Name: _Solutions_

1. (15 points) Circle True or False and **explain.** (three points each)

   (a) T or (F): The area on each side of the average on a histogram is always equal.

   The area on each side of the _median_ is always equal.

   (b) T or (F): Investigators eliminate chance error with sophisticated equipment.

   Chance error always exists and can be understood through repeated measurements.

   (c) (T) or F: Association does not imply causation.

   Two variables might both have the same underlying cause, but may not influence each other (Ex: Shoe size vs. spelling in middle school)

   (d) T or (F): An event occuring at a 10% chance means that out of every 10 trials that event will always occur exaclty once.

   This is only true on average. It is possible that it will occur 0 or even 10 times out of 10 trials

   (e) T or (F): Since the chance of rolling a pair of ones is $1/36$ we conclude that the chance of rolling two die and not getting any ones is $35/36$.

   Rolling ⊡ ⊡ is neither of these, so they are not opposites.

2. (15 points) Explain the difference. (three points each)

   (a) Standard Error(SE) and Standard Deviation(SD)

SD is for a list of data and SE is for a chance process.

   (b) Probability histogram and the "usual" histogram (sometimes called a density histogram).

Area in a probability histogram = chance and area in "usual" histogram = percent. The probability histogram is theoretical and the usual histogram is empirical (collected data)

   (c) Chance error and bias

Chance error is the error that always exists when taking a measurement and bias is some systematic error which is the result of some mistake.

   (d) Average and median

The median is the "middle value" with 50% of the data on each side (uneffected by outliers) and the average is the balence point (effected by outliers)

   (e) Controlled experiment and observational study

In a controlled experiment the investigators decide who will get the treatment being studied and in an observational study they can only watch.

3. (16 points) Suppose that the average score for the Math portion of the SAT is 500 and the average score for the Language portion is 550. The SD of the math portion is 50 and the SD of the Language portion is 100. It is found that the correlation coefficient between this data is $r = 0.25$. (four points each)

(a) Suppose that Fred scored a 600 on the Math portion. What would you guess that his Language score would be?

600 on Math $= 500 + 2(50) \rightarrow +2$ in standard units.

$2 \cdot r = 2 \cdot \frac{1}{4} = \frac{1}{2} \rightarrow +\frac{1}{2}$ su (Language).

$550 + \frac{1}{2}(100)$

$= \boxed{600}$

(b) Suppose that George got a ~~560~~ on the Language portion. What would you guess that his ~~Language~~ Math score would be?

$600 = 550 + \frac{1}{2}(100) \rightarrow +\frac{1}{2}$ su.

$+\frac{1}{2} \cdot r = \frac{1}{2} \cdot \frac{1}{4}$ su $= \frac{1}{8}$ su $\boxed{500 + \frac{1}{8}(50)} = $ ~~XXX~~ $\boxed{506.25}$

(c) What do parts (a) and (b) show us about how regression works?

Predicting x from y is not the opposite of predicting y from x. ("Two regr. lines")

(d) Suppose that Harley got a perfect 800 in Math. Would you expect her to get a perfect score in Language?

$800 = 500 + 6(50) \rightarrow +6$ su (Math) $\boxed{\text{NO}}$

$\frac{1}{4} \cdot 6 = +\frac{3}{2}$ su (Lang.) $550 + \frac{3}{2}(100) = \boxed{700}$

4. (4 points) The regression line. (two points each)

(a) What is the graph of averages?

For each x average all y's paired with it. Plot that point.

(b) What is the relationship between the regression line and the graph of averages?

The regression line is a linear approx of the graph of averages.

5. (10 points) Consider the following data:

x	y
1	5
1	3
1	5
1	7
2	3
2	3
2	1
3	1
3	1
4	1

(a) Just by looking at the data (without doing any computations) what can you say about the value of r? (two points)

· negative

· not close to 0

(b) Compute r. (eight points)

x

$$\text{Ave} = \frac{1+1+1+1+2+2+2+3+3+4.}{10} = 2$$

$$\text{SD} = \sqrt{\frac{1^2+1^2+1^2+1^2+0^2+0^2+0^2+1^2+1^2+2^2}{10}} = 1$$

y

$$\text{Ave} = \frac{5+3+5+7+3+3+1+1+1+1}{10} = 3$$

$$\text{SD} = \sqrt{\frac{2^2+0^2+2^2+4^2+0^2+0^2+2^2+2^2+2^2+2^2}{10}} = 2$$

x (su)	y (su)	xy
−1	+1	−1
−1	0	0
−1	+1	−1
−1	+2	−2
0	0	0
0	0	0
0	−1	0
+1	−1	−1
+1	−1	−1
+2	−1	−2

$$r = \frac{-1-1-2-1-1-2}{10}$$

$$\boxed{r = -0.8}$$

6. (10 points) Consider the following box of numbers:

2	4	5	6	8

Suppose that 100 draws are taken from this box (with replacement) and the sum of the drawn numbers is computed.

(a) What is the expected value? (three points)

$$\#draws \times Ave\ of\ box$$
$$= 100 \times \frac{2+4+5+6+8}{5} = \boxed{500}$$

(b) What is the standard error? (three points)

$$\sqrt{\#draws} \times SD\ of\ box$$
$$= \sqrt{100} \times \sqrt{\frac{3^2+1^2+0^2+1^2+3^2}{5}} = 10 \times \sqrt{4} = \boxed{20}$$

(c) What are the chances that the sum will lie between ~~483~~ 470 and ~~515~~ 530? (four points)

$$470 \quad \frac{4\!\!/\!8\!\!/\!5}{} = 500 - 1.5(20)$$
$$530 = 500 + 1.5(20)$$

Normal Approx.
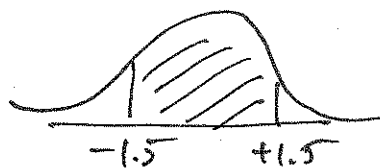
$$z = 1.5$$

$$\boxed{86.64\%} \approx \boxed{87\%}$$

7. (10 points) Consider the following box of numbers:

0	0	3	3	3	3	3	6	6

Suppose that 100 draws are taken from this box (with replacement) and the sum of the drawn numbers is computed. What are the chances that the sum will be greater than 330?

$$\text{expected} = 100 \times \text{Ave}$$
$$= 100 \times \frac{0 + 0 + 3 + 3 + 3 + 3 + 3 + 6 + 6}{9} = \boxed{300}$$

$$SE = \sqrt{100} \times SD$$
$$= 10 \times \sqrt{\frac{3^2 + 3^2 + 3^2 + 3^2}{9}} = 10 \times \sqrt{\frac{36}{9}} = 10 \times \sqrt{4}$$
$$= \boxed{20}$$

$$330 = 300 + 1.5(20) \qquad \boxed{+1.5\ \text{su}}.$$

$$= \frac{1}{2}\left(\quad \overset{-1.5 \quad 1.5}{} \quad \right)$$

$$= \frac{1}{2}\left(100 - \frac{}{-1.5 \quad 1.5} \right)$$

$$= \frac{1}{2}\left(100 - 87 \right)$$

$$= \boxed{\frac{13}{2}\% = 6.5\%}$$

8. (12 points) Circle True or False and **explain**. (three points each)

(a) Since the probability of getting a H when flipping a coin is $1/2$ we know that the probability of getting three heads out of three coin tosses is $(1/2)^3 = 1/8$.

~~True~~

mult. rule
Coin tosses are indep.

(b) Since the proability of getting a 6 when tossing a die is $1/6$, and the probability of getting a 5 is also $1/6$ we know that the probability of rolling exactly three 5's and one 6 out of four rolls is

$$\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{6^4}.$$

~~False~~ this ↗ is prob of getting 6,5,5,5.
But there are 4 different orders
which get one 6 and three 5's.

(c) If the dice you are using are fair, the probability of rolling two dice and getting a sum of 7 is the same as the probability of rolling two dice and getting a sum of 2.

~~False~~ There is only one way to get
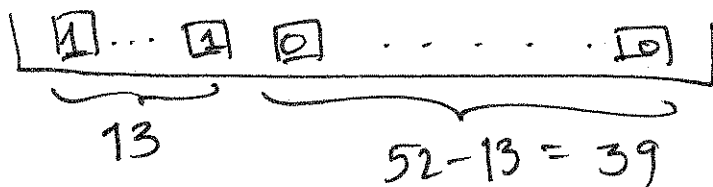a 2 (⚀ ⚀), but there are
many ways to get a 7 (⚃ ⚂, ⚄ ⚁ etc...)

(d) If you draw two cards from a standard deck the probability of getting a Queen of hearts for the first card and an Ace of Spades for the second is

$$\frac{1}{52} \times \frac{1}{52}$$

~~False~~

It is $\frac{1}{52} \times \frac{1}{51}$. Remember
to use conditional probability
for mult rule.

9. (8 points) Suppose you randomly draw from a deck of cards (with replacement) 100 times. What is the expected value and standard error for the number of spades you draw? (Remember that there are 52 cards in a deck and 13 of these are spades).

$$\underbrace{\boxed{1} \cdots \boxed{1}}_{13} \quad \underbrace{\boxed{0} \cdot \cdot \cdot \cdot \boxed{0}}_{52-13 \ = \ 39}$$

$$\text{expected value} = 100 \times \text{Ave}$$

$$= 100 \times \frac{13}{52} = \boxed{25}$$

$$SE = \sqrt{100} \times SD$$

shortcut for SD

$$= 10 \times 1 \cdot \sqrt{\frac{13}{52} \cdot \frac{39}{52}}$$

$$= 10 \sqrt{\frac{1}{4} \cdot \frac{3}{4}}$$

$$= \frac{10}{4} \sqrt{3} = \boxed{\frac{5\sqrt{3}}{2}}$$