

In almost all problems, I have given the answers to four significant digits. If your answer is slightly different from one of mine, consider that to be roundoff error and mark the closely matching one. If your answer differs from the closest one of mine by more than one percent (meaning the ratio of yours to mine is less than 0.99 or greater than 1.01), then mark “J) None of the preceding”.

1. A start-up company is building a database of customers and sales information. For each customer it records name, ID number, region of the country (1=East, 2=South, 3=Midwest, 4=West), date of last purchase, amount of purchase, and item purchased. Identify the “Whats” in this database and say how many of them there are.

A) 0 B) 1 C) 2 D) 3 E) 4 F) 5 G) 6 H) 7 I) 8 J) None of the preceding

2. A survey of autos parked in student and staff lots at a large university classified the brands by country of origin, as seen in the table.

	<b>Student</b>	<b>Staff</b>
<b>American</b>	324	123
<b>European</b>	105	29
<b>Asian</b>	237	97

What percent of the students owned American cars?

A) 33.63% B) 36.64% C) 39.64% D) 42.64% E) 45.65% F) 48.65% G) 51.65% H) 54.65% I) 57.66% J) None of the preceding

3. A 1975 article in the magazine Science examined the graduate admissions process at Berkeley for evidence of gender bias. The table below shows the number of applicants accepted to each of four graduate programs.

	<b>Males accepted (of applicants)</b>	<b>Females accepted (of applicants)</b>
<b>Program 1</b>	511 of 825	89 of 108
<b>Program 2</b>	352 of 560	17 of 25
<b>Program 3</b>	137 of 407	132 of 375
<b>Program 4</b>	22 of 373	24 of 341

What is the percentage of males admitted overall minus the percentage of females admitted overall?

A) 10.26% B) 11.13% C) 12.00% D) 12.87% E) 13.74% F) 14.61% G) 15.48% H) 16.35% I) 17.22% J) None of the preceding

4. What phenomenon is the data in Problem 3 an example of?

A) Fechner's Law B) Regression towards Mediocrity C) Pearson's Correlation D) Galton's Correlation E) Venn's Diagrams F) Newton's Method G) Playfair's Postulate H) Simpson's Paradox I) Area Principle J) None of the preceding

5. In one of the exercises on displaying numeric data, a histogram of singers' heights showed what feature?

- A) Uniformity B) Leptokurtosis C) Platykurtosis D) Right skewness E) Left skewness F) Bimodality G) Trimodality  
H) Unimodality I) Homoskedasticity J) None of the preceding

6. In one of the exercises involving the plotting of numeric data, a histogram had a gap in the middle corresponding to a very noticeable drop in the time series plot. What numeric variable was being graphed in both cases?

- A) Atmospheric  $\text{N}_2\text{O}$  B) Atmospheric  $\text{CO}_2$  C) Ozone layer thickness D) Global warming temperatures E) Hurricanes  
F) Inflation rate G) Deflation rate H) Drunk driving deaths I) Prime rate J) None of the preceding

7. Find the interquartile range of the following five boys' reading scores:

27.3, 46.9 12.8 33.3 45.2

Use the definition given in the book, not the calculator's definition.

A) 4.40 B) 7.10 C) 9.80 D) 12.50 E) 15.20 F) 17.90 G) 20.60 H) 23.30 I) 26.00 J) None of the preceding

8. An incoming freshman took her college's placement exams in Urdu and mathematics. In Urdu, she scored 81, and in math, 87. The overall results on the Urdu exam had a mean of 72, and a standard deviation of 7, while the mean math score was 68, with a standard deviation of 11. To see on which exam she did better relative to the other freshmen, she calculates the two z-scores and compares them. Find the difference between the larger z-score and the smaller z-score.  
A) 0.4416 B) 0.4773 C) 0.5130 D) 0.5487 E) 0.5844 F) 0.6201 G) 0.6558 H) 0.6915 I) 0.7272 J) None of the preceding

9. Fifty-three men qualified for the men's alpine downhill race in Salt Lake City. The gold medal winner finished in 1 minute 39.13 seconds. All competitors' times were given in a table for which the mean was 102.71 seconds, with a standard deviation of 3.01 seconds. If the Normal model is appropriate, what percent of times will be less than 100 seconds?
- A) 12.35% B) 13.56% C) 14.77% D) 15.98% E) 17.19% F) 18.40% G) 19.61% H) 20.82% I) 22.03% J) None of the preceding

10. Every Normal model is defined by its parameters. For the normal model with  $\sigma = 250$  and 80% above 1000, what is the value of  $\mu$ ?
- A) 684.4 B) 789.6 C) 894.8 D) 1000.0 E) 1105.2 F) 1210.4 G) 1315.6 H) 1420.8 I) 1526.0 J) None of the preceding

11. The top men's and women's 500-m speed skating times in the 2002 Winter Olympics were given to us. We were told that the mean finishing time was 73.46 seconds with a standard deviation of 3.33 seconds. If the Normal model is appropriate, what percent of the times should be within 1.05 seconds of 73.46?  
A) 20.41% B) 22.58% C) 24.75% D) 26.92% E) 29.09% F) 31.26% G) 33.43% H) 35.60% I) 37.77% J) None of the preceding

12. Does how long children remain at the lunch table help predict how much they eat? The data below are for seven toddlers observed over several months at a nursery school:

Time (minutes):	21.4	30.8	37.7	33.5	32.8	39.5	22.8
Calories	472	498	465	456	423	437	531

Find the correlation for these data if the time is changed from minutes to hours.

A) -0.8811 B) -0.6322 C) -0.3833 D) -0.1344 E) 0.1145 F) 0.3634 G) 0.6123 H) 0.8612 I) 1.1101 J) None of the preceding

13. In a set of data on SAT scores, the correlation between the SAT math score and the SAT verbal score was 0.8134. The verbal scores averaged 500 with a standard deviation of 90, and the math scores averaged 550 with a standard deviation of 80. A few students every year score a perfect 1600. Based on this model, and assuming Normality, what would such a student's residual be for her math score?
- A) 17.34 B) 21.28 C) 25.22 D) 29.16 E) 33.10 F) 37.04 G) 40.98 H) 44.92 I) 48.86 J) None of the preceding

14. In the Western Hemisphere study of female life expectancy and number of children a woman has, one country yielded an extremely influential point in the scatterplot. Which country was it, and why was its point so influential?
- A) Costa Rica, high births B) Brazil, low births C) Uruguay, short life D) Paraguay, long life E) Panama, high births F) Guatemala, short life G) Peru, long life H) Bolivia, long life I) Canada, long life J) None of the preceding

15. Here are summary statistics for Olympic long jumps and high jumps, in inches:

Long jump: Mean = 304.10 StdDev = 23.71

High jump: Mean = 86.04 StdDev = 8.26

Correlation between long and high:  $r = 0.853$

What is the intercept of the line of regression for estimating high jump from long jump?

A) -6.484 B) -5.945 C) -5.406 D) -4.867 E) -4.328 F) -3.789 G) -3.250 H) -2.711 I) -2.172 J) None of the preceding

16. Here are summary statistics for Olympic long jumps and high jumps, in inches:

Long jump: Mean = 304.10 StdDev = 23.71

High jump: Mean = 86.04 StdDev = 8.26

Correlation between long and high:  $r = 0.853$

What is the slope of the line of regression for estimating long jump from high jump?

A) 1.744 B) 1.979 C) 2.214 D) 2.449 E) 2.684 F) 2.919 G) 3.154 H) 3.389 I) 3.624 J) None of the preceding



17. In Chapter 3 there was a misleading picture portraying the relative numbers of first, second, and third class passengers and crew aboard the Titanic on its fateful maiden voyage. Why is the picture misleading?

A) Fechner's Law B) Regression towards Mediocrity C) Pearson's Correlation D) Galton's Correlation E) Venn's Diagrams F) Newton's Method G) Playfair's Postulate H) Simpson's Paradox I) Area Principle J) None of the preceding

18. What is a cumulative percentage graph (like the one of the cholesterol values from the Framingham study) called?

A) Ogive B) Minimax C) Regression D) Probit E) Logistic F) Pedigree G) Spline H) Lowess I) Density J) None of the preceding

19. If the following three numbers are converted to z-scores, what does the smallest one become?

1, 3, 10

A) -1.1731 B) -1.0407 C) -0.9083 D) -0.7759 E) -0.6435 F) -0.5111 G) -0.3787 H) -0.2436 I) -0.1139 J)  
None of the preceding

20. A student scores 90 on the first exam in a chemistry course. Assume the correlation between exams 1 and 2 is 0.7, and that both exams have means of 70 and standard deviations of 10. According to regression, the student should not expect to do as well as 90 on the second exam. What score should the student expect?

A) 72 B) 74 C) 76 D) 78 E) 80 F) 82 G) 84 H) 86 I) 88 J) None of the preceding

21. In a biology course, the regression slope for the second exam on the first exam is 0.75, while the regression slope for the first exam on the second is 0.90. What is the correlation coefficient between the two exams?

A) 0.5762 B) 0.6171 C) 0.6580 D) 0.6989 E) 0.7398 F) 0.7807 G) 0.8216 H) 0.8625 I) 0.9034 J) None of the preceding

22. Which federal agency had the highest rate for both assaults per 1000 and killed-or-injured per 1000?

A) FBI B) DEA (Drug Enforcement Agency) C) IRS (Internal Revenue Service) D) Postal Service E) Capitol Police  
F) Customs G) National Park Service H) U.S. Marshal Service I) Secret Service J) None of the preceding

23. Assume that the distribution of SAT math scores is Normal with mean 500 and standard deviation 100. What percent of the population is expected to score 630 or more on the SAT math exam? (Assume there is no rounding, meaning that the scores have a continuous distribution.)

A) 6.681% B) 7.030% C) 7.379% D) 7.728% E) 8.077% F) 8.426% G) 8.775% H) 9.124% I) 9.473% J) None of the preceding

24. John Tukey, inventor of the boxplot, the stem-and-leaf plot, and scads of other simple and effective ways to look at data, also had the knack of coining words and phrases that have become part of standard English. Which one of the following words did Tukey invent?

A) Software B) Byte C) Nybble D) Mouse E) Bluetooth F) Median G) Standard deviation H) Laser I) Correlation J) None of the preceding

25. Which one of the following people was Francis Galton's cousin?

A) Morse B) Edison C) Pearson D) Einstein E) Darwin F) Huxley G) Mendel H) Faraday I) Carnegie J) None of the preceding

26. (The Bonus Question) In a Foxtrot cartoon printed in our book, what did Peter Fox expect his teacher to do with his incomplete homework assignment?

A) Stem-and-Leaf plot B) Boxplot C) Standardize D) Correlate E) Regress F) Extrapolate G) Calculate residuals H) Normal Plot I) Toss it down the stairs J) None of the preceding