

## Stat 301 Review (Test 1)

*The following concepts are all fair game for the first exam. Use the following check list and the review problems to help you in your studying. (Most of the review problems come from old tests.)*

Concept	Check List
Individuals (units)	
Categorical and quantitative variables	
Bar graphs and pie charts	
Stemplots (standard, back-to-back, and split)	
Histogram	
Examining distributions: <ul style="list-style-type: none"> <li>• Outliers and influential observations</li> <li>• Shape (number of modes and symmetric versus skewed)</li> <li>• Center of distribution</li> <li>• Spread of distribution</li> </ul>	
Center (mean versus median)	
Spread (range, IQR, variance and standard deviation)	
Resistant measure	
5-Number summary	
1.5(IQR) rule for outliers	
Boxplots (modified and side-by-side)	
Unit/subject	
Population versus sample	
Census	
Experiment versus observational study	
Non random / convenience sampling versus random probability based sampling	
Types of samples: <ul style="list-style-type: none"> <li>• Voluntary response sample</li> <li>• Simple random sample (SRS)</li> <li>• Stratified random sample</li> <li>• Multistage random sample</li> <li>• Capture-recapture sample</li> </ul>	
Anecdotal evidence	
Types of sampling bias: <ul style="list-style-type: none"> <li>• Undercoverage</li> <li>• Nonresponse</li> <li>• Response bias</li> </ul>	
Sampling variability	
Parameter versus statistic	
Sampling distribution	
Treatments	
Factors and factor levels	
Explanatory variable versus response variable	
Control group	
Placebo	
Bias	

3 principals of experimental design	
Type of experimental designs <ul style="list-style-type: none"> <li>• Completely randomized</li> <li>• Block</li> <li>• Matched-pairs</li> </ul>	
Double blind experiment	
Lurking variables	
Confounding	
Common response	
Causation	
Normal distribution <ul style="list-style-type: none"> <li>• 68-95-99.7 rule</li> <li>• Standard normal distribution (converting to Z)</li> <li>• Using the normal table</li> <li>• “Forward” normal distribution problems</li> <li>• “Backward” normal distribution problems</li> </ul>	
Central Limit Theorem	
Normal approximations with sample means	
Statistical process control <ul style="list-style-type: none"> <li>• Examples of processes</li> <li>• Reading , creating, and understanding process control charts</li> <li>• Statistical control</li> <li>• Control versus capability</li> </ul>	
Ethics of doing experiments with humans <ul style="list-style-type: none"> <li>• Review board</li> <li>• Informed consent</li> <li>• Individual data kept confidential</li> </ul>	
Ethics of doing experiments with animals (3 Rs) <ul style="list-style-type: none"> <li>• Reduce</li> <li>• Refine</li> <li>• Replace</li> </ul>	

*The problems below have been taken from old tests.*

1. Choose the letter for the graph that is most appropriate for displaying/analyzing the data for the following questions. Each letter may be used once, more than once, or not at all. For some questions, 2 or more graphs may be appropriate.

**A. Stem Plot      B. Side-by-side boxplot      C. Pie Chart**

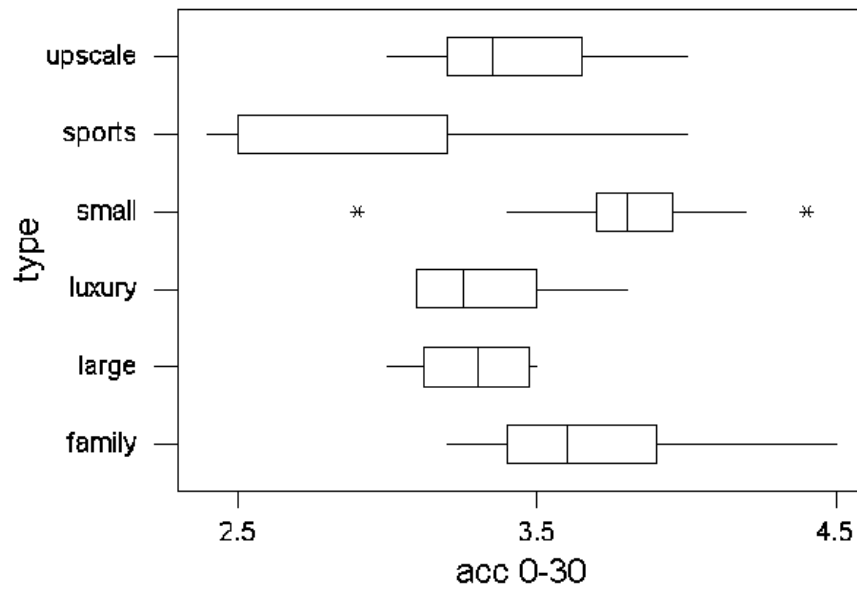
- \_\_\_\_\_ a. What % of golfers use Top-Flight, Nike, Titleist, Callaway, and all other golf balls?
- \_\_\_\_\_ b. How far does a Titleist golf ball go? Ten Titleist balls were hit by a mechanical golfer and the distances recorded.
- \_\_\_\_\_ c. Which type of ball goes farther? Ten Titleist and Ten Nike balls were hit by a mechanical golfer and the distances recorded.

2. George has a taxi company with 17 taxicabs. George keeps records of many things including the miles driven per year for each taxi. Below are the miles driven in 2004 (rounded to the nearest thousand miles) for each of his 17 taxicabs.

28 30 42 34 29 31 52 28 32 29 28 45 39 27 36 25 29

- a. Generate a stemplot of the data.
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
- b. Looking at the stemplot you developed, how would you describe the shape of the data distribution?
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
- c. The mean is 33.2 and the standard deviation is 7.2. What is the five-number summary?
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
- d. Are there any outliers? Justify your answer.
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
- e. Make a modified boxplot. Be sure to label the plot.
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
- f. What measures of center and spread are most appropriate for this data? Why?

3. The 1999 Consumer Reports New Car Buying Guide reported on the number of seconds required for a variety of cars to accelerate from 0 to 30 mph. The cars were also classified in six categories according to type. The following boxplots display these distributions:



- a. Which type of car tends to accelerate the fastest?
- b. Which type of car has the smallest IQR ?
4. The following data represent the number of insect eggs per leaf that were recorded on each of 14 randomly selected birch leaves:  
 22, 10, 12, 17, 19, 24, 25, 26, 30, 34, 26, 21, 20, 19
- a. Construct a split stem plot for these data.
- b. Is the distribution above symmetric, skewed left, or skewed right?
- c. Find the **median** and the **interquartile range** for these data.

5. An education researcher wishes to study the association between GPA and whether a student takes the SAT as a sophomore for Harrison High School students. There are 800 sophomores students in the school, and the researcher randomly selects 150 of these students to record their sophomore-year cumulative GPA and whether they took the SAT.
  - a. What is the unit/subject for this study?
  - b. What is the population for this study?
  - c. What is the sample for this study?
  - d. What are the response variables for this study? Are they categorical or quantitative?
  - e. What type of sampling design did the researcher use?
  
6. You are in need of surgery for an abdominal hernia. Consequently, you are interested in comparing Hospital A and Hospital B, which serve your community. Hospital A loses 3% of its surgery patients, whereas Hospital B loses only 2% of its surgery patients. Is it clear that you definitely should use Hospital B? Explain your answer.
  
7. Suppose the local Childhood Lead Poisoning Prevention Council in a metropolitan area in western Tennessee undertakes the responsibility of determining the proportion of homes in their city that have unsafe lead levels. Because of the great expense involved in performing spectrometric testing they decide to test only some of the homes. The Council assumes that houses built prior to 1970 are more likely to have unsafe lead levels. Consequently, they divide their population into homes built prior to 1970 and homes built after 1970. They then take a random sample of 100 homes built prior to 1970 and 100 homes built after 1970 and record the lead levels for each of the 200 homes. This is an example of a
  - a. block design.
  - b. simple random sample.
  - c. stratified random sample.
  - d. completely randomized design.
  - e. multistage random sample.
  
8. Suppose in the process of obtaining the population of all homes in problem #2, several neighborhoods were mistakenly omitted. What type of bias will this study potentially suffer from?
  - a. Nonresponse
  - b. Response bias.
  - c. Double blind.
  - d. Placebo effect.
  - e. Undercoverage.

9. Do piano lessons improve the spatial-temporal reasoning of preschool children? Researchers gave one group of 34 preschool children 6 months of piano lessons and another group of 34 preschool children 6 months of computer lessons. The children were given pre-tests and post-tests in reasoning, and the improvement was recorded (integers ranging from  $-4$  to  $9$ ). The average improvement for the piano lesson children was 3.618 points. The average improvement for the computer lesson children was 0.386 points.
- What is the population?
  - What is the sample?
  - What is the response variable? Is it categorical or quantitative?
  - Is this an observational study or an experiment? Why?
  - What is the design used here? Give a detailed outline.
10. George W. Bush claims that he won more votes (in the 2004 presidential election) than any other president in U.S. history. Does this mean that he is by far the most popular president ever? Suggest at least one lurking variable (perhaps another number from the election) that might be important to know when answering this question and why this variable might be important to your understanding of the situation.

11. Hospital floors are usually covered by bare tiles. Carpets would cut down on noise but might be more likely to harbor germs. To study this possibility, investigators randomly assigned 8 of 16 available hospital rooms to have carpet installed. The others were left bare. Later, air from each room was pumped over a dish of agar. The dish was incubated for a fixed period, and the number of bacteria colonies were counted.

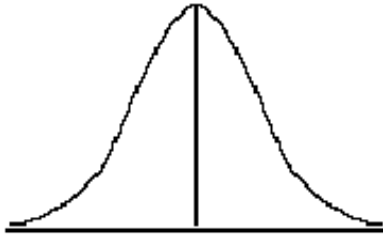
- a. Select the appropriate statistical term for the 16 hospital rooms.
  1. Experimental Units.
  2. Treatments.
  3. Response.
  4. Control Group.
  
- b. Select the appropriate statistical term for number of colonies in a dish.
  1. Experimental Units.
  2. Treatments.
  3. Response.
  4. Control Group.

12. The sampling distribution of a statistic is

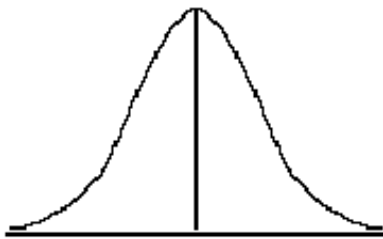
- a. the probability that we obtain the statistic in repeated random samples.
- b. the mechanism that determines whether randomization was effective.
- c. the distribution of values taken by a statistic in all possible samples of the same size from the same population.
- d. the extent to which the sample results differ systematically from the truth.

13. Weights of many dog breeds are normally distributed. West Highland Terrier (Westie) weights are normally distributed with a mean of 16.3 pounds and a standard deviation of 2.1 pounds. **For each question below, show your work, answer the question, clearly label the graph, and shade in the area that represents the percentage.**

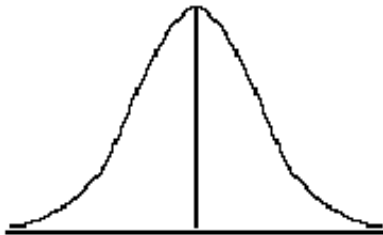
a. What is the range for the central 95% of Westie weights?



b. A certain Westie breeder is accused of underfeeding his dogs. A sample of 25 dogs is taken from the breeder, and the dogs are weighed. What is the probability the sample mean is less than 14 pounds?



c. Tom has a Westie that weighs 20 pounds. Tom thinks his dog might be eating too well. What percentage of Westies weigh as much as Tom's or more?



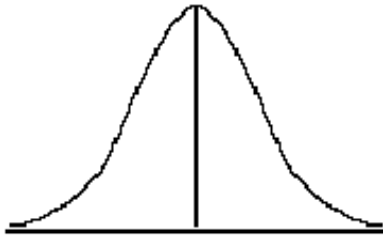
d. How much do the highest 10% of Westies weigh?



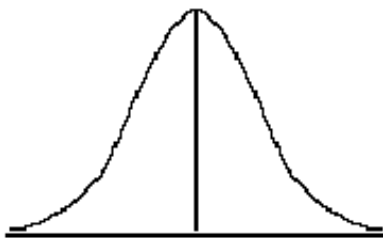


14. According to Harper's magazine, the time spent by kids in front of the television set per year can be modeled by a normal distribution with a mean of 1500 hours and a standard deviation of 250 hours.

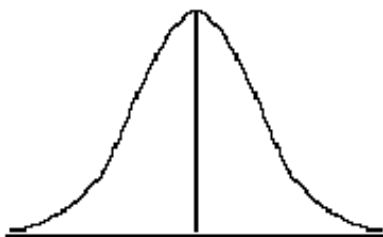
- a. What percent of kids watch television for less than 1200 hours per year? (Answer the question, clearly label the graph, and shade in the area that represents the percentage).



- b. How many hours of t.v. do the highest 10% of kids watch? (Answer the question, clearly label the graph, and shade in the area that corresponds to the probability.)



- c. A researcher followed a random sample of 49 children for one year and calculated the mean number of hours that these 49 children watched television. Approximate the probability that the sample mean is between 1400 and 1600 hours. (Answer the question, clearly label the graph, and shade in the area that corresponds to the probability.)



15. A student believes his well-organized cheat sheet helped him to do better on his STAT 301 exam. His teacher believes that the student's hard work on the homework and lab assignments and the many hours of studying are what helped him do well on the exam. This is an example of:
- response bias.
  - causation.
  - nonresponse.
  - common response.
  - confounding.
16. What is the difference between control and capability?
17. A maker of freezers checks a sample of 10 thermostatic controls from each hour's production. The thermostats are set at  $29^{\circ} F$  and then placed in a freezer. The temperature at which the thermostat turns on the motor of the freezer is recorded. The process mean should be  $\mu = 29^{\circ} F$ . Past experience indicates that the response temperature of properly adjusted thermostats varies with  $\sigma = 0.5^{\circ}$ . The mean response temperature  $\bar{x}$  for each hour's sample is plotted on an  $\bar{x}$  control chart. Calculate the center line and control limits (UCL and LCL) for this chart. Draw the center line, UCL, and LCL on a control chart with labeled axes.
18. Since 1976 the Nurses' health study has followed more than 100,000 nurses. Every two years, the nurses fill out a questionnaire about their habits and their health. Results from this study indicated that post-menopausal women have a reduced risk of heart disease if they take a hormone replacement drug. Which of the following is **not** true concerning this study?
- It is obvious from this study that the hormone-replacement drug reduces the risk of heart disease.
  - A possible lurking variable may have caused the relationship, making it not so obvious that hormone-replacement drug reduces the risk of heart disease.
  - A randomized experiment would be a better approach to testing the effects of the hormone-replacement drug.
  - This study is an observational study.

19. Suppose that it was found that hormone replacement drugs do not affect heart disease risk. A possible lurking variable that caused the two to appear related is that women who took the drug tended to be more conscientious about their personal health than those who did not take it. This is an example of
- A. confounding.
  - B. double blind.
  - C. common response.
  - D. Inference.
20. The magazine called Literary Digest conducted a poll to predict the results of the 1936 Presidential election between Franklin Roosevelt (Democrat and incumbent) and Alf Landon (Republican). They mailed questionnaires to 10 million people and asked how they planned to vote. The list that they sampled from was constructed from telephone directories, country club memberships, and automobile registrations. At the time, the United States was in the Great Depression; those who had cars, phones, and country club memberships and thus received questionnaires tended to be relatively wealthy. Consequently, they had sampling bias due to
- A. Response bias.
  - B. Nonresponse.
  - C. Undercoverage.
  - D. None of the above.

**For each story in questions 21-26, choose which type of sampling or experimental design was used in the data collection process.**

<b>A.</b> Simple Random Sample	<b>B.</b> Stratified Random Sample	<b>C.</b> Multistage Random Sample	<b>D.</b> Completely Randomized Design
<b>E.</b> Randomized Block Design.	<b>F.</b> Voluntary Response Sample	<b>G.</b> Matched Pairs Design	<b>H.</b> Anecdotal Evidence

21. A marketing department of a major oil company wants to investigate whether cars get better mileage using their gas (Brand A) than using gas from another company (Brand B). They use 20 identical Ford Escorts for the study. They randomly assign 10 cars to Brand A and 10 to Brand B and record the mileage from cars in each group.
22. A marketing department of a major oil company wants to investigate whether cars get better mileage using their gas (Brand A) than using gas from another company (Brand B). They use 10 identical Ford Escorts and 10 identical Toyota Camrys for the study. They randomly select 5 Fords to receive Brand A and independently randomly select 5 Toyotas to receive Brand A. The remaining cars receive Brand B. They then compare the results.
23. You plan to sample residents of registered nursing homes in your state. You randomly select 5 counties from your state. From the 5 counties you obtain a list of all nursing homes in those counties and randomly select 5 from each county. You then obtain a list of all residence in the 25 selected nursing homes and randomly select 5 individuals from each nursing home to sample.
24. In Fall 1995, the BBC in Britain requested viewers to call the network and indicate their favorite poem.
25. In order to determine students' opinions of STAT 301, Ellen Gundlach randomly selects 5 of her students, 5 of Tom Howell's students, 5 of Joan Brenneman's students and 5 of Christa Sorola's students to survey.

Answer questions 26 and 27 based on this story. To determine if a new gunpowder is superior to the standard gunpowder, a scientist randomly assigns the new powder to 180 bullets and the standard powder to 100 bullets. The bullets are then fired and the velocity of each bullet is measured.

26. The unit in this study is

- A. A bullet.
- B. All bullets.
- C. The two types of gun powder.
- D. The gun used to shoot the bullets.

27. The response variable in this study is

- A. The two types of gun powder.
  - B. A bullet.
  - C. All bullets.
  - D. The velocity of the bullet.
- 

28. Which of the following will **not** necessarily decrease the bias?

- A. Increase the sample size.
- B. Use a probability based sampling technique.
- C. Make sure all members of the population have a non-zero chance of being selected into our sample.
- D. Write our survey questions carefully to avoid one answer appearing to be more favorable.

29. Which of the following is/are examples of a process?

- A. Reviewing job applications for some new positions that have opened up at a local company.
- B. Balancing a check book.
- C. Making scrambled eggs.
- D. Making a standard screw to be sold at hardware stores.
- E. All of the above.

30. If a process is in control,

- A. nothing needs to be changed in the process.
- B. the product will be of “satisfactory quality”.
- C. the end product will meet or exceed the requirements placed on it.
- D. we know what to expect in the finished product.
- E. All of the above.

For questions 31 – 36, choose A, B, C, D, E, or F from the table below. Each answer may be used once, more than once or not at all.

A	B	C																																								
<pre> 0   1 1   2   3   4   5   2 6   1 4 7   0 3 7 8   1 2 3 9   4 5 9 9 10   4 5 5 5 7 11   3 4 7 8 9 9 12   0 1 1 4 5       1 2 3 4           </pre>	<table border="1"> <caption>Data for Graph B</caption> <thead> <tr> <th>Group</th> <th>Value Frequency</th> </tr> </thead> <tbody> <tr><td>A</td><td>80</td></tr> <tr><td>B</td><td>110</td></tr> <tr><td>C</td><td>90</td></tr> <tr><td>D</td><td>60</td></tr> <tr><td>E</td><td>40</td></tr> <tr><td>F</td><td>20</td></tr> <tr><td>G</td><td>10</td></tr> <tr><td>H</td><td>5</td></tr> </tbody> </table>	Group	Value Frequency	A	80	B	110	C	90	D	60	E	40	F	20	G	10	H	5	<table border="1"> <caption>Data for Graph C</caption> <thead> <tr> <th>Value</th> <th>Frequency</th> </tr> </thead> <tbody> <tr><td>1</td><td>4</td></tr> <tr><td>2</td><td>5</td></tr> <tr><td>3</td><td>4</td></tr> <tr><td>4</td><td>3</td></tr> <tr><td>5</td><td>2</td></tr> <tr><td>6</td><td>1</td></tr> <tr><td>7</td><td>1</td></tr> <tr><td>8</td><td>0</td></tr> <tr><td>9</td><td>0</td></tr> <tr><td>10</td><td>1</td></tr> </tbody> </table> <p>Mean = 3.33 Std. Dev. = 2.123 N = 21</p>	Value	Frequency	1	4	2	5	3	4	4	3	5	2	6	1	7	1	8	0	9	0	10	1
Group	Value Frequency																																									
A	80																																									
B	110																																									
C	90																																									
D	60																																									
E	40																																									
F	20																																									
G	10																																									
H	5																																									
Value	Frequency																																									
1	4																																									
2	5																																									
3	4																																									
4	3																																									
5	2																																									
6	1																																									
7	1																																									
8	0																																									
9	0																																									
10	1																																									
D	E	F																																								
<table border="1"> <caption>Data for Graph D</caption> <thead> <tr> <th>Value</th> <th>Frequency</th> </tr> </thead> <tbody> <tr><td>0</td><td>1</td></tr> <tr><td>2.5</td><td>2</td></tr> <tr><td>5</td><td>3</td></tr> <tr><td>7.5</td><td>4</td></tr> <tr><td>10</td><td>3</td></tr> <tr><td>12.5</td><td>2</td></tr> <tr><td>15</td><td>1</td></tr> </tbody> </table> <p>Mean = 7.5 Std. Dev. = 4 N = 16</p>	Value	Frequency	0	1	2.5	2	5	3	7.5	4	10	3	12.5	2	15	1	<p>Mean = 16.5 Std. Dev. = 4 N = 16</p>	<p style="text-align: center;"><b>None of the Above / No graph could tell you this information</b></p>																								
Value	Frequency																																									
0	1																																									
2.5	2																																									
5	3																																									
7.5	4																																									
10	3																																									
12.5	2																																									
15	1																																									

31. Which graph shows you data in which the mean is greater than the median?
32. Which graph clearly shows the 25<sup>th</sup> percentile?
33. Which graph should be used only for categorical data?
34. Which graph has a distribution that is similar to the distribution in graph D? Do not choose D as your answer!
35. Which graph shows a distribution that is skewed right?
36. Which graph shows you data that has undercoverage bias?

Answer questions 37 and 38 based on this story. Assume that the development time for a particular type of photographic printing paper when it is exposed to a light source for five seconds has a normal distribution with mean,  $\mu = 202$  seconds, and standard deviation,  $\sigma = 4$  seconds.

37. What is the probability that a particular print will require more than 205 seconds to develop?
- A. 0.0329
  - B. 0.2266
  - C. 0.75
  - D. 1.84
  - E. 0.7734
38. Suppose you wanted to find the probability that the mean of a random sample of 20 trials is more than 205 seconds. Which distribution below would you use to standardize your probability statement?
- A.  $X \sim N(202, 4)$
  - B.  $\bar{X} \sim N(202, 4/\sqrt{20})$
  - C.  $X \sim N(205, 4)$
  - D.  $\bar{X} \sim N(205, 4/\sqrt{20})$
  - E. None of the above.