SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

General Information about the ***"Investigative Task:"***

- There are ____ free response questions on the AP Statistics Exam.
- Students have ______ minutes to complete these six questions.
- #____ is called the Investigative Task.
- Students should reserve about ______ minutes (MAX!) to complete this problem.
- It is worth ____ of the free response section, or _____ of the entire test.
- It typically covers several ________ ________________.
- It typically introduces something _________.
- Students should have a strategy:

    1. S___________/s________ the entire test, and r_______ problems.
    2. Find the two _________________ and do those first (≈20 minutes)
    3. **Then do #6 (up to 25 minutes)**
    4. Then spend the rest of your time (≈45 minutes) on the last three.

The investigative Task is designed to ________________ you. Try to ________________ read, think and give your best answers.

________ credit earned on the Investigative Task will help your score, so start s______________

Mean scores on Investigative Tasks:
(8-year avg: 1.52)

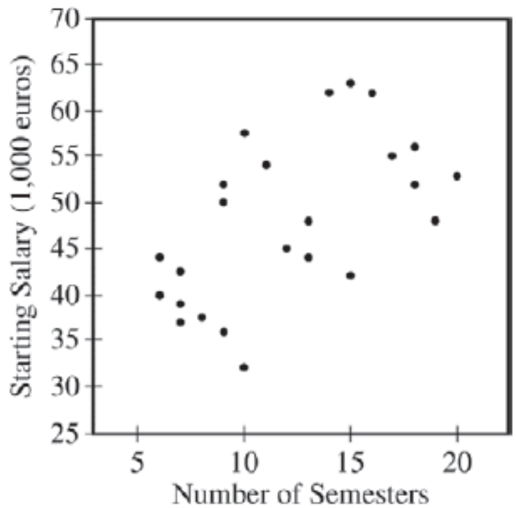
Year	2016	2015	2014	2013	2012	2011	2010	2009
Mean	1.61	1.08	1.29	2.14	1.50	1.31	1.92	1.32

6, 90, 6, 25, ¼, 1/8, big ideas, new, scout/scan, rate, easiest, stretch, carefully, ANY, somewhere

Investigative Task Samples

2016 #6

6. A newspaper in Germany reported that the more semesters needed to complete an academic program at the university, the greater the starting salary in the first year of a job. The report was based on a study that used a random sample of 24 people who had recently completed an academic program. Information was collected on the number of semesters each person in the sample needed to complete the program and the starting salary, in thousands of euros, for the first year of a job. The data are shown in the scatterplot below.



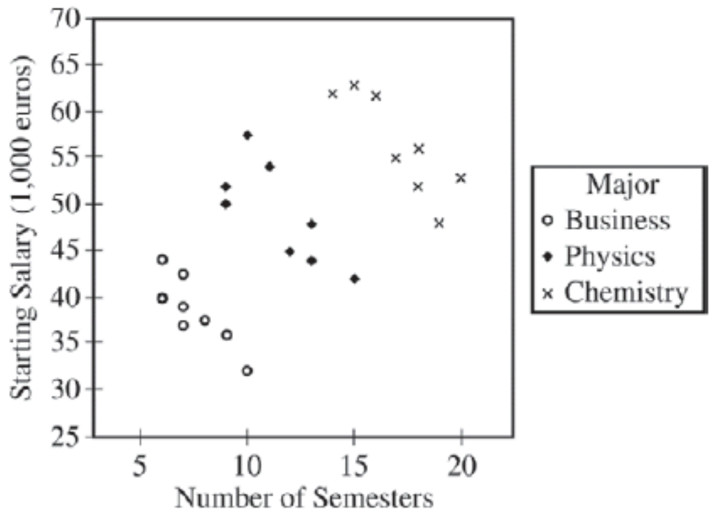
(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

The table below shows computer output from a linear regression analysis on the data.

Predictor	Coef	SE Coef	T	P
Constant	34.018	4.455	7.64	0.000
Semesters	1.1594	0.3482	3.33	0.003
S = 7.37702	R-Sq = 33.5%		R-Sq(adj) = 30.5%	

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

An independent researcher received the data from the newspaper and conducted a new analysis by separating the data into three groups based on the major of each person. A revised scatterplot identifying the major of each person is shown below.



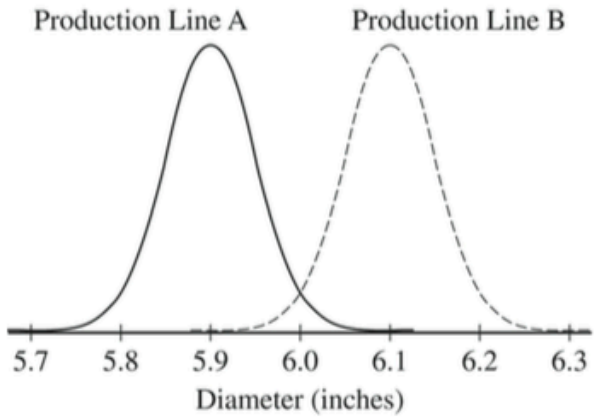
(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the <u>business</u> majors.

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

2015 #6:

6. Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.



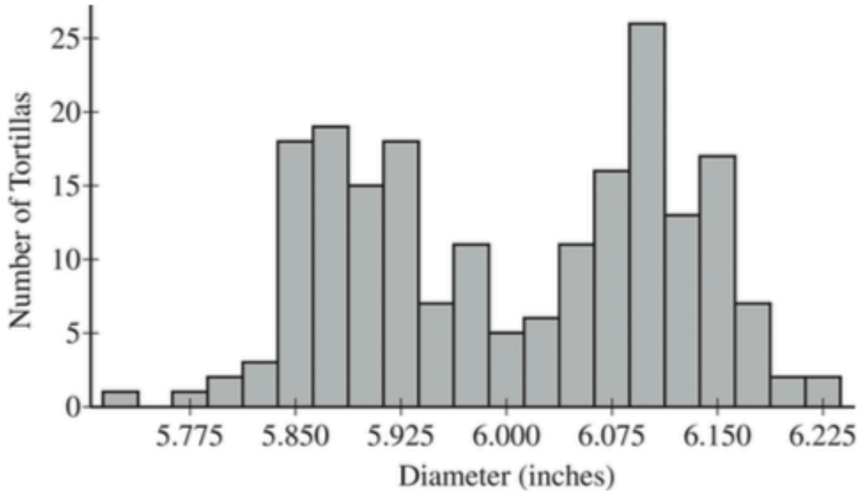
The tortillas produced at the factory are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected and the diameters are measured. From the sample of 200 tortillas, the manager of the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

(a) Will a sample obtained using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Explain why or why not.

(b) The figure below is a histogram of 200 diameters obtained by using one of the two sampling methods described. Considering the shape of the histogram, explain which method, Method 1 or Method 2, was most likely used to obtain a such a sample.



(c) Which of the two sampling methods, Method 1 or Method 2, will result in less variability in the diameters of the 200 tortillas in the sample on a given day? Explain.

Each day, the distribution of the 200,000 tortillas made that day has mean diameter 6 inches with standard deviation 0.11 inch.

(d) For samples of size 200 taken from one day's production, describe the sampling distribution of the sample mean diameter for samples that are obtained using Method 1.

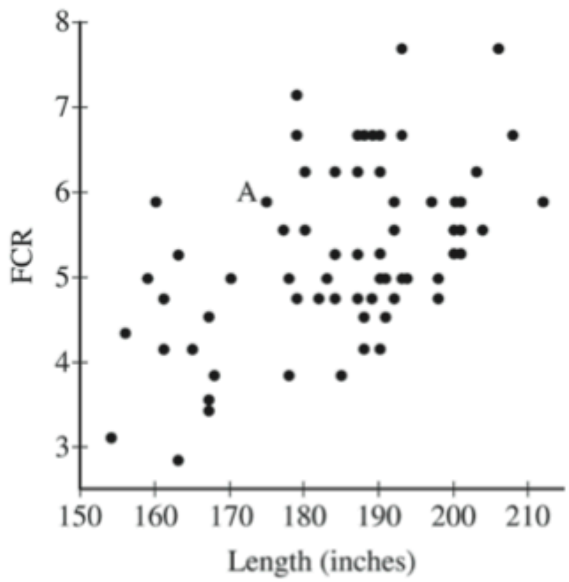
(e) Suppose that one of the two sampling methods will be selected and used every day for one year (365 days). The sample mean of the 200 diameters will be recorded each day. Which of the two methods will result in less variability in the distribution of the 365 sample means? Explain.

(f) A government inspector will visit the facility on June 22 to observe the sampling and to determine if the factory is in compliance with the advertised mean diameter of 6 inches. The manager knows that, with both sampling methods, the sample mean is an unbiased estimator of the population mean. However, the manager is unsure which method is more likely to produce a sample mean that is close to 6 inches on the day of sampling. Based on your previous answers, which of the two sampling methods, Method 1 or Method 2, is more likely to produce a sample mean close to 6 inches? Explain.

2014 #6:

6. Jamal is researching the characteristics of a car that might be useful in predicting the fuel consumption rate (FCR); that is, the number of gallons of gasoline that the car requires to travel 100 miles under conditions of typical city driving. The length of a car is one explanatory variable that can be used to predict FCR. Graph I is a scatterplot showing the lengths of 66 cars plotted with the corresponding FCR. One point on the graph is labeled A.

GRAPH I



Jamal examined the scatterplot and determined that a linear model would be a reasonable way to express the relationship between FCR and length. A computer output from a linear regression is shown below.

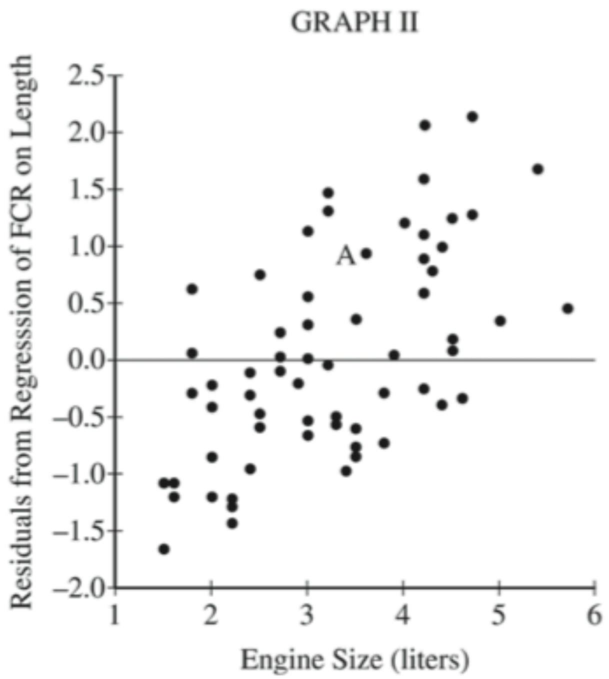
Linear Fit
FCR = −1.595789 + 0.0372614 * Length

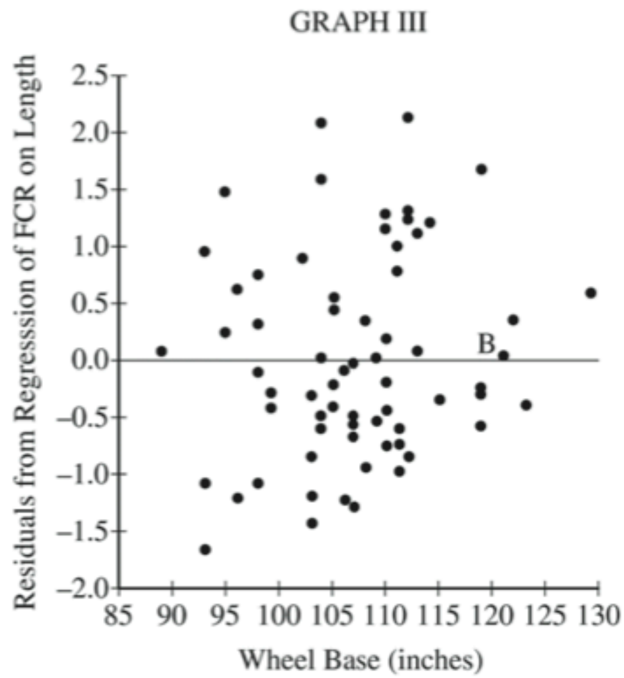
Summary of Fit

RSquare	0.250401
Root Mean Square Error	0.902382
Observations	66

(a) The point on the graph labeled A represents one car of length 175 inches and an FCR of 5.88. Calculate and interpret the residual for the car relative to the least squares regression line.

Jamal knows that it is possible to predict a response variable using more than one explanatory variable. He wants to see if he can improve the original model of predicting FCR from length by including a second explanatory variable in addition to length. He is considering including engine size, in liters, or wheel base (the length between axles), in inches. Graph II is a scatterplot showing the engine size of the 66 cars plotted with the corresponding residuals from the regression of FCR on length. Graph III is a scatterplot showing the wheel base of the 66 cars plotted with the corresponding residuals from the regression of FCR on length.





(b) In graph II, the point labeled A corresponds to the same car whose point was labeled A in graph I. The measurements for the car represented by point A are given below.

FCR	Length (inches)	Engine Size (liters)	Wheel Base (inches)
5.88	175	3.6	93

(i) Circle the point on graph III that corresponds to the car represented by point A on graphs I and II.

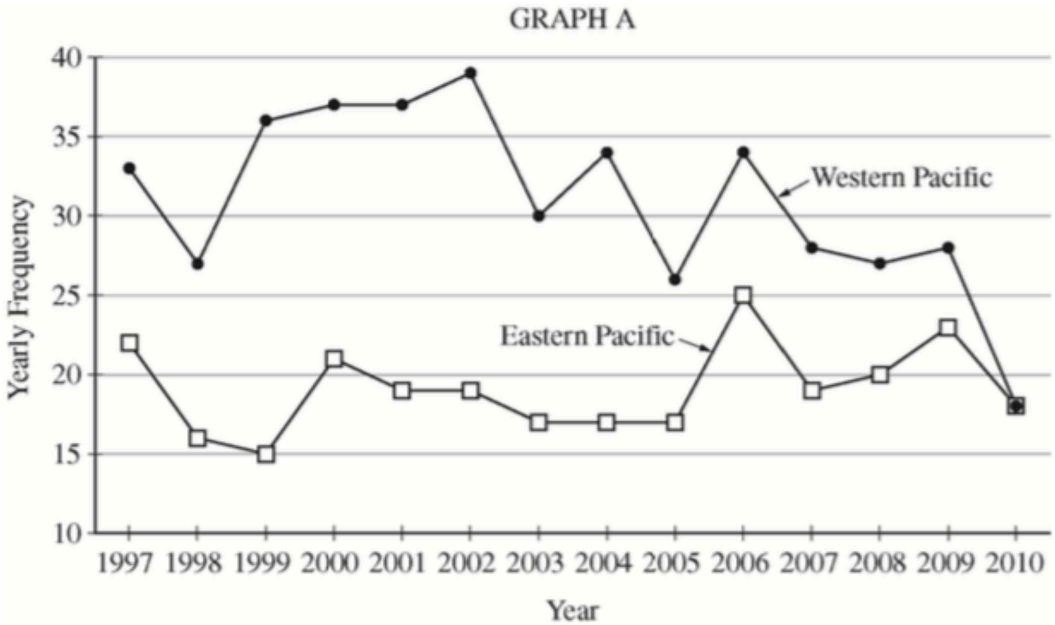
(ii) There is a point on graph III labeled B. It is very close to the horizontal line at 0. What does that indicate about the FCR of the car represented by point B?

(c) Write a few sentences to compare the association between the variables in graph II with the association between the variables in graph III.

(d) Jamal wants to predict FCR using length and one of the other variables, engine size or wheel base. Based on your response to part (c), which variable, engine size or wheel base, should Jamal use in addition to length if he wants to improve the prediction? Explain why you chose that variable.

2013 #6:

6. Tropical storms in the Pacific Ocean with sustained winds that exceed 74 miles per hour are called typhoons. Graph A below displays the number of recorded typhoons in two regions of the Pacific Ocean—the Eastern Pacific and the Western Pacific—for the years from 1997 to 2010.



(a) Compare the distributions of yearly frequencies of typhoons for the two regions of the Pacific Ocean for the years from 1997 to 2010.

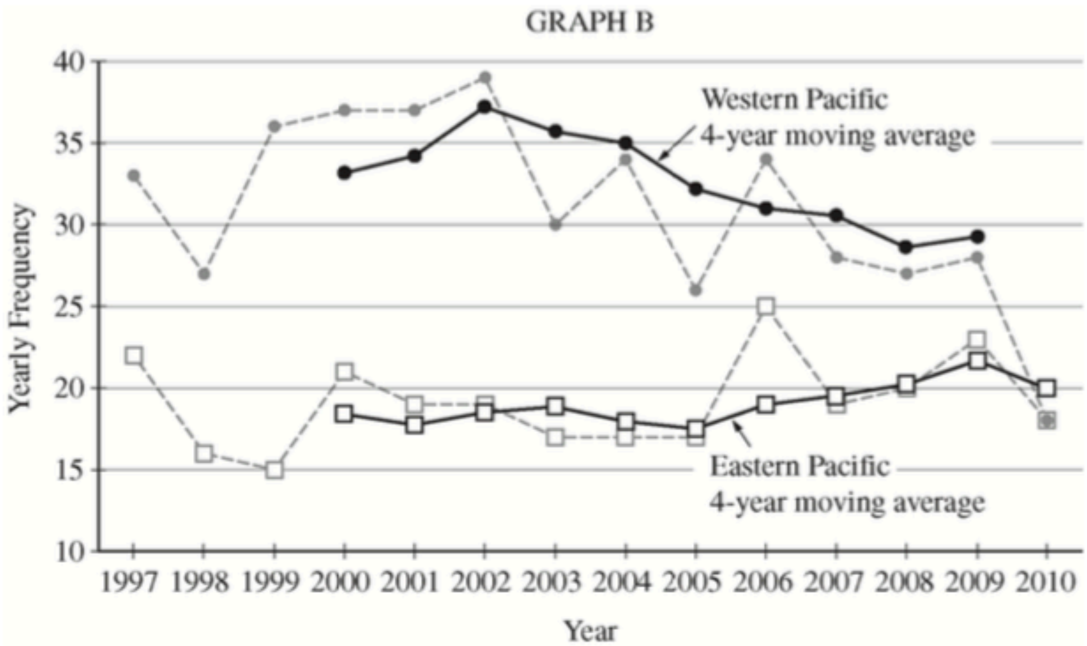
(b) For each region, describe how the yearly frequencies changed over the time period from 1997 to 2010.

A moving average for data collected at regular time increments is the average of data values for two or more consecutive increments. The 4-year moving averages for the typhoon data are provided in the table below. For example, the Eastern Pacific 4-year moving average for 2000 is the average of 22, 16, 15, and 21, which is equal to 18.50.

Year	Number of Typhoons in the Eastern Pacific	Eastern Pacific 4-year moving average	Number of Typhoons in the Western Pacific	Western Pacific 4-year moving average
1997	22		33	
1998	16		27	
1999	15		36	
2000	21	18.50	37	33.25
2001	19	17.75	37	34.25
2002	19	18.50	39	37.25
2003	17	19.00	30	35.75
2004	17	18.00	34	35.00
2005	17	17.50	26	32.25
2006	25	19.00	34	31.00
2007	19	19.50	28	30.50
2008	20	20.25	27	28.75
2009	23	21.75	28	29.25
2010	18	20.00	18	

(c) Show how to calculate the 4-year moving average for the year 2010 in the Western Pacific. Write your value in the appropriate place in the table.

(d) Graph B below shows both yearly frequencies (connected by dashed lines) and the respective 4-year moving averages (connected by solid lines). Use your answer in part (c) to complete the graph.



(e) Consider graph B.

i) What information is more apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

ii) What information is less apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

2012 #6:

6. Two students at a large high school, Peter and Rania, wanted to estimate μ, the mean number of soft drinks that a student at their school consumes in a week. A complete roster of the names and genders for the 2,000 students at their school was available. Peter selected a simple random sample of 100 students. Rania, knowing that 60 percent of the students at the school are female, selected a simple random sample of 60 females and an independent simple random sample of 40 males. Both asked all of the students in their samples how many soft drinks they typically consume in a week.

(a) Describe a method Peter could have used to select a simple random sample of 100 students from the school.

Peter and Rania conducted their studies as described. Peter used the sample mean X as a point estimator for μ. Rania used $\overline{X}_{\text{overall}} = (0.6)\overline{X}_{\text{female}} + (0.4)\overline{X}_{\text{male}}$ as a point estimator for μ, where $\overline{X}_{\text{female}}$ is the mean of the sample of 60 females and $\overline{X}_{\text{male}}$ is the mean of the sample of 40 males.

Summary statistics for Peter's data are shown in the table below.

Variable	N	Mean	Standard Deviation
Number of soft drinks	100	5.32	4.13

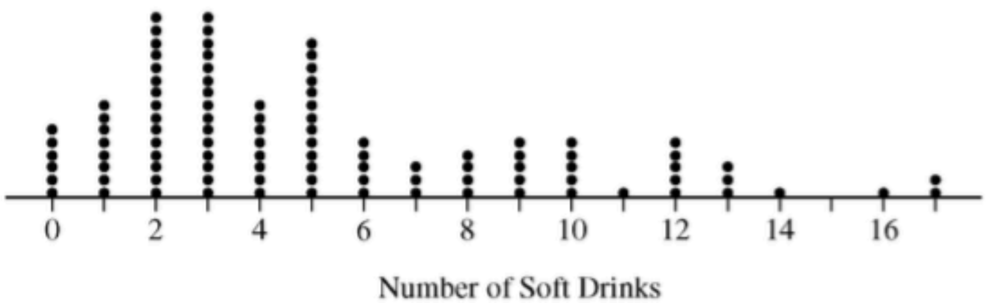
(b) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution (sometimes called the standard error) of Peter's point estimator $\overline{X}$.

Summary statistics for Rania's data are shown in the table below.

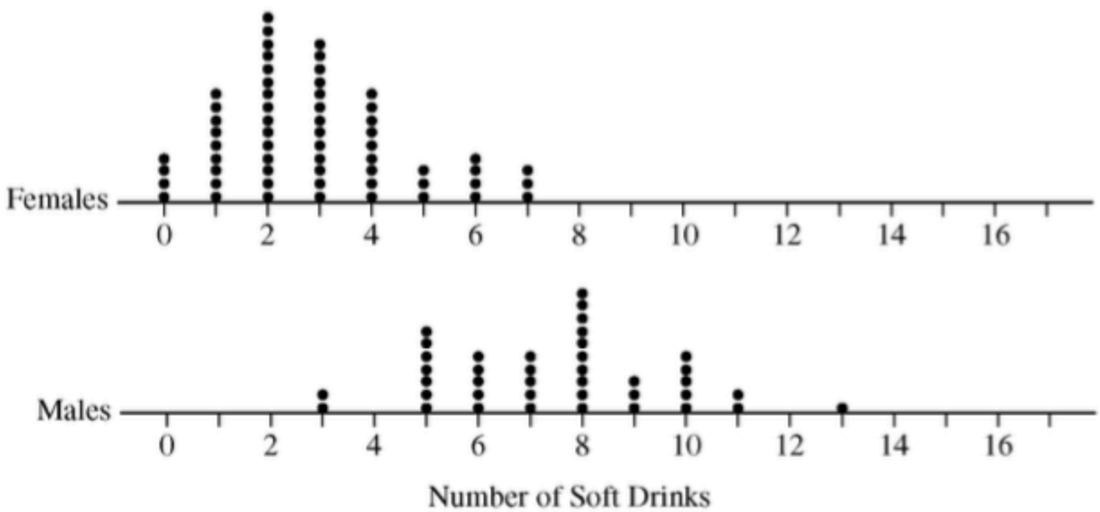
Variable	Gender	N	Mean	Standard Deviation
Number of soft drinks	Female	60	2.90	1.80
	Male	40	7.45	2.22

(c) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution of Rania's point estimator $\overline{X}_{\text{overall}} = (0.6)\overline{X}_{\text{female}} + (0.4)\overline{X}_{\text{male}}$.

A dotplot of Peter's sample data is given below.



Comparative dotplots of Rania's sample data are given below.



(d) Using the dotplots above, explain why Rania's point estimator has a smaller estimated standard deviation than the estimated standard deviation of Peter's point estimator.

2010 #6:

6. Hurricane damage amounts, in millions of dollars per acre, were estimated from insurance records for major hurricanes for the past three decades. A stratified random sample of five locations (based on categories of distance from the coast) was selected from each of three coastal regions in the southeastern United States. The three regions were Gulf Coast (Alabama, Louisiana, Mississippi), Florida, and Lower Atlantic (Georgia, South Carolina, North Carolina). Damage amounts in millions of dollars per acre, adjusted for inflation, are shown in the table below.

HURRICANE DAMAGE AMOUNTS IN MILLIONS OF DOLLARS PER ACRE

	Distance from Coast				
	< 1 mile	1 to 2 miles	2 to 5 miles	5 to 10 miles	10 to 20 miles
Gulf Coast	24.7	21.0	12.0	7.3	1.7
Florida	35.1	31.7	20.7	6.4	3.0
Lower Atlantic	21.8	15.7	12.6	1.2	0.3

(a) Sketch a graphical display that compares the hurricane damage amounts per acre for the three different coastal regions (Gulf Coast, Florida, and Lower Atlantic) and that also shows how the damage amounts vary with distance from the coast.

(b) Describe differences and similarities in the hurricane damage amounts among the three regions.

Because the distributions of hurricane damage amounts are often skewed, statisticians frequently use rank values to analyze such data.

(c) In the table below, the hurricane damage amounts have been replaced by the ranks 1, 2, or 3. For each of the distance categories, the highest damage amount is assigned a rank of 1 and the lowest damage amount is assigned a rank of 3. Determine the missing ranks for the 10-to-20-miles distance category and calculate the average rank for each of the three regions. Place the values in the table below.

ASSIGNED RANKS WITHIN DISTANCE CATEGORIES

	Distance from Coast					
	< 1 mile	1 to 2 miles	2 to 5 miles	5 to 10 miles	10 to 20 miles	Average Rank
Gulf Coast	2	2	3	1		
Florida	1	1	1	2		
Lower Atlantic	3	3	2	3		

(d) Consider testing the following hypotheses.

H_0: There is no difference in the distributions of hurricane damage amounts among the three regions.

H_a: There is a difference in the distributions of hurricane damage amounts among the three regions.

If there is no difference in the distribution of hurricane damage amounts among the three regions (Gulf Coast, Florida, and Lower Atlantic), the expected value of the average rank for each of the three regions is 2. Therefore, the following test statistic can be used to evaluate the hypotheses above:

$$Q = 5\left[\left(\bar{R}_G - 2\right)^2 + \left(\bar{R}_F - 2\right)^2 + \left(\bar{R}_A - 2\right)^2\right]$$

where $\bar{R}_G$ is the average rank over the five distance categories for the Gulf Coast (and $\bar{R}_F$ and $\bar{R}_A$ are similarly defined for the Florida and Lower Atlantic coastal regions).

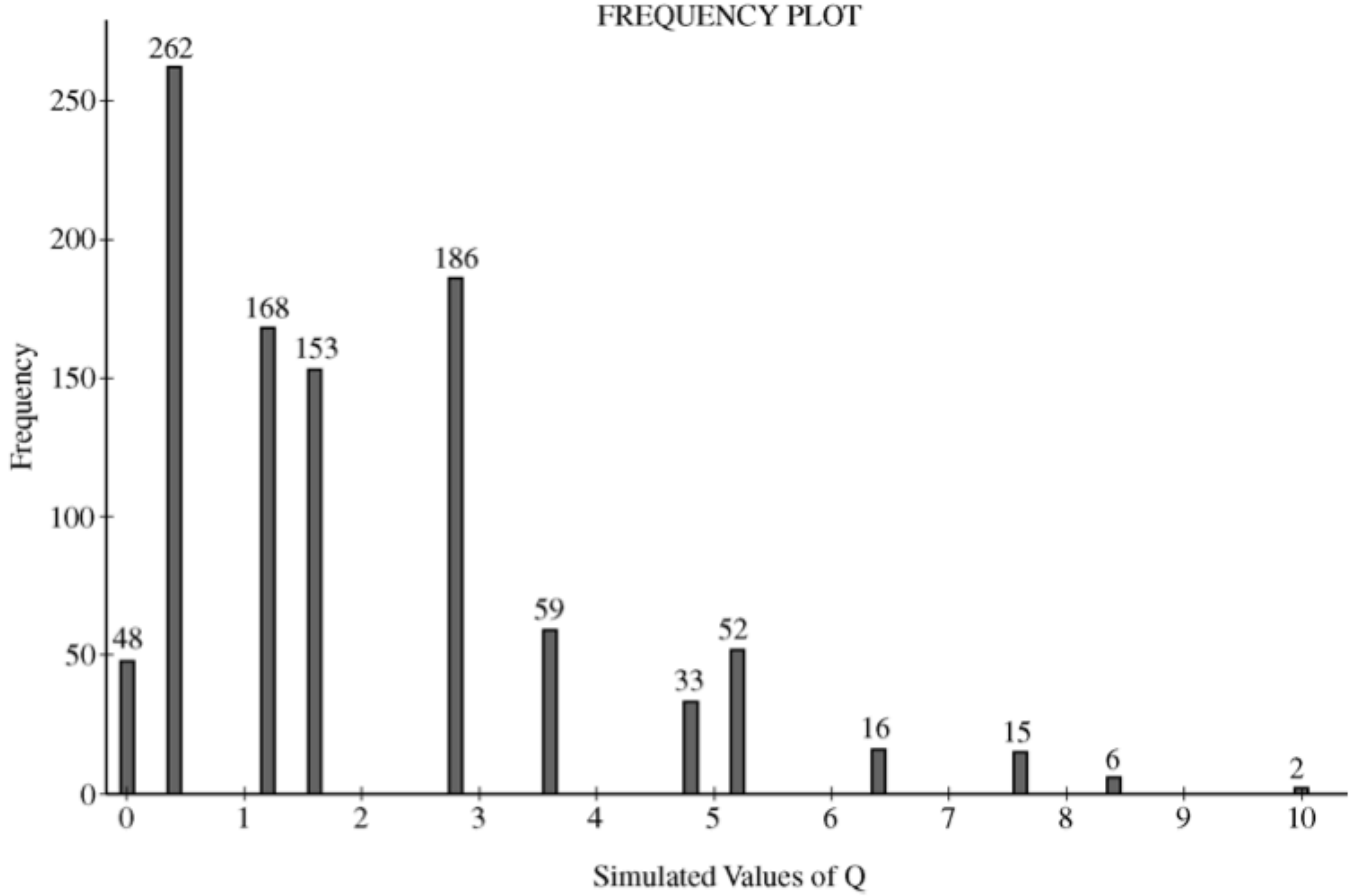
Calculate the value of the test statistic Q using the average ranks you obtained in part (c).

(e) One thousand simulated values of this test statistic, Q, were calculated, assuming no difference in the distributions of hurricane damage amounts among the three coastal regions. The results are shown in the table below. These data are also shown in the frequency plot where the heights of the lines represent the frequency of occurrence of simulated values of Q.

Frequency Table for Simulated Values of Q

Q	Frequency	Cumulative Frequency	Percent	Cumulative Percent
0.0	48	48	4.80	4.80
0.4	262	310	26.20	31.00
1.2	168	478	16.80	47.80
1.6	153	631	15.30	63.10
2.8	186	817	18.60	81.70
3.6	59	876	5.90	87.60
4.8	33	909	3.30	90.90
5.2	52	961	5.20	96.10
6.4	16	977	1.60	97.70
7.6	15	992	1.50	99.20
8.4	6	998	0.60	99.80
10.0	2	1000	0.20	100.00

FREQUENCY PLOT



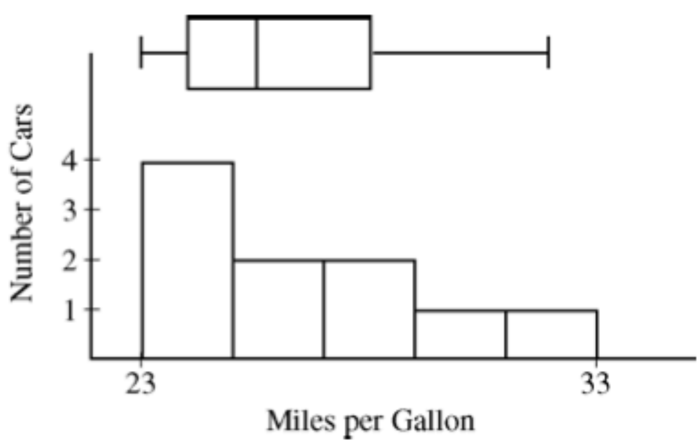
Use these simulated values and the test statistic you calculated in part (d) to determine if the observed data provide evidence of a significant difference in the distributions of hurricane damage amounts among the three coastal regions. Explain.

2009 #6:

6. A consumer organization was concerned that an automobile manufacturer was misleading customers by overstating the average fuel efficiency (measured in miles per gallon, or mpg) of a particular car model. The model was advertised to get 27 mpg. To investigate, researchers selected a random sample of 10 cars of that model. Each car was then randomly assigned a different driver. Each car was driven for 5,000 miles, and the total fuel consumption was used to compute mpg for that car.

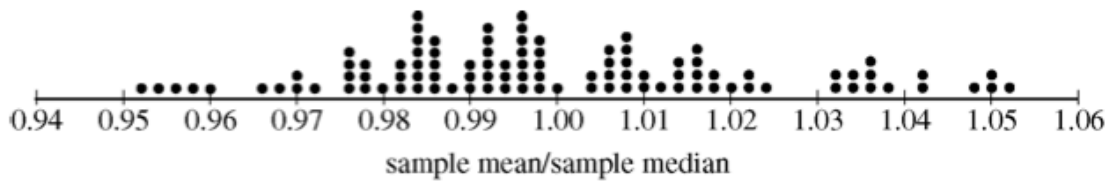
   (a) Define the parameter of interest and state the null and alternative hypotheses the consumer organization is interested in testing.

One condition for conducting a one-sample *t*-test in this situation is that the mpg measurements for the population of cars of this model should be normally distributed. However, the boxplot and histogram shown below indicate that the distribution of the 10 sample values is skewed to the right.



(b) One possible statistic that measures skewness is the ratio $\frac{\text{sample mean}}{\text{sample median}}$. What values of that statistic (small, large, close to one) might indicate that the population distribution of mpg values is skewed to the right? Explain.

(c) Even though the mpg values in the sample were skewed to the right, it is still possible that the population distribution of mpg values is normally distributed and that the skewness was due to sampling variability. To investigate, 100 samples, each of size 10, were taken from a normal distribution with the same mean and standard deviation as the original sample. For each of those 100 samples, the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was calculated. A dotplot of the 100 simulated statistics is shown below.



In the original sample, the value of the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was 1.03. Based on the value of 1.03 and the dotplot above, is it plausible that the original sample of 10 cars came from a normal population, or do the simulated results suggest the original population is really skewed to the right? Explain.

(d) The table below shows summary statistics for mpg measurements for the original sample of 10 cars.

Minimum	Q1	Median	Q3	Maximum
23	24	25.5	28	32

Choosing only from the summary statistics in the table, define a formula for a different statistic that measures skewness.

What values of that statistic might indicate that the distribution is skewed to the right? Explain.

SOLUTIONS:

2016:

Part (a):

The scatterplot supports the newspaper report about number of semesters needed to complete an academic program and starting salary because it shows a positive association between these two variables.

Part (b):

The slope is 1.1594. For each additional semester needed to complete an academic program, the predicted starting salary increases by €1,159.40.

Part (c):

For the business majors alone, there is a strong, negative, linear association between number of semesters and starting salary. Business majors who need a greater number of semesters to complete an academic program tend to have lower starting salaries.

Part (d):

Business majors have the lowest median starting salary at around €38,000, followed by physics majors at around €48,000, and then chemistry majors with the highest median starting salary at around €55,000.

Part (e):

The newspaper report should be modified to account for major. Overall, majors that take longer to complete tend to have higher starting salaries, with chemistry the highest, physics the next highest, and business the lowest. However, within a major, students who take a greater number of semesters tend to have lower starting salaries.

2015:

Part (a):

No, a sample obtained using Method 2 will not be representative of all tortillas made that day. The sample obtained using Method 2 will only represent the tortillas from one production line, not from the entire population because the distributions of diameters for the two production lines are different.

Part (b):

Method 1 was most likely used to select this sample. The bimodal shape in the histogram of sample data indicates that tortillas were selected from both production lines, which is what would happen using Method 1. Method 2 would be likely to produce a unimodal distribution of diameters centered at either 5.9 inches or 6.1 inches.

Part (c):

Method 2 would result in less variability in the sample of 200 tortillas on a given day because the sample comes from only one production line. Because the distributions of diameters are not the same for the two production lines, selecting tortillas from both lines as in Method 1 would result in more variable sample data.

Part (d):

The sampling distribution of the sample mean diameter for samples obtained using Method 1 would be approximately normal with mean 6 inches and standard deviation $\frac{0.11}{\sqrt{200}} \approx 0.0078$ inch.

Part (e):

Method 1 would result in less variability in the sample means over the 365 days, because with Method 2, roughly half of the sample means will be around 5.9 inches and the other half will be around 6.1 inches. With Method 1, however, the sample means will all be very close to 6 inches, as indicated by the standard deviation in part (d).

Part (f):

Method 1 is more likely to produce a sample mean close to 6 inches. Because the sample mean is an unbiased estimator for both methods, the manager should pick the method that would result in less variability in the distribution of the sample mean. Based on the answer to part (e), Method 1 results in less variability in the distribution of the sample mean.

2014:

Part (a):

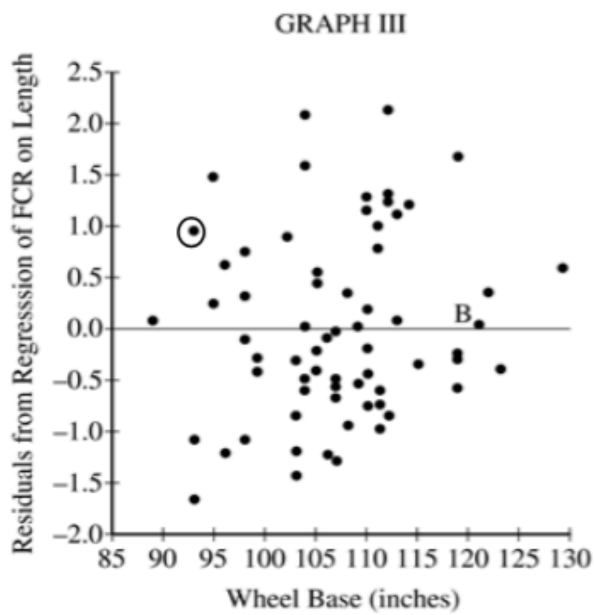
For a car with length 175 inches, the predicted value for the car's FCR, based on the least squares regression line, is

$$\text{predicted FCR} = -1.595789 + 0.0372614(175) \approx 4.92 \text{ gallons per 100 miles.}$$

The actual FCR for the car is 5.88, so the residual is $5.88 - 4.92 = 0.96$. The residual value means that the car's FCR is 0.96 gallons per 100 miles greater than would be predicted for a car of its length.

Part (b):

(i) The point with a wheel base of 93 inches and a residual of 0.96 gallons per 100 miles is circled in graph III below.



(ii) Point B corresponds to a car with an actual FCR that is very close to the FCR that would be predicted for a car with its length by the regression model which predicts FCR using the explanatory variable length.

Part (c):

Graph II reveals a moderate association that is positive and linear. In contrast, there is a weak association that is positive and linear in graph III. The association between engine size and residual (from predicting FCR based on length) is stronger than the association between wheel base and residual (from predicting FCR based on length).

Part (d):

Engine size is a better choice than wheel base for including with length in a regression model for predicting FCR. The stronger association between engine size and residual (from predicting FCR based on length) indicates that engine size is more useful than wheel base for reducing the variability in FCR values that remains unexplained (as indicated by residuals) after predicting FCR based on length.

2013:

Part (a):

The Western Pacific Ocean had more typhoons than the Eastern Pacific Ocean in all but one of these years. The average seems to have been about 31 typhoons per year in the Western Pacific Ocean, which is higher than the average of about 19 typhoons per year in the Eastern Pacific Ocean. The Western Pacific Ocean also saw more variability (in number of typhoons per year) than the Eastern Pacific Ocean; for example, the range of the frequencies for the Western Pacific is about 21 typhoons and only 10 typhoons for the Eastern Pacific.

Part (b):

The Western Pacific Ocean had a decreasing trend in number of typhoons per year over this time period, especially from about 2001 through 2010. In contrast, the Eastern Pacific Ocean was fairly consistent in the number of typhoons per year over this time period, with a slight increasing trend in the later years from 2005 through 2010.

Part (c):

The four-year moving average for the year 2010 in the Western Pacific Ocean is

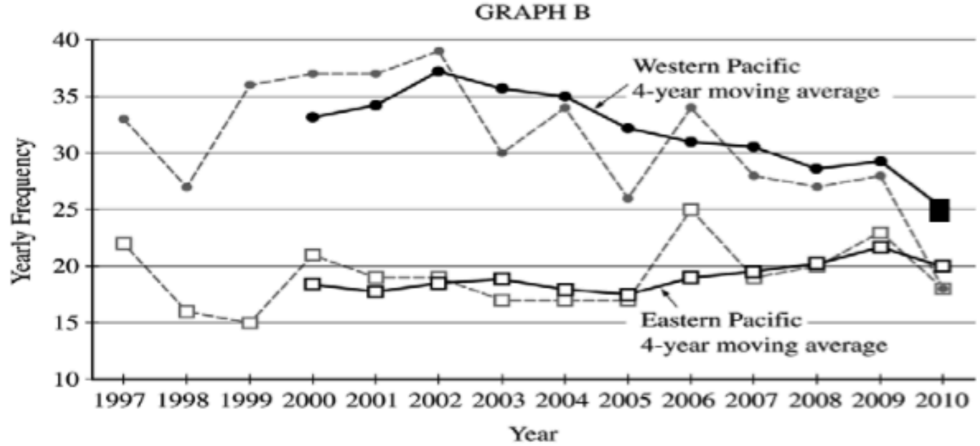
$$\frac{28 + 27 + 28 + 18}{4} = 25.25.$$

The value is written in the table as follows.

⋮	⋮	⋮	⋮	⋮
2008	20	20.25	27	28.75
2009	23	21.75	28	29.25
2010	18	20.00	18	**25.25**

Part (d):

GRAPH B



Part (e):

(i) The overall trends across this time period were more apparent with the moving averages than with the original frequencies. The moving averages reduce variability, making more apparent the overall decreasing trend in number of typhoons in the Western Pacific Ocean and the slight increasing trend in the number of typhoons in the Eastern Pacific Ocean.

(ii) The year-to-year variability in number of typhoons is less apparent with the moving averages than with the original frequencies.

2012:

Part (a):

Peter can number the students from 1 to 2,000 and then use a calculator or computer to generate 100 unique random numbers between 1 and 2,000 without replacement. If non-unique numbers are generated, the repeated numbers are ignored until 100 unique numbers are obtained. The students whose numbers correspond to the randomly generated numbers are then selected for the sample.

Part (b):

The estimated standard deviation of the sampling distribution of the sample mean is:
$\frac{s}{\sqrt{n}}$, or $\frac{4.13}{\sqrt{100}} = 0.413$.

Part (c):

The variance of Rania's estimator is $(0.6)^2 \text{Var}(\bar{X}_f) + (0.4)^2 \text{Var}(\bar{X}_m)$, where $\text{Var}(\bar{X}_f) = \frac{\sigma_f^2}{n_f}$ represents the variance of the point estimator for females and $\text{Var}(\bar{X}_m) = \frac{\sigma_m^2}{n_m}$ represents the variance of the point estimator for males.

The estimated standard deviation is the square root of the variance. Using the respective sample standard deviations s_f and s_m for the population parameters, Rania's estimate is calculated as:

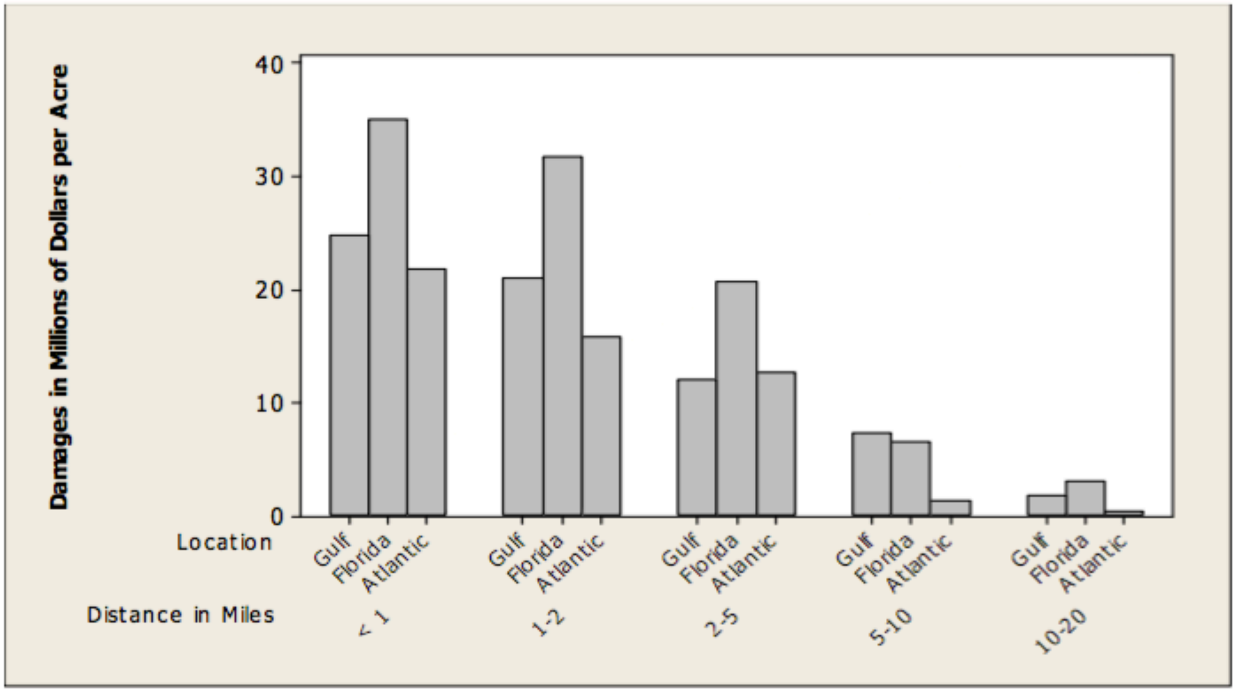
$$\sqrt{(0.6)^2 \frac{s_f^2}{n_f} + (0.4)^2 \frac{s_m^2}{n_m}} = \sqrt{(0.6)^2 \frac{(1.80)^2}{60} + (0.4)^2 \frac{(2.22)^2}{40}} \approx \sqrt{0.01944 + 0.01972} \approx 0.198.$$

Part (d):

The comparative dotplots from Rania's data reveal that the distribution of the number of soft drinks for females appears to be quite different from that of males. In particular, the centers of the distributions appear to be significantly different. Additionally, the variability of values around the center *within* gender in each of Rania's dotplots appears to be considerably less than the variability displayed in the dotplot of Peter's data. Rania's estimator takes advantage of the decreased variability within gender because her data were obtained by sampling the two genders separately. Peter's estimator has more variability because his data were obtained from a simple random sample of all the high school students.

2010:

Part (a):



Part (b):

In all three regions (Gulf Coast, Florida, Lower Atlantic) the hurricane damage amounts tend to decrease as distance from the coast increases. For almost all given distances from the coast, the Florida region has the largest damage amounts. Also, for any given distance, the Gulf Coast and Lower Atlantic regions have similar damage amounts but with the Lower Atlantic damage amounts generally smaller.

Part (c):

For the "10 to 20 miles" distance category: The Florida region has the most damage (3.0 million dollars per acre) and so has rank 1. The region with the second-most damage is the Gulf Coast (1.7 million dollars), obtaining rank 2. The Lower Atlantic region has the least damage (0.3 million dollars) and so has rank 3. The last columns of the table should be filled in as follows:

	10 to 20 miles	Average Rank
Gulf Coast	2	2.0
Florida	1	1.2
Lower Atlantic	3	2.8

The average ranks are computed $\frac{2+2+3+1+2}{5}=2.0$ for the five Gulf Coast damage ranks, $\frac{1+1+1+2+1}{5}=1.2$ for the five Florida damage ranks and $\frac{3+3+2+3+3}{5}=2.8$ for the five Lower Atlantic damage ranks.

Part (d):

The calculated value of the test statistic Q is

$$Q = 5\left[(2.0-2)^2 + (1.2-2)^2 + (2.8-2)^2\right] = 5[0+0.64+0.64] = 6.4.$$

Part (e):

A Q value of 6.4 or larger occurred in $\frac{39}{1{,}000} = 0.039$ (or 3.9 percent) of the 1,000 repetitions. All 1,000 repetitions of the simulation assumed there was no difference in the distribution of damage amounts among the three regions. This is a fairly small (approximate) p-value (less than 0.05), indicating that a test statistic as large or larger than the observed test statistic of $Q = 6.4$ would be fairly unlikely to occur by chance alone if there really was no difference among the regions for each distance category. The sample data therefore provide reasonably strong evidence that there is a difference in the distributions of hurricane damage amounts among these three regions.

2009:

Part (a):

The parameter of interest is $\mu =$ population mean miles per gallon (mpg) of a particular car model.

The null and alternative hypotheses are as follows:

H_0: $\mu = 27$
H_a: $\mu < 27$

Part (b):

If the distribution is right-skewed, one would expect the mean to be greater than the median.

Therefore the ratio $\dfrac{\text{sample mean}}{\text{sample median}}$ should be large (greater than 1).

Part (c):

Because we are testing for right-skewness, the estimated p-value will be the proportion of the simulated statistics that are greater than or equal to the observed value of 1.03. The dotplot shows that 14 of the 100 values are more than 1.03. Because this simulated p-value (0.14) is larger than any reasonable significance level, we do not have convincing evidence that the original population is skewed to the right and conclude that it is plausible that the original sample came from a normal population.

Part (d):

One possible statistic is $\dfrac{\text{maximum - median}}{\text{median - minimum}}$

If the distribution is right-skewed, one would expect the distance from the median to the maximum to be larger than the distance from the median to the minimum; thus the ratio should be greater than 1.

STUDENT SOLUTIONS (all earned a score of "4")

2016:

(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

Yes, because there is a positive correlation shown btw semesters in the program and starting salary, even if the correlation is not extremely strong.

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

$\hat{y} = 1.1594x + 34.08$

1.1594 is the slope of the least squares regression line. It says ~~that~~ for every additional semester starting salary increases by 1.1594 thousand Euros (on average).

(c) Based on the people in the sample, describe the association between starting salary and number of semesters for the business majors.

There is a moderately strong negative correlation btw the number of semesters in a program and starting salary for buisness majors

(d) Based on the people in the sample, compare the median starting salaries for the three majors.

Chemistry has the highest median starting salary, followed by Physics and then buisness.

(e) Based on the analysis conducted by the independent researcher, how could the newspaper report be modified to give a better description of the relationship between the number of semesters and the starting salary for the people in the sample?

Majors that typically take longer to complete tend to come with higher starting salaries, however, taking longer than average than other peers in your major tends to lead to a lower starting salary.

Sample: 6A
Score: 4

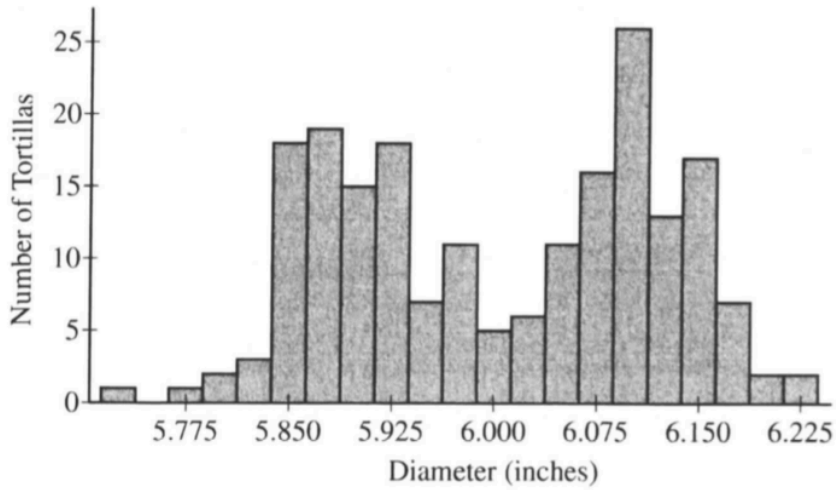
In part (a) the response addresses the positive association ("positive correlation") and uses the positive association to justify that the scatterplot supports the newspaper report ("Yes"), satisfying components 1 and 2 of section 1. The comment about the correlation not being extremely strong is ignored because the intention of part (a) is to address the direction of the association. In part (b) the response identifies the numerical value of the slope in the interpretation, interprets the slope correctly as the change in starting salary for each additional semester, and includes nondeterministic language in the interpretation ("on average"), satisfying components 3, 4, and 5. Because the response satisfies all five components, section 1 was scored as essentially correct. In part (c) the response states that the association is negative, states that the association is moderately strong, and uses both variable names (number of semesters, starting salary), satisfying components 1, 2, and 3 of section 2. In part (d) the response correctly compares the three majors and refers to "median starting salary," satisfying components 4 and 5. Because the response satisfies all five components, section 2 was scored as essentially correct. In part (e) the response states that there is a negative association for each major ("taking longer than average … in your major tends to lead to a lower starting salary") and notes the overall positive association ("majors that typically take longer to complete tend to come with higher starting salaries"). Section 3 was scored as essentially correct. Because all three sections were scored as essentially correct, the response earned a score of 4.

2015:

(a) Will a sample obtained using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Explain why or why not.

No. Method 2 suffers from selection bias because the sample will not be obtained from the entire population of tortillas, and the tortillas that are not sampled from will tend to have a different diameter.

(b) The figure below is a histogram of 200 diameters obtained by using one of the two sampling methods described. Considering the shape of the histogram, explain which method, Method 1 or Method 2, was most likely used to obtain a such a sample.

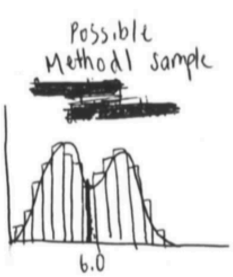


Method 1

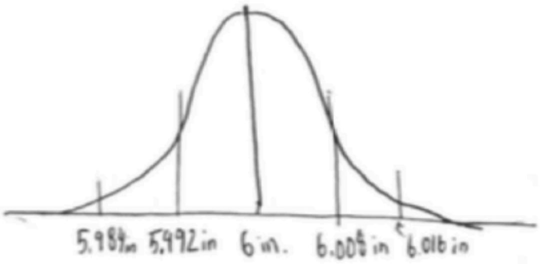
The histogram is bimodal, with a peak at ≈5.9 and a peak at ≈6.1 inches. This means the sample was likely obtained from both production lines, or Method 1.

(c) Which of the two sampling methods, Method 1 or Method 2, will result in less variability in the diameters of the 200 tortillas in the sample on a given day? Explain.

Method 2. Because Method 2 obtains the sample from just one production line, the sample will only have one peak, while Method 1 will have 2 peaks. Thus, Method 2 produces samples with smaller spreads/standard deviations.



(d) For samples of size 200 taken from one day's production, describe the sampling distribution of the sample mean diameter for samples that are obtained using Method 1.



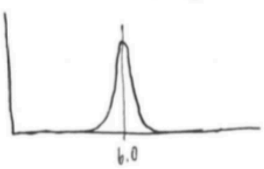
- Because $n = 200 \geq 30$, the central limit theorem says that the sampling distribution will be approximately normal.
- The mean of the sampling distribution is equal to the population mean.
- the standard deviation is:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.11}{\sqrt{200}} = 0.00778 \text{ in}$$

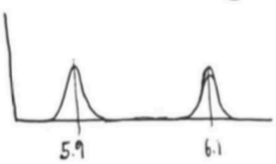
(e) Suppose that one of the two sampling methods will be selected and used every day for one year (365 days). The sample mean of the 200 diameters will be recorded each day. Which of the two methods will result in less variability in the distribution of the 365 sample means? Explain.

Method 1. The sampling distribution for Method 2 is bimodal, with a peak at 5.9 in and a peak at 6.1 in, because the samples have means close to either of these two values. This results in much greater variability in sample means than the unimodal Method 1 sampling distribution.

Method 1 sampling distribution



Method 2 sampling distribution



(f) A government inspector will visit the facility on June 22 to observe the sampling and to determine if the factory is in compliance with the advertised mean diameter of 6 inches. The manager knows that, with both sampling methods, the sample mean is an unbiased estimator of the population mean. However, the manager is unsure which method is more likely to produce a sample mean that is close to 6 inches on the day of sampling. Based on your previous answers, which of the two sampling methods, Method 1 or Method 2, is more likely to produce a sample mean close to 6 inches? Explain.

Method 1.

Method 2 is likely to produce a sample mean close to either 5.9 inches or 6.1 inches, depending on the chosen production line, but not close to 6.0 inches. Method 1 ~~tends~~ selects tortillas from BOTH production lines, so the tortilla diameter tends to average out very close to 6.0 inches.

Commentary on scoring (2015 #6, Student Sample A):

Section 1: Parts a, b, and c.
Section 2: Part d (shape, center, spread)
Section 3: Parts e and f.

A complete response (score = 4) requires E, E, E on all three sections.)

In part (a) the response says no and justifies the choice by stating that “the sample will not be obtained from the entire population of tortillas, and the tortillas that are not sampled from will tend to have a different diameter,” satisfying the first component of section 1. In part (b) the response says Method 1 because “the histogram is bimodal,” satisfying the second component of section 1. In part (c) the response says Method 2 because the sample is “from just one production line,” satisfying the third component of section 1. The picture in part (c), although not required, strengthens the explanation. Because the response satisfies all three components, **section 1 was scored as essentially correct.** In part (d) the response states that the sampling distribution will be “approximately normal,” satisfying the first component of section 2. The justification using the Central Limit Theorem is appropriate but not required. The response also correctly identifies the mean and standard deviation of the sampling distribution of the sample mean, satisfying the second and third components of section 2. The picture in part (d), although not required, is a nice way to summarize the response. Because the response satisfies all three components, **section 2 was scored as essentially correct.** In part (e) the response says Method 1 and correctly describes the sampling distribution of the sample mean for Method 2 as having some means close to 5.9 and other means close to 6.1, satisfying the first component of section 3. The picture in part (e), although not required, strengthens the already correct explanation. In part (f) the response says Method 1 because on a single day, “Method 2 is likely to produce a sample mean close to either 5.9 inches or 6.1 inches” while the average for Method 1 will be “very close to 6.0 inches.” Because the response satisfies both components, **section 3 was scored as essentially correct**. Because all three sections were scored as essentially correct, the response earned a score of 4.

2014:

(a) The point on the graph labeled A represents one car of length 175 inches and an FCR of 5.88. Calculate and interpret the residual for the car relative to the least squares regression line.

$\widehat{FCR} = -1.596 + .0373(\text{length (in.)})$

$= -1.596 + .0373(175)$

$= 4.925$ predicted FCR

residual = actual - predicted

$= 5.88 - 4.925$

$= .955$

A residual of .955 shows an underestimate by the least squares regression line. This residual shows that the predicted FCR is .955 gallons/100 miles lower than car A's actual consumption rate.

(b) In graph II, the point labeled A corresponds to the same car whose point was labeled A in graph I. The measurements for the car represented by point A are given below.

FCR	Length (inches)	Engine Size (liters)	Wheel Base (inches)
5.88	175	3.6	93

(i) Circle the point on graph III that corresponds to the car represented by point A on graphs I and II.

(93, .955) resid from part A

(ii) There is a point on graph III labeled B. It is very close to the horizontal line at 0. What does that indicate about the FCR of the car represented by point B?

This indicates that the predicted FCR for that particular car when using the least squares regression based on the car's length was very accurate, as it resulted in a very small residual. In other words, the prediction made using Jamal's initial least squares regression line was very close to the car's true FCR.

(c) Write a few sentences to compare the association between the variables in graph II with the association between the variables in graph III.

The association between engine size and residuals in Graph II shows a positive, roughly linear association, with weak to moderate strength. Graph III, between wheel base and residual shows no apparent pattern. Both graphs contain no obvious outliars. Graph III appears to have a larger scatter that Graph II.

(d) Jamal wants to predict FCR using length and one of the other variables, engine size or wheel base. Based on your response to part (c), which variable, engine size or wheel base, should Jamal use in addition to length if he wants to improve the prediction? Explain why you chose that variable.

Jamal should choose to use engine size in addition to length to improve his prediction. The scatter in Graph II, between engine size and residuals from FCR on length, appears to have a stronger association than Graph III. This means that more of the variation in the residuals will be able to be accounted for if engine size is added than if wheel base were to be added. The more variation that can be accounted for, the better Jamal will be able to make predictions.

In part (a) the residual is calculated correctly as 0.955, and it is stated that a residual of 0.955 shows that the predicted FCR is 0.955 gallons per 100 miles lower than car A's actual consumption rate. The response includes supporting calculations for the residual, and a correct interpretation of the residual value of 0.955 in context. Part (a) was scored as essentially correct. In part (b) the correct point was circled and labeled "A" on graph III, satisfying the first component. It is reported that the predicted FCR for the car corresponding to point B was very accurately predicted by the least squares regression based on the car's length. The response also states that the prediction made using Jamal's initial least squares regression line was very close to the car's true FCR, and the second component is satisfied. Part (b) was scored as essentially correct. In part (c) the association between engine size and residuals in graph II is described as positive, roughly linear with weak to moderate strength. No apparent pattern is reported for the association between engine size and residuals for graph III. Graph III is indicated to have a larger scatter than graph II. The stronger association of engine size and residuals than wheel base and residuals is specifically stated and used in part (d) in the choice of engine size. Thus, there is a description of form, direction and strength of association for both graphs and a comparison of strength of association. Part (c) was scored as essentially correct. In part (d) the correct choice of engine size is made. The choice is justified by both the stronger association in graph II than in graph III and by a greater reduction in the variation of the residuals when engine size is added to the model. Part (d) was scored as essentially correct. Because all four parts were scored as essentially correct, the response earned a score of 4.

2013:

(a) Compare the distributions of yearly frequencies of typhoons for the two regions of the Pacific Ocean for the years from 1997 to 2010.

The distribution of the western Pacific is centered higher than Eastern Pacific. There is more variability for western Pacific than for Eastern Pacific. The only time the two were the same was in 2010. The range of western is about 22 which is higher than Eastern rage of about 10.

(b) For each region, describe how the yearly frequencies changed over the time period from 1997 to 2010.

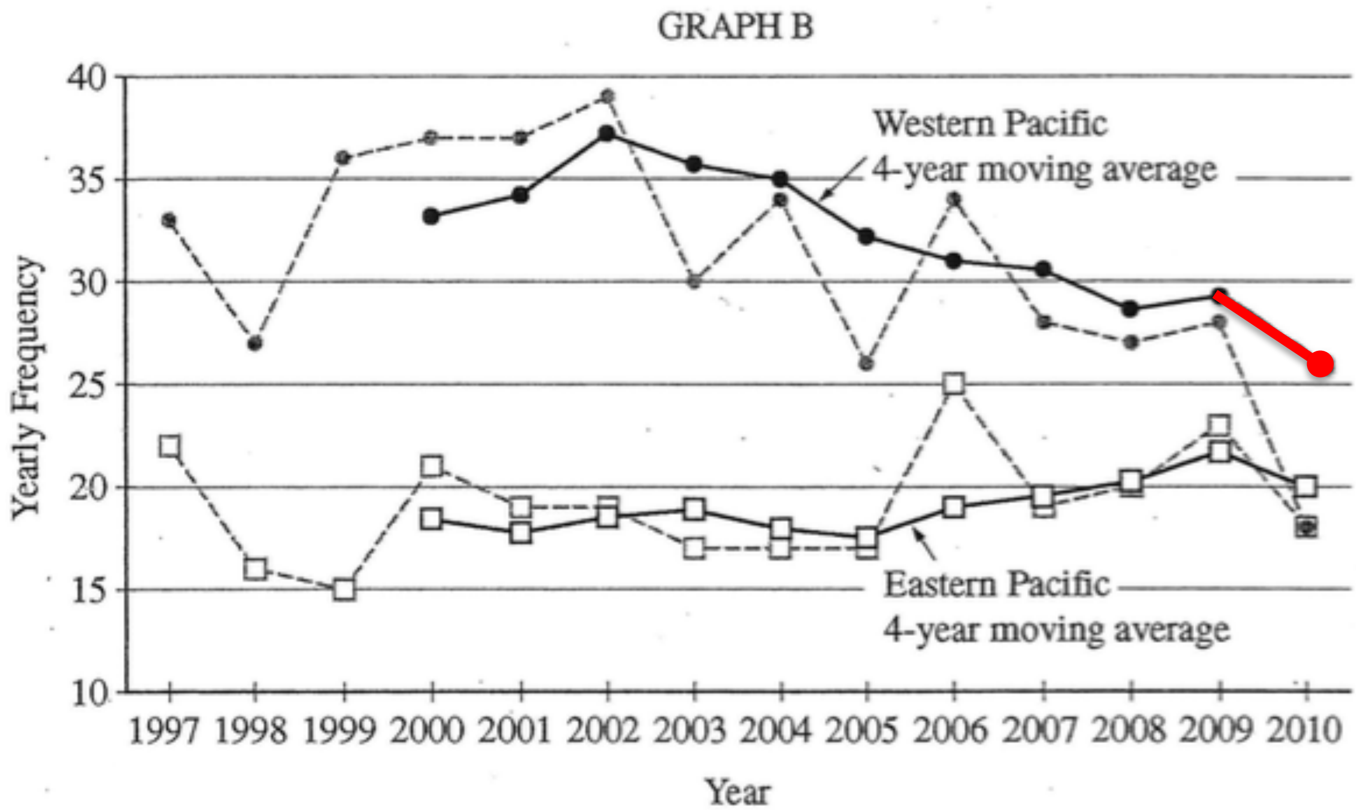
Western: The yearly frequencies for Western Pacific gradually got smaller over time.

Eastern: The yearly frequencies for western Pacific pretty much stayed the same, possibly getting slighthy bigger.

(c) Show how to calculate the 4-year moving average for the year 2010 in the Western Pacific. Write your value in the appropriate place in the table.

Take that year and the 3 previous year's number of typhoons and get the average of it. That's the 4-year moving average.

$$\frac{18+28+27+28}{4} = 25.25$$



(e) Consider graph B.

i) What information is more apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

The trends in the data are more apparent. Western is becoming less frequent and Eastern is becoming more frequent (not as drastically as western).

ii) What information is less apparent from the plots of the 4-year moving averages than from the plots of the yearly frequencies of typhoons?

Big spikes in the data are less apparant than in the regular yearly frequencies.

In part (a) it is stated that the distribution of the Western Pacific Ocean is centered higher than the distribution of the Eastern Pacific Ocean. The greater variability in the Western Pacific Ocean frequencies is stated and illustrated with calculated values for the two ranges. Section 1 was scored as essentially correct. The appropriate trends for the frequencies for both the Western Pacific Ocean and the Eastern Pacific Ocean are given in part (b), and section 2 was scored as essentially correct. In parts (c) and (d) the correct 4-year moving average for 2010 is calculated, appropriate calculations are shown, and the calculated value is used to appropriately complete the table and Graph B for the Western Pacific Ocean moving averages. Section 3 was scored as essentially correct. In part (e) it is indicated that trends for both the Western Pacific Ocean and the Eastern Pacific Ocean are more apparent in the plots of 4-year moving averages than in the plots of the yearly frequencies. It is also indicated that big changes from year to year are less apparent in the plots of 4-year moving averages than in the plots of the yearly frequencies. The description of declining trend in the Western Pacific Ocean and the slight increasing trend in the Eastern Pacific Ocean provides good linkage of the response to Graph B. Section 4 was scored as essentially correct. Because all four sections were scored as essentially correct, the response earned a score of 4.

2012:

6. Two students at a large high school, Peter and Rania, wanted to estimate μ, the mean number of soft drinks that a student at their school consumes in a week. A complete roster of the names and genders for the 2,000 students at their school was available. Peter selected a simple random sample of 100 students. Rania, knowing that 60 percent of the students at the school are female, selected a simple random sample of 60 females and an independent simple random sample of 40 males. Both asked all of the students in their samples how many soft drinks they typically consume in a week.

(a) Describe a method Peter could have used to select a simple random sample of 100 students from the school.

For Peter to select a simple random sample of 100 students from the school he could get a list of all 2000 students that attend the school. Number the students from 0000 to 1999. use a random integers table, starting from the first line, move across the table four digits at a time. Record the number skip any number over 1999, and ignore repeats. once the first 100 numbers are selected the students corresponding to those numbers will be the 100 students of the sample.

(b) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution (sometimes called the standard error) of Peter's point estimator $\overline{X}$.

$$\frac{4.13}{\sqrt{100}} = .413$$

(c) Based on the summary statistics, calculate the estimated standard deviation of the sampling distribution of Rania's point estimator $\overline{X}_{\text{overall}} = (0.6)\overline{X}_{\text{female}} + (0.4)\overline{X}_{\text{male}}$.

$$\sqrt{(.6)^2\left(\frac{1.8}{\sqrt{60}}\right)^2 + (.4)^2\left(\frac{2.22}{\sqrt{40}}\right)^2} = .198$$

(d) Using the dotplots above, explain why Rania's point estimator has a smaller estimated standard deviation than the estimated standard deviation of Peter's point estimator.

Rania's point estimator has a smaller estimated standard deviation than the estimated standard deviation of Peter's point estimator because Rania used two independent random samples. each random sample had a different standard deviation and sample size, both which are smaller than peter's sample size and standard deviation. Because of the rule for variances needed to calculate Rania's overall sample standard deviation, the small values would result in small estimated standard deviation.

2012--Sample: 6A Score: 4
In part (a) the 2,000 students in the high school are assigned numbers from 0000 to 1,999. A random selection of 100 unique numbers from a random integers table is clearly described by ignoring repeated four-digit numbers and by ignoring the four-digit numbers greater than 1,999. The 100 unique random numbers are used to select the 100 students with the corresponding numbers. Thus, part (a) was scored as essentially correct.

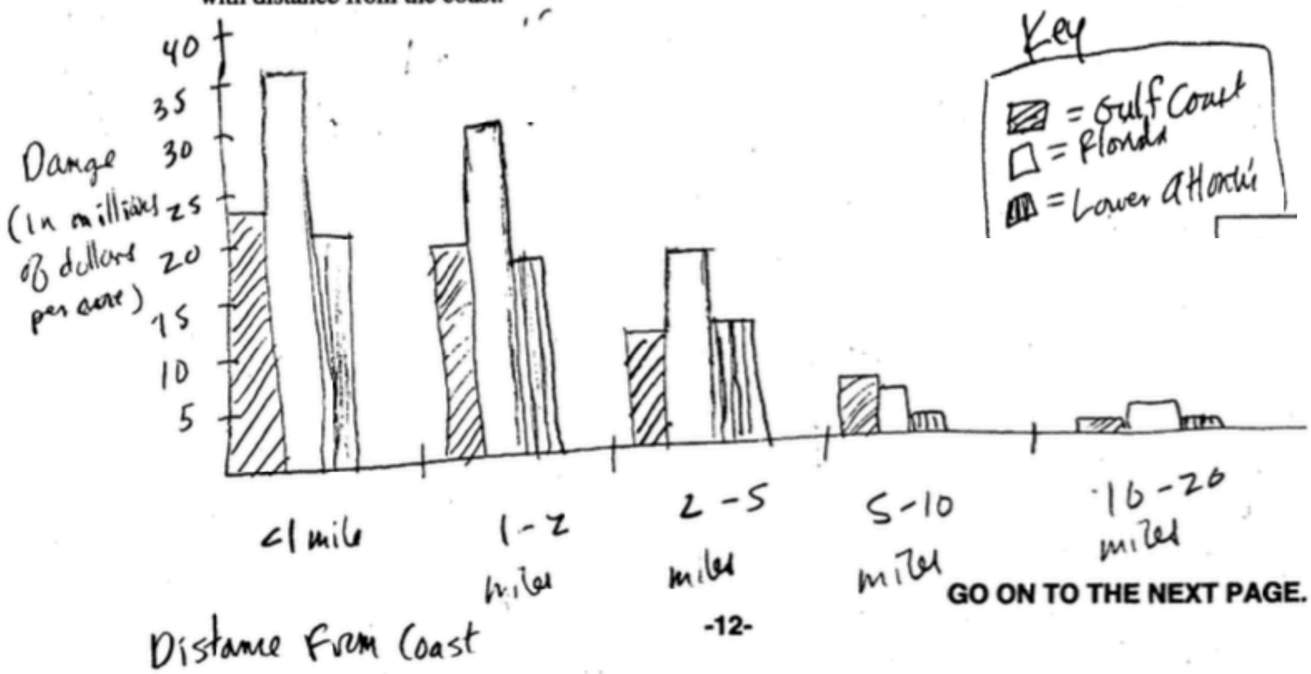
In part (b) the formula for the standard deviation of the sample mean is correctly identified by the appropriate statistics from the sample, and these statistics are used to calculate the correct standard error of Peter's point estimator. Thus, part (b) was scored as essentially correct.

The variances, weights, and sample sizes for female responses and male responses in Rania's samples are correctly combined in part (c) to produce the correct estimated standard deviation of the stratified sample mean. Thus, part (c) was scored as essentially correct.

In part (d) the standard deviations for each gender are identified as being smaller in Rania's sample data than the standard deviation in Peter's sample data. The two smaller variances for each gender are explicitly linked to the smaller sample standard deviation for Rania's point estimator with the phrase "the rule for variances needed to calculate Rania's." Hence, part (d) was scored as essentially correct. With all four parts scored as essentially correct, the response earned a score of 4.

2010:

(a) Sketch a graphical display that compares the hurricane damage amounts per acre for the three different coastal regions (Gulf Coast, Florida, and Lower Atlantic) and that also shows how the damage amounts vary with distance from the coast.



GO ON TO THE NEXT PAGE.

-12-

(b) Describe differences and similarities in the hurricane damage amounts among the three regions.

Florida had the most hurricane damage in 4 of the 5 strata. (Gulf Coast had more damage 5-10 miles from the coast).

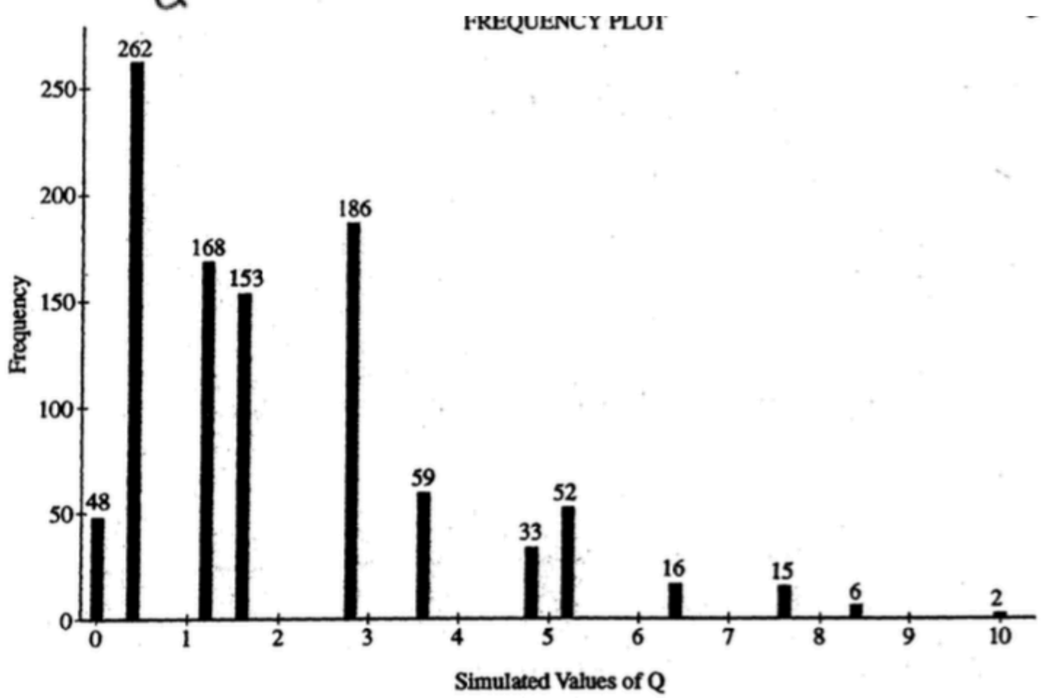
A similarity for all three regions, as expected, the damage decreased as we went farther from the coast, and it was greatest for all 3 regions when we went less than one mile from the coast.

ASSIGNED RANKS WITHIN DISTANCE CATEGORIES

	Distance from Coast					
	< 1 mile	1 to 2 miles	2 to 5 miles	5 to 10 miles	10 to 20 miles	Average Rank
Gulf Coast	2	2	3	1	2	2
Florida	1	1	1	2	1	1.2
Lower Atlantic	3	3	2	3	3	2.8

Calculate the value of the test statistic Q using the average ranks you obtained in part (c).

$$5[(2-2)^2+(1.2-2)^2+(2.8-2)^2]$$
$$5[0+.64+.64]$$
$$Q=6.4$$



Use these simulated values and the test statistic you calculated in part (d) to determine if the observed data provide evidence of a significant difference in the distributions of hurricane damage amounts among the three coastal regions. Explain.

The Q statistic for the observed data was 6.4.
a Q-value of 6.4 or greater only occurred 39 times out of the 1000 simulations.
Our P-value would be $\frac{39}{1000}=.039$.
at the significance level of $\alpha=.05$, we reject the null hypothesis with a P-value of .039. Based on this data, there is evidence to suggest that there is a difference in the distribution of hurricane damage in these 3 regions.

In part (a) the student draws a well-labeled bar graph in which the bars are grouped by the distance categories, and a key is provided for ease in identifying the three regions. Section 1, consisting of part (a), was scored as essentially correct. In part (b) Florida's having "the most hurricane damage in 4 of the 5 strata" is correctly identified as a significant difference. Although it was not necessary to describe the categorical distances as strata (as provided in the stem of the question), this is a nice detail. Also, the fact that damage decreases as the distance from the coast increases is correctly identified as a major similarity among the regions. Section 2, consisting of part (b), was scored as essentially correct. In parts (c) and (d) the ranks, average ranks and test statistic, Q, are correctly calculated, so section 3, consisting of parts (c) and (d), was scored as essentially correct. In part (e) a simulated p-value of 0.039 is correctly identified for the test statistic value of $Q = 6.4$ and is used to provide an appropriate conclusion in context. Section 4, consisting of part (e), was scored as essentially correct. With all four sections essentially correct, the response earned a score of 4.

2009:

(a) Define the parameter of interest and state the null and alternative hypotheses the consumer organization is interested in testing.

μ = ~~[illegible]~~ true population mean mpg for this model

$H_0: \mu = 27$

$H_a: \mu < 27$

(b) One possible statistic that measures skewness is the ratio $\frac{\text{sample mean}}{\text{sample median}}$. What values of that statistic (small, large, close to one) might indicate that the population distribution of mpg values is skewed to the right? Explain.

If the statistic is large, then the population distribution of mpg values is skewed to the right (assuming that the sample is a simple random sample). Since the mean is more sensitive/less resistant to extreme values than the median, it will be higher than the median when the skew is to the right, and thus the ratio would be large.

In the original sample, the value of the statistic $\frac{\text{sample mean}}{\text{sample median}}$ was 1.03. Based on the value of 1.03 and the dotplot above, is it plausible that the original sample of 10 cars came from a normal population, or do the simulated results suggest the original population is really skewed to the right? Explain.

The simulated results fail to ~~suggest~~ suggest that the original ~~pop~~ sample came from a skewed population. There are 14 points out of 100 that have a statistic $\frac{\text{sample mean}}{\text{sample median}}$ greater than 1.03, which means that a value of 1.03 or greater could occur around 14% of the time by chance in a normally distributed population. Generally, a ~~value~~ probability of less than 10% is needed to reject a null hypothesis, so this simulation does not provide enough evidence to say that it isn't plausible that the original sample came from a normal population.

(d) The table below shows summary statistics for mpg measurements for the original sample of 10 cars.

Minimum	Q1	Median	Q3	Maximum
23	24	25.5	28	32

Choosing only from the summary statistics in the table, define a formula for a different statistic that measures skewness.

$$\frac{Q_3 - \text{Median}}{\text{Median} - Q_1}$$

What values of that statistic might indicate that the distribution is skewed to the right? Explain.

Values substantially greater than 1 would indicate a distribution skewed to the right, since it would indicate that the distance between Q_3 and the median is much greater than the distance between the median and Q_1, and thus that the right tail is longer than the left, which indicates skewness.

In part (a) the student correctly defines the parameter of interest, including the concepts of mean and population in context, and correctly states the null and alternative hypotheses using standard notation.
Part (a) was scored as essentially correct. In part (b) the student correctly states that large values of the statistic indicate that the distribution is skewed to the right and gives a justification that describes the relationship between the mean and median when the distribution is skewed right. **Part (b) was scored as essentially correct**. In part (c) the student correctly states that there is not convincing evidence that the original sample came from a skewed population and gives specific numerical evidence from the dotplot ("14 points out of 100"). Although not necessary, the student goes on to correctly explain when there would be convincing evidence that the distribution was skewed right. **Part (c) was scored as essentially correct.** In part (d) the student provides a reasonable statistic to measure skewness. The student then correctly identifies the values of the statistic that indicate right-skewness and justifies the response by discussing how the components of the statistic are affected by right-skewness. **Part (d) was scored as essentially correct.** The entire answer, based on all four parts, was judged a complete response and earned a score of 4 points.