28. Two independent random samples of size $n_1 = 45$ and $n_2 = 38$, with respective standard deviations $\sigma_1 = 2.3$ and $\sigma_2 = 1.8$, are drawn from two normally distributed populations. Which of the following represents an estimate of the standard deviation of the sampling distribution corresponding to $\bar{x}_1 - \bar{x}_2$?

(A) 0.25  
(B) 0.31  
(C) 0.45  
(D) 0.50  
(E) 2.05

Answer: D

29. SRS refers to a simple random sample. Suppose an SRS of size $n$ is drawn from a population. Which of the following accurately characterizes this sample?

I. It is a sample drawn in such a manner so that every set of $n$ subjects in the population has an equal chance to be the sample chosen.

II. It is a sample drawn in such a manner so that every subject has some chance of being chosen.

III. It is a sample drawn in such a manner that every subject in the population has an equal chance of being selected.

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

Answer: E
30. The points \((x, y)\) on a scatterplot form an ellipse. As the ellipse becomes thinner, what can be concluded about the correlation \(r\) between the variables \(x\) and \(y\)?

(A) \(r\) decreases in absolute value  
(B) \(r\) increases in absolute value  
(C) \(r\) remains constant  
(D) the value of \(r\) changes sign  
(E) no conclusion can be drawn

Answer

31. The histogram below displays a set of measurements. Which of the boxplots below displays the same set of measurements?

![Histogram and Boxplots]

(A)  
(B)  
(C)  
(D)  
(E)

Answer
32. The correlation $r$ between the magnitude of an earthquake and the depth below the surface of the earth at which the quake occurs has been determined experimentally to be about 0.51. Suppose we use the magnitude of the earthquake ($x$) to predict the depth below the surface ($y$) at which the quake occurs. We can infer that

(A) the least squares regression line of $y$ on $x$ has slope equal to 0.51

(B) the fraction of the variation in depths explained by the least squares regression line of $y$ on $x$ is 0.26

(C) about 51% of the time, the magnitude of an earthquake will accurately predict the depth at which the earthquake occurs

(D) the numerical value of the depth is usually 51% of the numerical value of the earthquake

(E) twenty-six percent of the data values lie on the least squares regression line

Answer

33. A report from the Maine Department of Inland Fisheries and Wildlife indicates that there occurs on average one fatality per 100 collisions between cars and deer. In 300 collisions between a car and a deer, what is the expected number of fatalities and the standard deviation?

(A) mean = 0.33 standard deviation = 0.01

(B) mean = 1 standard deviation = 0.01

(C) mean = 3 standard deviation = 1.72

(D) mean = 3 standard deviation = 2.97

(E) mean = 30 standard deviation = 3.0

Answer
34. Johannes Kepler (1571–1630) was able to show that the square of the period of revolution of a planet is proportional to the cube of its mean distance from the sun. The equation $y = 0.001118x^{1.5}$, where $x$ is the distance of a planet from the sun (in millions of miles) and $y$ is the period of revolution (in years) of a planet around the sun, is a power regression model which represents this relationship. Which of the following will produce a linear representation of this model?

(A) plotting $y$ versus $\ln x$
(B) plotting $x$ versus $\ln x$
(C) plotting $x$ versus $\ln y$
(D) plotting $\ln x$ versus $\ln y$
(E) constructing the residual plot for the scatterplot of $y$ vs. $x$

Answer: D

35. Which of the following is an appropriate graph to display univariate categorical data?

(A) stemplot
(B) histogram
(C) boxplot
(D) pie chart
(E) scatterplot

Answer: D

36. A large university is considering introducing a new major in Economic Geography and wishes to poll the current student body for their opinion of the feasibility of introducing such a major. The Office of Public Relations mails a questionnaire on this issue to a SRS of 2000 students currently enrolled in the university. Of the 2000 questionnaires mailed, 532 have been returned of which 219 students support the new major. Which of the following represents the population for this study?

(A) the 2000 students receiving the questionnaire
(B) the 532 students who responded
(C) the 219 students who support the new major
(D) the 2000 students selected represent a sample of the population of all currently enrolled students
(E) all students who are currently enrolled and all past alumni of the university

Answer: D
37. A Physicians’ Study Group wishes to design a matched pairs test to determine the effects of a medication on reducing the elevated cholesterol levels in 200 adults in the 35–65 age group. Which of the following represents such a design?

(A) The Physicians’ Group compares cholesterol levels of each patient before he/she begins a daily regimen of the medicine to the patient’s cholesterol level after six weeks of using the medication.

(B) The Physicians’ Group compares the cholesterol levels of the patients using the medicine for six weeks with the cholesterol levels of 200 adults in the same age group who exhibited elevated cholesterol levels and did not take the medicine.

(C) The Physicians’ Group compared the mean cholesterol level of the 200 patients six weeks after using the medicine with a published national mean cholesterol level for all adults in this age group.

(D) The Physician’s group compares cholesterol levels of the 200 patients using the medicine for six weeks with the cholesterol levels of 200 adults in the same age group who were taking a placebo.

(E) All of the above are representative of a matched pairs design.

Answer: D

38. Let $X$ represent the number of independent rolls of a fair die required to obtain the first "2". What is $P(X = 5)$?

(A) $(1/6)^4 (5/6)$

(B) $(1/6)^4$

(C) $(5/6)^4 (1/6)$

(D) $(5/6)^5$

(E) $(5.6)^5$
39. A local university has a "drop-in center" located near the Student Union where students can "drop-in" to speak to a counselor on any issue or concern whether personal or academic. A researcher stations herself at a receptionist's desk during a few random evenings to collect data on whether men or women are more likely to use this service and whether a particular class (freshman, sophomore, junior, senior) is more likely to "drop-in." This study may be described as a

(A) controlled experiment
(B) matched pairs design
(C) census
(D) observational study
(E) a double blind completely randomized experiment

40. A researcher wishes to use a 95% confidence interval to estimate the proportion of Americans who have visited an entertainment theme park near Orlando within the last five years. The researcher wishes to choose a size that will insure a margin of error not to exceed 0.05. Which of the following is the smallest size that meets these criteria?

(A) 40
(B) 200
(C) 400
(D) 600
(E) 800
Section II

Part A

Time—65 minutes
Questions 1–5
Percent of Section II grade—75

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your method as well as on the accuracy of your results and explanations.

1. The table below shows the home run output of all American League and National League baseball teams during the 2000 season.

<table>
<thead>
<tr>
<th>American League</th>
<th>Home Runs</th>
<th>National League</th>
<th>Home Runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleveland</td>
<td>221</td>
<td>Colorado</td>
<td>161</td>
</tr>
<tr>
<td>Kansas City</td>
<td>150</td>
<td>San Francisco</td>
<td>226</td>
</tr>
<tr>
<td>Chicago White Sox</td>
<td>216</td>
<td>Houston</td>
<td>249</td>
</tr>
<tr>
<td>Texas</td>
<td>173</td>
<td>Cincinnati</td>
<td>200</td>
</tr>
<tr>
<td>Anaheim</td>
<td>236</td>
<td>Atlanta</td>
<td>179</td>
</tr>
<tr>
<td>New York Yankees</td>
<td>205</td>
<td>St. Louis</td>
<td>235</td>
</tr>
<tr>
<td>Detroit</td>
<td>177</td>
<td>Pittsburgh</td>
<td>168</td>
</tr>
<tr>
<td>Toronto</td>
<td>244</td>
<td>Montreal</td>
<td>178</td>
</tr>
<tr>
<td>Baltimore</td>
<td>184</td>
<td>Arizona</td>
<td>179</td>
</tr>
<tr>
<td>Minnesota</td>
<td>116</td>
<td>New York Mets</td>
<td>198</td>
</tr>
<tr>
<td>Oakland</td>
<td>239</td>
<td>Florida</td>
<td>160</td>
</tr>
<tr>
<td>Seattle</td>
<td>198</td>
<td>Los Angeles</td>
<td>211</td>
</tr>
<tr>
<td>Boston</td>
<td>167</td>
<td>Chicago Cubs</td>
<td>183</td>
</tr>
<tr>
<td>Tampa Bay</td>
<td>162</td>
<td>San Diego</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Philadelphia</td>
<td>144</td>
</tr>
</tbody>
</table>

(a) In the area provided below, construct one histogram displaying the distribution of homerun output for all teams (both American and National Leagues). Describe the nature of the distribution noting the center of the distribution, skewness, and outliers (if any).
(b) Draw box plots for the data in each of the two leagues and identify the five-number summary for each plot. What comparisons can you make between the homerun outputs of the two leagues?

(c) Construct a back-to-back stemplot displaying the homerun outputs of teams in the two leagues.

2. A telecommunications company uses three different systems to produce the print shaft holder for its most popular fax machine. System A produces 50% of the shaft holders, System B produces 30% of the shaft holders, and System C produces 20% of the shaft holders. The percentages of defective shaft holders produced by each system are respectively, 3%, 4% and 5%.

(a) If a shaft holder is selected at random, what is the probability that it is defective?

(b) If a shaft holder is selected at random and found to be defective, what is the probability that it was produced by System A?
3. A Television Monitoring Group (TMG) claims that at least three out of every four Americans believe that reality-based television shows are untruthful. They tested their claim by using data from a recent Gallup poll which indicated that in a sample of 1,015 Americans, 812 thought the shows were either somewhat or totally untruthful.

(a) Test the claim of the TMG using a significance level of 0.01. State clearly your assumptions and what test of hypothesis you are using. Identify an appropriate p-value and interpret your results within the context of the problem.

(b) Construct a 95% confidence interval to estimate the true proportion of Americans who think these shows are essentially untruthful.

4. A survey was conducted recently in ten large American cities to determine whether there is any relationship between average weekday hotel rates and average car rental rates. The following data was collected.

<table>
<thead>
<tr>
<th>Daily Hotel Rate (x) (in dollars)</th>
<th>Daily Car Rental Rate (y) (in dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>149</td>
<td>49</td>
</tr>
<tr>
<td>187</td>
<td>50</td>
</tr>
<tr>
<td>171</td>
<td>52</td>
</tr>
<tr>
<td>122</td>
<td>49</td>
</tr>
<tr>
<td>115</td>
<td>39</td>
</tr>
<tr>
<td>147</td>
<td>44</td>
</tr>
<tr>
<td>128</td>
<td>37</td>
</tr>
<tr>
<td>212</td>
<td>63</td>
</tr>
<tr>
<td>168</td>
<td>46</td>
</tr>
<tr>
<td>181</td>
<td>51</td>
</tr>
</tbody>
</table>

(a) Construct a scatterplot for this data.

(b) Use the scatterplot to determine if there is a linear relationship between the two variables.

(c) If there is a linear relationship, numerically describe the strength of this relationship and construct a least squares regression model.

(d) Find the residual associated with the point (168, 46).
5. A medical research group is interested in comparing the effectiveness of two new medications in reducing cholesterol levels in adult males ranging in age from 35–65 years of age who have exhibited elevated cholesterol levels for at least six months. Four hundred men with elevated cholesterol levels in this age group were randomly selected in New York City and 340 men were randomly selected in Philadelphia. It is suspected that results will differ greatly between the two cities due to a recent ‘heart health’ initiative in Philadelphia. Describe a randomized block design to conduct this experiment.
Section II

Part B  
Time—25 minutes

Question 6

Percent of Section II grade—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your method as well as on the accuracy of your results and explanations.

6. In 1908 W. S. Gosset wrote a paper titled, “The Probable Error of a Mean”. In the paper Gossett reported on the corn yield using two different kinds of seeds. The first seed was the usual seed used by farmers in England at that time. The second seed type was kilndried. Each type of seed was planted in adjacent plots, accounting for 11 pairs of “split” plots. The data are given below. The numbers are in pounds per acre.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Regular</th>
<th>Kiln-Dried</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1903</td>
<td>2009</td>
</tr>
<tr>
<td>2</td>
<td>1935</td>
<td>1915</td>
</tr>
<tr>
<td>3</td>
<td>1910</td>
<td>2011</td>
</tr>
<tr>
<td>4</td>
<td>2496</td>
<td>2463</td>
</tr>
<tr>
<td>5</td>
<td>2108</td>
<td>2180</td>
</tr>
<tr>
<td>6</td>
<td>1961</td>
<td>1925</td>
</tr>
<tr>
<td>7</td>
<td>2060</td>
<td>2122</td>
</tr>
<tr>
<td>8</td>
<td>1444</td>
<td>1482</td>
</tr>
<tr>
<td>9</td>
<td>1612</td>
<td>1542</td>
</tr>
<tr>
<td>10</td>
<td>1316</td>
<td>1443</td>
</tr>
<tr>
<td>11</td>
<td>1511</td>
<td>1535</td>
</tr>
</tbody>
</table>

The following computer outputs show the descriptive statistics of the two variables and a regression output which resulted from fitting a least squares regression line using yield from “regular” seed to predict yield from “kiln-dried” seed.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>11</td>
<td>1841</td>
<td>1910</td>
<td>1827</td>
<td>343</td>
<td>103</td>
</tr>
<tr>
<td>Kiln-Dri</td>
<td>11</td>
<td>1875</td>
<td>1925</td>
<td>1858</td>
<td>333</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>1316</td>
<td>2496</td>
<td>1511</td>
<td>2060</td>
</tr>
<tr>
<td>Kiln-Dri</td>
<td>1443</td>
<td>2463</td>
<td>1535</td>
<td>2122</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>120.4</td>
<td>116.7</td>
<td>1.03</td>
<td>0.329</td>
</tr>
<tr>
<td>Regular</td>
<td>0.95293</td>
<td>0.06241</td>
<td>15.27</td>
<td>0.000</td>
</tr>
</tbody>
</table>

S=67.65  R-Sq=96.3%  R-Sq(adj) =95.9%
Below is a residual plot and a normal quantile plot of the residuals resulting from the regression.

(a) Use side-by-side box plots to compare the corn yield for the two groups. Write a few sentences commenting on your display.

(b) Can yield of "regular seed" be used to predict the yield of "kiln dried seed"? Give statistical justification to support your response.

(c) Is there a difference in yield between the two types of seeds? Give statistical justification to support your response.