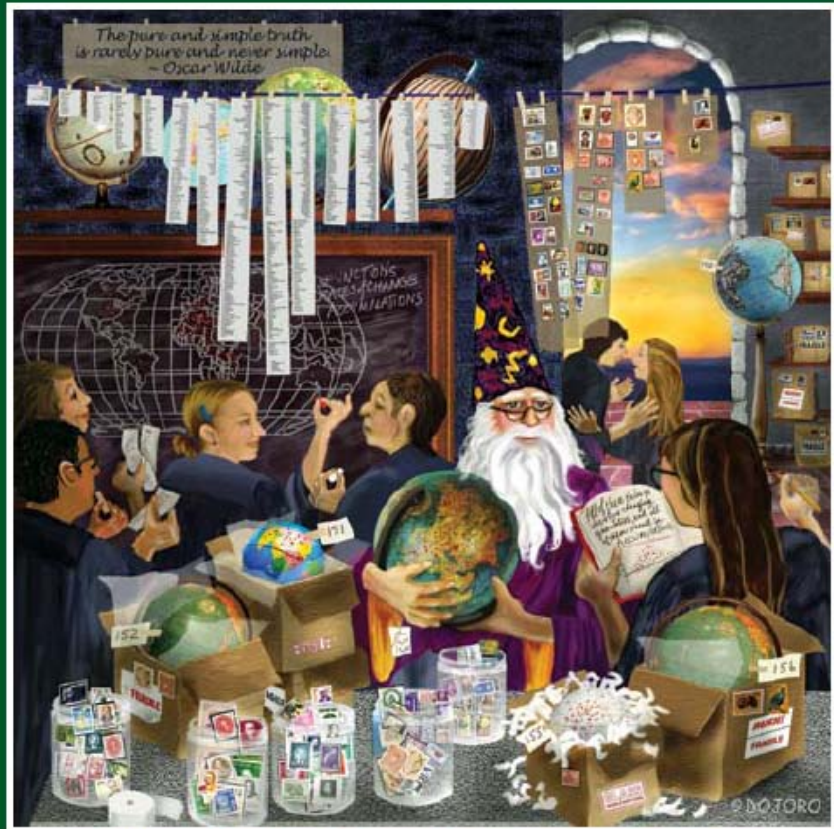


# AP STATISTICS



## QUESTION CATALOGUE

\*AP is a registered trademark of the College Entrance Examinations Board, which is not affiliated with this book.

# AP Statistics

## Table of Contents

### I. EXPLORING DATA

#### 1. Univariate Data

##### A. Summarizing and comparing Distributions

<b>i. Median and Mean</b> .....	1
ii. Range, Interquartile Range.....	8
iv. Variance, Standard Deviation.....	14
iv. Position, Percentile Ranking, Z-Score.....	21
v. Graphical Representations of Data.....	28

#### 2. Bivariate Data

<b>A. Correlation and Linearity</b> .....	37
B. Least Squares Regression.....	50
C. Residuals, Outliers.....	59
D. Transformations to Linearity.....	67

<b>3. Categorical Data</b> .....	70
----------------------------------	----

### II. SAMPLING AND EXPERIMENTATION

<b>1. Overview of Methods of Data Collection</b> .....	75
<b>2. Planning and Conducting a Survey</b> .....	83
3. Planning and Conducting an Experiment.....	101
<b>4. Generalizability of Results and Conclusions Drawn</b> .....	110

### III. ANTICIPATING PATTERNS

#### 1. Probability

A. Interpreting Probability.....	115
B. Counting.....	118
C. Rules of Probability	
<b>i. Law of Large Numbers</b> .....	124
ii. "OR" Statements.....	127
<b>iii. "AND" Statements</b> .....	131
iv. Independence.....	135
v. Conditional Probability.....	140
D. Probability Distributions.....	149

<b>2. Random Variables</b>	
<b>A. One Random Variable</b>	
i. Expected Value.....	152
<b>ii. Variation and Standard Deviation.....</b>	<b>157</b>
iii. Linear Transformations.....	160
<b>B. Two Random Variables</b>	
i. Independence vs. Dependence.....	163
ii. Mean for Sums and Differences.....	164
iii. Variance for Sums and Differences.....	165
<b>C. Sampling Distributions</b>	
i. Normal Distributions.....	167
ii. Binomial Distributions.....	173
iii. Gemoetric Distributions.....	179

## IV. STATISTICAL INFERENCE

<b>1. Tests of Significance</b>	
<b>A. Logic of Significance Tests</b>	
i. Null and Alternate Hypothesis.....	183
ii. P-Values, Significance Level.....	187
<b>iii. One-sided vs. Two-sided.....</b>	<b>189</b>
iv. Power.....	191
v. Type I and Type II Errors.....	193
vi. Sampling Distributions.....	197
<b>B. Z-Tests</b>	
i. For a Single Mean.....	198
ii. For a Single Proportion.....	204
iii. For Difference of Two Proportions.....	210
iv. For Difference Between Two Means.....	217
<b>C. T-Tests</b>	126
i. T Distribution.....	221
ii. Single Sample T-Test.....	224
iii. Two Independent Sample T-Test.....	228
iv. Mathed Pair T-Test.....	231
v. Slope in LSRL.....	234
<b>D. Chi-Squared Test</b>	
<b>i. Chi-Squared Distribution.....</b>	<b>236</b>
ii. Goodness of Fit.....	239
iii. Test for Homogeneity.....	243
iv. Test for Independence.....	246
<b>2. Confidence Intervals</b>	
<b>A. Logic of Confidence Intervals</b>	
<b>i. Definition and Interpretation.....</b>	<b>253</b>
ii. Sample Size and Interval Length.....	260
<b>B. Understanding Conference Intervals</b>	
i. One Mean (Z-Test).....	267
ii. One Mean (T-Test).....	271
iii. Two Mean (Z-Test).....	274
iv. Two Mean (T-Test).....	278

2127. The mean of 12 numbers is  $-2$ .

If all the numbers are added, what will the sum be?

- (A)  $-24$  (D)  $-6$   
 (B)  $-14$  (E)  $-2$   
 (C)  $-10$

2126.  $(5, 3, -2, -12, 29, 99)$

What is the median of the above data?

- (A)  $-2$  (D) **4**  
 (B)  $1$  (E)  $20$   
 (C)  $2$

Base your answers to questions **2124** and **2125** on the following information.

The median of six data points is  $12$ . The mean of those six points is also  $12$ . The data set is:

$(y, 8, 9, x, 16, 18)$

2124. What is the value of  $x$ ?

- (A)  $8$  (D)  $12$   
 (B)  $10$  (E) **15**  
 (C)  $11$

2125. What is the value of  $y$ ?

- (A)  $3$  (D)  $9$   
 (B) **6** (E)  $12$   
 (C)  $8$

2123.  $(-5, -2, -15, -6, -9, x)$

The median of the above data is  $-7$ . What is the value of  $x$ ?

- (A)  $-1$  (D)  $-10$   
 (B)  $-3$  (E)  $-11$   
 (C)  **$-8$**

2122.  $(12, 4, 3, 9, x)$

The median of the above data is  $4$ . Which of the following might be the value of  $x$ ?

- (A) **2** (D)  $8$   
 (B)  $5$  (E)  $11$   
 (C)  $7$

2121.  $(-4, -7, 12, -9, x)$

The mean of the above data is  $0$ . What is the value of  $x$ ?

- (A)  $-3$  (D) **8**  
 (B)  $0$  (E)  $12$   
 (C)  $3$

2120.  $(18, 7, 9, 2, x)$

The mean of the above data is  $24$ . What is the value of  $x$ ?

- (A)  $17$  (D) **84**  
 (B)  $24$  (E)  $102$   
 (C)  $54$

1876. The median of a set is sometimes more useful than the mean because the median

- (A) is more resistant to outliers  
 (B) does not reflect spread  
 (C) does reflect spread  
 (D) **is less resistant to outliers**  
 (E) is more often an actual member of the data set

1832. A teacher has two math classes. In one class there are  $80$  students and the average grade on a test was  $30$ . In her other class there are  $40$  students and the average grade on the same test is  $50$ .

What is the average grade of all her students?

- (A)  $24.2$  (D) **36.7**  
 (B)  $32.8$  (E)  $56.8$   
 (C)  $34.6$

1643. Base your answer to the following question on the following information.

Below is a table composed by a school dean, as he tried to determine if there was a correlation between amount of detentions and GPA.

Detention	1	1	1	2	3	5	6	7
GPA	3.1	2.6	3.5	3.9	2.8	3.2	2.7	2.7

What is the correlation coefficient of this data?

- (A)  $-0.6921$  (D)  $0.1613$   
 (B)  **$-0.4016$**  (E)  $0.3191$   
 (C)  $-0.2031$

1620. Base your answer to the following question on the following information.

Below are the grades of a college student for the year, along how much work the class required.

Grades	3.67	3.67	3.67	3.33	4.00	4.00	4.33	3.67
Work Load	10	2	5	6	8	6	4	5

Which of the following is the correlation coefficient for this data?

- (A)  **$-0.0800$**  (D)  $0.2100$   
 (B)  $-0.0320$  (E)  $0.4561$   
 (C)  $0.0064$

1183. A perfectly negative correlation

- (A) **implies a coefficient of determination of  $-1.0$ .**  
 (B) means that there must be a slope of  $-1.0$ .  
 (C) says that for every one additional unit of the variable on the  $x$ -axis, the variable on the  $y$ -axis must move down one additional unit.  
 (D) means that the  $r$ -value is exactly  $1.0$ .  
 (E) implies that all data points lie on the least squares regression line.

1199. Four data points  $(x,y)$  have a correlation coefficient of  $0.6$ .

- a) Give an example of four data points with a correlation of about  $0.6$ .  
 b) Demonstrate by adding a fifth point, that the correlation can become negative.

**a) There are infinite examples where this true. An example is  $(22,99)$ ,  $(32,66)$ ,  $(5,10)$ ,  $(17,60)$  and  $(32,56)$ . Their correlation coefficient is  $0.60$ .**

**b) Adding in the data point  $(10,200)$  makes the correlation coefficient  $-0.1398$ .**

1197. A survey was conducted to observe age and salary in a certain company. The correlation between the two was  $r = 0.4309$ . The survey was done again the very next year. No employee left or was hired. The data has changed as everyone has become one year older and has had their salary doubled.

- a) What is the new correlation coefficient?  
 b) What is the new coefficient of determination?

**a) The correlation coefficient is still  $0.4309$ . Adding or multiplying by a constant does not effect the correlation.**

**b) The same goes with the coefficient of determination. It remains  $0.1857$ .**

1180. The more closely together points on a scatterplot lie

- (A) the weaker the linear fit is.  
 (B) the more positive the correlation is.  
 (C) the closer the correlation coefficient is to  $0$ .  
 (D) **the higher the coefficient of determination.**  
 (E) the higher the slope is on the least squares regression line.

1151. Which of these is the correlation coefficient, if the coefficient of determination is  $0.81$ ?

- (A)  $0.18$  (D)  $0.81$   
 (B)  $0.53$  (E)  **$0.90$**   
 (C)  $0.66$

1934. What is the relative frequency distribution of Turkey?

- (A) 7.2% (D) 77.4%  
 (B) 18.2% (E) 100.0%  
 (C) **36.2%**

1931. Categorical Data is always which of the following?

- I. Qualitative Data  
 II. Quantitative Data  
 III. Continuous Data
- (A) **I only** (D) I and II only  
 (B) II only (E) I and III only  
 (C) III only

1930. Two vegetable producers both sell tomatoes and peppers. The first producer claims that the mean weight of his vegetables is higher than the mean weight of his competitors vegetables. The second producer claims that the mean weight of his tomatoes and the mean weight of his peppers are both higher than the respective means from his competitor.

Which of the following are true?

- I. If the first producer is correct, the second producer must be wrong.  
 II. If the second producer is correct, the first producer must be wrong.  
 III. It is possible for both to be correct.
- (A) I only  
 (B) II only  
 (C) **III only**  
 (D) I and II  
 (E) None of the above choices are correct.

1911. What type of hypothesis test is used to measure categorical data?

- (A) one-sample  $z$ -test (D) two-sample  $t$ -test  
 (B) two-sample  $z$ -test (E) **chi-square test**  
 (C) one-sample  $t$ -test

1929. Below is a two-way contingency table.

36	$x$
24	18

If the table shows perfect independence, what is the value of  $x$ ?

- (A) 16 (D) 32  
 (B) 24 (E) 36  
 (C) **27**

1928. Below is a two-way contingency table

6	9
18	$x$

If the table shows perfect independence, what is the value of  $x$ ?

- (A) 6 (D) 18  
 (B) 9 (E) **27**  
 (C) 12

Base your answers to questions 1920 through 1922 on the following information.

A waiter asks all his customers before they order their entrees if they would like soup or salad. Each customer can order both, one or neither. Below is the two-way contingency test of the data.

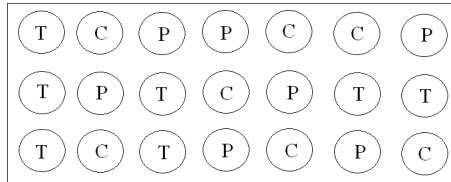
	Soup	No Soup
Salad	242	296
No Salad	307	231

1920. What percentage of the people served by the waiter ordered a salad?

- (A) 22.49% (D) **51.02%**  
 (B) 29.19% (E) 78.83%  
 (C) 44.08%

1902. Below is a drawing of a garden and the vegetables growing in it.

The gardener wants to study her garden to see how well her vegetables grow, in terms of inches grown.



**Key**  
 C = Carrots  
 T = Tomatoes  
 P = Peppers

- a) Explain how the gardener could conduct a census to determine how many inches each vegetable grew since the last survey.
- b) If the gardener wants to breakdown the study by vegetable, would a systematic or stratified sample be more appropriate?
- c) Design the study chosen in part b).

**a) The gardener should measure each and every individual vegetable in the garden. Then she could add all the data together and find the average amount grown. A confidence interval should possibly be created.**

**b) A stratified random sample would be appropriate. A systematic sample would take 1 out of every  $n$  vegetables. That would not, however, divide that data by vegetables. A stratified sample would work much better.**

**c) The gardener could take a simple random sample of each of the three vegetables. For example, she could 3 vegetables from each category. Then she could compare the data from each of the vegetable types separately.**

1138. A high school principal wants to conduct an experiment on many of his students in his school. In order to choose his experimental students, he rolls a die until the numbers 1, 2, 3 or 4 show up. The number 1 correlates to 9th grade, 2 to 10th grade, 3 to 11th grade and 4 to 12th grade. Hence, if he rolls a 3, a simple random sample of all of the eleventh graders will be in the experiment.

- a) What type of data collection is this?
- b) Why is this good or not good for this situation?

**a) This is a form of a stratified sample. A group is chosen and then groups are chosen from that group.**

**b) This is a bad way to get a random sample. Different grades share very distinct characteristics, so that choosing one grade will not be representative of the whole school. Therefore, the principal could do a random sample of the whole school by assigning everyone a number and choosing numbers randomly. Or if the principal wants to have a certain amount of students per grade in the study, he or she could block by grade, taking a random sample of those in each grade.**

1128. In which of the following examples would an experiment be preferred to a survey?

- (A) Finding a correlation between first degree burns and long-term skin problems.
- (B) Determining if there is a relationship between political revolution and class conflict in societies.
- (C) Finding if there is a relationship between a new exercise routine and long-term weight loss.**
- (D) Determining a correlation between heavy lifting jobs and long-term health conditions.
- (E) Finding a correlation between divorce rates and depression rates in various countries.

2037. Figure out how many people are in the sampling frame and divide that number by the amount of units you would like to have in your study. Call that number  $x$ . Then study one out of every  $x$  units.

What form of sampling is being described?

- (A) Cluster Sampling
- (B) Convenience Sampling
- (C) Simple Random Sample
- (D) Stratified Sample
- (E) Systematic Sample**

2036. Go to a place where you can get information from the population you are looking to obtain data on. Get information from as much of the population you need, until you have gotten enough data for your sample size. Get data from the most available part of the population.

What form of sampling is being described?

- (A) Cluster Sampling
- (B) Convenience Sampling**
- (C) Simple Random Sample
- (D) Stratified Sample
- (E) Systematic Sample

2035. Choose one subject randomly from the population. Keep choosing without replacement, until you have the correct number of people for your sample.

What form of sampling is being described?

- (A) Cluster Sampling
- (B) Convenience Sampling
- (C) Simple Random Sample**
- (D) Stratified Sample
- (E) Systematic Sample

2029. Which of these are an example of undercoverage bias?

I. A survey is to be done on the availability of satisfactory child care in America. A stratified random sample was taken through 10 random day cares and 15 people in each.

II. A survey is done to see if people at a theater show, enjoyed the performance. On the way out of the show, every fifth person was asked, "Did you enjoy the show?"

III. A survey is done to see if people a supermarket's prices are fair. Exiting the store, one out of every tenth person was asked, "Do you feel that the prices you were charged were fair?"

- (A) I only
- (B) II only
- (C) III only
- (D) I and III only**
- (E) I, II and III

2028. Which of the following are a probability sample?

I. Simple Random Sample

II. Stratified Sample

III. Systematic Sample

- (A) I only
- (B) II only
- (C) III only
- (D) I and II only
- (E) I, II and III**

2027. Which of the following are not a probability sample?

I. Voluntary Response

II. Convenience Sampling

III. Stratified Sampling

- (A) I only
- (B) II only
- (C) III only
- (D) I and II**
- (E) I, II and III

1141. A scandal erupted in your state governor's office. The newspaper asks readers to write in their opinions of the scandal. How and why do you feel that the opinions in the newspaper will compare the the population's true opinion?

**The newspaper's response would most likely be more severely negative than the general population's response. Since the newspaper is asking for responses, it is conducting a voluntary response poll. By asking for volunteers, it is getting the opinions of those who feel strongly about the issue. In the wake of a scandal, those who feel the strongest are most often those who are mad. Though there might be many people who support the governor, they might no right in tot he newspaper about it.**

1126. a) Give a detailed explanation, demonstrating step by step, how you would conduct a stratified random survey, to find out how many people in your town would support reducing the speed limit from 40 mph to 35 mph.

b) Give an example of the question you would ask the sample. Pay attention to possible bias.

c) Explain which forms of bias are present in your survey or how you managed to avoid them if they are not present.

**a) The student should show that he or she divided the town into groups and then choose a random sample from each of those groups. The student should explain once they choose the sample, and how they would administer the survey.**

**b) The question should not lean one way or the other and should avoid bias.**

**c) The student should give examples of bias in his or her work, or explain way they avoided bias.**

1139. A researcher wanted to conduct a 10,000 household survey to find out how many people in a town use the county parks.

Design this survey for each type of sampling:

a) Simple random sample

b) Cluster sample

c) Stratified sample

d) Systematic sample

e) Convenience sample.

**a) Simple Random Sample: As sign each house in the county a number. Randomly selects 10,000 numbers, and send questionnaires to each of them.**

**b) Cluster Sample: Divide the county up into all the towns that it encompasses. If there were 5,000 households in each of the towns, randomly choose two towns and send questionnaires to everyone in those two towns.**

**c) Stratified Sample: Divide up the town into all the towns it encompasses. If there are 50 towns, then randomly choose 200 households in each of the towns, to send the questionnaires to.**

**d) Systematic Sample: Give all the house in the county a number. Assume that there is 250,000 households in the county. First, choose a random household from 1 to 25. Then choose every 25<sup>th</sup> household after the one randomly selected.**

**e) Convenience Sample: Go to a place where a lot of people are on a given day. For example go to a large concert, a voting poll, or a super market. Give questionnaires to the first 10,000 people you see. It is important however, to make sure that no two people out of the 10,000 are in the same household.**

1943. In a real-life study on kidney stones, two different treatments were tested to see their effectiveness on removing them. For treatment A, 273 out of 350 people had a successful treatment (78%), and for treatment B, 289 out of 350 people had a successful treatment (83%).

Smaller stones		Larger stones	
Treatment A	Treatment B	Treatment A	Treatment B
Group 1	Group 2	Group 1	Group 2
81/87 (93%)	234/270 (87%)	192/263 (73%)	55/80 (69%)

- (a) What kind of grouping method is this known as?
- (b) What could possibly account for this?
- (c) What is this phenomenon known as?
  - (a) **Block grouping**
  - (b) **confounding variables**
  - (c) **simpson's paradox**

1873. A researcher wants to determine the effectiveness of a new sleeping drug by measuring how long it takes subjects to fall asleep with or without the medication.

With 100 volunteers, how would the experiment best be implemented?

- (A) Randomly select 50 volunteers to take a placebo, and 50 to take the drug. Then compare the average sleeping times.
- (B) Randomly select 50 volunteers to take the drug, and compare the sleeping time with results from the 50 volunteers who did not take the drug.
- (C) **Randomly select 50 people to get the drug the first night and placebo the second night, and 50 people to get the placebo the first night and the drug the second night. Then compare results for each person.**
- (D) Have volunteers randomly choose one of two identical pills, the placebo or the real drug.
- (E) None of these is effective.

1869. A clothing company feels that it makes the most comfortable sweaters.

Detail the procedures for an experiment, where the company would compare its sweater with a leading competitor. Include what type of test is needed to prove if the data collected would be significant.

**There are many possible variations of the experiment that can be done. The student should make sure that they choose a fairly high amount of test subjects (ex. five would not be enough). Both types of sweaters must be tested, and it must be done randomly.**

**For example, the company can take fifty people and randomly give them both sweaters. They can randomly be told which one to wear first. After they wear the first sweater they can report from a 1 - 10 scale how comfortable it feels. They could then repeat for the second sweater.**

**The type of test will either be a z-test or t-test. That depends on the amount of observations. If the amount of observations is 30 or more, choose a z-test, otherwise choose a t-test.**

1693. Which of the following is the best way to determine a cause-and-effect relationship?

- (A) A high correlation on the least squares regression line
- (B) A simple random sample
- (C) A census
- (D) **A controlled experiment**
- (E) A well-designed survey.

2112. Suppose that a certain lab wishes to test the effects of a certain chemical on a very common species of toad in the United States. The tests were conducted on toads from 2 different locations on site: Florida, and Louisiana. The chemical was applied to 250 toads in Louisiana, and 400 toads in Florida.

It was found that 83% of the toads in Louisiana and 72% of the toads in Florida grew warts as a result of exposure to the chemical. It was concluded that the toads in Louisiana were more susceptible to the chemical.

Which of the following about this experiment is true?

- (A) There were too many toads in the experiment, resulting in too much variability.
- (B) Because the Florida experiment had a much larger sample size than the Louisiana experiment, the results are biased.
- (C) Using two proportions to compare these is not valid.
- (D) The data is not consistent because the data was conducted in two different locations. Lurking variables are not entirely ruled out.**
- (E) There were too few toads used overall. A sample of several hundred for the population of a very common toad is much too small.

2061. A child wishes to see what burns fastest in a fire. He throws in a ball of yarn, paper, some wood chips, some cotton balls, straw, and some peanuts. Which of these are true?

- (A) There is one explanatory variable and six response variables.
- (B) There are six explanatory variables and one response variable.
- (C) There are six levels of a single explanatory level.**
- (D) Each explanatory level has a level of response associated with it.
- (E) This is not an experiment because it retains few elements of one and is merely done by a child.

2106. Two rival companies that have split the United States in equal halves in terms of population for sales boundaries wish to test who is doing better in respect to their half. One company decided to obtain a simple random sample of people through the use of stratified sampling and gave them a survey, while the other company did the same thing, but used a random number table.

Which of the following is true?

- (A) A stratified sampling method on so large a population would have a tendency to exclude certain people.
- (B) The random number table would not work because eventually, the numbers would repeat and that would create a biased simple random sample.
- (C) Both methods are incorrect for this instance and there will be no correct methods. Splitting the country into two halves will only result in two different halves and there will be a certain bias because these two halves of people will be completely different.
- (D) Both companies should be using the same sampling method because using two different sampling methods for two different populations would not result in two samples suitable for the same purpose and would result in a lack of consistency.
- (E) As long as the method used was correct, there is nothing truly wrong with the actual survey. In this case, the sampling method is correct.**

2050. Is a large sample more preferable or a small sample?

- (A) Large sample because it is more diverse
- (B) Large sample because they show less sampling variability
- (C) Large sample because they will contain less bias.
- (D) Small sample because a larger sample gives too much room for error.**
- (E) Small sample because it is more compact and allows for convenience when working with large numbers

2219. A child plays a game 1000 times. He loses 12 times.

With a high amount of certainty, what is the probability of the child losing?

- (A) **1.2%**                      (D) 88.0%
- (B) 12.0%                      (E) 98.0%
- (C) 24.0%

2218. There are 6 outcomes for a board game. The best outcome is a win that comes one out of every 15 times.

If the game is played 1,000 times, about how many times should one expect to win?

- (A) 15                              (D) 98
- (B) 48                              (E) 1000
- (C) **67**

2217. A man has developed a game where he wins \$24.00 ten percent of the time and he losses \$2.50 ninety percent of the time.

Which of the following amount of games does it make the most sense for the man to play?

- (A) 1                                (D) 50
- (B) 10                                (E) **100**
- (C) 25

2216. A gambler rolls a die 100 times.

Which of the following is the most likely amount of sixes rolled?

- (A) 6                                (D) 19
- (B) **16**                                (E) 22
- (C) 18

1375. Using the law of large numbers, which of the following choices is the most plausible outcome of flipping a coin five hundred times.

- (A) 236 tails                      (D) 299 heads
- (B) **237 tails**                      (E) 480 tails
- (C) 264 heads

2173. A batter wants to get a hit once every three at bats. His batting average is 0.280.

How many times should he come to bat over the season in order to have the best chance at his goal?

- (A) **250**                              (D) 470
- (B) 320                              (E) 600
- (C) 340

2049. The lottery should be won by an individual one out of every 1,237,824 times.

If a lottery player would like to win the lottery one out of every 7 times he plays, how many times should the lottery player play?

- (A) **7**                                (D) 707,328
- (B) 105                                (E) 1,237,824
- (C) 176,382

2048. The probability of a salesperson getting someone to purchase something is 0.33.

Knowing these odds, if a salesperson wishes to sell exactly one third of the products, what number of people should the salesman try to persuade to get closer to this number?

- (A) 3                                (D) 33
- (B) 18                                (E) **174**
- (C) 21

2047. The probability of throwing a baseball over two hundred feet for a certain athlete is 0.245.

If the athlete would like to throw the ball over two hundred feet, 1 out of every three times, how many times should he or she throw the ball?

- (A) **12**                                (D) 42
- (B) 21                                (E) 51
- (C) 33

Base your answers to questions **1407** through **1411** on the following chart depicting the amount of internet usage vs. waist size.

		Waist Size		
		Small	Medium	Large
Internet Usage	Constant	500	800	200
	Average	1100	1900	50
	Infrequent	1200	1300	5

1407. What is the probability that someone uses the internet constantly and has a large waist size?  
 (A) 0.008 (D) 0.249  
**(B) 0.028** (E) 0.671  
 (C) 0.125
1409. What is the probability that someone has a small or medium waist size and uses the internet infrequently?  
 (A) 0.287 (D) 0.549  
**(B) 0.354** (E) 0.965  
 (C) 0.368
1411. What is the probability that someone uses the internet constantly and has a small waist size?  
 (A) 0.010 (D) 0.132  
 (B) 0.059 (E) 0.429  
**(C) 0.071**

1368. The probability of getting an A in Ms. McGougal's statistic class is 0.07 and the probability of getting a C is 0.27. What is  $P(A \cup C)$ ?  
 (A) Based on the multiplication rule, we find  $P(A \cup C) = 0.0189$   
 (B) Based on the multiplication rule, we find  $P(A \cup C) = 0.2593$   
 (C) Based on the addition rule, we know that  $P(A \cup C) = 0.3211$   
**(D) Based on the addition rule, we know that  $P(A \cup C) = 0.3400$**   
 (E) Based on the addition rule, we know that  $P(A \cup C) = 0.6800$

1381. Base your answer to the following question on the table below, which measures the amount of hours students slept on average per night and their average grade.

	A	B	C-F	Total
< 6	23	67	120	210
6-8	29	117	14	160
> 8	21	13	6	40
Total	73	197	140	410

- What is the probability of having a C-F and getting less than 6 hours of sleep?  
**(A) 0.2927** (D) 0.5610  
 (B) 0.3415 (E) 0.8537  
 (C) 0.5122
1359. If two events A and B are mutually exclusive then which of the following is true?  
 (A)  $P(A \cup B) = 1$  (D)  $P(A \cup B) = 0$   
 (B)  $P(A \cap B) = 1$  (E) None of these  
**(C)  $P(A \cap B) = 0$**
1357. An asteroid of sufficient size to cause unprecedented regional devastation for a land impact or the threat of a major tsunami for an ocean impact impacts the earth roughly every 50,000 years. An asteroid capable of causing global climatic catastrophe that may threaten the future of civilization as we know it, whether impacting land or ocean impacts the earth roughly every 100,000 years.

- What is the probability that both events will occur in the next year?  
 (A) one in 50000  
 (B) one in 100000  
**(C) one in 5 billion**  
 (D) There is too little information to solve the problem  
 (E) None of the above

A. One Random Variable

ii. Variation and Standard Deviation

1962.

Right	Probability
0	0.12
1	0.02
2	0.11
3	0.22
4	0.53

A teacher gave a four question pop quiz. Below is the distribution of scores.

What is the standard deviation of the distribution?

- (A) 0.98
- (B) **1.67**
- (C) 2.79
- (D) 3.02
- (E) 4.33

1961. A restaurant is curious to know how many sides of a sandwich their patrons actually eat. Below is the table of data they found.

Sides	Probability
0	0.06
1	0.42
2	0.52

What is the standard deviation of the distribution?

- (A) 0.37
- (B) **0.61**
- (C) 0.74
- (D) 1.12
- (E) 1.46

1958. A college class was polled about how many times they ate in a dining center over the past 24 hours. The results were:

- 0: 17%
- 1: 12%
- 2: 42%
- 3: 21%
- 4: 8%

What is the standard deviation of the distribution?

- (A) 1.05
- (B) **1.15**
- (C) 1.32
- (D) 1.58
- (E) 1.67

1959. A gym teacher asked his students how many days they worked out in the past week. The results were:

- 0: 39%
- 1: 4%
- 2: 8%
- 3: 31%
- 4: 12%
- 5: 4%
- 6: 1%
- 7: 1%

What is the standard deviation of the distribution?

- (A) **1.78**
- (B) 1.84
- (C) 1.94
- (D) 2.09
- (E) 2.45

1406. Base your answer to the following question on the table below, which gives the probability that a salesman makes a certain commission.

Commission	5%	7%	10%
Probability	0.65	0.33	0.02

What is the standard deviation of the commission variable?

- (A) 0.621%
- (B) 1.032%
- (C) **1.111%**
- (D) 2.516%
- (E) 2.759%

1035. For a certain professional bowler, the odds of making a spare is 0.963. Over a 3 game series, the bowler needs to make 12 spares. What is the expected number of spares the bowler will make and the standard deviation of the distribution?

- (A) **11.56; 0.43**
- (B) 11.56; 0.65
- (C) 12.00; 0.44
- (D) 12.00; 0.67
- (E) The standard deviation cannot be determined with this data.

1864. George and Georgina used the same data for a significance test but George found that the results were significant at the 0.10 level and Georgina found that they were not. It turns out the only difference was that one of them used a two-sided test and the other a one-sided test.
- a) Which test (one-sided or two-sided) makes it harder to reject the null hypothesis?  
 b) Who used the two sided test?
- a) **Using a two sided test makes it harder to reject the null hypothesis as the critical region on a specific side of the distribution is smaller.**
- b) **Since Georgina found that the results were not significant, she would be the one to use a two-sided test.**
1820. If the alternate hypothesis contains a " $\neq$ " symbol in it, then \_\_\_\_\_.
- (A) there is a one side hypothesis test, with the critical region on the left.  
 (B) there is a one side hypothesis test, with the critical region on the right.  
 (C) **there is a two side hypothesis test, with the critical region on both sides.**  
 (D) the critical region is larger than it would be if there would be a " $>$ " in the same spot  
 (E) the critical region is smaller than it would be if there would be a " $>$ " in the same spot
1819. If the alternate hypothesis contains a " $>$ " symbol in it, then
- (A) there is a one side hypothesis test, with the critical region on the left.  
 (B) **there is a one side hypothesis test, with the critical region on the right.**  
 (C) there is a two side hypothesis test, with the critical region on both sides.  
 (D) the critical region is larger than it would be if there would be a " $\neq$ " in the same spot  
 (E) the critical region is smaller than it would be if there would be a " $\neq$ " in the same spot
1818. If the alternate hypothesis contains a " $<$ " symbol in it, then
- (A) **there is a one side hypothesis test, with the critical region on the left.**  
 (B) there is a one side hypothesis test, with the critical region on the right.  
 (C) there is a two side hypothesis test, with the critical region on both sides.  
 (D) the critical region is larger than it would be if there would be a " $\neq$ " in the same spot  
 (E) the critical region is smaller than it would be if there would be a " $\neq$ " in the same spot
684. A two-sided hypothesis test should be used to test the alternative hypothesis that
- (A) **the claimed population mean is not equal to the actual population mean**  
 (B) the claimed mean is less than the actual population mean  
 (C) the claimed mean is greater than the actual population men  
 (D) either choice (a) or (c) based on the situation  
 (E) all of the above
682. A one sided significance test should be used to check the alternative hypothesis that
- (A) the claimed population mean is not equal to the actual population mean  
 (B) the claimed population mean is less than the actual population mean  
 (C) the claimed population mean is greater than the actual population mean  
 (D) **either choices (b) or (c) depending on the circumstances**  
 (E) the claimed population standard deviation matches the actual population standard deviation.
673. A particular significance test is designed to determine is the difference between two means is statistically significant.
- Should a one or two sided test be used?
- (A) One sided.  
 (B) **Two sided.**

D. Chi-Squared Test

655. A study was conducted in which researchers wished to find out whether the order of birth of a child in his family had any correlation with whether that child would be prone to turn delinquent at an earlier age. The table below shows the data collected during the study.

	Oldest	Mid-range	Youngest	Lone Child
Not Delinquent	277	254	223	100
Delinquent	22	33	43	27

An appropriate null hypothesis for this would be

- (A) Older children in families tend to grow more delinquent.
- (B) Younger children in families tend to grow more delinquent.
- (C) The distribution of the number of delinquent children is the same for each birth
- (D) The distribution of the order of birth is same in delinquent and non-wayward behavior.
- (E) **Birth order and delinquent behavior are independent of each other.**

1861. Which of the following are true statements about the chi-squared distribution?

- I. For fixed degrees of freedom, the variance is greater than for a *t*-distribution.
  - II. For fixed degrees of freedom, the variance is greater than for a *z*-distribution
  - III. The distribution is skewed.
- (A) I and II only
  - (B) II and III only
  - (C) I and III only
  - (D) **I, II, III**
  - (E) None of these is the complete set of responses

1762. As the degrees of freedom in the chi-squared distribution increases, what happens to the symmetry of the curve?

- (A) **Increases**
- (B) Decreases
- (C) Increases then decreases
- (D) Decreases then increases
- (E) Remains the same

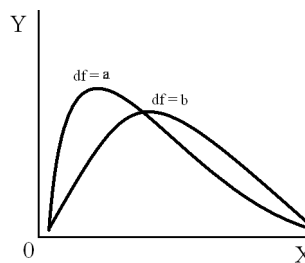
1754. The degrees of freedom in a 3 by 2 two-way table are

- (A) 1
- (B) **2**
- (C) 4
- (D) 6
- (E) 7

1519. The Chi-square distribution describes which of the following types of data?

- I. Univariate Data
  - II. Bivariate Data
  - III. Categorical Data
- (A) I only
  - (B) II only
  - (C) **III only**
  - (D) I and II only
  - (E) I, II and III

1435.



In these distributions, what are likely values of *a* and *b*?

- (A) *a* = 5 ; *b* = 2
- (B) *a* = 4 ; *b* = 3
- (C) *a* = 6 ; *b* = 6
- (D) ***a* = 7 ; *b* = 18**
- (E) *a* = 500 ; *b* = 520

1420. Which of the following  $\chi^2$  statistics is closest to the critical line at the 95% significance level, when there are 17 degrees of freedom?

- (A) **27.58**
- (B) 28.31
- (C) 28.38
- (D) 28.77
- (E) 29.54

1422. A school cafeteria has two different entrees and wants to find out if they all are ordered equally by boys and girls. The lunch director conducts a study of how the entrees are selected to see if there is a difference in entree between both genders.

Which chi-square test should be conducted for the above situation?

- (A) Chi-square test for goodness of fit.
- (B) Chi-square test for homogeneity.
- (C) Chi-square test for independence.
- (D) Any of these tests would be appropriate.
- (E) All chi-square tests would not be appropriate.**

1391. Base your answer to the following question on the below table which displays the claimed percentage of the student's favorite colors in a certain school, and someone survey of the same thing.

	Red	Blue	Green	Total
Claimed %	65%	31%	4%	100%
Observed	50	23	2	75

What is the  $\chi^2$  statistic for this data?

- (A) 0.003
- (B) 0.032
- (C) 0.333
- (D) 0.368**
- (E) 0.370

1349. There are 132 people observed during a specific part of a chi-square goodness-of-fit test. The  $\chi^2$  statistic for that section was 6.38. What was the possible expected value for this part of the chi-square goodness-of-fit test?

- (A) 97
- (B) 106
- (C) 158
- (D) 161
- (E) More than one of these is possible.**

1348. What does a high  $\chi^2$  statistic imply about expected and observed frequencies?

- (A) It implies that the sum of observed and expected frequencies is high.
- (B) It implies that both observed and expected values are of low numbers.
- (C) It implies that the observed and expected frequencies are close together.
- (D) It implies that there is a lot of difference between observed and expected frequencies.**
- (E) It implies nothing about expected and observed values.

1227. As the degrees of freedom in a  $\chi^2$  distribution increases, then

- (A) the curve remains the same.
- (B) the curve becomes more symmetric.**
- (C) the curve skews more to the right.
- (D) the curve skews more to the left.
- (E) the curve skews more to either the right or left.

1226. What makes the  $\chi^2$  score of  $-999.73$  impossible?

- (A) The score is so far from zero, that he becomes to unlikely to be considered true.
- (B) Decimals are not allowed in the  $\chi^2$  distribution.
- (C) A score of that size would imply that the distribution is not symmetric.
- (D) It is a negative number and negative numbers are not allowed under the  $\chi^2$  distribution.**
- (E) The degrees of freedom chart does not reach  $-999.73$ .

1224. Which of the following does the shape of a  $\chi^2$  curve depend on?

- (A) degrees of freedom**
- (D) z score
- (B) observed mean
- (E) expected mean
- (C) confidence level

## A. Logic of Confidence Intervals

## i. Definition and Interpretation

1569. A German language class in high school created a 95% confidence interval for the amount of people that will attend its end of the year picnic, based on the past twenty years of attendance data from the end of the year picnic. The confidence interval was (69.32, 85.71).
- a) If 95 people came to the picnic, what could be said about the confidence interval?  
 b) If it was found that the true mean was 95, what could be said of the procedure taken to create the confidence interval.
- a) If 95 people came to the picnic, the confidence interval would still be valid. The confidence interval explains the true mean and not each and every data point. If, for example, the true mean was 85.71, then 95 attendees at the picnic would not be so unlikely.**
- b) A true mean of 95 might imply that the confidence interval was done correctly. However, it does not mean that it was done wrong. There was a 5% chance that the true mean was outside the confidence interval, so the confidence interval might have been done correctly.**
1203. Which of the following does a confidence interval contain?  
 (A) Sample mean of the data  
**(B) Population mean of the data**  
 (C) Sample standard deviation of the data  
 (D) Population standard deviation of the data  
 (E) None of the above
1202. With a confidence level of 99%, what does the confidence interval contain 99% of the time?  
 (A) Sample mean of the data  
**(B) Population mean of the data**  
 (C) Sample standard deviation of the data  
 (D) Population standard deviation of the data  
 (E) None of the above
1568. A 95% confidence interval is  $(-3.25, 7.21)$ .
- Which of the following statements is true?  
 (A) Given that this was a difference between two proportions or two means, we could be 5% certain that the difference between both variables greater than 7.21.  
 (B) Given that this was a difference between two proportions or two means, it could be 95% sure that the difference between both variables is significant.  
 (C) Given that this was a one sample  $z$  or  $t$  confidence interval, a data point outside that confidence interval is 5% likely.  
**(D) Given that this was a one sample  $z$  or  $t$  confidence interval, a true mean not in that confidence interval is 5% likely.**  
 (E) Given that this was a one sample  $z$  or  $t$  confidence interval, the sample mean of this data is 10.46.
1565. A confidence interval is created between the difference of two proportions or two means, which contains the value of 0 in the confidence interval.
- a) What is the significance of containing the value of zero?  
 b) How can this be related to significance tests?  
**a) When the value of zero is a confidence interval, it shows that the two means or proportions are statistically significantly different from each other.**
- b) Similar to significance tests, the confidence intervals will show you whether something is significance or not. If the same significance level is used, then there is no difference between performing a two-tailed significance test and a confidence interval. Both will prove the same thing.**