

**GROUND RULES:**

- **Print** your full name clearly at the top of this page. Use the name that appears on university records.
- This is a closed-book and closed-notes exam. You can not use external notes of any kind.
- You may use a calculator. You may not use your phone as a calculator.
- Moore and Notz's Table B is provided on the last page of this exam.
- This exam consists of 30 multiple choice questions.
  - Each question is worth 2 points.

This exam is worth **60 points**.

- Any discussion or inappropriate communication between you and another examinee, as well as the appearance of any unnecessary material, is not allowed. All violations will be reported to the Student Conduct and Academic Integrity Office immediately.
- You have **50 minutes** to complete this exam.

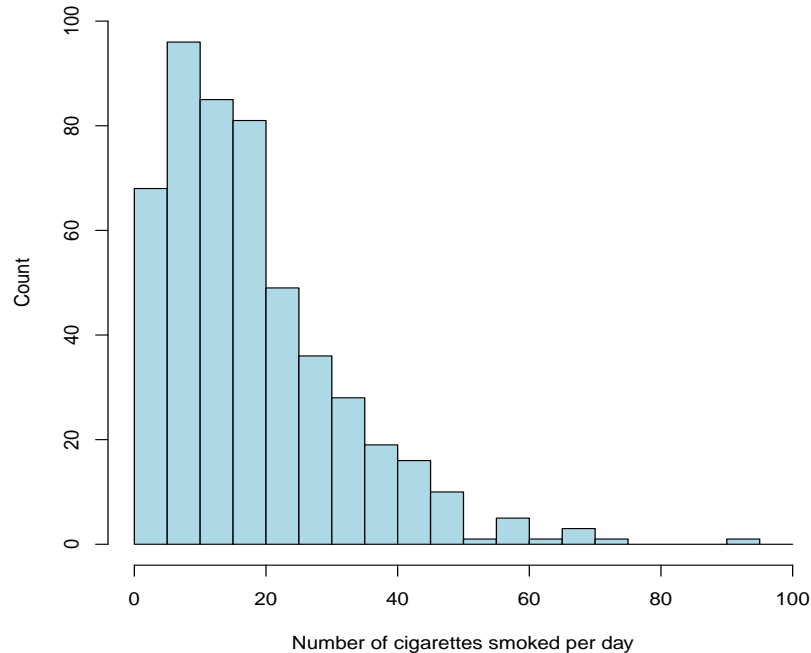
**HONOR PLEDGE FOR THIS EXAM:**

After you have finished the exam, please read the following statement and sign your name below it.

*I promise that I did not discuss any aspect of this exam with anyone other than the instructor, that I neither gave nor received any unauthorized assistance on this exam, and that the work presented herein is entirely my own.*

**MULTIPLE CHOICE.** Circle the best answer. Make sure your answer is clearly marked. Ambiguous responses will be marked wrong.

1. A histogram of 500 observations is below:



How do the mean and median compare for these 500 observations?

- (a) The mean and median are the same.
- (b) The mean is smaller than the median.
- (c) The mean is larger than the median.

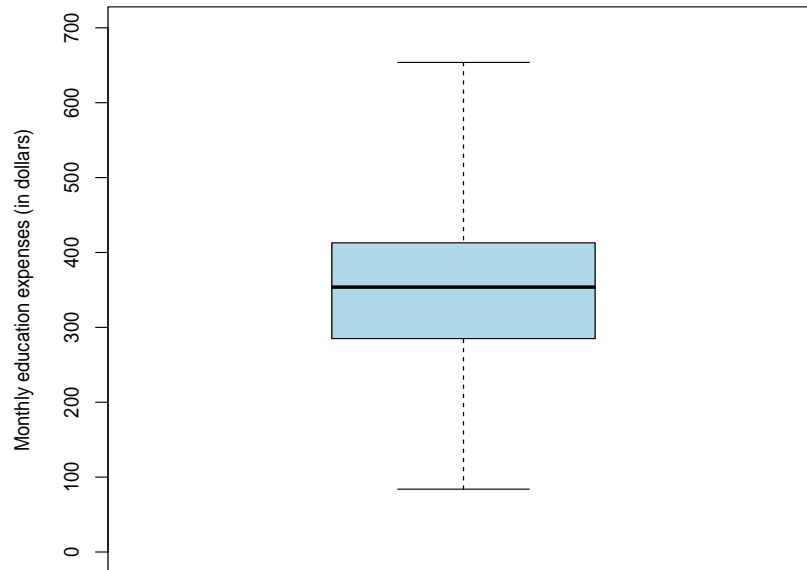
2. True or False. A normal population density curve is symmetric about its mean.

- (a) True
- (b) False

3. Among college-bound high-school seniors, SAT scores are normally distributed with mean 1060 and standard deviation 200. What percentage of these students will score **above** 1100? Round to the nearest percent.

- (a) 20%
- (b) 42%
- (c) 58%
- (d) 80%

4. An intern in the Office of Student Advocacy observes a sample of 100 USC freshman and sophomore students and records their education expenses last month. He prepares a boxplot of the observations:



The **interquartile range** (IQR) is closest to which value?

- (a) 60
- (b) 120
- (c) 350
- (d) 550

5. Which statement regarding the correlation  $r$  does **not** contain a mistake?

- (a) “We found a negative linear relationship between a car’s weight and the year it was manufactured ( $r = -0.34$ ).”
- (b) “The correlation between a car’s weight and a car’s horsepower ( $r = 0.51$ ) shows as one variable increases, the other variable tends to decrease.”
- (c) “There is a high correlation ( $r = 0.92$ ) between a car’s manufacturer and the gas mileage of the car.”
- (d) “We found a very strong correlation ( $r = 1.09$ ) between the horsepower of a car and the gas mileage of the car.”

6. A new study found that people with colorectal cancer who drink three to four cups of coffee a day can reduce their risk of disease recurrence. Furthermore, people who drink more coffee had fewer incidents of the cancer returning. The study's lead investigator commented,

*"It's intriguing this study suggests drinking coffee may reduce the recurrence of colorectal cancer. However, our study has revealed an association between colorectal cancer and coffee consumption, and not a direct effect."*

What is the investigator trying to convey in this passage?

- (a) Direct evidence should be summarized mathematically using regression equations.
- (b) Statistical methods for quantitative variables can be applied to categorical variables, but the analyses should be interpreted with caution.
- (c) Correlation does not imply causation.
- (d) Statistically significant results should not be disseminated until the studies are published.

7. True or False. In a scatterplot of observations, if the correlation  $r$  is positive ( $r > 0$ ), then the slope of the least-squares regression line  $b$  must also be positive.

- (a) True
- (b) False

8. As part of a waste removal project, a new compression machine for processing sewage sludge was studied. Engineers were interested in the two variables:

$$\begin{aligned}x &= \text{machine filtration rate (measured in kg/m/hr)} \\y &= \text{moisture of compressed pellets (measured as a \%)}.\end{aligned}$$

They collected observations from a sample of  $n = 20$  specimens and calculated the correlation to be  $r = 0.89$ . In this example,  $r$  is measured in which units?

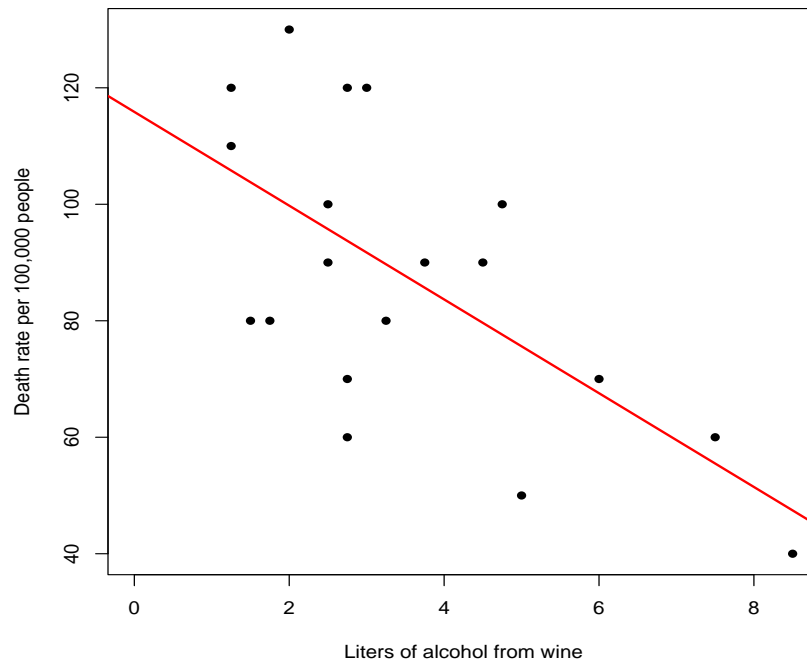
- (a) kg/m/hr
- (b) moisture percentage (out of 100%)
- (c) kg/m/hr  $\times$  moisture percentage (out of 100%)
- (d) The correlation does not have units attached to it.

9. In the homework, you examined the relationship between

$x$  = amount of alcohol consumed from wine per person (in liters)

$y$  = death rate (per 100,000 people)

for 19 developed countries. A scatterplot of these observations is below.



After rounding, the least-squares regression line (shown above) is

$$y = 116 - 8x.$$

What death rate (per 100,000 people) would you predict for country whose alcohol consumption from wine was 4.5 liters per person?

- (a) 80 deaths
- (b) 92 deaths
- (c) 108 deaths
- (d) 152 deaths

10. Resting pulse rates for healthy adults follow a normal distribution with mean 70 beats per minute and standard deviation 10 beats per minute. My resting pulse rate is 80. Which statement is correct?

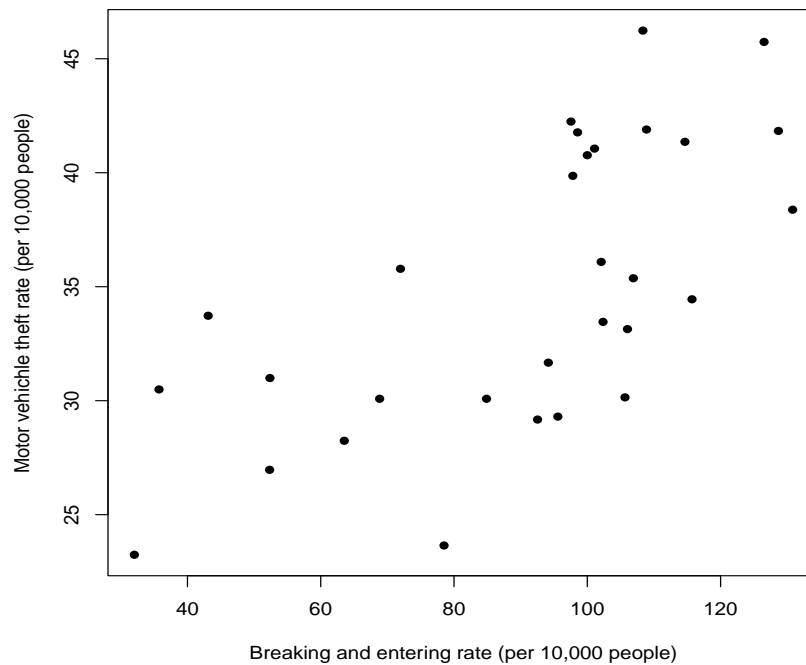
- (a) Approximately 2.5% of all healthy adults have a resting pulse higher than mine.
- (b) The standard score of my pulse rate is  $-2$ .
- (c) My pulse rate is higher than the pulse rate of 68% of all healthy adults.
- (d) None of the above.

11. The South Carolina State Law Enforcement Division publishes crime statistics in South Carolina. A recent publication shows observations of

$x$  = breaking and entering rate (per 10,000 residents)

$y$  = motor vehicle theft rate (per 10,000 residents)

recorded each year from 1993-2022. Here is a scatterplot of the yearly observations:



The correlation  $r$  for these data is closest to

- (a)  $-0.26$
- (b)  $-0.02$
- (c)  $0.65$
- (d)  $0.99$

12. In a scatterplot of data points, we defined the “best-fit line” for regression as the one that makes the

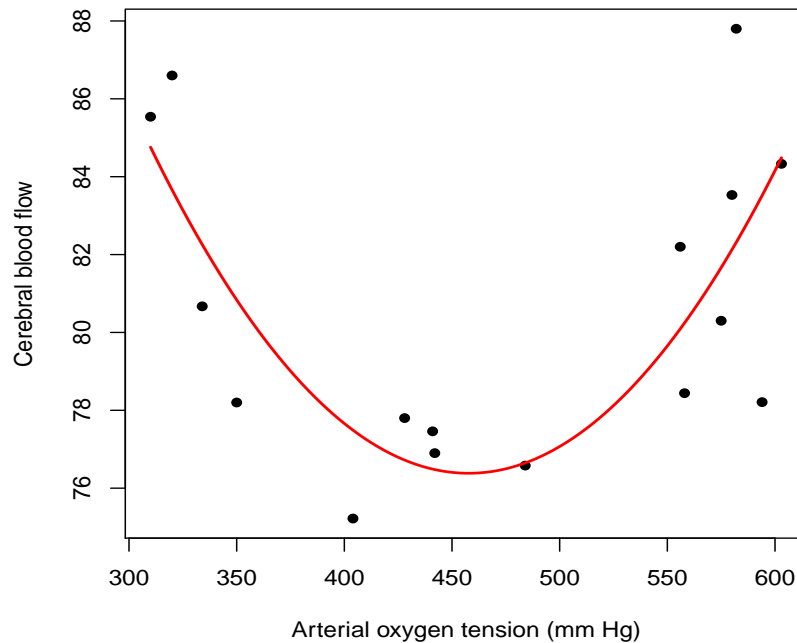
- (a) sum of the vertical distances of the data points from the line as small as possible.
- (b) sum of the vertical distances of the data points from the line as large as possible.
- (c) sum of the squared vertical distances of the data points from the line as small as possible.
- (d) sum of the squared vertical distances of the data points from the line as large as possible.

13. A study was conducted to determine if there was a relationship between

$x$  = arterial oxygen tension

$y$  = cerebral blood flow.

Sixteen adults participated in the study. A scatterplot of the data is below. A quadratic curve has been added.



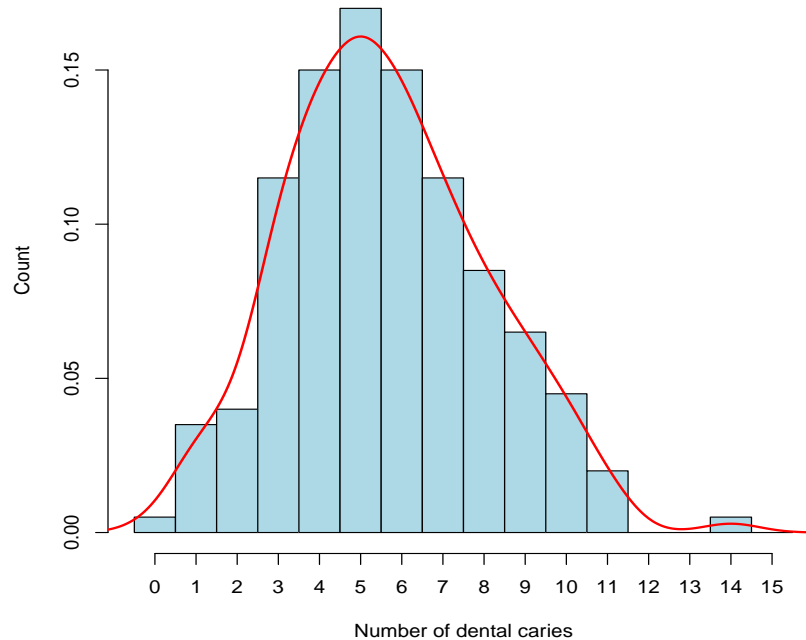
The correlation is  $r = 0.07$ . Your friend says, “*these variables are not related because the correlation is so close to zero.*” How should you respond?

- (a) “You have to square the correlation to see how strongly the variables are related.”
- (b) “The correlation only measures straight-line relationships.”
- (c) “Seven percent of the variation being explained by arterial oxygen tension is not an insignificant amount.”
- (d) “If you switch the variables, then the correlation will increase to  $r = 0.93$ .”

14. True or False. The correlation can only be used in randomized comparative experiments where there is a well-defined response variable ( $y$ ) and an explanatory variable ( $x$ ) that was measured in the experiment.

- (a) True
- (b) False

15. A researcher selects a sample of 200 South Carolina adults (aged 18 and older) who have not received dental care in the past 10 years. The researcher records the number of dental caries for each adult in the sample. A histogram of the 200 observations is shown below. An estimate of the population density curve has been added.



What group of individuals does the population density curve describe?

- (a) the 200 adults in the sample
- (b) those adults in the sample who have one or more dental caries (excluding the outlier)
- (c) all South Carolina adults who have not received dental care in the past 10 years
- (d) all South Carolina adults

16. A high school principal employs five teachers with salaries listed below:

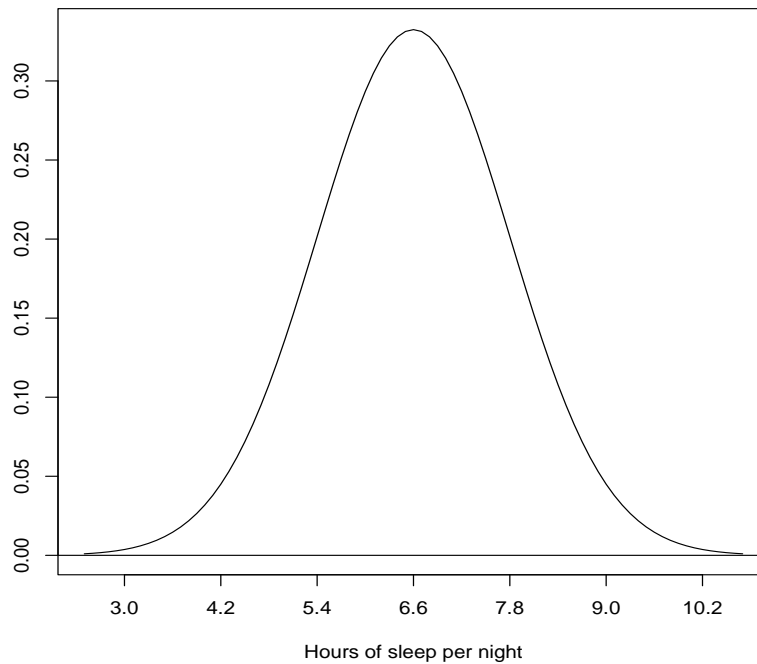
\$42000   \$45000   \$48000   \$51000   \$54000

If she decides to give each teacher a \$3000 increase in salary, then

- (a) the mean will increase by \$3000, and the standard deviation will increase by  $(\$3000)^2$ .
- (b) the standard deviation will increase by \$3000, but the mean will remain the same.
- (c) the mean and standard deviation will both increase by \$3000.
- (d) the mean will increase by \$3000, but the standard deviation will remain the same.



17. The distribution of hours of sleep per school night, among high school seniors, is normally distributed with mean 6.6 hours and standard deviation 1.2 hours. This distribution is shown below.



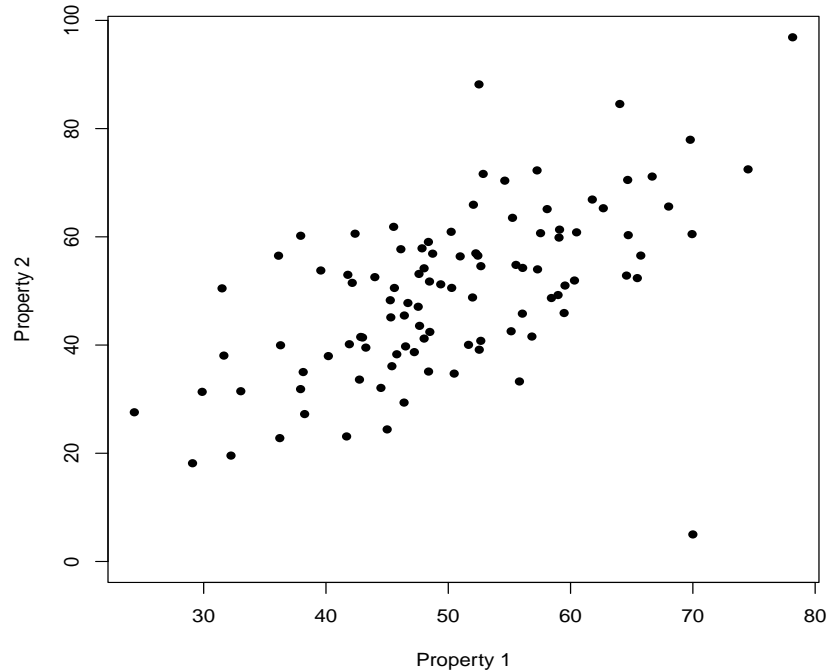
A student in this population slept 8.1 hours last night. What is his **standard score**?

- (a) 0.9032
- (b) 0.0968
- (c) 1.25
- (d)  $-1.25$

18. If the observations in a scatterplot are very tightly clustered around a straight line, then the correlation must be

- (a) close to  $+1$
- (b) close to either  $+1$  or  $-1$
- (c) close to  $0$
- (d) close to  $-1$

19. Lab chemists measure two quantitative properties of automotive paint. A total of  $n = 100$  paint specimens are measured. A scatterplot of these observations is below:



The observation where Property 1 = 70 (in the lower right) is an outlier. If we removed this outlier, then

- (a) the correlation  $r$  would be close to 0.
- (b) the correlation  $r$  would increase.
- (c) the correlation  $r$  would remain the same.
- (d) the correlation  $r$  would be close to  $-1$ .

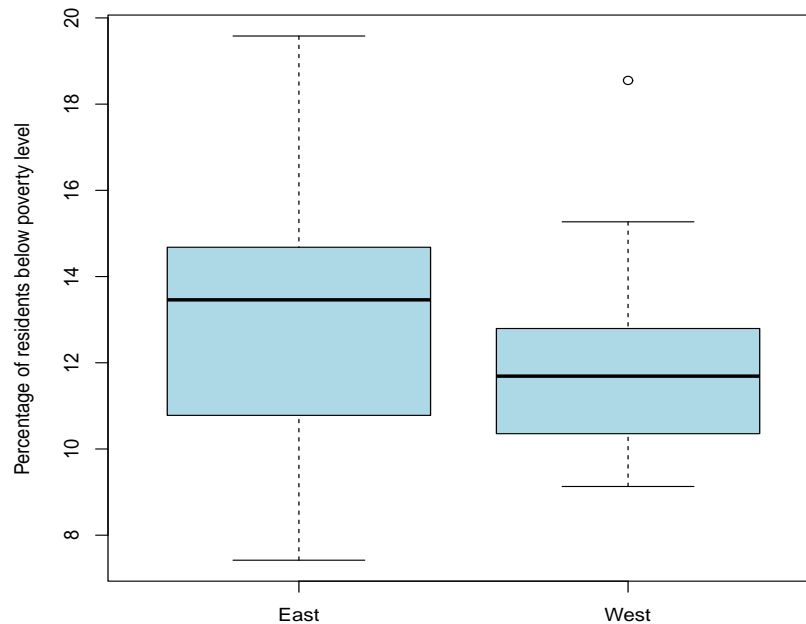
20. Here are observations of the phosphate level in the blood (measured in mg/dl) for six patients:

5.6 5.2 4.6 4.9 5.7 6.4

The mean is  $\bar{x} = 5.4$  and the standard deviation is  $s = 0.64$ . How do we interpret the **standard deviation**?

- (a) It measures the cutoff where we would classify an observation to be an outlier on the low side.
- (b) It measures the observations' average distance from the mean.
- (c) It measures the range where the middle 50% of the observations fall.
- (d) It measures the difference between the mean and median.

21. The US Census Bureau records the percentage of residents living below the poverty level in each of the 50 states. The plot below gives side-by-side boxplots of these percentages for states east of the Mississippi River and the states west of it.



The outlier among the western states is New Mexico (18.55%). If we removed this outlier, how would the regions compare in terms of their means and standard deviations?

- (a) East would have a smaller mean. East would have a smaller standard deviation.
- (b) East would have a smaller mean. West would have a smaller standard deviation.
- (c) West would have a smaller mean. East would have a smaller standard deviation.
- (d) West would have a smaller mean. West would have a smaller standard deviation.

22. In class, we watched a short video featuring Hans Rosling, who was an internationally known expert in health statistics. The highlight of the video was his use of visual displays to show data in multiple dimensions. What was he was talking about?

- (a) life expectancy and income for every country in the world
- (b) the long-term impact of American Red Cross intervention in different regions in the Middle East and Africa
- (c) different measurements recorded for new breast cancer treatments
- (d) increased radon and arsenic levels in drinking water in developing nations and their effects on tumor development

23. Physicians measured the concentration of calcium (in nM) in blood samples from 15 healthy patients. Here are the data:

95 112 122 88 66 104 90 110 100 122 126 102 122 96 135

I used R to calculate the five-number summary:

```
> quantile(calcium,type=2)
 0%  25%  50%  75% 100%
 66   95  104  122  135
```

How many **outliers** are in the data set? Use the 1.5(IQR) rule of thumb we talked about in class.

- (a) 0
- (b) 1
- (c) 2
- (d) 3

24. Consumer Reports magazine reported data on the number calories in a hot dog for each of 17 brands of meat hot dogs. Here is the stemplot of these data:

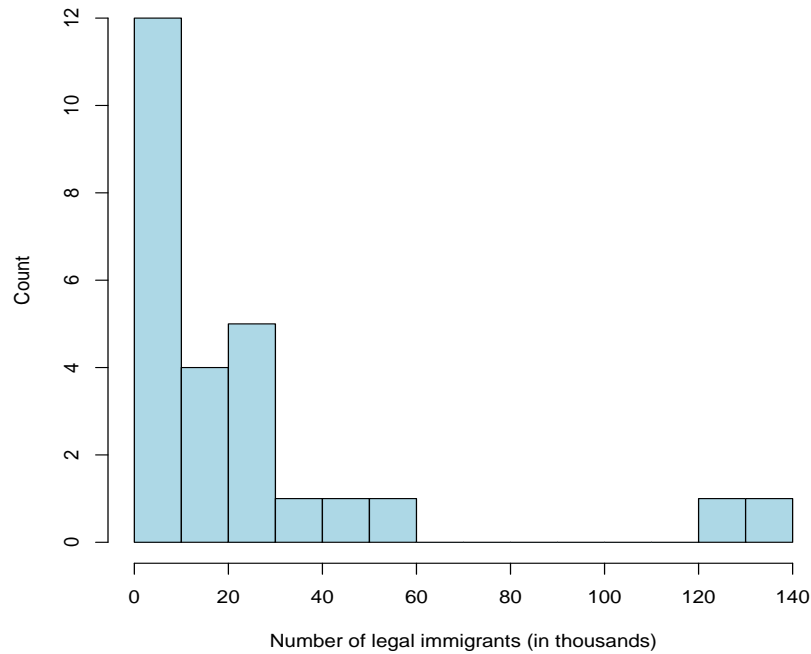
```
> stem(calories,scale=2)

 10 | 7
 11 |
 12 |
 13 | 5689
 14 | 067
 15 | 3
 16 |
 17 | 2359
 18 | 2
 19 | 015
```

What is the **median** of this data set?

- (a) 138.5
- (b) 153
- (c) 158.7
- (d) 182

25. The histogram below shows the number of legal immigrants who settled in each state east of the Mississippi River in 2017.



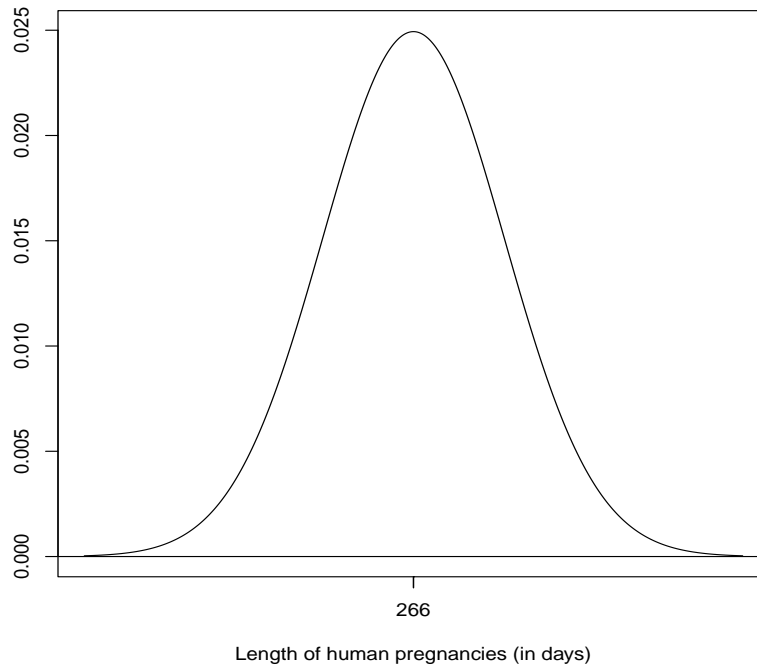
What is the best description of the **shape** of this distribution?

- (a) symmetric
- (b) skewed right
- (c) skewed left
- (d) uniform

26. A standard normal distribution is a normal distribution with mean 0 and standard deviation 1. This is the distribution shown in Table B. Which value is closest to the **third quartile**  $Q_3$  of this distribution?

- (a)  $-0.7$
- (b)  $0.7$
- (c)  $1.7$
- (d)  $2.7$

27. The length of human pregnancies from conception to birth has a normal distribution with mean  $\mu = 266$  days and standard deviation  $\sigma = 16$  days. This distribution is shown below.



**Ninety-five percent (95%)** of all human pregnancies will be between which two values?

- (a) 250 and 282 days
- (b) 234 and 298 days
- (c) 218 and 314 days
- (d) 202 and 330 days

28. The US Census Bureau Current Population Survey reports the median household income for all US households is \$74,580. This number is a

- (a) parameter
- (b) statistic

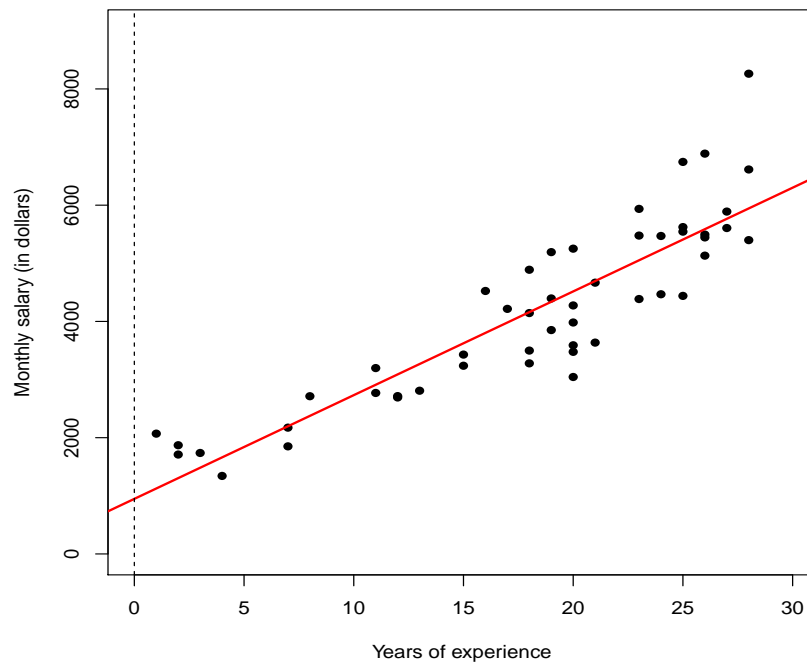
Use the information on this page to answer Questions 29-30 (next page).

A researcher in the College of Social Work at USC reports on salaries for social workers in South Carolina. With a sample of  $n = 50$  social workers, she wants to describe the relationship between

$x$  = years of experience

$y$  = monthly salary (in dollars).

Here is a scatterplot of the data with the least-squares regression line superimposed. A vertical line at  $x = 0$  has been added.



I used R to calculate the correlation  $r$ :

```
> cor(years.of.experience,salary)
[1] 0.89
```

I used R to calculate the least-squares regression line:

```
> fit = lm(salary~years.of.experience)
> fit
Coefficients:
  (Intercept)  years.of.experience
        947.4             178.4
```

The equation of the least-squares regression line is

$$y = 947.4 + 178.4x.$$

29. In this study, the square of the correlation is

$$r^2 = (0.89)^2 \approx 0.79 \text{ (or 79\%).}$$

What does this percentage tell us?

- (a) the percentage of observations that are within one standard deviation of the regression line
- (b) the percentage change between the slope and the intercept
- (c) the percentage of variability in the monthly salary data that is explained by the straight-line regression with years of experience
- (d) the percentage of predictions that will be correct

30. In this study, what would be an example of **extrapolation**?

- (a) predicting the monthly salary for a social worker with 5 years of experience
- (b) predicting the monthly salary for a social worker with 15 years of experience
- (c) predicting the monthly salary for a social worker with 25 years of experience
- (d) predicting the monthly salary for a social worker with 35 years of experience



642 Table B

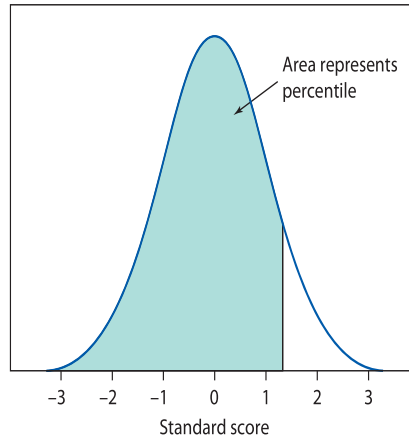


Table B Percentiles of the Normal distributions

Standard score → Percentile	Standard score → Percentile	Standard score → Percentile			
-3.4	0.03	-1.1	13.57	1.2	88.49
-3.3	0.05	-1.0	15.87	1.3	90.32
-3.2	0.07	-0.9	18.41	1.4	91.92
-3.1	0.10	-0.8	21.19	1.5	93.32
-3.0	0.13	-0.7	24.20	1.6	94.52
-2.9	0.19	-0.6	27.42	1.7	95.54
-2.8	0.26	-0.5	30.85	1.8	96.41
-2.7	0.35	-0.4	34.46	1.9	97.13
-2.6	0.47	-0.3	38.21	2.0	97.73
-2.5	0.62	-0.2	42.07	2.1	98.21
-2.4	0.82	-0.1	46.02	2.2	98.61
-2.3	1.07	0.0	50.00	2.3	98.93
-2.2	1.39	0.1	53.98	2.4	99.18
-2.1	1.79	0.2	57.93	2.5	99.38
-2.0	2.27	0.3	61.79	2.6	99.53
-1.9	2.87	0.4	65.54	2.7	99.65
-1.8	3.59	0.5	69.15	2.8	99.74
-1.7	4.46	0.6	72.58	2.9	99.81
-1.6	5.48	0.7	75.80	3.0	99.87
-1.5	6.68	0.8	78.81	3.1	99.90
-1.4	8.08	0.9	81.59	3.2	99.93
-1.3	9.68	1.0	84.13	3.3	99.95
-1.2	11.51	1.1	86.43	3.4	99.97