Geraldo Braho  
MATH 1312 Statistics  
Prof. Suat Namli  
Homework #2

**Chapter #3**

**Section 1 and 2**

53. **Food Choice.** As you discovered earlier, *ordinal data* are data about order or rank given on a scale such as 1,2,3,... or A, B, C, . . . . Most statisticians recommend using the median to indicate the center of an ordinal data set, but some researchers also use the mean. In the paper “Measurement of Ethical Food Choice Motives” (*Appetite*, Vol. 34, pp. 55–59), research psychologists M. Lindeman and M. Va ̈a ̈na ̈nen of the University of Helsinki published a study on the factors that most influence people’s choice of food. One of the questions asked of the participants was how important, on a scale of 1 to 4 (1 = not at all important, 4 = very important), is ecological welfare in food choice motive, where ecological welfare includes animal welfare and environmental protection. Here are the ratings given by 14 of the participants.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 1.21.58 PM.png**a.** Compute the mean of the data  
**b.** Compute the median of the data  
**c.** Decide which of the two measures of center is best

a) x= (2+4+1+2+4+3+3+2+2+1+2+4+2+3)/14 = 35/14 = 5/2

b) median (middle one) 🡪 DS in increasing order   
1,1,2,2,2,2,2,2,3,3,3,4,4,4 🡪 (7th + 8th)/2 = (2+2)/2 = 2 (median)  
c)

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 1.37.39 PM.png54. **Outliers and Trimmed Means.** Some data sets contain *outliers,* observations that fall well outside the overall pattern of the data. (We discuss outliers in more detail in Section 3.3.) Sup- pose, for instance, that you are interested in the ability of high school algebra students to compute square roots. You decide to give a square-root exam to 10 of these students. Unfortunately, one of the students had a fight with his girlfriend and cannot concentrate—he gets a 0. The 10 scores are displayed in increasing order in the following table. The score of 0 is an outlier.

Statisticians have a systematic method for avoiding extreme observations and outliers when they calculate means. They compute *trimmed means,* in which high and low observations are deleted or “trimmed off” before the mean is calculated. For in- stance, to compute the 10% trimmed mean of the test-score data, we first delete both the bottom 10% and the top 10% of the ordered data, that is, 0 and 80. Then we calculate the mean of the remaining data. Thus the 10% trimmed mean of the test-score data is

(58+61+63+67+69+70+71+78) /8 = 67.1

The following table displays a set of scores for a 40-question algebra final exam.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 1.40.32 PM.png

a. Do any of the scores look like outliers?  
b. Compute the usual mean of the data.  
c. Compute the 5% trimmed mean of the data.  
d. Compute the 10% trimmed mean of the data.  
e. Compare the means you obtained in parts (b)–(d). Which of the three means provides the best measure of center for the data?

a) Yes 2 and 4 look like outlines  
b) x = (2+4+15+15+16+16+16+17+10+20+21+21+21+24+25+25+26+27+27+28)/20 = 385/20  
c) 5% trimmed mean (omit one lowest and one highest)  
(4+15+15+16+16+16+17+10+20+21+21+21+24+25+25+26+27+27)/18 = 355/18   
d) 10% trimmed mean (omit two lowest and one highest)  
(15+15+16+16+16+17+10+20+21+21+21+24+25+25+26) /16 = 324/16   
e) 10% trimmed mean provides the best measure of the center for the data

57. Explain the purpose of a measure of variation.   
Measure of variation is used to determine an amount how much is data spread.

58. Why is the standard deviation preferable to the range as a measure of variation?  
Because standard deviation gives us an idea how much every element of our population varies from the mean (average).

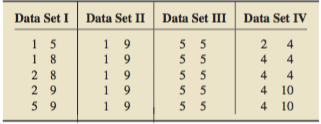
61. Consider the data set: 1,2,3,4,5,6,7,8,9.

a. Use the defining formula to obtain the sample standard deviation.  
b. Replace the 9 in the data set by 99, and again use the defining formula to compute the sample standard deviation.  
c. Compare your answers in parts (a) and (b). The lack of what property of the standard deviation accounts for its extreme sensitivity to the change of 9 to 99?

a) SD =   
x = (1+2+3+4+5+6+7+8+9)/9 = 45/9 = 5  
SD = =

b) x = (1+2+3+4+5+6+7+8+99)/9 = 15  
SD = =

c) Keeping 99 in the second case, is actually keeping an outliner in our data, and that is way we see the extreme change between the SD in part a) and in part b); that is why it is recommended to trim our data before analyzing it.

62. Consider the following four data sets.

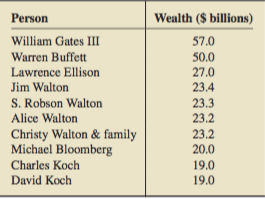
a. Compute the mean of each data set  
b. Although the four data sets have the same means, in what respect are they quit different  
c. which data set appears to have the least variation? The greatest variation?  
d. Compute the range of each data set

a) DS I; x = (1+1+2+2+5+5+8+8+9+9)/10 = 50/10 = 5  
 DS II; using the same method as we used in DS I = 5  
 DS III; -||- = 5  
 DS IV; -||- = 5

b) Data distribution and data spread is different, (median and mode)

c) Data Set I, since it has 4 different modes.

d) DS I; Range = max –min = 9-1= 8  
 DS II; Range = max –min = 9-1= 8  
 DS III; Range = max-min = 5-5 = 0  
 DS IV; Range = max-min = 10-2 = 8

In Exercises 3.71–3.78, determine the range and sample standard deviation for each of the data sets. For the sample standard deviation, round each answer to one more decimal place than that used for the observations.

75. Billionaires’ Club. Each year, Forbes magazine compiles a list of the 400 richest Americans. As of September 17, 2008, the top 10 on the list are as shown in the following table.

Range = max-min = 57.0 – 19.0 = 38  
Sample Standard Deviation will be calculated using the same formula from previous questions. First we calculate the sample mean (x) and then apply the formula used to calculate sample SD:

SD = (rounding to two decimal points)

73. Tornado Touchdowns. Each year, tornadoes that touch down are recorded by the Storm Prediction Center and published in Monthly Tornado Statistics. The following table gives the number of tornadoes that touched down in the United States during each month of one year. [SOURCE: National Oceanic and Atmospheric Administration.]

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 3.35.37 PM.png

Range = max-min = 204 -2 = 202  
Sample Standard deviation will be calculated using the same method as in previous question.

77. Fuel Economy. Every year, Consumer Reports publishes a magazine titled New Car Ratings and Review that looks at vehicle profiles for the year’s models. It lets you see in one place how, within each category, the vehicles compare. One category of interest, especially when fuel prices are rising, is fuel economy, measured in miles per gallon (mpg). Following is a list of overall mpg for 14 different full-sized and compact pickups.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 3.38.06 PM.png

Range = max –min = 17-11 = 6  
Sample Standard deviation – same calculation as explained in the previous questions related to the standard deviation.

78. Router Horsepower. In the article “Router Roundup” (Popular Mechanics, Vol. 180, No. 12, pp. 104–109), T. Klenck reported on tests of seven fixed-base routers for performance, features, and handling. The following table gives the horsepower for each of the seven routers tested.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 3.40.16 PM.png Range = max –min = 2.25 – 1.75 – 0.5  
Sample Standard deviation – same calculation (just rounding to three decimal points)

79. Medieval Cremation Burials. In the article “Material Culture as Memory: Combs and Cremations in Early Medieval Britain” (Early Medieval Europe, Vol. 12, Issue 2, pp. 89–128), H. Williams discussed the frequency of cremation burials found in 17 archaeological sites in eastern England. Here are the data.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 3.44.28 PM.png

1. Obtain the sample standard deviation of these data.
2. Do you think that, in this case, the sample standard deviation provides a good measure of variation? Explain your answer.
3. Sample standard deviation will be calculated using previously used formula after computing the sample mean of the given data set:

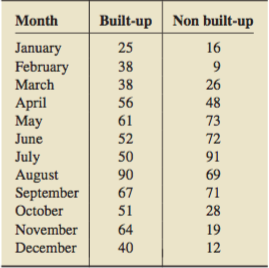
SD =

1. Not within this data set, because outlines weren’t trimmed before processing the data.

80. Monthly Motorcycle Casualties. The Scottish Executive, Analytical Services Division Transport Statistics, compiles data on motorcycle casualties. During one year, monthly casualties resulting from motorcycle accidents in Scotland for built-up roads and non–built-up roads were as follows.

a. Without doing any calculations, make an educated guess at which of the two data sets, built-up or non-built up, has the greater variation.

b. Find the range and sample standard deviation of each of the two data sets. Compare your results here to the educated guess that you made in part (a).

a)  According to my observation of data, I think that non build-up data has the greater variation.

b) Range for built-up = max – min = 90-25 = 65   
 Range got non built-up = max – min = 91-9 = 82

Sample Standard Deviation for built-up data:   
SD =

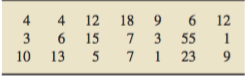
Sample Standard Deviation for non built-up data:  
SD =   
Result I got, goes along with my educational guess I made, that non built-up data has the greater deviation.

**Section 3 and 4**

In Exercises 3.121–3.128,

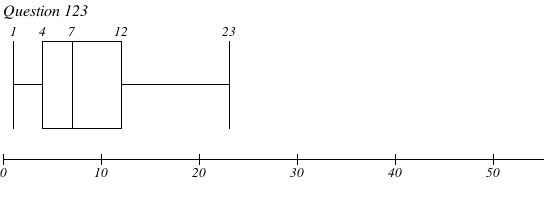
a. Obtain and interpret the quartiles.  
b. Determine and interpret the interquartile range.   
c. Find and interpret the five-number summary.  
d. Identify potential outliers, if any.  
e. Construct and interpret a boxplot.

123. Hospital Stays. The U.S. National Center for Health Statistics compiles data on the length of stay by patients in short- term hospitals and publishes its findings in Vital and Health Statistics. A random sample of 21 patients yielded the following data on length of stay, in days.

1. In order to interpret the quartiles, we have to sort data in increasing order:  
   1,1,3,3,4,4,5,6,6,7,7,9,9,10,12,12,13,15,18,23,55  
   Q2 = (21+1)st / 2 = 11th number is Q2 = 7

Q1 = {1,1,3,3,4,4,5,6,6,7,7} = (11+1)st / 2 = 6th number is Q1 = 4  
Q3 = {7,7,9,9,10,12,12,13,15,18,23,55} = (6+7)th /2 = Q3 = 12

b) IQR = Q3-Q1 = 12 – 4 = 8  
c) min = 1  
 max=55  
 Q1 = 4  
 Q2 = 7  
 Q3 = 12  
d) Outliers:   
upper limit: Q3 + 1.5x(Q3-Q1) = 12 + 1.5x8 = 12 + 12 = 24 Outliers {55}

e) 

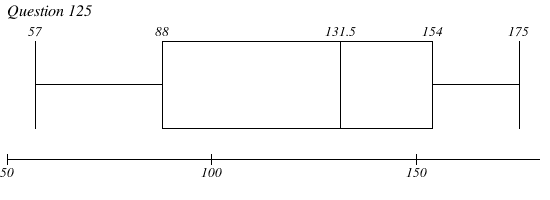
125. Hurricanes. An article by D. Schaefer et al. (Journal of Tropical Ecology, Vol. 16, pp. 189–207) reported on a long- term study of the effects of hurricanes on tropical streams of the Luquillo Experimental Forest in Puerto Rico. The study shows that Hurricane Hugo had a significant impact on stream water chemistry. The following table shows a sample of 10 ammonia fluxes in the first year after Hugo. Data are in kilograms per hectare per year.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 10.55.39 PM.png

a) 57,66,88,96,116,147,147,154,154,175  
Q2 = (5+6)th / 2 = (116+147) / 2 = 131.5  
Q1 = {57,66,88,96,116} = (5+1)st / 2 = 3rd number = 88  
Q3 = {147,147,154,154,175} = (5+1)st / 2 = 3rd number = 154

b) IQR = Q3 – Q1 = 154 – 88 = 66  
c) min = 57  
 max = 175  
 Q1 = 88  
 Q2 = 131.5  
 Q3 = 154

d) Outliers:   
Lower: Q1- 1.5x IQR = 88 – 1.5 x 66 = -11 / no lower outlier  
Upper: Q3 + 1.5x IQR = 154 + 1.5x 66 = 253 / no upper outlier

e)   


126. Sky Guide. The publication California Wild: Natural Sciences for Thinking Animals has a monthly feature called the “Sky Guide” that keeps track of the sunrise and sunset for the first day of each month in San Francisco. Over several issues, B. Quock from the Morrison Planetarium recorded the following sunrise times from July 1 of one year through June 1 of the next year. The times are given in minutes past midnight.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 11.14.24 PM.png

a) b) c) d) e) answers would be found using exactly the same principle of calculation as in the previous two questions.

In Exercises 3.153–3.158, we have provided simple data sets for you to practice the basics of finding a

a) Population mean  
b) Population standard deviation

157. Data Set: 1,9,8,4,3   
 = (1+9+8+4+3)/ 5 = 5 🡪 a) population mean

b) Population Standard Deviation:

SD(p) = =

158. Data Set: 4,2,0,2,2  
a) and b) would be calculated using the same calculation methods as in the previous question.

160. Back to Pinehurst. In the June 2005 issue of Golf Digest is a preview of the 2005 U.S. Open, titled “Back to Pinehurst.” Included is information on the course, Pinehurst in North Carolina. The following table lists the lengths, in yards, of the 18 holes at Pinehurst.

medina:Users:Medina:Desktop:Screen Shot 2016-02-03 at 11.41.19 PM.png

a. Obtain and interpret the population mean of the whole lengths at Pinehurst.   
b. Obtain and interpret the population standard deviation of the whole lengths at Pinehurst.

a) population mean and b) population standard deviation would be calculated using the same calculating principle from the previous two questions.

165. Doing Time. According to Compendium of Federal Jus- tice Statistics, published by the Bureau of Justice Statistics, the mean time served to first release by Federal prisoners is 32.9 months. Assume the standard deviation of the times served is 17.9 months. Let x denote time served to first release by a Federal prisoner.

a. Find the standardized version of x.  
b. Find the mean and standard deviation of the standardized variable.  
c. Determine the z-scores for prison times served of 81.3 months and 20.8 months. Round your answers to two decimal places.  
d. Interpret your answers in part (c).  
e. Construct a graph similar to Fig.3.15 on page 134 that depicts your results from parts (b) and (c).

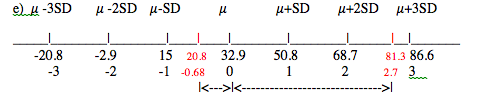
a) z = =

b) A standardized variable always has mean 0 and standard deviation 1

c) for x = 81.3   
z = = 2.70

for x = 20.8  
z =

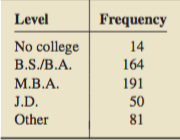
d) time for 81.3 months is 2.7 SD over the mean of 32.9 months  
 time for 20.8 months is 0.68 SD below the mean of 32.9 months



**Chapter 4**

**Section 1**

In Exercises 4.12–4.22, express your probability answers as a decimal rounded to three places.

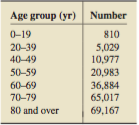
**12.** Educated CEOs. Reporter D. McGinn discussed the changing demographics for successful chief executive officers (CEOs) of America’s top companies in the article, “Fresh Ideas” (Newsweek, June 13, 2005, pp. 42–46). The following frequency distribution reports the highest education level achieved by Standard and Poor’s top 500 CEOs.

Find the probability that a randomly selected CEO from Standard and Poor’s top 500 achieved the educational level of

a. B.S./B.A. = 164/total = 164/500 = 0.328

b. either M.B.A. or J.D. = (191 + 50)/ 500 = 0.482

c. at least some college. = 1 – no college = 1 – 14/500 = 1- 0.028 = 0.972



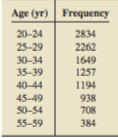
14. Cardiovascular Hospitalizations. From the Florida State Center for Health Statistics report Women and Cardiovascular Disease Hospitalization, we obtained the following table showing the number of female hospitalizations for cardiovascular dis- ease, by age group, during one year.

One of these case records is selected at random. Find the probability that the woman was

One of these case records is selected at random. Find the probability that the woman was

**a.** in her 50s.  
 **b.** less than 50 years old.   
**c.** between 40 and 69 years old, inclusive.   
**d.** 70 years old or older.

Total number = 208867  
a) P(50-59) = 20983/208867 = 0.100  
b) P(x<50) = (810+5029+10977)//208867 = 0.081  
c) P(40 <=x=>69) = (10977+20983+36884)/208867 = 0.330  
d) P(x=>70) = (65017+69167)/208867 = 0.642

16. Murder Victims. As reported by the Federal Bureau of Investigation in Crime in the United States, the age distribution of murder victims between 20 and 59 years old is as shown in the table.

A murder case in which the person murdered was between 20 and 59 years old is selected at random. Find the probability that the murder victim was

a. between 40 and 44 years old, inclusive.  
b. at least 25 years old, that is, 25 years old or older.   
c. between 45 and 59 years old, inclusive.  
d. under 30 or over 54.

Total = 11226  
a) P(40-44) = 1194/11226 = 0.106  
b) P(x=> 25) = 1-P(x<25) = 1-2834/11226 = 1 - 0.252 = 0.748  
c) P(45<=x=>59) = (938+708+384)/11226 = 0.181  
d) P(x<30 & x>54) = (2834+2262+384)/11226 = 0.488

21. Dice. Two balanced dice are rolled. Refer to Fig. 4.1 on page 147 and determine the probability that the sum of the dice is   
a. 6.   
b. even.  
c. 7 or 11.   
d. 2, 3, or 12

a) P(6)= 5/36  
b) P(even) = 18/36  
c) P(7 or 11) = 8/36  
d) P(2, 3, or 12) = 4/36

medina:Users:Medina:Desktop:Screen Shot 2016-02-04 at 11.39.53 AM.png22. Coin Tossing. A balanced dime is tossed three times. The possible outcomes can be represented as follows.

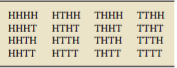
Here, for example, HHT means that the first two tosses come up heads and the third tails. Find the probability that

a. exactly two of the three tosses come up heads.  
b. the last two tosses come up tails.  
c. all three tosses come up the same.   
d. the second toss comes up heads.

a) P(exactly 2 heads) = 6/8  
b) P(at least 2 tails) = 4/8  
c) P(all 3 same tosses) = 2/8  
d) P(2nd toss head) = 4/8

**Section 2**

44. Coin Tossing. When a dime is tossed four times, there are the following 16 possible outcomes.



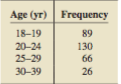
Here, for example, HTTH represents the outcome that the first toss is heads, the next two tosses are tails, and the fourth toss is heads. List the outcomes constituting each of the following four events.

A = event exactly two heads are tossed, (TTHH, HTHT, THHT, HTTH, THTH, HHTT)  
B = event the first two tosses are tails, (TTHH, TTHT, TTTH, TTTT)  
C = event the first toss is heads, (HHHH, HTHH, HHHT, HTHT, HHTH, HTTH, HHTT, HTTT)  
D = event all four tosses come up the same. (HHHH, TTTT)

48. Coin Tossing. Refer to Exercise 4.44. For each of the following events, list the outcomes that constitute the event, and describe the event in words.

a. (notB) (Every event where the first two tosses are not tails – HTHT, THHT, HTTH, THTH, HHTT, HHHH, HTHH, HHHT, HTHT, HHTH, HTTH, HHTT, HTTT, HHHH)  
b. (A&B) (Everything where we have two heads and where the first two tosses are tails – TTHH)  
c. (CorD) (All events where the first toss is head or ones with all four tosses same – HTHH, HHHT, HTHT, HHTH, HTTH, HHTT, HHHH, TTTT)

50. Family Planning. The following table provides a frequency distribution for the ages of adult women seeking pregnancy tests at public health facilities in Missouri during a 3-month period. It appeared in the article “Factors Affecting Contraceptive Use in Women Seeking Pregnancy Tests” (Family Planning Perspectives, Vol. 32, No. 3, pp. 124–131) by M. Sable et al.

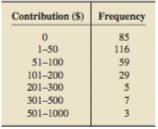
For one of these woman selected at random, let

A = event the woman is at least 25 years old,  
B = event the woman is at most 29 years old,  
C = event that the woman is between 18 and 29 years old, and  
D = event that the woman is at least 20 years old.

Describe the following events in words, and determine the number of outcomes (women) that constitute each event.

a. (not D) (Event where the woman is older than 20 years old)  
b. (B & D) (Woman between the age of 18 and 29 years old)  
c. (C or A) (Women that are 25 and up till 29 years old)  
d. (A & B) (Woman in age between or equal to 25-29 years old)

54. Protecting the Environment. A survey was conducted in Canada to ascertain public opinion about a major national park region in the Banff-Bow Valley. One question asked the amount that respondents would be willing to contribute per year to protect the environment in the Banff-Bow Valley region. The following frequency distribution was found in an article by J. Ritchie et al. titled “Public Reactions to Policy Recommendations from the Banff-Bow Valley Study” (Journal of Sustainable Tourism, Vol. 10, No. 4, pp. 295–308).



For a respondent selected at random, let

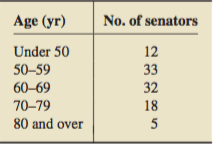
A = event that the respondent would be willing to contribute at least $101,  
B = event that the respondent would not be willing to contribute more than $50,  
C = event that the respondent would be willing to contribute between $1 and $200, and  
D= event that the respondent would be willing to contribute at least $1.

Describe the following events in words, and determine the number of outcomes (respondents) that make up each event.

a. (not D) (Everything except 0% contribution) (219)  
b. (A & B) (All respondents who are willing to contribute $101 and more)  
c. (C or A) (Respondents who are willing to contribute $101 and up to $200)  
d. (B & D) (Respondents who are willing to contribute $1 and up t0 $50)

**Section 3**

69. Ages of Senators. According to the Congressional Directory, the official directory of the U.S. Congress, prepared by the Joint Committee on Printing, the age distribution for senators in the 109th U.S. Congress is as follows.

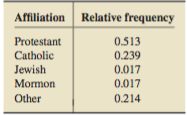


Suppose that a senator from the 109th U.S. Congress is selected at random. Let

A = event the senator is under 50,   
 B = event the senator is in his or her 50s,    
C = event the senator is in his or her 60s, and   
S = event the senator is under 70.

a. Use the table and the f/N rule to find P(S).  
b. Express event S in terms of events A, B, and C.  
c. Determine P(A), P(B), and P(C).  
d. Compute P(S), using the special addition rule and your answers from parts (b) and (c). Compare your answer with that in part (a).

a) P(S) = 77/100 = 0.77  
b)   
c) P(A) = 12/100 = 0.12  
 P(B) = 33/100 = 0.33  
 P(C) = 32/100 = 0.32  
d) P(S) = 77/100 = 0.77

74. Religion in America. According to the U.S. Religious Landscape Survey, sponsored by the Pew Forum on Religion and Public Life, a distribution of religious affiliation among U.S. adults is as shown in the following table.

Find the probability that the religious affiliation of a randomly selected U.S. adult is

a. Catholic or Protestant.  
b. not Jewish.  
c. not Catholic, Protestant, or Jewish.

a) P(C or P) = 0.752  
b) P(not J) = 0.983  
c) P(not C, P or J) = 0.231

75. Ages of Senators. Refer to Exercise 4.69. Use the complementation rule to find the probability that a randomly selected senator in the 109th Congress is

a. 50 years old or older. P(x=>50) = 1-P(x<50) = 1 – 12/100 = 1 – 0.12 = 0.88  
b. under 70 years old. P(x<70) = 1 – P(x=>70) = 1 – 23/100 = 1 – 0.23 = 0.77

79. Craps. In the game of craps, a player rolls two balanced dice. Thirty-six equally likely outcomes are possible, as shown in Fig. 4.1 on page 147. Let

A = event the sum of the dice is 7,   
B = event the sum of the dice is 11,  
C =event the sum of the dice is 2,  
D = event the sum of the dice is 3,  
E = event the sum of the dice is 12,  
F =event the sum of the dice is 8, and  
G = event doubles are rolled.

a. Compute the probability of each of the seven events.  
b. The player wins on the first roll if the sum of the dice is 7 or 11. Find the probability of that event by using the special addition rule and your answers from part (a).  
c. The player loses on the first roll if the sum of the dice is 2, 3, or 12. Determine the probability of that event by using the special addition rule and your answers from part (a).  
d. Compute the probability that either the sum of the dice is 8 or doubles are rolled, without using the general addition rule  
e. Compute the probability that either the sum of the dice is 8 or doubles are rolled by using the general addition rule, and compare your answer to the one you obtained in part (d).

a) P(A) = 6/36  
 P(B) = 2/36  
 P(C) = 1/36  
 P(D) = 2/36  
 P(E) = 1/36  
 P(F) = 5/36  
 P(G) = 6/36

b) P(7 or 11) = 8/36  
c) P(2 or 3 or 120 = 4/36  
d) P(8 or doubles) = 10/36  
e) P(8 or doubles) = 10/36

80. Gender and Divorce. According to Current Population Reports, published by the U.S. Census Bureau, 51.6% of U.S. adults are female, 10.4% of U.S. adults are divorced, and 6.0% of U.S. adults are divorced females. For a U.S. adult selected at random, let

F = event the person is female, and  
D = event the person is divorced.

a. Obtain P(F), P(D), and P(F & D).  
b. Determine P(F or D), and interpret your answer in terms of percentages.  
c. Find the probability that a randomly selected adult is male.

a) P(F) = 0.516  
 P(D) = 0.104  
 P(F&D) = 0.06

b) P(F or D) = 0.620 = 62%

c) P(male) = 0.484