The next three questions refer to the Harvard Nurses Health Study, which began in 1976. The study included more than 120,000 women who in successive years were asked to answer questions about overall health, smoking, diet, use of birth control pills, and postmenopausal estrogen supplements. Epidemiologists followed up each interview with a review of the women's medical records. Today after more than a decade the research team found that women on an extremely high-fat diet (nearly 50% of energy intake as fat) are at no more risk of breast cancer than those who adhere to an extremely lean diet (less than 29% fat). They did find a clear association between high fat diets and colon cancer, however.

1. What type of study is this?
   a) randomized experiment  
   b) observational study  
   c) sample survey  
   d) case study  
   e) opinion poll

2. What response variables were mentioned in the description?
   a) nurses and medical records  
   b) high fat diet and low fat diet  
   c) breast cancer and colon cancer  
   d) smoking, diet, use of birth control pills, and postmenopausal estrogen supplements  
   e) epidemiologists and Harvard University Nurses

3. Amount of fat in a person's diet is,
   a) a confounding variable  
   b) a response variable  
   c) an explanatory variable  
   d) a placebo  
   e) a unit

4. The evidence that smoking causes lung cancer, while strong, is not completely convincing. The most important reason for this is,
   a) Studies were conducted by doctors.  
   b) Randomized human smoking experiments can't be done.  
   c) Not all smokers die of lung cancer.  
   d) Volunteer samples are unreliable.  
   e) Not all lung cancer victims were smokers.  

5. Glamour magazine asked readers to fill out a questionnaire printed in the magazine and mail it in. The questionnaire asked about childhood sexual abuse. Forty-one percent of the respondents said they had been sexually abused as children. This figure is probably biased because,
   a) the sample was stratified.  
   b) the sample was voluntary.  
   c) the sample size was too small.  
   d) both a) and b)  
   e) None of the above, there is no reason to doubt the accuracy of the results.

6. In the Salk vaccine field trial of 1954 participating children received vaccine or placebo in a randomized, placebo-controlled, double-blindfolded experimental design. The response variable was whether or not the child was diagnosed with polio over the following summer. The main purpose of blindfolding was,
   a) to ensure an equal mix of rich and poor kids in the placebo and vaccine groups  
   b) the blindfold made the child less frightened during the injection  
   c) to avoid biasing parents' behavior and doctors' diagnoses  
   d) to avoid anger at not getting the real vaccine  
   e) to save money on vaccine

7. In a case-control study of smoking and death by lung cancer, the cases are ___ and the controls are ______.
   a) smokers / non-smokers  
   b) people who died of lung cancer / people who died of other causes  
   c) real cigarettes / placebo cigarettes  
   d) annual revenues of cigarette manufacturers / number of new lung cancer cases  
   e) randomized / blindfolded.

8. Right before a presidential election, the Gallup Poll increases its sample size to 4000 instead of the usual 1800. The reason Gallup does this is,
   a) to reduce bias.  
   b) to get a random sample.  
   c) to reduce the margin of error.  
   d) to get better coverage of special interest groups.  
   e) to use better trained interviewers.
9. A common source of bias in sample surveys is,
   a) use of random digits to select the sample
   b) failure to follow-up not at homes
   c) use of poorly worded questions
   d) both b) and c)
   e) all of the above

10. A market research company did a study for the National Automobile Dealers Association (NADA). They interviewed a scientific sample of 100 adult women who are permanent US residents (AWPR's) and found that 60% of them would prefer to buy a car from a "No Dicker" dealership (one price, no sales commissions).

   Identify the population:
   a) "No Dicker" dealerships
   b) Adult women who are permanent US residents
   c) AWPR's who buy from "No Dicker" dealerships.
   d) NADA members
   e) Two or more of the above.

11. Children living in homes with gas cook stoves have slightly more respiratory infections than children living in homes with electric cook stoves. However, families with gas cook stoves more often have one or more cigarette smokers than families with electric cook stoves. Without further investigation, we can't say if it's the gas or the cigarettes that causes the increase in respiratory symptoms. In other words, type of cooking fuel and parental smoking habits are,
   a) common effects.
   b) confounded variables.
   c) randomized assignments.
   d) response variables.
   e) treatment and placebo.

12. The Iowa Poll of farmers conducted April 3-8 of 1988 found that 55% of farmers blamed local lenders for their credit problems. Here is a description of how the poll was conducted.

   "The IOWA POLL is based on 304 interviews with Iowa farmers 18 years of age or older. Professionally trained interviewers contacted households with telephone numbers randomly selected by a computer, eliminating interviewers' choices in selecting persons to be included. Percentages based on the full poll sample are subject to a maximum margin of error of plus or minus _____ percentage points."

   Express 55% as a proportion. What is the margin of error of the proportion?
   Proportion:__________  Margin of Error:__________

13. A researcher wants a sample of Iowa Farms of 200 or more acres. She selected 10 Iowa counties at random. She then sent a research assistant to each of the 10 counties with instructions to prepare a list of 200+ acre farms from the county plat books. She then took a random sample of 20 farms from each list. What type of sample is this?
   a) A simple random sample
   b) A multistage cluster sample
   c) A systematic sample
   d) A convenience sample
   e) A census

14. The quality engineer for a large grocery store chain wants to know what percent of eggs shipped in cardboard cartons and what percent of eggs shipped in Styrofoam cartons get cracked. The number of eggs shipped in cardboard is 10,000 per day and the number shipped in Styrofoam is 100,000 per day. He plans to take a sample of each type from one day's shipments and determine the percent of cracked eggs in the sample. Which of the following samples would have a smaller margin of error?

   A: a 2% sample of cardboard, n=200
   B: a 1% sample of Styrofoam, n=1000

   a) the cardboard sample (A) will have a smaller margin of error than the Styrofoam sample (B) because it is a bigger proportion of its population.
   b) the Styrofoam sample (B) will have a smaller margin of error than the cardboard sample (A) because the sample size is larger.
   c) neither sample will have a small enough margin of error to be of any practical use because they are such small fractions of the populations
   d) both a) and c)
   e) a), b) and c)

15. Scores on a statistics exam were approximately normally distributed with mean 70 and standard deviation 10. The instructor wants to give A's to the top 16 percent of the class. Use the 68/95 rule to determine which scores should get A's?
   a) 70 and up  b) 75 and up  c) 80 and up
   d) 84 and up  e) 90 and up
The next two items refer to these 14 observations:

| 4.1 | 6.4 | 1.1 | 5.5 | 4.6 | 3.2 | 5.2 |
| 8.7 | 7.5 | 4.8 | 5.8 | 7.1 | 3.5 | 2.5 |

16. Here are some descriptive statistics for these data:

- n = 14
- Mean = 5.0
- Min = 1.1
- Q1 = 2.5
- Median = 5.0
- Q3 = 6.4
- Max = 8.7

Which of the statistics was/were incorrectly calculated:

a) the first and/or third quartile
b) the mean
c) the min and/or max
d) the median
e) all are correct

17. A set of data has median 9 and mean 25. Which of the following stemplots shows the shape of the distribution of these data (the stems are not shown, there is nothing to calculate, base your answer on the shapes).

18. Diastolic blood pressure readings for a sample of 200 men are approximately normally distributed with mean 70. One hundred and thirty-six (136) of the men had readings between 55 and 85. Use the 68/95 rule to decide whether the following is closest to the standard deviation of the blood pressure readings.
a) 1 b) 15 c) 30 d) 45 e) 90

The UNM/SIP Study.
Newborns in Albuquerque with non-smoking parents were recruited for an 18-month study of the effect of indoor nitrogen dioxide on respiratory disease. Parents agreed to keep a diary of their infant's coughs, colds and other respiratory symptoms and agreed to have an air-sampling device installed in the child's bedroom. Nitrogen dioxide is a colorless and odorless gas. Air sampling results were not reported to the parents until the end of the study.

19. The UNM/SIP Study was a comparison of two groups of infants: those living in high nitrogen dioxide homes and those living in low nitrogen dioxide homes. It was an observational study rather than an experiment because,

a) infants could not be randomly assigned to high- or low-nitrogen dioxide households
b) the parents were not blindfolded, they knew what group their infant was in
c) there is no outcome variable and no explanatory variable
d) the explanatory variable was confounded with cigarette smoking
e) both a) and b)

20. Which of the following should not be used with highly skewed data?

a) the 68/95 rule b) a five number summary
c) a stem plot d) the median
e) both a) and d)

21. Scores on a commonly used IQ test are approximately normally distributed with mean 100 and standard deviation 15. In a group of 1000 people, about how many would have IQ's above 115?
a) about 5 b) about 25 c) about 50
d) about 160 e) some other number

The next question refers to these tables:

<table>
<thead>
<tr>
<th>A</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Yes</td>
<td>32</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>Yes</td>
<td>10</td>
<td>30</td>
</tr>
</tbody>
</table>

Chi Square 0.607 15.238

22. Which table shows that the presence of the risk factor (the "Yes" row) significantly increases the probability of the adverse outcome (the "yes" column)?

a) A b) B c) Both A and B d) Neither A nor B

a) .105 b) .210 c) .350
d) .400 e) .650
23. An article in the New England Journal of Medicine in June of 1994 reported the results of a study of several years of medical records of 750 women with silicone breast implants and a matching group of 1500 women who did not have breast implants. Rates per 10,000 per year for three diseases which might have a statistical association with breast implants are shown in this table:

<table>
<thead>
<tr>
<th>Condition</th>
<th>Women with Implants</th>
<th>Women without Implants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connective Tissue Disease</td>
<td>8.6</td>
<td>8.1</td>
</tr>
<tr>
<td>Hashimoto's Thyroiditis</td>
<td>17.1</td>
<td>17.0</td>
</tr>
<tr>
<td>Cancer other than breast</td>
<td>22.2</td>
<td>20.2</td>
</tr>
</tbody>
</table>

The relative risk for connective tissue disease is,

a) less than 1  
b) between 1 and 1.10  
c) between 1.11 and 1.20  
d) between 1.21 and 1.30  
e) greater than 1.30

24. What is the risk factor?

a) cancer other than breast  
b) connective tissue disease  
c) Hashimoto's Thyroiditis  
d) silicone breast implant  
e) 750 women

25. There is a negative statistical association between heart attacks and amount of hair. Guys with lots of hair are less likely to have heart attacks than bald guys (less hair, more heart attacks). But of course getting older is the real cause. Old guys have less hair and more heart attacks than young guys. What idea does this illustrate?

a) Women don't go bald.  
b) Heart attacks cause baldness.  
c) Both variables result from a common cause  
d) The association is a coincidence  
e) Baldness causes heart attacks.