Chapter 9: Section 9.1-3

1. The 94 students in a statistics class are categorized by gender and by the year in school. The numbers obtained are displayed below:

	Year in school					
Gender	Freshman	Sophomore	Junior	Senior	Graduate	Total
Male	1	2	9	17	2	31
Female	23	17	13	7	3	63
Total	24	19	22	24	5	94

Suppose we wish to test the null hypothesis that there is no association between the year in school and gender. Under the null hypothesis, what is the expected number of male sophomores?

A) 2
B) 6
C) 6.27
D) 9.5

Use the following to answer questions 2-4:

Are avid readers more likely to wear glasses than those who read less frequently? Three hundred men in Ohio were selected at random and characterized as to whether they wore glasses and whether the amount of reading they did was above average, average, or below average. The results are presented in the following table:

	Glasses?	
Amount of reading	Yes	No
Above average	47	26
Average	48	78
Below average	31	70
Total	126	174

2. This is an $r \times c$ table. What is the value of the number r?

A) 2
B) 3
C) 4
D) 6

3. Suppose we wish to test the null hypothesis that there is no association between the amount of reading and wearing glasses. Under the null hypothesis, what is the expected number of above average readers who wear glasses?

A) 27.2
B) 30.7
C) 47
D) 81.1

4. Suppose we wish to test the null hypothesis that there is no association between the amount of reading and wearing glasses. Under the null hypothesis, what is the numerical value of the chi-square statistic?

A) 2
B) 8.65
C) 21
D) 30.7

Use the following to answer questions 5–8

When a police officer responds to a call for help in a case of spousal abuse, what should the officer do? A randomized controlled experiment in Charlotte, North Carolina, studied three police responses to spousal abuse: advise and possibly separate the couple, issue a citation to the offender, and arrest the offender. The effectiveness of the three responses was determined by re-
arrest rates. The table below shows these rates:

Number of subsequent arrests	Assigned treatment: Arrest	Citation	Advise/separate
0	175	181	187
1	36	33	24
2	2	7	1
3	1	1	0
4	0	2	0

5. What is the proportion of all cases in which the assigned treatment was "Advise/separate"?

A) 0.29
B) 0.33
C) 0.84
D) 0.88

6. What proportion of times were there no subsequent arrests out of all cases in which the assigned treatment was "Advise/separate"?

A) 0.29
B) 0.33
C) 0.84
D) 0.88

7. Suppose we wish to test the null hypothesis that the proportions of subsequent arrests are the same regardless of the treatment assigned. Under the null hypothesis, what is the expected number of times no subsequent arrest would occur for the treatment "Advise/separate"?

A) 177
B) 181
C) 187
D) 543

8. Suppose we wish to test the null hypothesis that the proportions of subsequent arrests are the same regardless of the treatment assigned. Which of the following statements is true?

A) We cannot test this hypothesis, because the police officers did not record the expected counts.
B) The test of the null hypothesis will have a very small *P*-value (below 0.0001), because the counts in each row are not identical.
C) We cannot test this hypothesis, because the expected cell counts are less than five in too many of the cells.
D) The test of the null hypothesis will have a very small *P*-value (below 0.0001), because there were so few cases where there were more than one re-arrest.

Use the following to answer questions 9–13:

Even though Puerto Rico is a commonwealth of the United States, there are many cultural differences between the states on the continent and the island of Puerto Rico. These differences include the way consumers handle problems with purchases. Two researchers surveyed owners of VCRs in the Northeast U.S. and in Puerto Rico. They asked those who had experienced problems with their VCRs whether they complained. The results are given in the table below:

Complained?	Location: United States	Puerto Rico
No	94	33
Yes	330	64

The SPSS output for the above table is given below. The output includes the cell counts, the expected cell counts, and the chi-square statistic. Expected counts are printed below observed counts:

Complain? * Location Crosstabulation

			Location		
			United States	Puerto Rico	Total
Complain?	No	Count	94	33	127
		Expected Count	103.4	23.6	127.0
	Yes	Count	330	64	394
		Expected Count	320.6	73.4	394.0
Total		Count	424	97	521
		Expected Count	424.0	97.0	521.0

Chi-square = 0.855 + 3.744 + 0.276 + 1.204 = 6.079

9. What proportion of Puerto Ricans did not complain?
A) 0.063
B) 0.26
C) 0.34
D) 0.516

10. What proportion of Americans did not complain?
A) 0.18
B) 0.194
C) 0.22
D) 0.74

11. What are the appropriate degrees of freedom for the chi-square statistic?
A) 1
B) 2
C) 3
D) 4

12. Which cell contributes most to the chi-square statistic?
A) Americans in the United States who did not complain.
B) Americans in the United States who complained.
C) Puerto Ricans who did not complain.
D) Puerto Ricans who did complain.

13. What do we know about the *P*-value for testing the null hypothesis that the probability of complaining is the same for the United States and Puerto Rico?
A) *P*-value < 0.010
B) 0.010 < *P*-value < 0.025
C) 0.025 < *P*-value < 0.05
D) This cannot be determined, because these are not the hypotheses being tested by the chi-square test.

14. Which of the following statements about $r \times c$ tables is (are) FALSE?
A) The null hypothesis is tested using the chi-square statistic with $(r\text{-}1)(c\text{-}1)$ degrees of freedom.
B) The statistic used to test the null hypothesis is denoted by $X^2 = \sum \frac{(\text{observed-expected})^2}{\text{observed}}$.
C) The *P*-value for the test is $P(\chi^2 \geq X^2)$ where χ^2 is a random variable having the χ^2 (df) distribution with df = $(r-1)(c-1)$.
D) Under the null hypothesis the X^2 statistic has approximately a chi-square distribution with $(r-1)(c-1)$ degrees of freedom.
E) The chi-square approximation is adequate when the average expected cell count is 5 or greater and all individual expected counts are 1 or greater, except in the case of 2×2 tables.

Use the following to answer questions 15-18:

A study was performed to examine the personal goals of children in grades 4, 5, and 6. A random sample of students was selected from each of the grades 4, 5, and 6 from schools in Georgia. The students received a questionnaire regarding achieving personal goals. They were asked what they would most like to do at school: make good grades, be good at sports, or be popular. Results are presented by the gender of the child in the table below:

	Gender	
Goal	Boys	Girls
Make good grades	96	295
Be popular	32	45
Be good in sports	94	40

15. What would be the null hypothesis for a chi-square test based on these data?
A) Personal goals and gender are independent.
B) The mean personal goal is the same for boys and girls.
C) The distribution of personal goals is different for boys and girls.
D) The distribution of gender is different for the three different personal goals.

16. The data are going to be summarized by computing the conditional distributions of personal goals for boys and girls. What would be the entry for boys who like to make good grades?
A) 0.22
B) 0.43
C) 0.58
D) 0.70

17. What is the value of the chi-square statistic based on the data presented in this table?
A) 3.84
B) 5.99
C) 16.105
D) 89.966

18. Are the data statistically significant at the 5% significance level?
A) Yes
B) No
C) This cannot be determined from the information given.

Use the following to answer questions 19-22:

A particular paperback mystery book is published with a choice of three different pictures on the cover: a photograph of the actor playing the main character in the movie version of the book, a drawing of the mansion where the story in the book takes place, or an embossed graphic of the murder weapon. A certain bookstore keeps copies of this book with each of the pictures on the cover on its racks. To test the hypothesis that sales of this book are equally divided among the three choices, a simple random sample of 120 purchases of this book is obtained. The numbers are displayed in the table below:

Picture on the cover			
Photograph	Drawing	Embossed graphic	Total
31	47	42	120

19. Under the null hypothesis that sales are equally divided among the three different covers, what is the value of the expected count for sales of the book with the photograph of the actor on the cover?
A) 0.258
B) 0.333
C) 31
D) 40

20. What is the value of the chi-square statistic?
A) 1.225
B) 3
C) 3.35
D) 40

21. What are the appropriate degrees of freedom for the chi-square statistic?

A) 1
B) 2
C) 3
D) 4

22. At the 10% significance level, do we reject the null hypothesis?

A) Yes
B) No
C) This cannot be determined from the information given.

Use the following to answer questions 23 and 24:

There have been many studies that looked at the incidence of heart attack on the different days of the week. Studies in Japan and Scotland seemed to find that there was a substantial "spike" in heart attack on Mondays, perhaps as many as 20% more. A researcher studied a random selection of 200 heart attack victims and recorded the day of the week that their attack occurred. The following table summarizes the results:

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Count	24	36	27	26	32	26	29

The researcher was interested in whether the distribution of heart attacks was the same or different across the days of the week.

23. What are the null and alternative hypotheses the researcher should establish?

A) H_o: The distribution of heart attacks is independent of the days of the week.
H_a: The distribution of heart attacks depends on the days of the week.

B) H_o: The probability of heart attack on any day is highest on Monday.
H_a: The probability of heart attack on any day is not highest on Monday.

C) H_o: The probability of heart attack is the same for each day of the week.
H_a: The probability of heart attack is not the same for at least one or more days.

D) H_o: The probability of heart attack is the same for each day of the week.
H_a: The probability of heart attack is different for each day of the week.

E) H_o: The distribution of heart attack has a known distribution across the week.
H_a: The distribution of heart attack has an unknown distribution across the week.

24. Under the appropriate null hypothesis, what is the expected number of heart attacks that occur on Monday?

A) 34.3
B) 36.0
C) 40.2
D) 28.6
E) Not within ± 1 of any of the above.