STATISTICS
Midterm Review
[Solutions are at the end]

The following problems are designed to help you study for the midterm examination. In addition to these problems, you may wish to check out the terminology study guide if your instructor indicates that some portion of the exam will cover terminology (definitions or multiple choice questions).

1. For each of the following, indicate whether the data is measured on a nominal, ordinal, interval, or ratio scale.

_____ a) waist sizes of 65-year-old men
_____ b) amount of alcohol in a can of Budweiser beer
_____ c) countries of origin of CUNY students
_____ d) temperature in centigrade of Starbuck's coffee
_____ e) religions of CUNY college students
_____ f) social class of CUNY students
_____ g) weights of 65-year-old women
_____ h) IQ of CUNY students
_____ i) scores on the CPA exam for CUNY students
_____ j) speeds of fastballs thrown by American League pitchers
_____ k) strength categories of hurricanes (1, 2, …, 5)
_____ l) genders of Wal-Mart employees
_____ m) rank of college professors (Instructor, Assistant Professor, Associate Professor, Full Professor)
_____ n) class standing of CUNY students (Freshman, Sophomore, …)

2. Indicate which of the following are discrete measurements and which are continuous measurements:

Discrete	Continuous	
___	___	a) the number of defective laptops in batches of 30
___	___	b) life of a Duracell battery
___	___	c) mileage of Toyota Prius cars
___	___	d) weight of guinea pigs
___	___	e) span of butterfly wings
___	___	f) number of left-handed people on basketball teams
___	___	g) time to complete the New York City Marathon
___	___	h) amount of alcohol in a can of beer
___	___	i) number of foreign students in each statistics class
___	___	j) height of basketball players
___	___	k) number of pens in backpacks of college students

3. Indicate which of the following is a parameter and which is a statistic:

Parameter	Statistic	
___	___	a) sample mean
___	___	b) population standard deviation
___	___	c) sample standard deviation
___	___	d) population variance
___	___	e) population median
___	___	f) population mean
___	___	g) a mean obtained from the U.S. census

4. For each of the following, indicate the appropriate statistical measures that may be used for analysis (e.g., proportions, median, quantiles, mode, mean, …). List as many as are appropriate.
a) For data measured on a nominal scale, you may use: _____
b) For data measured on an ordinal scale, you may use: _____
c) For data measured on an interval scale, you may use: _____

5. CUNY has 200,000 students. A researcher wants a sample of 2,000 students. Students are assigned numbers and then random numbers are used so that every student has an equal chance of being selected (1%).
a) This is known as a: _____.
b) Measurements obtained from this are known as: _____.
c) Suppose the researcher decides to survey all 200,000 students. This is called a _____.
d) Any measurement obtained is known as a _____.

6. A researcher has converted all grades on this year's CPA exams into Z-scores.
a) The average Z-score will be: _____
b) If CPA exam scores are normally distributed, about _____ % of scores will be between +1 and -1.
c) If CPA exam scores are very skewed and not normally distributed, we still would expect at least _____ % of the Z-scores to be between +2 and -2.
d) You find out that the exam scores are normally distributed, and that your score on the exam is exactly +1.96. This means you scored higher than _____ % of the individuals who took the exam.

7. The life of a Hyundai Dred car is normally distributed with a mean life of 15 years, and a population standard deviation of 2 years. What proportion of Hyundai Dreds will die within 10 years?

8. a) If P(A and B) = 0, then A and B are: _____
b) If P(A|B) = P(A), then A and B are: _____
c) Is P(A|B) always equal to P(B|A)? _____
d) Is P(A and B) always equal to P(B and A)?____

9. Define each of the following: (a) parameter and statistic (b) population mean (c) mutually exclusive (d) interquartile range (e) histogram and bar chart (f) skewness (g) symmetric (h) independence (i) permutation (j) combination (k) probability density function (l) percentile (m) nominal, ordinal, interval, and ratio scales. (n) discrete and continuous measurements (o) frequency distribution.

10. A manufacturer of computers has lowered prices for her product. A sample of 16 stores selected randomly indicate the following sales (in units) during the past week:
0, 10, 2, 10, 3, 10, 3, 10, 6, 6, 8, 8, 4, 4, 5, 6
(a) Calculate: The mean, median, Q1, Q3, mode, range, IQR, standard deviation, variance, and the coefficient of variation.
(b) Standardize the 10 (i.e., convert the 10 into a z-score).

11. A researcher wants to determine the average wage of American steel workers. A sample of 12 workers was selected randomly and indicated the following wages:
$20, $14, $10, $11, $12, $40, $30, $21, $22, $20, $20, $20.
(a) Calculate: The mean, median, Q1, Q3, mode, range, IQR, standard deviation, variance, and the coefficient of variation.
(b) Standardize the $40 (i.e., convert the $40 into a z-score).

12. A study of smoking and sex found the following:

	Male	Female
Smokes	150	130
Does Not Smoke	250	470

(a) Compute the following probabilities: P(male) ; P(female (and) smoker); P(female or smoker); P(male|smoker); P(smoker/male).
(b) Prove that smoking and sex are not independent.

13. A study of tattoos and crime found the following:

	Tattoo (T)	No Tattoo (T')
Criminal Record (C)	100	440
No Criminal Record (C')	100	1360

(a) Compute these probabilities: P(C AND T) ; P(C); P(C|T); P(C|T'); P(C OR T).
(b) Prove that tattoos and criminal records are not independent.

14. The average wage of plumbers is normally distributed with a mean of $20 and a standard deviation of $5.50. Calculate the following:
(a) The proportion of plumbers earning between $18 and $22.
(b) The proportion of plumbers earning more than $28.
(c) The proportion of plumbers earning less than $15.
(d) Compute the 70th Percentile.

15. The mileage of cars is normally distributed with a mean of 20 mpg and a standard deviation of 4 mpg. (mpg is miles per gallon). Calculate the following:
(a) The percentage of cars under 15 mpg:
(b) The percentage of cars above 22 mpg:
(c) The probability that a car will have a mileage between 22 and 28 mpg:
(d) Compute the mpg of the top 10% of cars (i.e., the 90th percentile)
(e) Compute the 9th percentile.

[16-21 are part of the topic on Binomial Distribution. Check with your instructor to see if you are responsible for that topic on your exam.]

16. Suppose 20% of the students at a well-known college in NYC are business majors. A researcher selects a sample of 6 students from this college, what is the probability that none of the selected students are business majors?

17. Forty percent of people above the age of 80 have Alzheimer's disease. A researcher investigates the incidence of Alzheimer's disease in a certain island in the South Pacific where people eat a certain kind of fish every day. She takes a random sample of 30 people above the age of 80. What is the probability of finding that exactly 3 people have Alzheimer's disease?

18. Ten percent of the population is left-handed. What is the probability of 2 lefties in a group of 10?

19. The probability of a computer being a Dell is 15%. A company has 26 computers. What is the probability that 7 are Dells?

20. There are 18 people. (a) How many ways can you assign them to 18 different jobs? (b) How many ways can you assign them to 3 different jobs? (c) How many committees of 3 can you create from the 18?

21. A machine has 280 parts. The probability that any part is defective is .01. What is the probability that the machine works, i.e., 0 defective parts? Assume independence among the parts. (b) What is the probability that the machine does not work?

22. The following is a frequency distribution showing the amount of time it took a sample of employees to complete a certain job:

NUMBER OF DAYS	FREQUENCY
2	10
4	17
5	18
7	12
9	40
10	3

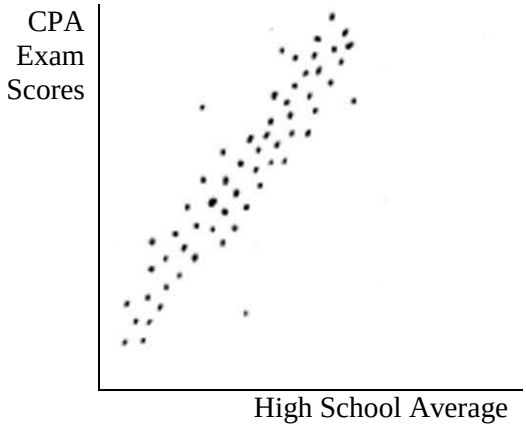
(a) Calculate the mean, median, Q1, Q3, mode, and range.

23. The following is a frequency distribution showing the hourly wages of a sample of coal workers:

HOURLY WAGE	FREQUENCY
$10	900
$12	2100
$15	3400
$18	1600
$20	1800
$25	200

(a) Calculate the mean, median, Q1, Q3, mode, and range.

24) [Check with your instructor to see if you are responsible for this topic – scatter plots – on the exam.] A researcher wants to determine whether there is a relationship between high school average and scores on the CPA exam.



a) The chart above is known as a _____ . b) The data suggest what kind of relationship, if any: _____

SOLUTIONS

1. a) ratio b) ratio c) nominal d) interval e) nominal f) ordinal g) ratio h) interval i) interval j) ratio k) ordinal l) nominal m) ordinal n) ordinal

2. a) Discrete b) Continuous c) Continuous d) Continuous e) Continuous f) Discrete g) Continuous h) Continuous i) Discrete j) Continuous k) Discrete

3. a) Statistic b) Parameter c) Statistic d) Parameter e) Parameter f) Parameter g) Parameter

4. a) frequencies, proportions b) frequencies, proportions, median, percentiles and other quantiles c) frequencies, proportions, median, percentiles and other quantiles, mean, standard deviation

5. a) simple random sample b) statistics c) census d) parameter
6. a) 0 b) 68.26% c) 75% d) 97.5%
7. Z = -2.5 Proportion of Hyundai Dreds that will die within 10 years is 0.62%.
8. a) mutually exclusive b) independent c) no d) yes
9. See Definitions

10. mean =5.938 units, median = 6 units, Q1=3.5 units, Q3=9 units, mode =10 units, range = 10 units, IQR = 5.5 units, s.d. = 3.17 units, C.V. = 53.4%,
Z-score for 10 is 1.28 [note that the Z-score is a "pure" number]

11. mean=$20, median=$20, Q1=$13, Q3=$21.50, mode=$20, range=$30, IQR=$8.50, s.d.=$8.45, C.V.=42.25%, Z-score for $40 is 2.37

12. P(male) = .40, P(female and smoker) = .13, P(female or smoker) = .75, P(male|smoker) = 53.57% , P(smoker|male) = .375,
Smoking and sex are not independent since P(S|M) is .375 and P(S|F) is .217. This indicates that the probability of being a smoker is higher for men than for women.

13. (a) P(C and T)=.05 ; P(C)= .27; P(C|T)=.50; P(C|T')= .244; P(C or T) =.32.
(b) since P(C), P(C|T), and P(C|T') are not equal. Indeed, 50% of those with tattoos have criminal records vs. 24.4% without.

14. (a) 28.12% (b) 7.35% (c) 18.14% (d) $22.86

15. (a) .1056 (b) .3085 (c) .2857 (d) 25.12 mpg [(x - 20) / 4 = 1.28, solve for x]
(e) 14.64 mpg [(x - 20) / 4 = -1.34, solve for x]

16. Answer: .2621
17. Answer: .000266
18. Answer: .19371
19. Answer: .0512
20. (a) 18! (b) 18P3 = 4,896. (c) 18C3 = 816
21. 6% that it works and 94% that it does not work.

22. mean = 6.52 days, median = 7 days, Q1= 4 days, Q3=9 days, mode = 9 days, range = 8 days.
23. mean=$15.50, median=$15, Q1=$12, Q3=$18, mode=$15, range=$15.
24. a) scatter plot b) linear and positive