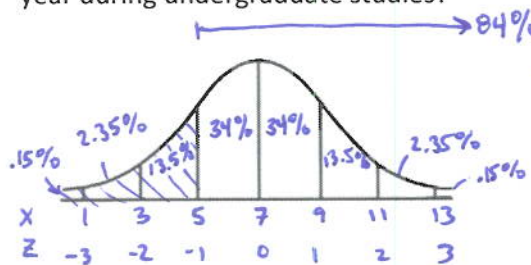
→ 84% of undergraduates averaged less than nine 'hook-up' partners per year.

- f. 84 percent of undergraduates averaged more than what number 'hook-up' partners per year during undergraduate studies?



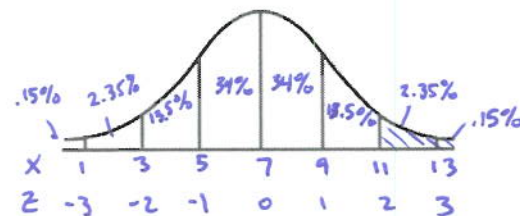
→ 84% of undergraduates averaged more than 5 'hook-up' partners per year during undergraduate studies.

- g. 16 percent of undergraduates averaged less than what number 'hook-up' partners per year during undergraduate studies?



→ 16% of undergraduates averaged less than 5 'hook-up' partners per year during undergraduate studies

- h. 2.5 percent of undergraduates averaged more than what number 'hook-up' partners per year during undergraduate studies?



→ 2.5% of undergraduates averaged more than 11 'hook-up' partners per year during undergraduate studies

- i. What is the z-score of an undergraduate that averages 10 'hook-up' partners per year during undergraduate studies? Is this an outlying observation?

$$z = \frac{x - \mu}{\sigma} = \frac{10 - 7}{2} = \frac{3}{2} = 1.5$$

This is not outlying as $-3 \leq 1.5 \leq 3$

- j. What is the z-score of a student that averages 16 'hook-up' partners per year during undergraduate studies? Is this an outlying observation?

$$z = \frac{x - \mu}{\sigma} = \frac{16 - 7}{2} = \frac{9}{2} = 4.5$$

This is outlying as $4.5 > 3$

- k. What is the z-score of a student that averages 5 'hook-up' partners per year during undergraduate studies? Is this an outlying observation?

$$z = \frac{5 - 7}{2} = -1$$

This is not outlying as $-3 \leq -1 \leq 3$

- l. What is the z-score of a student that averages 0 'hook-up' partners per year during undergraduate studies? Is this an outlying observation?

$$z = \frac{0 - 7}{2} = -3.5$$

This is outlying as $-3.5 < -3$

- m. What is the z-score of a student that averages 7 'hook-up' partners per year during undergraduate studies? Is this an outlying observation?

$$z = \frac{7 - 7}{2} = 0$$

This is not outlying as $-3 \leq 0 \leq 3$