

Take your time and make sure you follow all instructions. Where necessary, work must be shown in order to receive partial credit.

1. There are 20,000 households in a city that spend an average of \$250 a month on groceries. The local Chamber randomly contacts 100 households and finds they spend on average \$226 a month on groceries. [2 pts each]

- A. Parameter                      B. Population                      C. Sample                      D. Statistic
- C a. The 100 households is an example of a what.
- A b. The \$250 is an example of a what.
- B c. The 20,000 households is an example of a what.
- D d. The \$226 is an example of a what.

2. The state Retail Federation is interested in learning the spending habits for households across the state. Which sampling method is described in each of the following? [2 pts each]

- A. Census                      B. Simple Random Sample (SRS)                      C. Stratified Random Sample  
D. Cluster Sample                      E. Multistage Sample                      F. Systematic Sample  
G. Voluntary Response Sample                      H. Convenience Sample
- H a. Contact 100 residents of different households as they enter a local arena.
- F b. Contact every 35<sup>th</sup> household from a list of households in the state.
- C c. Randomly select 50 households each from those living in Rural, Suburban, or Urban areas.
- A d. Contact every household in the state.
- E e. Randomly select 10 counties in the state. Then randomly select 10 streets in the selected counties. Within the selected streets randomly contact 5 households.
- D f. Randomly select 5 counties in the state and contact every household in the selected counties.
- B g. Randomly select 200 households in the state.
- G h. Place an ad on the state's website asking households to respond to a survey.

3. Match each of the following.

[2 pts each]

A. Experiment

B. Prospective Study

C. Retrospective Study

- A a. A researcher testing a new drug randomly assigns 50 allergy sufferers to a group receiving a new drug and a different 50 to a group receiving a placebo.
- C b. A researcher studying blood pressure studies the exercise habits from the last year of 100 randomly selected patients.
- B c. A researcher identifies 100 e-cigarette users and follows them over the next 5 years to see how rates of lung disease compare to traditional cigarette smokers.

4. The price of a vehicle (in \$) and the age (in years) are linear associated. Computer software gives the regression line,  $\widehat{\text{price}} = \$35000 - \$4500 \text{ age}$  and  $R^2 = 0.72$ . Use the information to answer the following questions. [4 pts each]

- D a. Which is the best interpretation of the y-intercept.  
 A) The price is \$4500 when the age is 0 years.  
 B) The price is \$4500 when the age is 10 years.  
 C) The price is \$35000 when the age is 10 years.  
 D) The price is \$35000 when the age is 0 years.
- C (B) b. Which is the best interpretation of the slope of the regression line.  
 A) For every decrease in 1 year in age there is a decrease of \$4500 in price.  
 B) For every decrease in 1 year in age there is an increase of \$4500 in price.  
 C) For every increase in 1 year in age there is a decrease of \$4500 in price.  
 D) For every increase in 1 year in age there is an increase of \$4500 in price.
- B c. Which is the best interpretation of the value of  $R^2$ .  
 A) The line describes 28% of the variation between the vehicle's age and price.  
 B) The line describes 72% of the variation between the vehicle's age and price.  
 C) The vehicle's age describes 28% of the variation between the line and price.  
 D) The vehicle's age describes 72% of the variation between the line and price.
- d. What is the correlation for the linear association.

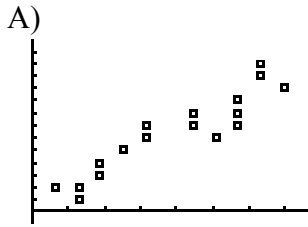
```

2√(2)-2√(3)
-.6356744904
Ans/4
-.1589186226
-√(.72)
-.8485281374

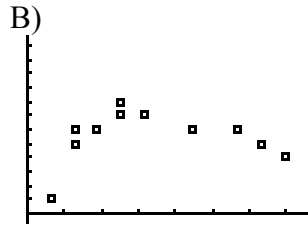
```

$$r = -\sqrt{.72} = -.849$$

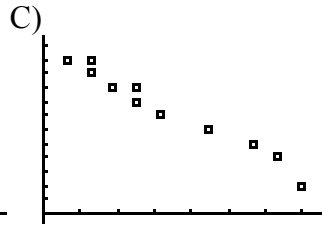
5. Match the graphs with the one of the following correlations:  $-0.983$ ,  $0.068$ ,  $0.824$ , and  $0.927$  [12 pts]



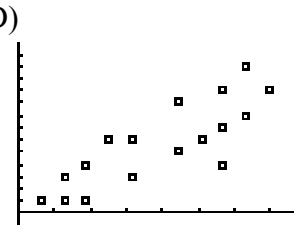
$.927$



$.068$



$-.983$



$.824$

6. Use the graphs in #5 to find the best answer to the following questions.

[3 pts each]

a) Which of the graphs has a moderately strong linear association?

D

b) Which of the graphs has negative association?

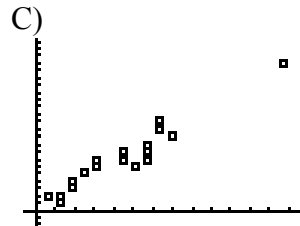
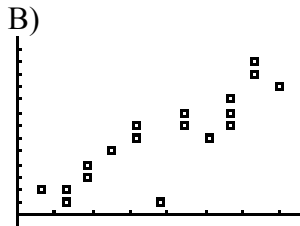
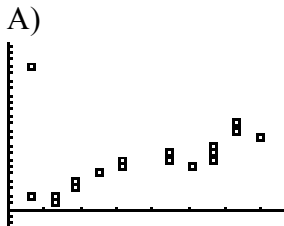
C

c) Which of the graphs has a strong association that has no direction?

B

7. Use the graphs below to find the best answer to the following questions.

[3 pts each]



a) Which of the graphs has a stray point that is the most influential?

A

b) Which of the graphs has a stray point that has high leverage but low residual?

C

c) Which of the graphs has a stray point that has low leverage but high residual?

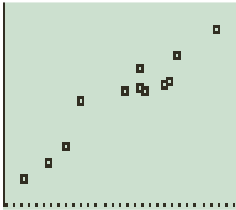
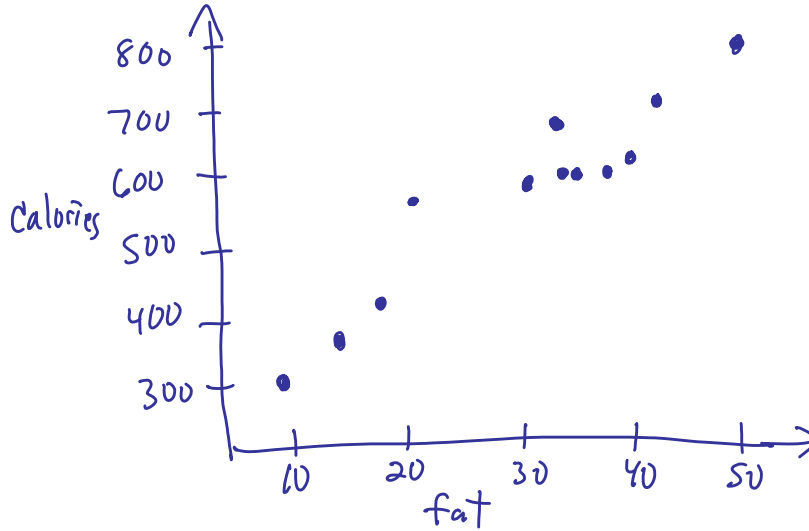
B

8. The following table lists the amount of fat (in grams) and calories for different foods at a restaurant.

[29 points]

a) Sketch a scatterplot (Label your axis!)

Fat (x)	Calories (y)
10	300
15	350
19	410
22	550
31	580
34	590
34	650
35	580
39	600
40	610
42	700
50	780



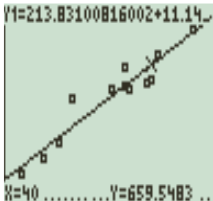
b) Determine the correlation and the equation of the least squares regression line ( $\hat{y} = a + bx$ ) using your calculator.

```
LinReg
y=a+bx
a=213.8310082
b=11.14293235
r^2=.909218605
r=.9535295512
```

$$r = .954$$

$$\hat{y} = 213.83 + 11.14x$$

c) Use the linear model to determine the residual for a food item with 40 g fat.



$$x = 40 \Rightarrow \hat{y} = 213.83 + 11.14(40) = 659.43$$

$$\text{Residual} = y - \hat{y} = 610 - 659.43 = -49.43$$

d) What is wrong with using the regression line to estimate the calories for a food item that has 0 g fat?

Extrapolation is being used and linear model may not be appropriate for values outside data range.

$$x = 0 \Rightarrow \hat{y} = 213.83$$