Problem 1

In each of the following situations, which is the explanatory variable and which is the response variable? Are they categorical or quantitative (quantitative means "numerical")?

1. The typical number of calories a person consumes per day and that person's percent of body fat.

a) Number of calories consumed per day: response, quantitative. Percent of body fat: explanatory, quantitative.
b) Number of calories consumed per day: explanatory, quantitative. Percent of body fat: response, quantitative.
c) Number of calories consumed per day: response, quantitative. Percent of body fat: explanatory, categorical.
d) Number of calories consumed per day: explanatory, categorical. Percent of body fat: response, categorical.

2. Water temperature controlled at different levels and growth (measured by weight) of corals in aquariums.

a. Water temperature: response, quantitative. Growth: explanatory, categorical.
b. Water temperature: explanatory, categorical. Growth: response, categorical.
c. Water temperature: response, categorical. Growth: explanatory, quantitative.
d. Water temperature: explanatory, quantitative. Growth: response, quantitative

Problem 2

Coffee is a leading export from several developing countries. When coffee prices are high, farmers often clear forest to plant more coffee trees. Here are data on prices paid to coffee growers in Indonesia and the rate of deforestation in a national park that lies in a coffee-producing region, for five years:

Price (cents per pound)	Deforestation (percent)
29	0.49
40	1.59
54	1.69
55	1.82
72	3.10

1 Duke, Çetinkaya-Rundel

3. Coffee is currently priced in dollars. If it were priced in euros, and the dollar prices in the above table were translated into the equivalent prices in euros, would the correlation between coffee price and percent deforestation change?
   a) The correlation would remain zero, because the two variables are independent
   b) Yes, units affect correlation
   c) No, units do not affect correlation
   d) It is impossible to calculate the correlation, because coffee price is categorical.

Problem 3

A study shows that there is a positive correlation between the size of a hospital (measured by its number of beds (x)) and the median number of days (y) that patients remain in the hospital.

4. What lurking variable could be present in this study?

   a) cost: its more expensive to run larger hospitals.
   b) severity of disease: since large hospitals have better facilities and more doctors to cope with severe illness.
   c) number of visitors: since larger hospitals receive more visitors.
   d) facilities: since larger hospitals have better facilities, patients choose to stay longer

Problem 4

Milk use is positively correlated to cancer rates. While this is not a popular finding within the milk industry, there is a moderately positive correlation with drinking milk and getting cancer (Paulos, 1990). Milk consumption is greater in wealthier countries. In wealthier countries people live longer. Greater longevity means people live long enough to eventually get some type of cancer.

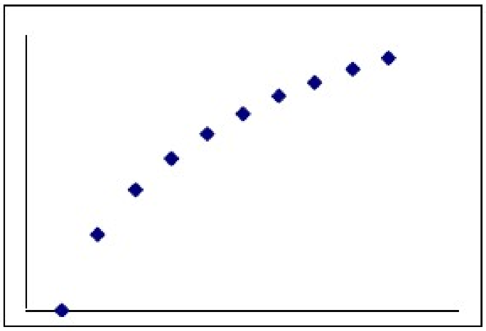
5. Which is the response and explanatory variable? **Response – cancer. Explanatory- milk consumption**
6. Which is a lurking variable? **Longevity**
7. Will you conclude that drinking more milk increases the chance of getting cancer? Explain your reasoning. **No because it is not a randomized experiment.**

Problem 5

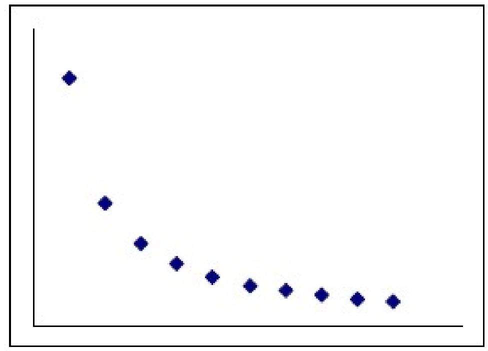
1 Duke, Çetinkaya-Rundel

8. Which of the following plots will have a correlation coefficient of .85? D

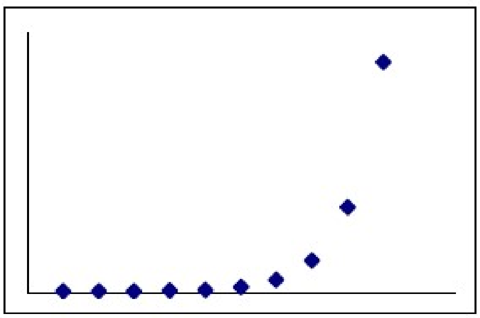
A.



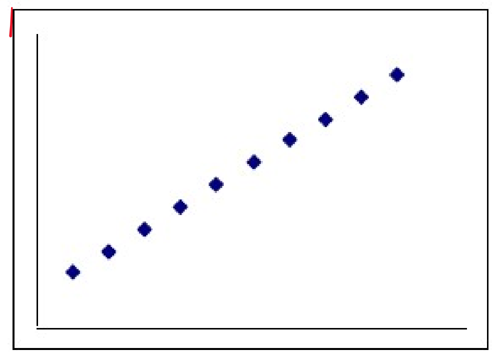
B.



C.



D.



1 Define A = event exactly 9 people say yes.

1. The complement of A is the event that:
   a) anything other than 9 people say yes
   b) anything other than 9 people say no
   c) exactly 6 people say yes
   d) exactly 6 people say no
   e) more than 6 people say no

Problem 2 At a large university, the probability that a student takes calculus and statistics in the same semester is 0.0125. The probability that a student takes statistics is 0.125; the probability that a student takes calculus is 0.3

2. Find the probability that a student is taking calculus, given that he or she is taking statistics.
a) 0.1
b) 0.1125
c) 0.0016
d) 0.1375
e) 0.4800

1 Duke, Çetinkaya-Rundel

3. Is the event of taking calculus independent of the event of taking statistics? Justify your answer numerically.

Using the answer in question two, P(C|S)=.1 and it is not equal to P(C)=.3. Therefore the first condition for independence is not satisfied.

Problem 3

11) If you flip a coin three times, the possible outcomes are HHH HHT HTH HTT THH THT TTH TTT. What is the probability of getting at least one head?

A) $\frac{1}{2}$ B) $\frac{7}{8}$ C) $\frac{1}{4}$ D) $\frac{3}{4}$

12) When a quarter is tossed four times, 16 outcomes are possible.

HHHH	HHHT	HHTH	HHTT
HTHH	HTHT	HTTH	HTTT
THHH	THHT	THTH	THTT
TTHH	TTHT	TTTH	TTTT

Here, for example, HTTH represents the outcome that the first toss is heads, the next two tosses are tails, and the fourth toss is heads. The events A and B are defined as follows.

Event A = the first two tosses are heads
Event B = the first and last tosses are the same

Are the events A and B mutually exclusive?

A) Yes B) No

A and B are not mutually exclusive because there are outcomes that are both in A and B like {HHHH,HHTH}.

Problem 4

The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English (foreign language) at home, and 4.2% fall into both categories.

Consider P(Poverty)=0.146; P(Foreign)=0.207 and P(Poverty ∩ Foreign) = 0.042

4. Are living below the poverty line and speaking a foreign language at home disjoint?
   No, because there are Americans that live below poverty and speak a foreign language.

5. What percent of Americans live below the poverty line and only speak English at home?
   0.146 – 0.042= 0.104 (Hint – use a venn diagram)

1 Duke, Çetinkaya-Rundel

6. What percent of Americans live below the poverty line or speak a foreign language at home?
   P(Poverty ∪ Foreign) = P(Poverty) + P(Foreign) - P(Poverty ∩ Foreign) = 0.146 + 0.207 – 0.042 = 0.311

7. What percent of Americans live above the poverty line and only speak English at home

   0.207-0.042

8. Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

   They are not independent because

   P(Poverty ∩ Foreign) = 0.042 ≠ P(poverty) × P(foreign) = 0.146 x 0.207 = 0.0302

Problem 5

Consider P(A) = 0.3, P(B) = 0.7

9. Can you compute P(A and B) if you only know P(A) and P(B)?
   If A and B are independent then $P(A \cap B) = P(A) \times P(B)$

10. Assuming that events A and B arise from independent random processes,

    a. What is P(A and B)?

    $P(A \cap B) = 0.3 \times 0.7$=.21

    b. What is P(A or B)?

    $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.7 -$0.21 = 0.79

    c. What is P(A|B)?

    $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.21}{0.7} = 0.3$$

11. If we are given that P(A and B) = 0.1, are the random variables giving rise to events A and B independent?

    A and B are not independent because $P(A \cap B) = 0.1 \neq P(A) \times P(B) = 0.21$

12. If we are given that P(A and B) = 0.1, what is P(A|B)?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.7} = 0.1428$$

Problem 6

1 Duke, Çetinkaya-Rundel

A quiz consists of 3 multiple choice questions with 4 possible answers each. If a student guesses the answer to each question, what is the probability that they get all the answers correct?

$$P(C_1 \; and \; C_2 \; and \; C_3) = P(C_1) \times \; P(C_2) \; \times P(C_3) = 0.25 \; \times 0.25 \; \times 0.25 \; \times 0.25$$

Problem 7[1]

Topics: contingency table, conditional probability, joint probability, marginal probability.

The family college data set contains a sample of 792 cases with two variables, teen and parents, and is summarized in the following Table:

		parents		
		degree	not	Total
teen	college	231	214	445
	not	49	298	347
	Total	280	512	792

The teen variable is either college or not, where the college label means the teen went to college immediately after high school. The parent's variable takes the value degree if at least one parent of the teenager completed a college degree.

13. If at least one parent of a teenager completed a college degree, what is the chance the teenager attended college right after high school?

P(C|D) = 231/280

14. Probability of a random teenager from the study went to college rightafter highschool:

P(C)= 445/792

Problem 8[1]

Topics: general multiplication rule

A smallpox data set provides a sample of 6,224 individuals from the year 1721 who were exposed to smallpox in Boston. Doctors at the time believed that inoculation, which involves exposing a person to the disease in a controlled form, could reduce the likelihood of death.

Suppose we are given only two pieces of information: 96.08% of residents were not inoculated, and 85.88% of the residents who were not inoculated ended up surviving.

15. How could we compute the probability that a resident was not inoculated and lived?

P (Not inoculated and survived) = P (survived | not inoculated) x P (not inoculated)
= .8588 x .9608

1 Duke, Çetinkaya-Rundel