Quiz 11.1AAP StatisticsName:

1. A grocery store sells four different sizes of a popular brand of corn flakes. For the past few years the proportion of boxes they sell of each size has been quite stable: 10% Small, 15% Medium, 60% Large, and 15% Jumbo. They decide to change the pricing of the four sizes and want to see if this changes the proportion of boxes they sell of each size. To test this, a few weeks after changing the prices they take a simple random sample of 120 transactions involving corn flakes and count how many boxes of each size were sold. Here are the results.

01 1 1	01	110	1 1	•
()hearved number	of hoves	cold tor	each hov	C170
	UI UUACS	5010 101	Cacil DOA	SILC

Small	Medium	Large	Jumbo
8	24	61	27

- (a) We wish to carry out a test of significance to see if the distribution of sizes of cereal boxes sold has changed. State the null and alternative hypotheses for this test.
- (b) Find the expected counts for each size box under the assumption that the null hypothesis is true.

Expected number of boxes sold for each box size

Small	Medium	Large	Jumbo

(c) Discuss whether the conditions for this test have been met.

(d) Find the value of the test statistic and the *P*-value of the test, and make the appropriate conclusion. Use $\alpha = 0.05$.

(e) Based on your answer to (d), which error is it possible that you have made, Type I or Type II? Describe that error in the context of the problem.

(f) Use the components of the chi-square statistic to perform a follow-up analysis on the impact of the new prices on the sales of different sizes of cereal boxes.

Chapter 11 Solutions

Quiz 11.1A

1. (a) H_0 : The distribution of sizes of all boxes sold of this brand of cereal did not change when the prices changed. H_a : The distribution of sizes of all boxes sold of this brand of cereal changed when the prices changed. (b) Expected counts: Small: (0.10)(120) = 12,

Medium: (0.15)(120) = 18, Large: (0.60)(120) = 72, Jumbo: (0.15)(120) = 18. (c) *Random*: the data come from a simple random sample of sales records. *10%*: We must assume that there were more than 1200 boxes of cereal sold. *Large counts*: All the expected counts (see part (b)) are at least 5.

(d)
$$\chi^2 = \frac{(8-12)^2}{12} + \frac{(24-18)^2}{18} + \frac{(61-72)^2}{72} + \frac{(27-18)^2}{18} = 9.514; df = 3; P-value 0.023 (Using Table)$$

C, 0.02 < P-value < 0.025). Since the *P*-value is less than $\alpha = 0.05$, we can reject H_0 . There is convincing evidence that the distribution of size of boxes sold changed after the prices changed. (e) It's possible that we have made a Type I error, which is concluding that the distribution of box sizes sold has changed when it has not. (f) Components of the chi-square statistic: Small: 1.33, Medium: 2.00, Large: 1.68, Jumbo: 4.5. The observed count of jumbo-sized boxes was much larger than the expected count, so it appears than an increase in sales of jumbo-sized boxes was the biggest impact of the price changes.

Quiz 11.2AAP StatisticsName:

1. A radio station trying to determine what kind of music to play takes a simple random sample of 50 students at each of three locations: a local middle school, a high school, and a college. The students are asked to choose which of three different music genres they most enjoy hearing on the radio. Here are the results:

		Нір Нор	Alternative	Post-rock	Total
1 ~~	Middle School	28	18	4	50
Age	High School	22	22	6	50
level	College	16	20	14	50
	Total	66	60	24	150

(a) In the table below, provide the appropriate conditional distributions based on the data collected for comparing the music-listening preferences of the three age levels, based on the data above. Music Genre

			Whate Ochie	
		Нір Нор	Alternative	Post-rock
	Middle School			
Age level	High School			
	College			

(b) Make a graph that illustrates these conditional distributions effectively, and use the table from (a) and your graph to describe the relationship between age level and preferred music genre.

i i											: :
;;											;;
											: :
	1										
											: :
											+ i
	l i								•		
											i i
1		r			r			r			r ,
											: :
											: :
ii											ii
											: :
i i											i i
											!
											i i
!!			!							!	<u> </u>
											: :
											i i
! !											!!
											* !
i											: i
	l i								•		
											!
•	1	•	•	•	•	•	•	•	1	•	. i
											!!
'							!			!	<u>' '</u>

(c) Perform the appropriate statistical test to determine if there is a difference in the music preference of these three age groups.

(d) If you chose a chi-square test for homogeneity in part (c), explain how the data could have been obtained to make a chi-square test for independence appropriate. If you chose a test for independence, explain how the data could have been obtained to make a test for homogeneity appropriate.

Quiz 11.2A

1. (a) & (b) Since age level is the explanatory variable, we calculate conditional distributions for each age level. See table and graph below (two options for graphs are shown: parallel and segmented bar graphs.) The data suggest that middle school students strongly prefer Hip Hop, most high school students are evenly divided between Hip Hop and Alternative, and college students are more evenly divided between all three genres. (c) State: We are testing the hypothesis H_{a} : The distribution of music genre is the same for the populations in all three age groups, against H_a : The distribution of music genre is not the same for the populations in all three age groups. We will use a significance level of $\alpha = 0.05$. Plan: The procedure is a chi-square test for homogeneity. Conditions: *Random*: the data come from SRSs of each population. 10%: The radio station's listening audience is likely to be more than 10 times the sample size. Large counts: All expected counts are at least 5 (see expected counts table below). Do: Using a calculator, $\chi^2 = 10.673$; df = 4; P - value = 0.0305. (From Table C, *P*-value is between 0.025 and 0.05). Conclude: Since the *P*-value is smaller than $\alpha = 0.05$, we can reject H_0 . There is convincing evidence that the distribution of music genre is different for the populations in the three age groups. (d) The test in (c) was a test of homogeneity. If we had treated all of the station's listeners as a single population from which we took one SRS, and if we treated age level and preferred music genre as two categorical variables, then a test of independence would have been appropriate.

Conditional	distributions	for	part (a)
Contantional	anounounono	101	puit	u	,

		Нір Нор	Alternative	Post-rock
	Middle School	0.56	0.36	0.08
Age	High School	0.44	0.44	0.12
CVCI	College	0.32	0.40	0.28



Expected counts for part (c)

		Нір Нор	Alternative	Post-rock
	Middle School	22	20	8
Age	High School	22	20	8
level	College	22	20	8