Free-Response Questions

Do #1, 2, & 8

Directions: You must show all work and indicate the methods you use. You will be graded on the correctness of your methods and on the accuracy of your final answers.

Ten Open-Ended Questions

1. Average home attendance and number of home wins for the 2009–2010 NBA Pacific Division teams were as follows:

	Lakers	Suns	Clippers	Warriors	Kings
Average attendance	18,997	17,648	16,343	18,027	13,254
Home wins	34	32	21	18	18

   (a) Does a winning team bring out the fans? Can average attendance be predicted from number of wins? Find the equation of the best-fitting straight line.
   (b) Interpret the slope.
   (c) Predict the average attendance for a team with 25 home wins.
   (d) What number of home wins will predict an average of 17,000 fans?
   (e) What is the residual for the Lakers average attendance?

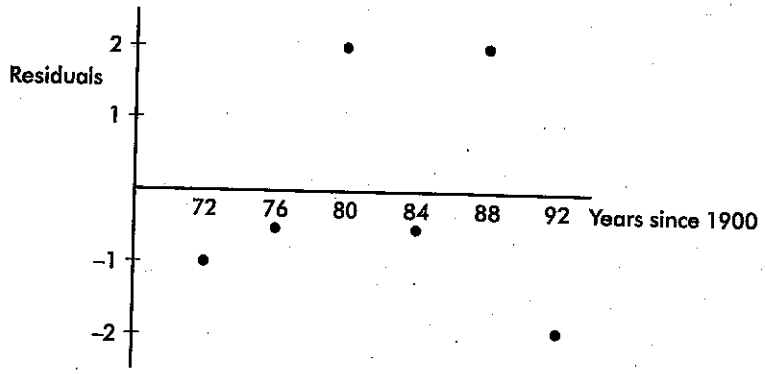
2. The shoe sizes and the number of ties owned by ten corporate vice presidents are as follows:

Shoe size, x:	8	9.5	9	11	9	9.5	8.5	9	9	9.5
Number of ties, y:	10	10	8	15	12	13	16	7	12	4

   (a) Draw a scatterplot for these data.
   (b) Find the correlation r.
   (c) Can we find the best-fitting straight-line approximation to the above data? Does it make sense to use this equation to predict the number of ties owned by a corporate executive who wears size 10 shoes? Explain.

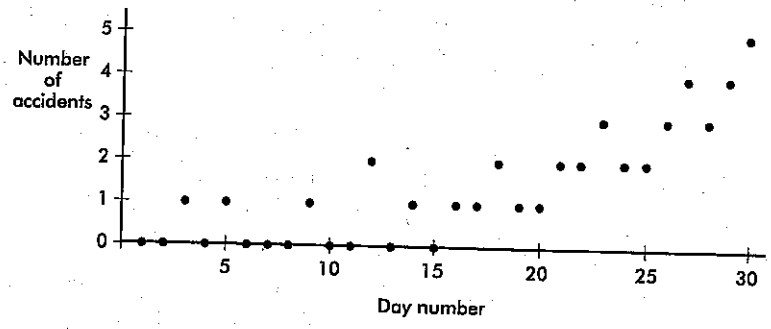
8. Data show a trend in winning long jump distances for an international competition over the years 1972–92. With jumps recorded in inches and dates in years since 1900, a least squares regression line is fit to the data. The computer output and a graph of the residuals are as follows:

```
R squared = 92.1%
Variable    Coefficient    SE of Coeff    t-ratio    Prob
Constant      256.576         11.59         22.1     0.0001
Year          0.95893         0.141         6.81     0.0024
```



(a) Does a line appear to be an appropriate model? Explain.
(b) What is the slope of the least squares line? Give an interpretation of the slope.
(c) What is the correlation?
(d) What is the predicted winning distance for the 1980 competition?
(e) What was the actual winning distance in 1980?

9. A scatterplot of the number of accidents per day on a particular interstate highway during a 30-day month is as follows:



(a) Draw a histogram of the frequencies of the number of accidents.
(b) Draw a boxplot of the number of accidents.
(c) Name a feature apparent in the scatterplot but not in the histogram or boxplot.
(d) Name a feature clearly shown by the histogram and boxplot but not as obvious in the scatterplot.

Free-Response Questions

DO #1

Directions: You must show all work and indicate the methods you use. You will be graded on the correctness of your methods and on the accuracy of your final answers.

Eleven Open-Ended Questions

1. The belief that sugar causes hyperactivity is the most popular example of how people believe that food influences behavior.
   (a) Many parents, witnessing the aftermath of cake and ice cream at birthday parties, attest to the relationship between sugar and hyperactivity. Are these observational studies or experiments? Explain.
   (b) Name a confounding variable to the above and explain how it is confounded with sugar.
   (c) Design a study to allow a parent to determine whether sugar causes hyperactivity in his/her child and explain why double blinding is so important here.

2. Suppose a new drug is developed that appears in laboratory settings to completely prevent people who test positive for human immunodeficiency virus (HIV) from ever developing full-blown acquired immunodeficiency syndrome (AIDS). Putting all ethical considerations aside, design an experiment to test the drug. What ethical considerations might arise during the testing that would force an early end to the experiment?

3. A new weight-loss supplement is to be tested at three different levels (once, twice, and three times a day). Design an experiment, including a control group and including blocking for gender, for 80 overweight volunteers, half of whom are men. Explain carefully how you will use randomization.

4. Two studies are run to measure the health benefits of long-time use of daily high doses of vitamin C. Researchers in the first study send a questionnaire to all 50,000 subscribers to a health magazine, asking whether they have taken large doses of vitamin C for at least a 2-year period and what they perceive to be the health benefits, if any. The response rate is 80%. The 10,000 people who did not respond to the first mailing receive follow-up telephone calls, and eventually responses are registered from 98% of the magazine subscribers. Researchers in a second study take a group of 200 volunteers and randomly select 100 to receive high doses of vitamin C while the others receive a similar-looking, similar-tasting placebo. The volunteers are not told whether they are receiving the vitamin, but their doctors know and are asked to note health changes during a 2-year period. Comment on the designs of the two studies, remarking on their good points and on possible sources of error.

Free-Response Questions

Do #1

Directions: You must show all work and indicate the methods you use. You will be graded on the correctness of your methods and on the accuracy of your final answers.

Two Open-Ended Questions

1. The time it takes Steve to walk to school follows a normal distribution with mean 30 minutes and standard deviation 5 minutes, while the time it takes Jan to walk to school follows a normal distribution with mean 25 minutes and standard deviation 4 minutes. Assume their walking times are independent of each other.

   (a) If they leave at the same time, what is the probability that Steve arrives before Jan?

   (b) How much earlier than Jan should Steve leave so that he has a 90% chance of arriving before Jan?

2. (a) Two components are in series so the failure of either will cause the system to fail. The time to failure for a new component is normally distributed with a mean of 3000 hours and a standard deviation of 400 hours. One of the components has already run for 2500 hours, while the other has run for 2800 hours. Assuming component failures are independent, what is the probability the system survives for 10 more days (240 hours)?

   (b) If the components are put in parallel, the system will fail only if both components fail. If the two components in (a) are put in parallel, what is the probability that the system survives for 10 more days?

Free-Response Questions

Do #1-3

Directions: You must show all work and indicate the methods you use. You will be graded on the correctness of your methods and on the accuracy of your final answers.

Four Open-Ended Questions

1. Car insurance policies for teenagers are typically higher than insurance rates for adult drivers because teenagers, with their lesser driving experience, are considered to be a "high risk" population. The average yearly cost of teenage auto insurance is \$2171 with a standard deviation of \$612. An SRS is taken of 100 teenage drivers.
   (a) Explain why there is insufficient information to determine the probability a randomly chosen teenage driver pays over \$2300 a year for auto insurance.
   (b) What are the mean and standard deviation for the sampling distribution for $\bar{x}$, the mean amount paid for insurance.
   (c) What is the probability that the average amount paid in the sample is over \$2300?
2. A health foods magazine conducts a large random sample of honey prices with the resulting data showing a roughly normal distribution with mean \$1.87/lb and standard deviation \$0.193/lb.
   (a) What is the interquartile range (IQR) of the distribution of honey prices?
   (b) What is the probability that a majority in a random sample of three honey prices are over \$2.00/lb?
   (c) What is the probability that the mean price in a random sample of three honey prices is over \$2.00/lb?
3. Suppose that the heights of college basketball players are normally distributed with a mean of 74 inches and a standard deviation of 4 inches.
   (a) What percentage of players are over 7 feet?
   (b) What is the probability that at least one of ten randomly selected players is over 7 feet?
   (c) What is the probability that the mean height in an SRS of size 10 is over 6 feet?
   (d) If an outlier is defined to be any value more than 1.5 interquartile ranges above the third quartile or below the first quartile, what percentage of heights of players are outliers?