

In almost all problems, I have rounded the answers to two decimal places. Delay rounding until the final step, unless otherwise instructed. Assume standard deviation refers to the sample standard deviation. If your answer is slightly different from one of mine, consider that to be round off error. Mark the closely matching one, as long as there is not a more precise option.

1. Which of the following best describes the data: classifications of unlikely, likely, or very likely to describe possible buying of a product?
- A) Qualitative B) Qualitative – Discrete C) Qualitative – Nominal **D) Qualitative – Ordinal**  
 E) Quantitative F) Quantitative – Nominal G) Quantitative – Ordinal H) Quantitative – Discrete  
 I) Quantitative – Continuous J) None of the preceding

Ordinal Variable (p.15): A qualitative variable that incorporates an ordered position, or ranking.

2. An office supply warehouse has boxes of pencils, 100 pencils to the box. Information about the entire warehouse as well as a sample of the boxes is shown below:

Number of defectives per box	Number of boxes (in warehouse)	Number of boxes (in sample)
0	1500	50
1	250	20
2	75	3
3	40	3
4	10	1

- i. A quality control technician is interested in the number of boxes with more than two defectives. What is the value of the parameter?
- ii. A quality control technician is interested in the proportion of boxes with no more than one defective pencil. What is the value of the statistic?

- A) 50, 0.91** B) 50, 70 C) 125, 0.91 D) 125, 70 E) 0.03, 7 F) 50, 0.09 G) 125, 7 H) 50, 0.93  
 I) 0.03, 0.91 J) None of the preceding

Parameter refers to population: with more than 2 defectives - are 40 with 3 defectives & 10 with 4, so a total of 50.

Statistic refers to sample: with no more than 1 defective - are 50 with 0 & 20 with 1, so 70 total. We are asked for proportion, so  $\frac{70}{77} = .9090... \approx \underline{.91}$

3. TRUE FALSE QUESTIONS: Determine which of the following eleven statements are true. Record the total number of true answers according to the following scale:

A) 2 B) 3 C) 4 D) 5  E) 6 F) 7 G) 8 H) 9 I) 10 J) None of the preceding

- When a proportional random sample is drawn, the sampling frame is subdivided into various strata, and then a subsample is drawn from each stratum. **TRUE**
- A representative sample is a sample obtained in such a way that all individuals had an equal chance to be selected. **FALSE**
- In a box-and-whisker display, the length of the "box" is the same as the interquartile range. **TRUE**
- A line graph of a cumulative frequency or cumulative relative frequency distribution is referred to as an ogive. **TRUE**
- The mean, median and mode for the set of data {3, 5, 3, 8, 6} are all the same value. **FALSE**
- The mean absolute deviation for any set of data is zero since  $\sum(x - \bar{x})$  is always zero. **FALSE**
- Chebyshev's theorem says that within two standard deviations of the mean, you will always find at least 89% of the data. **FALSE**
- On a test Jim scored at the 50<sup>th</sup> percentile and Jean scored at the 25<sup>th</sup> percentile; therefore, Jim's test score was twice Jean's test score. **FALSE**
- For a bell-shaped distribution, the range will be approximately equal to six standard deviations. **TRUE**
- A set of data for which  $r = -1$  or  $r = +1$  will be such that  $\sum(y - \bar{y})^2$  equals zero. **TRUE**
- The line of best fit is used to predict the average value of  $y$  that can be expected to occur at a given value of  $x$ . **TRUE**

4. Which of the following are correct statements? Use the following key, and only mark one selection on your answer card.

A) c, e B) d, h C) c, f D) d, e E) a, f F) a, b, d G) d, g, h H) a, e, g  I) b, d, h J) None of the preceding

- a. The interquartile range is found by taking the difference between the first and third quartiles and dividing that value by 2.
- b. The standard deviation is expressed in terms of the original units of measurement but the variance is not.
- c. The values of the standard deviation may be either positive or negative, while the value of the variance will always be positive.
- d. The standard deviation is a measure of dispersion.
- e. The range is a measure of central tendency.
- f. The median is a measure of dispersion.
- g. Chebyshev's theorem applies only to non-normal distributions.
- h. The sum of  $(x - \bar{x})$  will always be zero.

5. An aptitude test is known to have a mean score of 37.5 with a standard deviation equal to 3.5. A company requires a standard score of 1.5 for employment as one of its requirements. What must your test score be in order to be considered for employment? (Note: scores on this test can only take on integer values.)

A) 42 or larger   B) 42.75 or larger    C) 43 or larger   D) 33 or larger   E) 32.25 or larger   F) 17 or larger  
 G) 88 or larger   H) 57 or larger   I) 50 or larger   J) None of the preceding

$$Z = \frac{x - \bar{x}}{s}$$

$$1.5 = \frac{x - 37.5}{3.5}$$

$$(1.5)(3.5) = x - 37.5$$

$$5.25 = x - 37.5$$

$x = 42.75$   
 since must be integer value,

43 or better

6. A set of measurements has a mean equal to 35.5 and a standard deviation equal to 3.0. At least what percent of the data falls between 31.0 and 40.0?

A) 75%   B) 30%   C) 95.1%   D) 68.3%    E) 55.6%   F) 81.5%   G) 95%   H) 62.6%   I) 86.6%  
 J) None of the preceding

$$35.5 - 31.0 = 4.5$$

$$\frac{4.5}{3.0} = 1.5 \text{ SD's}$$

$$40 - 35.5 = 4.5$$

also 1.5 SD's

Question is:

at least what percent of data falls within 1.5 SD's of mean?

Chelyshev's Theorem says

$$1 - \frac{1}{k^2} = 1 - \frac{1}{1.5^2} = .555\bar{5}$$

$\approx 55.6\%$

of data

7. For a particular sample  $\bar{x} = 4.2$ . One item in the sample is  $x = 4.8$ . This item has a  $z$ -score at 2.50. Find the standard deviation.

A) 1.5 B) 0.38 C) 0.28 D) 4.17 E) 3.12 **F) 0.24** G) 0.06 H) 2.04 I) 1.77 J) None of the preceding

$$z = \frac{x - \bar{x}}{s}$$

$$2.50 = \frac{4.8 - 4.2}{s}$$

$$s = \frac{.6}{2.5} = \boxed{.24}$$

8. For a particular sample of size  $n = 10$ , the sample variance is 4.8 and  $\bar{x} = 0.5$ . For this sample, find  $\sum(x^2)$ .

A) 5.08 **B) 45.7** C) 48.2 D) 68.2 E) 25 F) 5.21 G) 11.41 H) 26.91 I) Can not be determined with the information provided J) None of the preceding

$$\text{Recall variance} = s^2 = \frac{\sum(x - \bar{x})^2}{n-1} = \frac{\sum(x^2) - \frac{(\sum x)^2}{n}}{n-1}$$

$$\text{Here, } \underline{s^2 = 4.8}, \underline{n = 10}$$

$$\bar{x} = \frac{\sum x}{10} = .5 \implies \underline{\underline{\sum x = 5}}$$

$$\text{So } s^2 = \frac{\sum(x^2) - \frac{(\sum x)^2}{n}}{n-1}$$

$$4.8 = \frac{\sum(x^2) - \frac{(5)^2}{10}}{9}$$

$$(4.8)(9) = \sum(x^2) - \frac{25}{10}$$

$$43.2 + 2.5 = \sum(x^2)$$

$$\sum(x^2) = \boxed{45.7}$$

9. A group of children had the following heights in inches: 45, 46, 42, 56, 37, 50, 51, 50, 47, 47. Find the sample standard deviation for the scores.

A) 47.39 B) 15.71 C) 14.90 D) 47 E) 47.5 F) 4.9 G) 47.1 **H) 5.22** I) 4.95 J) None of the preceding

$x$	$x^2$
45	2025
46	2116
42	1764
56	3136
37	1369
50	2500
51	2601
50	2500
47	2209
47	2209

Sums: 471    22,429

$$\begin{aligned}
 S &= \sqrt{\frac{SS(x)}{n-1}} \\
 &= \sqrt{\frac{\sum(x^2) - \frac{(\sum x)^2}{n}}{n-1}} \\
 &= \sqrt{\frac{22429 - \frac{471^2}{10}}{9}} \\
 &\approx 5.22
 \end{aligned}$$

or w/ calculator

10. If a sample with a mean of 10.5 and a standard deviation of 2.30 has every item multiplied by 10, find the variance of the new sample.

A) 72.73 B) 241.5 C) 52.9 **D) 529** E) 230 F) 2415 G) 45.65 H) 4.57 I) Can not be determined with the information given J) None of the preceding

Recall 
$$s^2 = \frac{\sum(x^2) - \frac{(\sum x)^2}{n}}{n-1}$$

Original Sample

$$(2.30)^2 = \frac{\sum(x^2) - \frac{(\sum x)^2}{n}}{n-1}$$

so

$$\frac{\sum(x^2) - \frac{(\sum x)^2}{n}}{n-1} = 5.29$$

notice that this is 100 · original variance!

New Sample

$$\begin{aligned}
 S^2 &= \frac{\sum[(10x)^2] - \frac{[\sum(10x)]^2}{n}}{n-1} \\
 &= \frac{\sum(100x^2) - \frac{[10\sum x]^2}{n}}{n-1} \\
 &= \frac{100 \sum(x^2) - \frac{100(\sum x)^2}{n}}{n-1} \\
 &= \frac{100 \left[ \sum(x^2) - \frac{(\sum x)^2}{n} \right]}{n-1} \\
 &= 100 \cdot (5.29) = \boxed{529}
 \end{aligned}$$

11. The following grouped frequency distribution gives the pay ranges (in thousands of dollars) for all middle management personnel in large company. Find the mean pay.

Class Boundaries	f
$20 < x < 30$	4
$30 \leq x < 40$	27
$40 \leq x < 50$	29
$50 \leq x < 60$	25
$60 \leq x < 70$	17

- A) 42.35 B) 45.00 C) 47.35 D) 52.35 E) 44.71 F) \$42,352.95 **G) \$47,352.94** H) \$52,352.94  
 I) 44,705.88 J) None of the preceding

Class	$x = \text{mark}$	frequency	$x \cdot f$
1	25	4	100
2	35	27	945
3	45	29	1305
4	55	25	1375
5	65	17	1105
		$\Sigma = 102$	$\Sigma = 4830$

$\bar{x} = \frac{4830}{102}$   
 $= 47.35294$   
 but mean pay is in \$, so  
**\$47,352.94**

12. The mean for 50 pressure readings equals 5.5, and the sum of the squares of the readings equals 1622.75. Find the standard deviation of these pressure readings.

- A) 33.01 B) 5.75 C) 2.40 D) 2.25 E) 2.21 F) 1.59 **G) 1.5** H) 1.26 I) 1.22 J) None of the preceding

$$\bar{x} = \frac{\Sigma x}{n}$$

$$5.5 = \frac{\Sigma x}{50}$$

$$275 = \Sigma x$$

$$s = \sqrt{\frac{SS(x)}{n-1}}$$

$$= \sqrt{\frac{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{1622.75 - \frac{275^2}{50}}{49}}$$

$$\approx 1.5$$

13. Consider the following sample of size  $n = 60$ , ordered from smallest to largest. Find the 35<sup>th</sup> percentile

10	20	27	33	37	45	59	63	66	79
11	21	28	34	38	47	59	63	68	80
13	22	28	<u>34</u>	39	48	61	63	69	86
13	24	28	<u>35</u>	41	50	61	64	70	98
17	25	29	36	43	51	62	64	73	99
19	27	33	37	44	56	62	66	79	99

- A) 50.5 B) 51 C) 34 D) 53.5 E) 44.5 **F) 34.5** G) 31.15 H) 27.5 I) 35 J) None of the preceding

$$n = 60$$

$$\frac{n \cdot P}{100} = \frac{60 \cdot 35}{100} = 21$$

$P_{35}$  will be at 21.5 depth

$$P_{35} = \frac{34 + 35}{2} = \boxed{34.5}$$

14. Consider the sample 9, 11, 17, 23, 26, 38, 47. Find the  $z$ -score for the data point of "11." Round any intermediate statistics to four decimal places, if necessary.

- A) -0.96** B) 0.79 C) -9.25 D) 0.7 E) -1.75 F) -0.12 G) -0.79 H) 1.75 I) -0.96  
J) None of the preceding

$x$	$x^2$
9	81
11	121
17	289
23	529
26	676
38	1444
47	2209

$$\Sigma = 171 \quad 5349$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{171}{7} \approx 24.4286$$

$$s = \sqrt{\frac{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{5349 - \frac{171^2}{7}}{6}} \approx 13.9745$$

$$z = \frac{x - \bar{x}}{s} = \frac{11 - 24.4286}{13.9745}$$

$$\boxed{z \approx -0.96}$$

15. A student computed the mean of a particular sample to be 40.0. After computing the mean, he discovered that he forgot to include the number 36 in the sample. When this number was included, the sample mean changed to 39.5. What is the sample size when the number 36 is correctly included in the sample?

A) 5 B) 6 C) 7 **D) 8** E) 9 F) 10 G) 12 H) 18 I) Can not be determined with the information given J) None of the preceding

Let  $n$  = correct sample size

Forgetting 36

$$\bar{x} = \frac{\sum x}{n-1} = 40.0$$

so

$$\sum x = 40.0(n-1)$$

substitute

With 36

$$\bar{x} = \frac{\sum x + 36}{n} = 39.5$$

so

$$\underline{\underline{\sum x}} + 36 = 39.5 \cdot n$$

$$\underline{\underline{40.0(n-1)}} + 36 = 39.5n$$

$$40n - 40 + 36 = 39.5n$$

$$\frac{.5n}{.5} = \frac{4}{.5} \implies \boxed{n = 8}$$

16. Starting with the data values 70 and 100, add three data values to your sample so that the sample has a mean of 95 and a mode of 80. One of the three numbers is:

A) 35 B) 65 **C) 145** D) 115 E) 40 F) 130 G) 160 H) Many possible answers I) Impossible J) None of the preceding

Since mode is 80, at least 2 of 3 must be 80.

So our values are 70, 80, 80, 100, &  $x$ .

$$\text{mean} = \bar{x} = \frac{70 + 80 + 80 + 100 + x}{5} = 95$$

$$330 + x = 95 \cdot 5$$

$$330 + x = 475$$

$$\boxed{x = 145}$$



17. A survey of 15 doctors and 15 nurses was conducted, and one question was related to their smoking habit. The following coding was used: Doctor (D), Nurse (N), Smoker (S), Nonsmoker (NS). The following results were obtained:

Respondent	D	D	D	N	N	D	D	D	N	D
Smoking Habit	S	NS	NS	S	S	NS	NS	NS	S	NS

Respondent	D	N	N	N	D	N	N	D	D	D
Smoking Habit	NS	NS	S	NS	S	NS	NS	NS	NS	NS

Respondent	D	N	N	N	D	D	N	N	N	N
Smoking Habit	NS	S	NS	NS	S	S	NS	S	NS	NS

Summarize the data into a  $2 \times 2$  cross-tabulation table of percentages based on the grand total (entire sample).

- A) 

	Doctor	Nurse
S	4	6
NS	11	9
- B) 

	Doctor	Nurse
S	4	11
NS	6	9
- C) 

	Doctor	Nurse
S	13.3%	20%
NS	36.7%	30%
- D) 

	Doctor	Nurse
S	13.3%	36.7%
NS	20%	30%
- E) 

	Doctor	Nurse
S	40%	60%
NS	55%	45%
- F) 

	Doctor	Nurse
S	40%	55%
NS	60%	45%
- G) 

	Doctor	Nurse
S	26.7%	40%
NS	73.3%	60%
- H) 

	Doctor	Nurse
S	26.7%	73.3%
NS	40%	60%

I) None of the preceding

	Doctor	Nurse
S	(4)	(5)
NS	(5)	(9)

$n = 11$

want percentages based on grand total,  
so divide each frequency by 30

	D	N
S	$\frac{4}{30} = 13.3\%$	$\frac{6}{30} = 20\%$
NS	$\frac{11}{30} = 36.7\%$	$\frac{24}{30} = 30\%$

18. Based on the following bivariate data, find the value of  $k$  so that the value of the coefficient of linear correlation  $r$  will be exactly zero.

$x$	2	4	7
$y$	3	5	$k$

- A) 8 B) 3 C) 4.33 D) 4  E) 3.25 F) 1.41 G) 2.52 H) 1.88 I) ) Can not be determined with the information given J) None of the preceding

$$r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}} = 0$$

so we need

$$SS(xy) = 0$$

$x$	$y$	$xy$
2	3	6
4	5	20
7	$k$	$7k$

$$\Sigma = 13 \quad (8+k) \quad (26+7k) = 0$$

$$SS(xy) = \Sigma(xy) - \frac{\Sigma(x)\Sigma(y)}{n}$$

$$= (26+7k) - \frac{(13)(8+k)}{3}$$

$$3 \left[ (26+7k) - \frac{(13)(8+k)}{3} \right] = [0] \cdot 3$$

$$3(26+7k) - 13(8+k) = 0$$

$$78+21k - 104 - 13k = 0$$

$$8k - 26 = 0$$

$$\frac{8k}{8} = \frac{26}{8} \Rightarrow k = 3.25$$

19. Consider the following bivariate data, extensions, and totals. Find the following the equation of the line of best fit.

x	y	x <sup>2</sup>	xy	y <sup>2</sup>
2	14	4	28	196
3	13	9	39	169
4	11	16	44	121
5	8	25	40	64
5	9	25	45	81
7	4	49	28	16
7	3	49	21	9
33	62	177	245	656

- A)  $\hat{y} = -0.44 + 863x$  B)  $\hat{y} = 8.86 + 1.88x$  C)  $\hat{y} = 8.63 - 0.44x$  D)  $\hat{y} = 73.52 - 337.42x$   
 E)  $\hat{y} = 21.33 - 19.26x$  F)  $\hat{y} = -2.21 + 19.26x$  G)  $\hat{y} = -337.42 + 73.52x$  H)  $\hat{y} = -91.60 + 21.33x$   
 I)  $\hat{y} = 19.26 - 2.21x$  J) None of the preceding

$$\hat{y} = b_0 + b_1x$$

$$b_1 = \frac{SS(xy)}{SS(x)}$$

$$SS(xy) = 245 - \frac{33 \cdot 62}{7} \approx -47.2857$$

$$SS(x) = 177 - \frac{33^2}{7} \approx 21.4286$$

$$b_1 \approx \frac{-47.2857}{21.4286} \approx -2.21$$

or w/ calculator

$$\bar{x} = \frac{33}{7} \approx 4.7143$$

$$\bar{y} = \frac{62}{7} \approx 8.8571$$

$$b_0 = \bar{y} - b_1\bar{x} = 8.8571 + 2.21(4.7143) \approx 19.26$$

$$\hat{y} = 19.26 - 2.21x$$

20. Find the value of  $\sum(y - \hat{y})^2$ , given  $\hat{y} = -0.6 + 3.2x$  is the equation of the line of best fit for the following data.

x	1	2	3	4	5
y	5	1	13	9	17

- A) -96 B) -6 C) 0 D) 45 E) 57.6 F) 64.8 G) 507.4 H) 931.43 I) 2617.6 J) None of the preceding

x	y	$\hat{y} = -0.6 + 3.2x$	$(y - \hat{y})$	$(y - \hat{y})^2$
1	5	2.6	2.4	5.76
2	1	5.8	-4.8	23.04
3	13	9	4	16
4	9	12.2	-3.2	10.24
5	17	15.4	1.6	2.56

$$\sum = 57.6$$

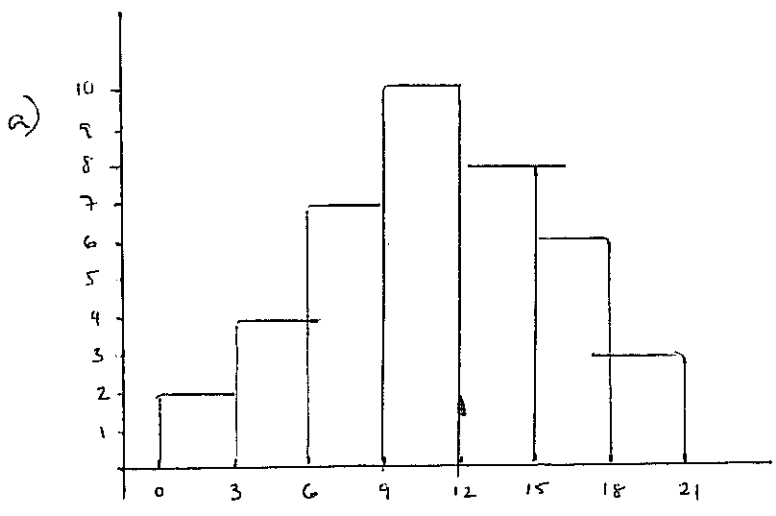
PRINT NAME: Key

The following two problems are each worth 10%. Please show all work to a) ensure full credit for work and b) receive partial credit.

1. The frequency distribution below gives the daily high temperature for 40 consecutive winter days in northern Wisconsin.

Class Boundaries	f
$0 \leq x < 3$	2
$3 \leq x < 6$	4
$6 \leq x < 9$	7
$9 \leq x < 12$	10
$12 \leq x < 15$	8
$15 \leq x < 18$	6
$18 \leq x \leq 21$	3

- a. Construct a histogram for this distribution.
- b. Convert the above frequency distribution to a relative frequency distribution.
- c. What is the proportion of days in which the daily high temperature was between 15 and 18?
- d. Construct the cumulative frequency distribution
- e. Construct the cumulative relative frequency distribution.



$$c) \frac{6}{40} = .15$$

Class boundaries	f	(b) relative freq.	(d) cumulative freq.	(e) Cumulative rel. freq.
$0 \leq x < 3$	2	$\frac{2}{40} = .05$	2	.05
$3 \leq x < 6$	4	.1	6	.15
$6 \leq x < 9$	7	.175	13	.325
$9 \leq x < 12$	10	.25	23	.575
$12 \leq x < 15$	8	.2	31	.775
$15 \leq x < 18$	6	.15	37	.925
$18 \leq x < 21$	3	.075	40	1.00

$\Sigma = 40$

2. Consider the following data, which give the weight (in thousands of pounds)  $x$  and gasoline mileage (miles per gallon)  $y$  for ten different automobiles.

$x$	2.0	2.4	2.6	2.9	3.2	3.5	3.8	4.2	4.6	5.2
$y$	45	40	42	39	44	36	34	28	18	13

- a. Calculate  $SS(x)$ ,  $SS(y)$ , and  $SS(xy)$   
 b. Find Pearson's product moment  $r$ , and interpret its meaning.  
 c. Find the linear regression line for this data, and use it to find the average gasoline mileage one would expect from a car weighing 3,000 pounds.

$x$	$y$	$x^2$	$y^2$	$xy$
2.0	45	4	2025	90
2.4	40	5.76	1600	96
2.6	42	6.76	1764	109.2
2.9	39	8.41	1521	113.1
3.2	44	10.24	1936	140.8
3.5	36	12.25	1296	126
3.8	34	14.44	1156	129.2
4.2	28	17.64	784	117.6
4.6	18	21.16	324	82.8
5.2	13	27.04	169	67.6

$$\Sigma = \begin{array}{l} 34.4 \\ 339 \\ 127.7 \\ 12575 \\ 1072.3 \end{array}$$

$$a) \quad SS(x) = \sum(x^2) - \frac{(\sum x)^2}{n} = 127.7 - \frac{34.4^2}{10} = \boxed{9.364}$$

$$SS(y) = 12575 - \frac{339^2}{10} = \boxed{1082.9}$$

$$SS(xy) = 1072.3 - \frac{(34.4)(339)}{10} = \boxed{-93.86}$$

$$b) \quad r = \frac{SS(xy)}{\sqrt{SS(x) \cdot SS(y)}} = \frac{-93.86}{\sqrt{9.364 \cdot 1082.9}} \approx \boxed{-0.93}$$

Answers will vary, but should include:

a) because  $|r|$  is close to 1, this bivariate data is very close to linear (strong correlation)

b) since it is negative, the linear regression will have negative slope (decreasing)

Extra Credit (5 points):

Use partial derivatives and calculus principles to develop the formulas for the coefficients  $b_0$  and  $b_1$  in the least squares line  $\hat{y} = b_0 + b_1x$ .

- see notes from Friday 2/3 for complete development:

Basis: minimize  $S(b_0, b_1) = \sum (y - b_0 - b_1x)^2$

$$\frac{\partial S}{\partial b_0} = \sum 2(y - b_0 - b_1x)(-1)$$

$$\frac{\partial S}{\partial b_1} = \sum 2(y - b_0 - b_1x)(-x)$$

Set both partial derivatives to zero & solve

$$\begin{cases} \sum (y - b_0 - b_1x) = 0 \\ \sum x(y - b_0 - b_1x) = 0 \end{cases}$$

$$= \begin{cases} \sum y - nb_0 - b_1 \sum x = 0 \\ \sum xy - b_0 \sum x - b_1 \sum x^2 = 0 \end{cases}$$

$$= \begin{cases} \boxed{b_0 = \frac{\sum y - b_1 \sum x}{n} = \bar{y} - b_1 \bar{x}} & \# \text{ one formula} \\ \sum xy - \left[ \frac{\sum y - b_1 \sum x}{n} \right] \sum x - b_1 \sum x^2 = 0 & \text{by substitution} \end{cases}$$

$$\Rightarrow \sum xy - \frac{\sum x \sum y}{n} + b_1 \frac{(\sum x)^2}{n} - b_1 \sum x^2 = 0$$

$$\sum xy - \frac{\sum x \sum y}{n} = b_1 \sum x^2 - b_1 \frac{(\sum x)^2}{n}$$

$$SS(xy) = b_1 \left[ \sum x^2 - \frac{(\sum x)^2}{n} \right]$$

$$\boxed{\frac{SS(xy)}{SS(x)} = b_1} \quad \# \text{ remaining formula}$$