Please provide computational details for questions and problems to get any credit.

The following problem is associated with questions 1 to 5.

Most presidential polling institutions focus on the likely voters' support for either President Bush or Senator Kerry in the 2004November 2nd election. A polling organization wishes to estimate the current likely voters' support for the two candidates. To do so, 1050 likely voters nationwide are selected randomly and found that their support for either candidate is tied at 47%.

- 1. What is the sample statistics?
 - a. The sample percentage of support from the 1050 likely voters for either candidate is at 47%.
 - b. The population percentage of support from all likely voters for either candidate.
 - c. Randomly selected 1050 likely voters national wide
 - d. The percentage of support from likely voters for either candidate..
- 2. What is the population parameter?
 - a. The sample percentage of support from the 1050 likely voters for either candidate is at 47%.
 - b. The population percentage of support from all likely voters for either candidate.
 - c. Randomly selected 1050 likely voters national wide
 - d. The percentage of support from likely voters for either candidate.
- 3. What is the sample of this study?
 - a. The sample percentage of support from the 1050 likely voters for either candidate is at 47%.
 - b. The population percentage of support from all likely voters for either candidate.
 - c. Randomly selected 1050 likely voters national wide
 - d. The percentage of support from likely voters for either candidate.
- 4. What is the variable of interest in the study?
 - a. The sample percentage of support from all likely voters for either candidate.
 - b. Randomly selected 1050 likely voters national wide
 - c. The percentage of support from likely voters for either candidate.
 - d. All likely voters national wide.
- 5. What is the population from which the study was sampled?
 - a. The population percentage of support from all likely voters for either candidate.
 - b. Randomly selected 1050 likely voters national wide

- c. The percentage of support from likely voters for either candidate.
- d. All likely voters national wide.

The following problem is associated with Questions 6 to 15

The Best Hotel asks its customers to fill out a questionnaire to rate the quality of meals served in the last 12 months. Two of the questions are listed here.

1) Would you be willing to pay additional \$10 for better meals? Yes

No

2) How would you rate the meals served? Good Fair Poor

As a consultant to Best Hotel, you are asked to evaluate the design of the questionnaire. Among many shortcomings are the following four potential problem areas. For each question below, please identify the problem area to which each error is best viewed as belonging. Pick one answer for each question. Answers may be used more than once, or not at all. Choices are:

> A). sampling error **B**). measurement error **C**). selection bias (or coverage error) **D**). nonresponse bias (or nonresponse error)

- What is the type of error it may suffer from if a questionnaire is handed out to each customer who stayed in Best 6 Hotel over the next 12 months?
 - A) sampling error
 - **B)** measurement error
 - C) selection bias (or coverage error)
 - D) nonresponse bias (or nonresponse error)
- 7 What is the type of error if Best Hotel is only interested in the people he calls "Business Travelers", but his survey may be filled out by customers who stay in Best Hotel for none business purposes?
 - A) sampling error
 - B) selection bias (or coverage error)
 - C) measurement error
 - D) nonresponse bias

What is the type of error it may cause if you decide to ask a sample of customers who have stayed in Best Hotel and 8

filed a complaint over the past year to fill out the survey?

- A) measurement error
- **B)** sampling error
- C) nonresponse bias
- D) selection bias (or coverage error)
- 9 What is the type of error if customers who do not stay in Best Hotel cannot provide information at all?
 - A) sampling error
 - **B)** measurement error
 - C) selection bias (or coverage error)
 - D) nonresponse bias

- 10 What is the type of error if customers lie about how much they would truly pay for better meals?
 - A) sampling error
 - B) measurement error
 - C) selection bias (or coverage error)
 - D) nonresponse bias
- 11 What is the type of error if only 90 people can fill out the survey? Even if the other problems are straightened out, there are many customers who will not be surveyed, so the results are only an approximation of the real interest?
 - A) sampling error
 - **B)** measurement error
 - C) selection bias (or coverage error)
 - D) nonresponse bias
- 12 The possible responses to the question "How many times have you stayed in Best Hotel?" are values from a
 - A). parameter.
 - B). categorical random variable.
 - C). continuous random variable.
 - D). discrete random variable.
- 13 What type of data does the variable for the response to the questions "Would you be willing to pay extra for better meals?" belong to (circle the best one you think fit)?
 - A) Categorical and On a nominal scale
 - B) Numerical and on a interval scale
 - C) Categorical and On a ordinal scale
 - D) Discrete and on a nominal scale
- 14 The possible responses to the question "How would you rate the meals served?" in this survey can be best described as
 - A). parameter.
 - B). discrete random variable.
 - C). continuous random variable.
 - D). categorical random variable.
- 15 If Best Hotel has two types of customers: business or none business, the best sampling strategy would be to use
 - A). a cluster sample.
 - B). a systematic sample.
 - C). a stratified sample.
 - D). a simple random sample.

The following problem is associated with Questions 16 to 25.

A City Bus Company has categorized complaints from riders for last August and September as follows with the corresponding ROW, Column and Total Percentages, respectively. Assume the sample is a good representation of the population.

Type of	Buses sometimes don't	Bus schedules make	Route maps not	Col
Complaints	stop at bus stops	transfer difficult	available	Total
August	25	35	5	65
September	50	70	15	135
RowTotal	75	105	20	200
Row				
Percentages	Buses sometimes don't	Bus schedules make	Route maps	Col
(%)	stop at bus stops	transfer difficult	not available	Total

i da la companya da l	i de la constante de	i i	i da la companya da l	
August	0.33	0.33	0.25	0.33
September	ember 0.67 0.67		0.75	0.68
RowTotal	al 1 1		1	1
Column Percentages (%)	Buses sometimes don't stop at bus stops	Bus schedules make transfer difficult	Route maps not available	Col Total
August	0.38	0.54	0.08	1
September	0.37	0.52	0.11	1
RowTotal	0.38	0.53	0.1	1
Total Percentages (%)	Buses sometimes don't stop at bus stops	Bus schedules make transfer difficult	Route maps not available	Col Total
August	0.13	0.18	0.03	0.33
September	0.25	0.35	0.08	0.68
RowTotal	0.38	0.53	0.1	1

Referring to Tables above when you answer the following questions. You must first identify the relationship between or among the events in question, such as, **Union, Joint or Intersection, Conditional, Simple or Marginal**, and so forth.

- 16 Referring to the Tables above, what is the probability that a random selected rider in the population who complained Route maps not available and the complaint was filed in August?
 - A) 0.250 B) 0.425 C) 0.400 D) 0.025 D
- 17 Referring to the Tables above, what is the probability that a random selected rider filed a complaint in September?
 - A) 0.675 B) 0.667 C) 0.525 D) 0.370
- 18 Referring to the Tables above, of those riders complained for Bus schedules make transfer difficult, what is the probability that a random selected rider filed the complaint in August?
 - A) 0.333 B) 0.325 C) 0.175 D) 0.538
- 19 Referring to the Tables above, what is the probability that a random selected rider complained for Buses sometimes don't stop at bus stops or filed the complaint in September?
 - A) 0.800 B) 0.667 C) 1.000 D) 0.250
- 20 Referring to the Tables above, what is the probability that a random selected rider complained for Buses sometime don't stop at bus stops or complained Route maps not available?
 - A) 0.100 B) 0.375 C) 0.475 D) 0.250
- 21 Referring to the Tables above, the events of complaining in August and complaining in September are: A). mutually exclusive.
 - B). collective exhaustive.
 - C). statistically dependent.
 - D). all of the above
- 22 Are the events of complaining Bus schedules make transfer difficult and complaining Route maps not available statistically independent?
 - A) Yes, because P(Map) = P(Map | Transfer)

- B) Yes, because P(Map) = P(Transfer)
- C) No, because $P(Map) \neq P(Map | Transfer)$
- D) No, $P(Map) \neq P(Transfer)$

The following problem is associated with Questions 26 to 40.

A city bus company receives complaints from its riders in the last few months as follows

	January	February	March	April	May	June	July	Total
Complaints X	11	22	18	8	9	27	24	119
$(X-\overline{X})$	-6	5	1	-9	-8	10	7	
$(X-\overline{X})^2$	36	25	1	81	64	100	49	356

Please answer the following questions:

23 What is the relative frequency of number of complaints in March?

A)	18/119=15.13%	B) 51	C) 18	D) 51/119=42.86%
/	10/11/ 10/10/0	2)01	0)10	2)01/11/ 12/00/0

24 What is the Cumulative Relative Frequency of number of complaints in March?

24	what is the Cumulative Relative requency of number of complaints in March.					
	A) 18	B) 51	C) 51/119=42.86%	D) 18/119=15.13%		
25	What is the arithme	etic mean (\overline{X}) or the ave	erage number of complain	nts?		
	A) 19.0	B) 20	C) 18.0	D) 17.0		
26	What is the Mode?					
	A) 7.7	B) 16.5	C) 13.0	D) Not available		
27	What is the First Q	puartile Q_1 ?				
	A) 9	B) 8	C) 11	D) 10		
28	What is the Median	n or Second Quartile Q ₂ ?				
	A) 20.0	B) 18.0	C) 17.0	D) 23.0		
29	What is the Excel e	equation to computer the	median?			
	I. =MEDIAN(range) II. =Q	UARTILE(range,3)	III. =QUARTILE(range,2)		
	A) I only	B) I and II only	C) I, II and III	D) I and III only		
30	What is the Interqu	artile Range?				
	A) 15.0	B) 23.0	C) 19.0	D) 13.0		

- 31 What is the Midhinge?
 - A) 13.0 B) 16.5 C) 19.0 D) 59.3

32 What is the Sample Standard Deviation ?

- A) 18.0 B) 19.0 C) 7.7 D) 59.3
- 33 What is the Coefficient of Variation?
 - A) 17.0 B) 0.4531 C) 7.7 D) 59.3

34 Construct a Box-and-Whisker Plot for the number of complaints from January to July? Is the data skewed to the right

or to the left _____?

- A) Skew to the right, because the mean is greater than the median
- B) Skew to the left, because the mean is greater than the median
- C) Skew to the left, because the mean is less than the median
- D) Skew to the right, because the mean is less than the median
- 35 In right-skewed distributions, which of the following is the correct statement?
 - A). The distance from the smallest observation to Q_1 is smaller than the distance from Q_3 to the largest observation.
 - B). The distance from the smallest observation to Q_2 is larger than the distance from Q_2 to the largest observation.
 - C). The distance from Q_2 to Q_3 is smaller than the distance from Q_1 to Q_2 .
 - D). The distance from Q_1 to Q_3 is twice the distance from Q_2 to Q_3
- 36 Construct a Steam and Leaf Graph for the number of complaints from January to July?

Stem	Leaf
(0)	89
(1)	18
(2)	247

Referring your Stem and Leaf display in part a) of this questions, what percentage of the observations had the

number of complaints between ten and twenty?

A) 2/7=28.6% B) 1/7 C) 20 D) 2

37 Among the values computed above, which ones are the measures for central tendency?

- A) Mean, Median, and Range
- B) Mean, Mode, and Median
- C) Mean, Range and Standard deviation
- D) Mode, Median and Coefficient of Variation
- 38 Among the values computed above, which ones are the measures for variation?
 - I. MeanII. Interquartile RangeIII. Standard deviationIV. Coefficient of Variation

39 Which of the values computed above is/are sensitive (not resistant) to extreme values?

	I. Mea	n	II. Median	
	III. Sta	andard deviation	IV. Coefficient of V	ariation
A) I and	II only	B) II and III only	C) I, III and IV only	D) III and IV only

40 Which of the following is NOT sensitive to extreme values?

- A). The range.
- B). The interquartile range.
- C). The coefficient of variation.
- D). The standard deviation.

8/15 Solution for COB191 Test 1, Fall 2004, 9/21

Please provide computational details for questions and problems to get any credit.

The following problem is associated with questions 1 to 5.

Most presidential polling institutions focus on the likely voters' support for either President Bush or Senator Kerry in the coming November 2nd election. A polling organization wishes to estimate the current likely voters' support for the two candidates. To do so, 1050 likely voters nationwide are selected randomly and found that their support for either candidate is tied at 47%.

- 1. Describe the sample statistics _____ $p_x = 47\% = 0.47$, s_{px} as standard deviation _____
- 2. Identify the sample of this study._____1,050 likely voters nationwide in the sample
- 3. Describe the population parameters.____p (%) of support for either candidate among all likely voters nationwide
- 4. Identify the variable of interest in the study____do you support Bush or Kerry as x = Yes/No
- 5. Identify the population from which the study was sampled.____nationwide all likely voters___

The following problem is associated with Questions 6 to 15

The Best Hotel asks its customers to fill out a questionnaire to rate the quality of meals served in the last 12 months. Two of the questions are listed here.

- 1). Would you be willing to pay additional \$10 for better meals? _____Yes _____No
- 3) How would you rate the meals served? _____Good _____Fair ____Poor

As a consultant to Best Hotel, you are asked to evaluate the design of the questionnaire. Among many shortcomings are the following four potential problem areas. For each question below, please identify the problem area to which each error is best viewed as belonging. Pick one answer for each question. Answers may be used more than once, or not at all. Choices are:

- (i) sampling error
- **B**). measurement error
- C). selection bias (or coverage error)
- D). nonresponse bias (or nonresponse error)
- 6. What is the type of error it may suffer from if a questionnaire is handed out to each customer who stayed in Best Hotel over the next 12 months? __D or nonresponse_____
- 7. What is the type of error if Best Hotel is only interested in the people he calls "Business Travelers", but his survey may be filled out by customers who stay in Best Hotel for none business purposes? ____C selection ____
- 8. What is the type of error it may cause if you decide to ask a sample of customers who have stayed in Best Hotel

and filed a complaint over the past year to fill out the survey? ____C selection______

9. What is the type of error if customers who do not stay in Best Hotel cannot provide information at all? __C

selection _____

10. What is the type of error if customers lie about how much they would truly pay for better meals? _B

measurement_____

- 11. What is the type of error if only 90 people can fill out the survey. Even if the other problems are straightened out, there are many customers who will not be surveyed, so the results are only an approximation of the real interest? ____A sampling_____.
- 12. The possible responses to the question "How many times have you stayed in Best Hotel?" are values from a A). parameter.

B). discrete random variable.

- C). categorical random variable.
- D). continuous random variable.
- 13. What type of data does the variable for the response to the questions "Would you be willing to pay extra for better meals?" belong to (circle as many as you think fit)?

Numerical	Categorical	On a nominal scale
On a ordinal scale	On a ratio scale	On a interval scale
Discrete	Continuous	

- 14. The possible responses to the question "How would you rate the meals served?" in this survey can be best described as ______
 - A). parameter.

B). discrete random variable.

C). continuous random variable.

D). categorical random variable.

15. If Best Hotel has two types of customers: business or none business, the best sampling strategy would be to use A). a cluster sample.

B). a stratified sample.

C). a systematic sample.

D). a simple random sample.

The following problem is associated with Questions 16 to 25.

A City Bus Company has categorized complaints from riders for last August and September as follows with the corresponding ROW, Column and Total Percentages, respectively. Assume the sample is a good representation of the population.

Type of Complaints	August	September	RowTotal
Buses sometimes don't stop at bus stops	25	50	75
Bus schedules make transfer difficult	35	70	105
Route maps not available	5	15	20
Col Total	65	135	200

Row Percentages (%)	August	September	RowTotal
Buses sometimes don't stop at bus stops	0.33	0.67	1.00
Bus schedules make transfer difficult	0.33	0.67	1.00
Route maps not available	0.25	0.75	1.00
Col Total	0.33	0.68	1.00

Column Percentages (%)	August	September	RowTotal
Buses sometimes don't stop at bus stops	0.38	0.37	0.38
Bus schedules make transfer difficult	0.54	0.52	0.53
Route maps not available	0.08	0.11	0.10

Col Total	1.00	1.00	1.00
Total Percentages (%)	August	September	RowTotal
Buses sometimes don't stop at bus stops	0.13	0.25	0.38
Bus schedules make transfer difficult	0.18	0.35	0.53
Route maps not available	0.03	0.08	0.10
Col Total	0.33	0.68	1.00

Referring to Tables above when you answer the following questions. You must first identify the relationship between or among the events in question, such as, Union, Joint or Intersection, Conditional, Simple or Marginal, and so forth.

16. Referring to the Tables above, what is the probability that a random selected rider in the population who complained Route maps not available and the complaint was filed in August?

a). What type of probability is it _____Joint or Intersect____?
B) What is the probability P(Map and August) = __5/200 = 0.025 = 0.03 = 2.5% = 3% ____?

- 17. Referring to the Tables above, what is the probability that a random selected rider filed a complaint in September?
 - a). What type of probability is it __Simple or Marginal ____?
 - b) What is the probability $\underline{P(\text{September})} = 135/200 = 0.68 = 68\%$?
- 18. Referring to the Tables above, of those riders complained for Bus schedules make transfer difficult, what is the probability that a random selected rider filed the complaint in August?

 - a). What type of probability is it __Conditional _____?
 b). What is the probability <u>P(August | Transfer)</u> = _35/105 = 0.33 = 33% ____?
- 19. Referring to the Tables above, what is the probability that a random selected rider complained for Buses sometimes don't stop at bus stops or filed the complaint in September?
 - a). What type of probability is it ____Union OR U_____? b). What is the probability <u>P(Stop OR September)</u> = $P(Stop) + P(September) - P(Stop \& September) = \frac{75}{200} + \frac{135}{200} - \frac{50}{200} = \frac{160}{200} = 0.80 = 80\%$
- 20. Referring to the Tables above, what is the probability that a random selected rider complained for Buses sometime don't stop at bus stops or complained Route maps not available?
 - a). What type of probability is it _Union for mutually exclusive events)
 - b). What is the probability P(Stop OR Map) = (75 + 20) / 200 = 95/200 = 0.475 = 47.5%?
- 21. Referring to the Tables above, the events of complaining in August and complaining in September are: A). mutually exclusive.
 - B). collective exhaustive.
 - C). statistically dependent.

D). all of the above

22. Are the events of complaining Bus schedules make transfer difficult and complaining Route maps not available

statistically independent?

- a). Are the two events statistically independent____NO_____
- b). What is your reasoning (or justification) for your answer in a) above _ $P(Transfer) = 105/200 \neq P(Map | Transfer) = 0_?$

23. Draw a Decision Tree. Starts with August and September, and then Don't stop, Transfer difficult, and Maps not available. Be sure to label each branch of the Tree clearly with the corresponding event and its corresponding probability. You may use notations P(X), P(X|Y) and P(X ∩ Y), etc, where it is appropriate, and provide the probability value for each of those probabilities.



24. Draw a Pie Chart to show the percentages of complaints for Buses sometimes don't stop at bus stops, Bus schedules make transfer difficult, and Route maps not available in August and September combined. You should label each portion clearly on your chart.



P(Stop) = 75/200, P(Transfer) = 105/200, P(Map) = 20/200

What is the probability that a randomly selected rider with complaints would complain the Bus schedule making transfer difficult?

Key is 0.52 so make up 0.38, 0.10 and 0.48

25. Draw a Histogram of Relative Frequency Distribution to show the percentages of complaints for Buses sometimes don't stop at bus stops, Bus schedules make transfer difficult, and Route maps not available in August and September combined. You should label each part clearly on your chart.



The following problem is associated with Questions 26 to 40.

A city bus company receives complaints from its riders in the last few months as follows

	January	February	March	April	May	June	July	Total
Complaints X	11	22	18	8	9	27	24	119
$(X-\overline{X})$	-6	5	1	-9	-8	10	7	
$(\mathbf{X} - \overline{X})^2$	36	25	1	81	64	100	49	356
Rel Freq	11/119	22/119	18/119	8/119	9/119	27/119	24/119	1
Cum Rel Freq	11/119	33/119	3/7	59/119	4/7	95/119	1	
8, 9, 11, 18, 22, 24, 27								

Please answer the following questions:

- 26. a). What is the number of complaints in March <u>18</u>?
- b). What is the Relative Frequency of number of complaints in March_18/119 = 15.13%____?,
- c). What is the Cumulative Relative Frequency of number of complaints in March _51/119 = 42.86% ____?
- 27. What is the arithmetic mean (\overline{X})_119/7 = 17_?
- 28. a) What is the Mode? ____NA____?
- b). What is the minimum number of complaints (X_{min}) 8 ??
- c). What is the maximum number of complaints (Xmax_t)_27____?
- d)). What is the Range 27 8 = 19 ____?
- 29. a) Where (location or index) is the First Quartile $Q_1 _ 2_(May) _ ?$
- b). What is the First Quartile $Q_1 __9 __(May) __?$
- 30. a) What is the Median or Second Quartile $Q_2 = 18$ (March)?
- b). What is the Third Quartile Q_3 ____24__(July)____?

- 31. a). What is the Interquartile Range 24 9 = 15?
- b). What is the Midhinge____(9 + 24) / 2 = 16.5___?
- 32. a) What is the Sample Standard Deviation $_S = sqrt(356/(7-1)) = 7.7028___?$
- b). What is the Coefficient of Variation____S/Xbar_= 7.7028/17 = 0.4531____?
- 33. Construct a Box-and-Whisker Plot for the number of complaints from January to July?



34. a). Is the data skewed to the right or to the left _____Left____?

- b). Why do you identify the data as skewed to the right or to the left _ Xbar = 17 < Md = 18____?
- 35. In right-skewed distributions, which of the following is the correct statement?

A). The distance from the smallest observation to Q_2 is larger than the distance from Q_2 to the largest observation. B). The distance from the smallest observation to Q_4 is smaller than the distance from Q_3 to the largest observation.

C). The distance from Q_2 to Q_3 is smaller than the distance from Q_1 to Q_2 .

D). The distance from Q_1 to Q_3 is twice the distance from Q_2 to Q_3

- 36. a). Construct a Steam and Leaf Graph for the number of complaints from January to July?
 - Stem Leaf () 8 9 (a) 1 8 (b) 2 4 7
- b). Referring your Stem and Leaf display in part a) of this questions, what percentage of the observations had the

number of complaints between ten and twenty 2/7 = 0.286 = 28.6%?

37. a). Among the values computed above, which ones are the measures for central tendency_Xbar, Mo, Md, Q1, Q2,

Q3, Midhinger _____?

b). Interpret the meaning of the central tendency measures______?

38. a). Among the values computed above, which ones are the measures for variation_Range, Std, S/Xbar, IQR __?

b). Interpret the meaning of the variation measures? _____?

39. a). Which of the values computed above is/are sensitive (not resistant) to extreme values__Xbar, Range, Std, CV?

b). Explain your answers in a) above_____?

- 40. Which of the following is NOT sensitive to extreme values? A). The range.
 - B). The interquartile range.
 - C). The coefficient of variation.

D). The standard deviation.

15/15	
1	А
2	В
3	С
4	С
5	D
6	D
7	В
8	D
9	С
10	С
11	В
12	D
13	А
14	D
15	С
16	D
17	A
18	A
19	A
20	C
21	D
22	C
23	A
24	C
25	D
26	D
27	A
28	В
29	
30 21	A D
31 22	Б
32 22	C D
33 34	D
34 25	
36	Δ
37	A B
38	B
30	С С
39 40	C B
4 0	D