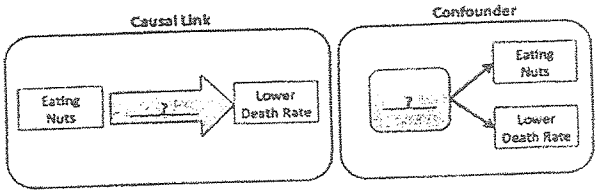
Stat 100 F1 Exam 1 page 1 side EXAM 1 KEY Spring 2014

**Question 1** (6 pts.)

Researchers tracked 119,000 men and women over a 30 year period and found that those who ate nuts every day were 20 percent less likely to die than those who never ate nuts. The risk of dying of heart disease dropped 29 percent and the risk of dying of cancer fell 11 percent among those who had nuts seven or more times a week compared with people who never ate them.

A-C below are either Causal Links that would fit in the Causal Link Arrow (in the diagram to the right) or Confounders that would fit in the Confounder Box, or Neither (other causes of a lower death rate that have nothing to do with eating nuts.)




A. Sex – Females have a lower death rate than males. (2 pts.)
**Choose one:** i) Causal Link ii) Confounder (iii) Neither

B. Wealth--Nuts are more expensive than other snack foods. People who are wealthier can afford to eat nuts every day and people who are wealthier have a lower death rate. (2 pts.)
**Choose one:** i) Causal Link (ii) Confounder iii) Neither

C. Antioxidants--Nuts are high in antioxidants that lower the risk of heart disease. (2 pts.)
**Choose one:** (i) Causal Link ii) Confounder iii) Neither

**Question 2** (6 pts.)

A recent study was done to test the effectiveness of acupuncture in relieving the side effects of menopause brought on by breast cancer drugs.The subjects were 48 breast cancer patients receiving the same cancer drug.

Half were **randomly** assigned to 8 weekly sessions of real acupuncture (treatment with needles in recognized acupoints believed to be helpful in relieving menopausal symptoms) and half were randomly assigned to 8 weekly sessions of sham acupuncture (treatment with non-penetrating needles placed in sham acupuncture points). Although the doctors who performed the acupuncture knew whether it was real or sham, neither the patients nor the researchers who evaluated them knew who was in which group.

All subjects kept daily diaries and filled out questionnaires rating the severity of their symptoms at 0 weeks, 1 week, 4 weeks and 8 weeks. Patient satisfaction with both the real and sham acupuncture was high. In fact, there was no significant difference between the 2 groups on any measure at any time period.

a) Which of the following best describes this study? (2 pts.) ***Choose one:***
- i) It's an observational study
- (ii) It's a randomized controlled double-blind experiment
- iii) It's a non-randomized controlled double-blind experiment.
- iv) It's a non-randomized experiment with a placebo.
- v) It's a non-randomized experiment with historical controls.

b) Which of the following statements is best? (2 pts.) ***Choose one:***
- ~~i)~~ This study is very strong evidence that traditional acupuncture works better than a placebo to ease the side effects of this particular cancer drug.
- ~~ii)~~ This study only shows an *association* between acupuncture and easing side effects. It does not prove or disprove that acupuncture *caused* a reduction in side effects since there's bound to be other differences between those who received the real and those who received the fake acupuncture that could confound the results.
- (iii) This study is strong evidence that real acupuncture works no better than sham acupuncture to ease the side effects of this cancer drug among this group of breast cancer patients.

c) Which of the following are likely to confound the results of this study? (2 pts.) ***Choose one:***
- i) Pain Tolerance- People who choose acupuncture may tolerate pain better and thus report fewer symptoms.
- ii) Alternative Medicine- People who choose acupuncture are more likely to be taking alternative therapies such as herbal cures and massage which could help alleviate their symptoms.
- iii) Those who were in the sham acupuncture group did not have penetrating needles so they probably knew they weren't getting the real thing and thus didn't feel the same relief as the treatment group.
- iv) All of the above are likely confounders.
- (v) None of the above are likely confounders. bc RCDB (ideal) study

**Question 3** (6 pts)
I've taught Stat 100 for 14 years and have always allowed students to sit anywhere they want. The same students sit in the same part of the lecture hall every class. Jackie believes there's probably an association between where students sit in class and how well they do, I doubt there's any such association. We can't check the records from past semesters because I never recorded where student sat.

a) How can I find out who's right? How can we see if there is some association between where Stat 100 students sit and their grades? ***Choose one:*** (2 pts.)

i) Observe and record what region of the lecture hall students sit in and see if there's any association between where they sit and the grade they get.
ii) Assign lecture seats by alphabetical order from front to back. (A's in the front, Z's in the back), then see if there's any association between where they sit and the grade they get.
iii) Assign lectures seats by time of arrival, filling up the lecture hall from front to back. Record where students sit each lecture and see if there's any association between where they sit and the grade they get.
iv) Assign lecture seats by an objective random procedure and see if there's any association between where they were randomly assigned to sit and the grade they get.

b) Now suppose we've seen that there IS an association—students who sit in the front rows get higher grades on the average than students who sit in the back rows. Does this mean we can conclude that sitting closer to the front *causes* students to get better grades? ***Choose one:*** (2 pts.)

i) Yes, once an association is found it's proof that it has to be at least one of the causes of students getting better grades, although there are bound to be others as well—hours spent studying, math ability, etc.
ii) Maybe, it could be one of the causes of students getting better grades, but we can't be sure since stronger and weaker students may choose to sit in different places for reasons that don't contribute to their success but just reflect it.
iii) No, it's not a cause of students getting better grades. Students sort themselves out; better students sit in the front because they're more serious. If you seat the better students in back they'd do just as well.

c) If you wanted to design a study to determine whether sitting closer to the front *causes* higher grades, which is best? *Choose one:* (2 pts.)

i) Encourage students to carefully consider where to sit based on all the factors that might contribute to their success in the course. The compare the grades of those in the front seats to those in the back.
ii) Give students a pre-test. Assign weaker students to the front and stronger students to the back. Then compare the grades of the 2 groups.
iii) Devise an interactive seating plan- one that evaluates and responds to student performance by re-assigning seats after each exam to maximize student potential, taking into account the fact that students have different responses to being in the front rows. Compare grades of those in the front to those in the back after each re-assignment.
iv) Assign seats by choosing a variety of students to sit in the front rows and then trying to match the students in the back rows to be as alike as possible on all characteristics relevant to doing well in the course (i.e., ACT scores, GPA, major, year in school, etc.) Then compare the grades of the 2 groups.
v) Assign lecture seats by an objective random procedure so that all students have the same chance of getting front row seats as back row seats. Compare the exam scores of those randomly assigned to the front rows to those randomly assigned to the back rows.

**Question 4** (8 pts.)
Two high schools offer courses to prepare students for both the Physics B (non-calculus based mechanics) and Physics C (calculus based mechanics) AP exams. For both exams, a score of 3 or higher is passing, and below 3 is failing. Here's the results for the past 5 years of all the students from School X and Y who took either exam.

| | School X | | | School Y | | |
|---|---|---|---|---|---|---|
| | # Pass | # Fail | % Pass | # Pass | # Fail | % Pass |
| **Physics B** | 900 | 100 | 90% | 850 | 150 | 85% |
| **Physics C** | 600 | 400 | 60% | 30 | 70 | 30% |
| **Total** | 1500 | 500 | 75% | 880 | 220 | 80% |

a) Which school has a higher passing rate for Physics B? 90% 85%
i) School X ii) School Y iii) not enough info

b) Which school has a higher passing rate for Physics C? 60% 30%
i) School X ii) School Y iii) not enough info

c) Which group has the higher overall passing rate (combining those taking Physics B and Physics C)?
i) School X ii) School Y iii) cannot be determined from the information given
75% 80%

d) Which conclusion is best supported based only on the data in the table? (see (a) & (b) above)
i) School X has better instruction for both Physics B and C.
ii) School Y has better instruction for both Physics B and C.
iii) The relative quality of the instruction at the two schools depends on which Physics course is being taught.

**Question 5 pertains to the following study on a counseling treatment for Type 2 Diabetes:** (10 pts.)

A study was done to see whether a counseling program could lower the death rate of patients suffering from Type 2 Diabetes. The subjects were 30,000 adults with Type 2 Diabetes who were all members of the same insurance plan. Half were **randomly** assigned to treatment and half to control. In the treatment group, the patients were given free weekly counseling sessions on diet and exercise. The control group was offered the usual health care, which did not include counseling. All patients were followed for 15 years and their death rates from diabetes were compared.

a) Is this study an observational study or a designed experiment? i) Observational Study (ii) Designed Experiment (2 pts)

b) Were there randomized controls? (i) Yes ii) No (2pts)

c) Was there a placebo? i) Yes (ii) No iii) Not enough information given ~~(1 pt.)~~ 2 pts

Even though everyone in the treatment group was encouraged to get counseling, only about 2/3 of the treatment group did. The other 1/3 refused counseling. The people who refused counseling tended to be poorer and less educated than those who accepted.

d) Should the researchers compare the death rate of **everyone** in the treatment group to the controls? Or should they just compare those who **accepted counseling** to the controls? (2 pts)

*Choose one:*

i) They should just compare those who accepted counseling to the controls, since counseling cannot help those who refuse to take it.

(ii) They should compare everyone randomly assigned to treatment to everyone randomly assigned to control, otherwise the treatment group will consist of a richer and better educated population which could confound the results.

iii) They should compare those in the treatment group who accepted counseling to those who refused counseling, since both groups were given the same opportunity for counseling.

sicker ?

e) Suppose Type 2 diabetes (like polio) afflicts the rich more than the poor. If just those who accept counseling (2 pts)
(a richer population) are compared to a control group which includes everyone (rich and poor) would that be a biased comparison?

i) Yes, it would be biased in favor of the counseling, making it appear more effective than it really is. 1 pt
(ii) Yes, it would be biased against the counseling, making it appear less effective than it really is. 2 pts
iii) No, it would not be biased. 0 pts

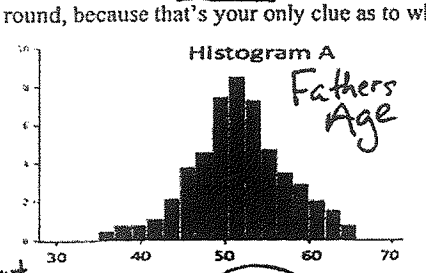
treatment group richer & more hi-fis ic for Diabetes

**Question 6** (8 pts.) Below are 3 histograms. Two of the histograms represent your responses to the Survey 1 questions: "How old is your father?" and "How much cash do you have in your pocket now? and one of the histograms represents Final Exam scores of Stat 100 students from last spring.

a) Which histogram matches which data set? **Fill in the 3 blanks below with A, B, or C.** (3 pts.)

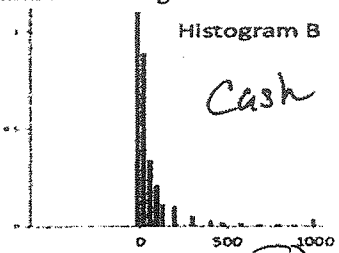
Histogram C depicts Final Exam scores, Histogram A depicts fathers' ages and Histogram B depicts cash.

b) Below each histogram are 2 numbers. One number is the average and the other is the median. For each histogram, circle the number that is the **average.** (Note: In Histogram B and Histogram C, both numbers are whole numbers. In Histogram A, I didn't round, because that's your only clue as to which is the average and which is the median. That's a big HINT! ) (3 pts.)



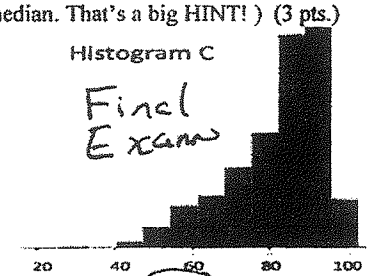

not whole no.

*Circle one:* (i) 51.79 ii) 52




*Circle one:* i) 25 (ii) 70

ave > med




*Circle one:* (i) 82 ii) 86

ave < med

c) For which histogram would it be appropriate to use the Normal Approximation? (2 pts.)

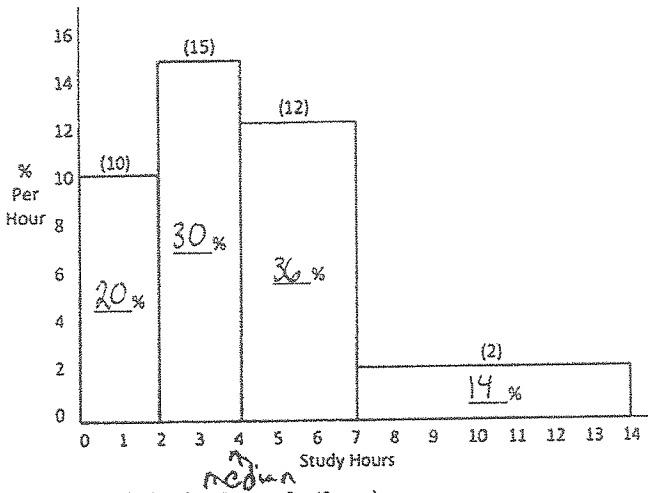
(i) Only A ii) Only B iii) Only C iv) All of them since the normal approximation converts the data to z scores, effectively making the data approximately normal.

PAGE 4 SIDE



**Question 7** (12 points)

The histogram to the right represents how much time students reported for their Sociology 100 midterm exam. The height of each block is given in parentheses. (Assume an equal distribution throughout each interval.)




a) Fill in the percentage for each block in the 4 blanks provided on the histogram. (4 pts.)

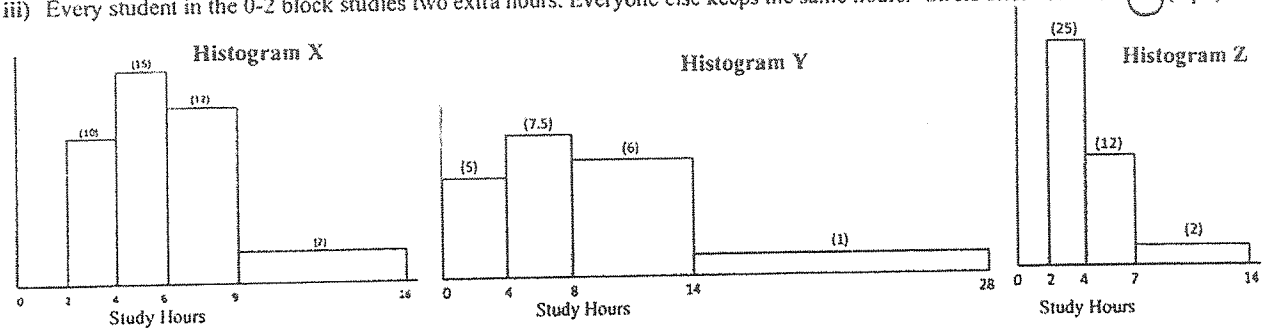
b) The median number of study hours is closest to ..... (2pts.)
i) 3 (ii) 4 iii) 5 iv) 7 v) 10

c) The average is > the median. (2 pts.)
*Fill in the blank with > (greater than), < (less than), or =.*

d) Did more students report studying between 0-4 hours or between 4-14 hours or are they the same? (1 pt.)
i) More between 0-4 ii) More between 4-14 (iii) Same

e) How would the histogram change if students changed their study hours in the following 3 ways? (3 pts.)
***Match the histogram below to the correct change by circling Histogram X, Y or Z.***

i) Every student doubles their study time *Circle one:* X (Y) Z (1 pt.)
ii) Every student studies two extra hours: *Circle one:* (X) Y Z (1 pt.)
iii) Every student in the 0-2 block studies two extra hours. Everyone else keeps the same hours. *Circle one:* X Y (Z) (1 pt.)




**Question 8** (12 points total)

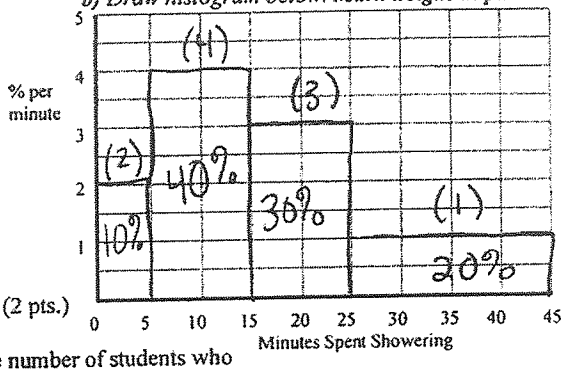
A large group of college students were asked: "How much time, in minutes, is your average shower?" . Don't deduct

*a) Fill in the 5 missing blanks in the distribution table below. (5 pts.)*

| Minutes In Shower | Width of Interval (min) | Area % | Height of Block (% per min) |
|---|---|---|---|
| 0-5 | 5 | 10 | 2 |
| 5-15 | 10 | 40 | 4 |
| 15-25 | 10 | 30 | 3 |
| 25-45 | 20 | 20 | 1 |

100

*b) Draw histogram below. **Mark height in parentheses** (2pts)*




c) What is the median number of minutes spent showering? 15 (2 pts.)
40%

d) The number of students who answered 5-15 minutes is ____ the number of students who answered 25-45 minutes. *Choose one:* i) less than (ii) more than iii) the same as (1 pt)
20%

e) Assuming an equal distribution throughout the interval, the percent of students who reported spending exactly 12 minutes taking a shower is closest to *Choose one:* i) 1% ii) 2% iii) 3% (iv) 4% v) 5% vi) 20% ( 1 pt.)

f) How many minutes corresponds to the 80th percentile? In other words 80% of the students said their average shower is less than 25 minutes long. **(Fill in the blank with a number.)** (1 pt.)

Stat 100 F1 Exam 1 Spring 2014

-3, -1, 1, 7

**Question 9 pertains to the following list of 5 numbers:** -1, 1, -3, 7 $\frac{-1+1+(-3)+7}{4} = 1$ (9 pts total)

*a)* The average is 1, The median is 0. -½ switched signs (4 pts.)

*b)* The deviations from the average are -4, -2, 0, 6, (List them in order from smallest to largest). (2 pts.) accept any order

no continued error

½ point for each deviation (1 pt)

*c)* The sum of the deviations from the average should = 0. *Fill in the blank with a number.*

Compute the Standard Deviation. Round your answer to 2 decimal places. ***Show your work. You may start with the deviations you found in part (b).***

$\frac{16+4+0+36}{4} = \frac{56}{4} = 14$ $\sqrt{14} = 3.74$

-1 for missing a step

continued from b

**Question 10** (12 pts.)

The number of calories consumed by a large group of Super Bowl partiers roughly followed the normal curve with an **average = 2500** calories and a **SD = 500** calories. ****Please round all middle areas given in the table to the nearest whole number.****

a) What percentage of people consumed *more* than **3000** calories on Super Bowl Sunday? (4 pts.)

CC for right hand tail

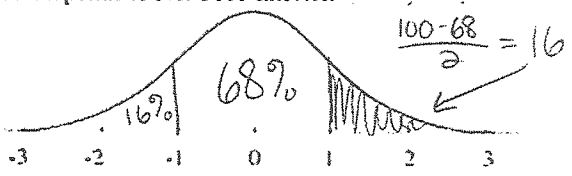
**i) First, convert 3000 calories to a Z-score. Show work for full credit.** (2 pts.)

$\frac{3000-2500}{500} = \frac{500}{500} = 1$

-1 for negative

Z-score = 1 1pt

**ii) Mark your Z-score accurately on the curve and divide the curve into a middle area and two tails. Write the % of the middle area inside the middle area. Shade the area which corresponds to *over* 3000 calories.**




1 pt. for correctly following ALL directions given in (ii) above

16 % of people who consumed over 3000 calories. (1pt)

b) What percent of people consumed between 1900 and 3000 on the day of the Super Bowl? (4 pts.)

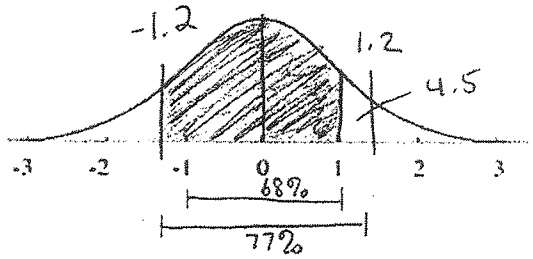
**i) First, convert 1900 calories to a Z-score. (You've already converted 3000 to a Z score above) Show work for full credit.** (2 pts.)

$\frac{1900-2500}{500} = \frac{-600}{500} = -1.2$

Z-score = -1.2

-1 for opposite Z

**ii) Mark both Z-scores on the curve & shade the area that corresponds to consuming between 1900 & 3000 calories.**




1 pt for correctly marking both Z scores and shading area in between shade Z-scores they have

72.5 % of people who consumed between 1900-3000 calories. (1pt)

**Show work for full credit.**

$68 + \frac{77-68}{2}$

$68 + 4.5 =$

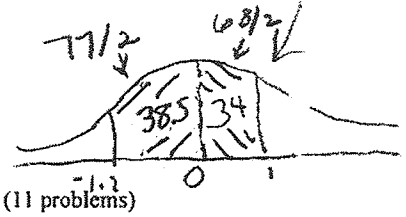
72.5

OR

c) If you're *above* average in consuming calories on Super Bowl Sunday, is your Z score positive or negative? ( 2 pts.)

**Choose one:** (i) positive ii) negative iii) not enough information given

d) If you're exactly at the 50th percentile in calorie consumption then your Z score = 0 & you consumed 2500 calories on Super Bowl Sunday. (Fill in the two blanks with numbers.) (2 pts)



PAGE 6 SIDE -1/2 for careless error

Stat 100 F1 Exam 1 Spring 2014

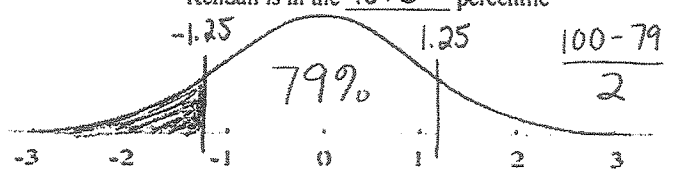
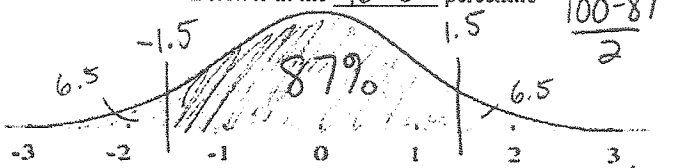
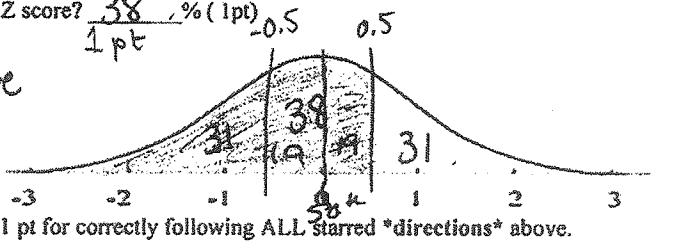
**Question 11 (11 pts.)**

Last semester, we took a survey asking the STAT 100 staff and their close friends how many 'likes' they had on their current Facebook profile picture. Our data happened to follow normal curve quite closely with an average = 32 likes and SD= 8 likes. In the table below, you're either given the number of likes, the Z-score, or the percentile for 4 people. Fill in the remaining two blanks.

***DIRECTIONS* FOR PERCENTILE column**, mark the Z-score on the histogram, divide the curve into a middle area and two tails. Then mark the area in each of the 3 sections and shade the percentile.

*(Round the middle area given in the table to the nearest whole number.)*

ave = 32
SD = 8

| Likes on Profile Pic | Z-Score | Percentile (% of people who had less likes) |
|---|---|---|
| Kendall had 22 likes on her profile picture.<br>$\frac{22-32}{8} = \frac{-10}{8} = -1.25$ | z = -1.25 (1pt) (show work)<br>-1/2 for wrong sign | Kendall is in the 10.5th percentile ← ce from z<br>  <br>$\frac{100-79}{2}$<br>1 pt for correctly following ALL starred *directions* above, middle area 1pt<br>1 pt for correct percentile. 1 pt |
| The number of likes on Derek's profile picture was 44 likes (1pt) (show work) -1/2<br>value = 32 + (1.5)(8) = 44<br>no partial credit | z = 1.5 | Derek is in the 93.5th percentile<br>  <br>$\frac{100-87}{2}$<br>1 pt for correctly following ALL starred *directions* above, middle area<br>1 pt for correct percentile<br>(accept rounded answer) |
| The number of likes on Jayden's profile picture was 36 likes (1pt) (show work) ce.<br>value = 32 + (0.5)(8) = 36 | z = 0.5 (1pt) ce from middle (no work necessary)<br>no credit for negative | Jayden is in the 69th percentile (69% of the people had less likes than her). What middle area should you look up on the normal table to find the correct Z score? 38 % (1pt) 1 pt<br>  <br>1 pt for correctly following ALL starred *directions* above. |
| The number of likes on Abby's profile picture was 28 likes (1pt) (show work) ce<br>value = 32 + (-0.5)(8) = 28 | z = -0.5 (1pt) (no work necessary) | Abby is in the 31th percentile.<br>(Hint: No work is necessary. Just use the histogram above.)<br>If the hint doesn't help, use the space below to draw a new histogram.<br>ce opposite of z from above |