Chapter 11: Problem Set

Numbers with an asterisk* have detailed solutions in the back of the book. These solutions display intermediate steps that involve rounded numbers. The numerical answers (Test Statistics and P-values) have been calculated at full precision using technology (**Excel and TI-84+**). As such, the intermediate steps may not fully align with final answers. Your answers may differ from those given. These differences will not change the conclusion of the test.

Chi-Squared Test for Goodness of Fit (11.1)

1.* **Answers to Multiple-Choice Problems:** A student wants to see if the correct answers to multiple choice problems are evenly distributed. She heard a rumor that if you don't know the answer, you should always *pick C*. In a sample of 100 multiple-choice questions from prior tests and quizzes, the distribution of correct answers are given in the table below. In all of these questions, there were four options {A, B, C, D}.

	Correct Answers ($n = 100$)			
	A	B	C	D
Count	12	24	31	33

The Test: Test the claim that correct answers for all multiple-choice questions are not evenly distributed. Test this claim at the 0.05 significance level.

2. **Choosing Lottery Numbers:** In the Super-Mega lottery there are 50 numbers (1 to 50), a player chooses ten different numbers and hopes that these get drawn. If the player's numbers get drawn, he/she wins an obscene amount of money. The table below displays the frequency with which classes of numbers are chosen (not drawn). These numbers came from a sample of 180 chosen numbers.

	Chosen Numbers ($n = 180$)				
	1 to 10	11 to 20	21 to 30	31 to 40	41 to 50
Count	54	42	34	27	23

The Test: Test the claim that chosen numbers are not evenly distributed across the five classes. Test this claim at the 0.05 significance level.

3. **Customer Distribution by Weekday:** A drop-in auto repair shop staffs the same number of mechanics on every weekday (weekends are not counted here). One of the mechanics thinks this is a bad idea because he suspects the number of customers is not evenly distributed across these days. For a sample of 289 customers, the counts by weekday are given in the table.

	Number of Customers ($n = 289$)				
	Monday	Tuesday	Wednesday	Thursday	Friday
Count	48	71	65	57	48

The Test: Test the claim that the number of customers is not evenly distributed across the five weekdays. Test this claim at the 0.05 significance level.

4.* **M&M's Color Distribution:** Suppose the makers of M&M candies give the following average percentages for the mix of colors in their bags of plain chocolate M&M's.

	Stated Distribution of Colors					
	Brown	Yellow	Red	Orange	Green	Blue
Percent	30%	20%	20%	10%	10%	10%

Now, you randomly select 200 M&M's and get the counts given in the table below. You expected about 20 blues but only got 10. You suspect that the maker's claim is not true.

	Observed Counts by Color ($n = 200$)					
	Brown	Yellow	Red	Orange	Green	Blue
Count	67	35	44	18	26	10

The Test: Test whether or not the color of M&M's candies fits the distribution stated by the makers (Mars Company). Conduct this test at the 0.05 significance level.

5. **Changes in Education Attainment:** According to the U.S. Census Bureau, the distribution of *Highest Education Attainment* in U.S. adults aged 25 - 34 in the year 2005 is given in the table below.

	Census: Highest Education Attainment - **2005**				
	No Diploma	High School Diploma	Associate's Degree	Bachelor's Degree	Graduate or Professional Degree
Percent	14%	48%	8%	22%	8%

In a survey of 4000 adults aged 25 - 34 in the year 2012, the counts for these levels of educational attainment are given in the table below.

	Survey ($n = 4000$): Highest Education Attainment - **2012**				
	No Diploma	High School Diploma	Associate's Degree	Bachelor's Degree	Graduate or Professional Degree
Count	483	1922	341	881	373

The Test: Test whether or not the distribution of education attainment has changed from 2005 to 2012. Conduct this test at the 0.05 significance level.

6. **Roulette:** In the casino game of roulette there is a wheel with 19 black slots, 19 red slots, and 2 green slots. In the game, a ball is rolled around a spinning wheel and it lands in one of the slots. It is assumed that each slot has the same probability of getting the ball. This results in the table of probabilities below (left). You watch the game for 100 rounds and count the number of black, red, and green results. The table on the right summarizes your observations.

	Fair Probabilities		
	black	red	green
Probability	19/40	19/40	2/40

	Outcomes ($n = 100$)		
	black	red	green
Count	39	54	7

The Test: Test the claim that this roulette table is not fair. That is, test the claim that the distribution of colors for all spins of the wheel does not fit the expected distribution from a fair table. Test this claim at the 0.05 significance level.

Chi-Squared Test of Independence (11.2)

7.* **Flu Vaccine:** The Center for Disease Control (CDC) claims that the flu vaccine is effective in reducing the probability of getting the flu. They conduct a trial on 3000 people. The results are summarized in the contingency table. Test for a dependent relationship between getting the vaccine and getting the flu. **Conduct this test at the 0.01 significance level.**

	Got Vaccine	No Vaccine	**Totals**
Got Flu	18	45	63
No Flu	1482	1455	2937
Totals	1500	1500	3000

8. **Beer and Diapers:** There is a popular story (among data miners) that there is a correlation between men buying diapers and buying beer while shopping. A student tests this theory by surveying 140 male shoppers as they left a grocery store. The results are summarized in the contingency table. Test for a dependent relationship between buying beer and buying diapers. Conduct this test at the 0.05 significance level.

	Bought Diapers	Did Not Buy Diapers	**Totals**
Beer	7	48	55
No Beer	9	76	85
Totals	16	124	140

9.* **Win/Loss and With/Without Joe:**
Joe plays basketball for the Wildcats and missed some of the season due to an injury. The win/loss record with and without Joe is summarized in the table to the right. Test for a significant dependent relationship between wins/losses and whether or not Joe played. Conduct this test at the 0.05 significance level.

	Wins	Losses	**Totals**
With Joe	15	9	24
Without Joe	6	10	16
Totals	21	19	40

10. **Win/Loss and Home/Visitor, Chicago Bulls:**
The home and visitor record for the Chicago Bulls 2012/2013 regular season is given in the contingency table to the right. Test whether or not there is a significant dependent relationship between wins/losses and the home/visitor status of the team. Conduct this test at the 0.05 significance level.

	Wins	Losses	**Totals**
Home	24	17	41
Visitor	21	20	41
Totals	45	37	82

11.* **Support of Background Checks by Political Party:** In April of 2013, the U.S. Senate did not pass a bill to expand background checks to all gun sales despite popular approval of the idea. Gallup conducted a poll on this issue with the question: *Would you vote for or against a law to require background checks for all gun sales?*. The results by political affiliation are summarized in the contingency table below.

	Republican	Independent	Democrat	**Totals**
For Background Checks	219	240	288	747
Against Background Checks	81	57	12	150
Totals	300	297	300	897

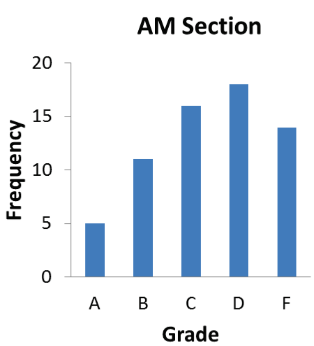
The Test: Test for a dependent relationship between party affiliation and opinion on expanded background checks. Test this claim at the 0.05 significance level.

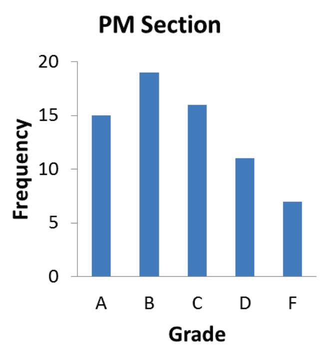
12. **Pro-choice/Pro-life and Region of the Country:** The results of a 2013 Gallup poll about people's position on abortion (pro-life or pro-choice) by region of the country are summarized in the contingency table below.

	East	Midwest	South	West	**Totals**
Pro-Choice	212	106	176	188	682
Pro-Life	184	94	231	215	724
Totals	396	200	407	403	1406

The Test: Test whether or not there is a dependent relationship between abortion stance and region. Conduct this test at the 0.05 significance level.

13.* **Grades and AM/PM Section of Stats:** There were two large sections of statistics this term at State College, an 8:00 (AM) section and a 1:30 (PM) section. The final grades for both sections are depicted in the bar graphs below and in the contingency table.





Corresponding Contingency Table

	Grades					
	A	B	C	D	F	**Totals**
AM	5	11	16	18	14	64
PM	15	19	16	11	7	68
Totals	20	30	32	29	21	132

The Test: Test for a significant dependent relationship between grades and the section of the course. Conduct this test at the 0.05 significance level.

14. **Hockey Teams:** There are five teams in the northeast conference of the NHL. Their win/loss records for the 2012/2013 season are depicted in the table below.

	Montreal Canadiens	Boston Bruins	Toronto Maple Leafs	Ottawa Senators	Buffalo Sabers	**Totals Totals**
Wins	29	28	26	25	21	129
Losses	19	20	22	23	27	111
Totals	48	48	48	48	48	240

The Test: Test for a significant dependent relationship between wins/losses and team for this season. Conduct this test at the **0.10 significance level.**

ANOVA Tests (11.3)

15.* **Wait-Times**: There are three registers at the local grocery store. I suspect the mean wait-times for the registers are different. The sample data is depicted below. The second table displays results from an ANOVA test on this data with software. I claim there is a difference in mean wait-times between the registers.

	Wait-times in Minutes								$\bar{x}$
Register 1	2.0	2.0	1.1	2.0	1.0	2.0	1.0	1.3	1.55
Register 2	1.8	2.0	2.2	2.2	1.8	2.1	2.2	1.7	2.00
Register 3	2.1	2.1	1.8	1.5	1.4	1.4	2.0	1.7	1.75

ANOVA Results	
F	P-value
3.285	0.0574

(a) Test my claim at the 0.10 significance level.

(b) Does your conclusion change at the 0.05 significance level?

16. **Tomato weights and Fertilizer**: Carl the farmer has three fields of tomatoes, on one he used no fertilizer, in another he used organic fertilizer, and the third he used a chemical fertilizer. He wants to see if there is a difference in the mean weights of tomatoes from the different fields. The sample data is given below. The second table gives the results from an ANOVA test. Carl claims there is a difference in the mean weight for all tomatoes between the different fertilizing methods.

	Tomato-Weight in Grams										$\bar{x}$
No Fertilizer	123	119	95	97	94	120	114	118	129	128	113.7
Organic Fertilizer	112	127	138	133	140	114	126	134	123	144	129.1
Chemical Fertilizer	115	141	143	134	129	134	135	129	113	148	132.1

ANOVA Results	
F	P-value
6.921	0.00375

(a) Test Carl's claim at the 0.05 significance level.

(b) Does your conclusion change at the 0.01 significance level?

17.* **Help Desk:** There are four student employees at the Computer Help Desk. The supervisor wants to determine if there is a difference in the mean number of clients served per hour between the four employees. The data from a random selection of hours is depicted below. The second table displays results from an ANOVA test on this data with software. The supervisor claims there is a difference in the mean number of clients served per hour for these employees.

	Clients Served in the Hour									$\bar{x}$
Alan	2	4	4	4	5	5				4.0
Beth	7	5	6	4	5	4	7	3	7	5.3
Carl	5	4	2	4	2	5	4			3.7
Denise	2	5	2	3	2	2	2	5	6	3.2

ANOVA Results	
F	P-value
3.556	0.0274

(a) Test the supervisor's claim at the 0.05 significance level.

(b) Does your conclusion change at the 0.01 significance level?

18. **Travel-Times**: There are three different ways I can go to work in the morning. I want to see if there is a difference in mean travel-times between the three different ways. The sample data is depicted below. The second table displays results from an ANOVA test on this data with software. I claim there is a difference in mean travel-times between the three different routes.

	Travel Time in Minutes						$\bar{x}$
Interstate	23	24	22	22	21	20	22.0
Route 15	17	21	22	19			19.8
Back Roads	22	19	19	24	18		20.4

ANOVA Results	
F	P-value
1.656	0.232

(a) Test my claim at the 0.05 significance level.

(b) Does your conclusion change at the 0.10 significance level?

19.* **Defects by Day of Week**: A bottling plant bottles a lot of soda. There are often errors that result in defective bottles. The plant manager wants to know if these defects happen more frequently on different days of the week. She has a lot of data (too much to put here) but the sample sizes and means are given in the table below. The second table displays results from an ANOVA test on the full data set. The manager claims there is a difference in the mean number of defects between production days.

Day	Sample Size	Mean Number of Defects ($\bar{x}$)
Mondays	31	82.3
Tuesdays	31	81.9
Wednesdays	32	79.2
Thursdays	32	80.2
Fridays	32	83.1

ANOVA Results	
F	P-value
12.054	1.6×10^{-8}

(a) Test this claim at the 0.05 significance level.

(b) Does your conclusion change at the 0.01 significance level?

20. **Income by State**: A student at a private college in New England wants to see if there is a difference in mean household incomes for students from the various New England states. The tables below give the sample sizes and means from her study and the results of an ANOVA test run on this data. She wants to test for a difference in mean household income for all students from the different states.

Student's State	Sample Size	Mean Household Income ($\bar{x}$)
New Hampshire	11	97,607
Connecticut	14	96,224
Massachusetts	19	85,790
Vermont	13	77,903
Maine	7	77,471

ANOVA Results	
F	P-value
2.798	0.0339

(a) Conduct the test at the 0.05 significance level.

(b) Does your conclusion change at the 0.01 significance level?