

Ch12 , 13 , and 14 Chi Square Tests and Two Independent Population Tests

2014

4. A parent advisory board for a certain university was concerned about the effect of part-time jobs on the academic achievement of students attending the university. To obtain some information, the advisory board surveyed a simple random sample of 200 of the more than 20,000 students attending the university. Each student reported the average number of hours spent working part-time each week and his or her perception of the effect of part-time work on academic achievement. The data in the table below summarize the students' responses by average number of hours worked per week (less than 11, 11 to 20, more than 20) and perception of the effect of part-time work on academic achievement (positive, no effect, negative).

		Average Time Spent on Part-Time Jobs		
		Less Than 11 Hours per Week	11 to 20 Hours per Week	More Than 20 Hours per Week
Perception of the Effect of Part-Time Work on Academic Achievement	Positive Effect	21	9	5
	No Effect	58	32	15
	Negative Effect	18	23	19

A chi-square test was used to determine if there is an association between the effect of part-time work on academic achievement and the average number of hours per week that students work. Computer output that resulted from performing this test is shown below.

CHI-SQUARE TEST

Expected counts are printed below observed counts

	<11	11-20	>20	Total
Positive	21 16.975	9 11.200	5 6.825	35
No effect	58 50.925	32 33.600	15 20.475	105
Negative	18 29.100	23 19.200	19 11.700	60
Total	97	64	39	200

Chi-Sq = 13.938, DF = 4, P-Value = 0.007

- State the null and alternative hypotheses for this test.
- Discuss whether the conditions for a chi-square inference procedure are met for these data.
- Given the results from the chi-square test, what should the advisory board conclude?
- Based on your conclusion in part (c), which type of error (Type I or Type II) might the advisory board have made? Describe this error in the context of the question.

2010

6. Hurricane damage amounts, in millions of dollars per acre, were estimated from insurance records for major hurricanes for the past three decades. A stratified random sample of five locations (based on categories of distance from the coast) was selected from each of three coastal regions in the southeastern United States. The three regions were Gulf Coast (Alabama, Louisiana, Mississippi), Florida, and Lower Atlantic (Georgia, South Carolina, North Carolina). Damage amounts in millions of dollars per acre, adjusted for inflation, are shown in the table below.

HURRICANE DAMAGE AMOUNTS IN MILLIONS OF DOLLARS PER ACRE

	Distance from Coast				
	< 1 mile	1 to 2 miles	2 to 5 miles	5 to 10 miles	10 to 20 miles
Gulf Coast	24.7	21.0	12.0	7.3	1.7
Florida	35.1	31.7	20.7	6.4	3.0
Lower Atlantic	21.8	15.7	12.6	1.2	0.3

- (a) Sketch a graphical display that compares the hurricane damage amounts per acre for the three different coastal regions (Gulf Coast, Florida, and Lower Atlantic) and that also shows how the damage amounts vary with distance from the coast.
- (b) Describe differences and similarities in the hurricane damage amounts among the three regions.

Because the distributions of hurricane damage amounts are often skewed, statisticians frequently use rank values to analyze such data.

- (c) In the table below, the hurricane damage amounts have been replaced by the ranks 1, 2, or 3. For each of the distance categories, the highest damage amount is assigned a rank of 1 and the lowest damage amount is assigned a rank of 3. Determine the missing ranks for the 10-to-20-miles distance category and calculate the average rank for each of the three regions. Place the values in the table below.

ASSIGNED RANKS WITHIN DISTANCE CATEGORIES

	Distance from Coast					Average Rank
	< 1 mile	1 to 2 miles	2 to 5 miles	5 to 10 miles	10 to 20 miles	
Gulf Coast	2	2	3	1		
Florida	1	1	1	2		
Lower Atlantic	3	3	2	3		

(d) Consider testing the following hypotheses.

H_0 : There is no difference in the distributions of hurricane damage amounts among the three regions.

H_a : There is a difference in the distributions of hurricane damage amounts among the three regions.

If there is no difference in the distribution of hurricane damage amounts among the three regions (Gulf Coast, Florida, and Lower Atlantic), the expected value of the average rank for each of the three regions is 2. Therefore, the following test statistic can be used to evaluate the hypotheses above:

$$Q = 5 \left[(\bar{R}_G - 2)^2 + (\bar{R}_F - 2)^2 + (\bar{R}_A - 2)^2 \right]$$

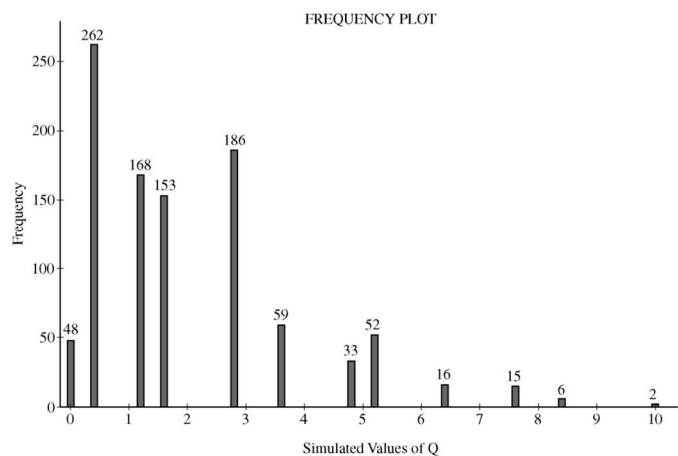
where \bar{R}_G is the average rank over the five distance categories for the Gulf Coast (and \bar{R}_F and \bar{R}_A are similarly defined for the Florida and Lower Atlantic coastal regions).

Calculate the value of the test statistic Q using the average ranks you obtained in part (c).

(e) One thousand simulated values of this test statistic, Q , were calculated, assuming no difference in the distributions of hurricane damage amounts among the three coastal regions. The results are shown in the table below. These data are also shown in the frequency plot where the heights of the lines represent the frequency of occurrence of simulated values of Q .

Frequency Table for Simulated Values of Q

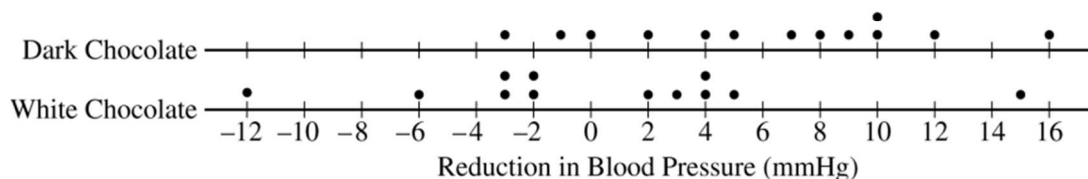
Q	Frequency	Cumulative Frequency	Percent	Cumulative Percent
0.0	48	48	4.80	4.80
0.4	262	310	26.20	31.00
1.2	168	478	16.80	47.80
1.6	153	631	15.30	63.10
2.8	186	817	18.60	81.70
3.6	59	876	5.90	87.60
4.8	33	909	3.30	90.90
5.2	52	961	5.20	96.10
6.4	16	977	1.60	97.70
7.6	15	992	1.50	99.20
8.4	6	998	0.60	99.80
10.0	2	1000	0.20	100.00



Use these simulated values and the test statistic you calculated in part (d) to determine if the observed data provide evidence of a significant difference in the distributions of hurricane damage amounts among the three coastal regions. Explain.

5. Studies have shown that foods rich in compounds known as flavonoids help lower blood pressure. Researchers conducted a study to investigate whether there was a greater reduction in blood pressure for people who consumed dark chocolate, which contains flavonoids, than people who consumed white chocolate, which does not contain flavonoids. Twenty-five healthy adults agreed to participate in the study and add 3.5 ounces of chocolate to their daily diets. Of the 25 participants, 13 were randomly assigned to the dark chocolate group and the rest were assigned to the white chocolate group. All participants had their blood pressure recorded, in millimeters of mercury (mmHg), before adding chocolate to their daily diets and again 30 days after adding chocolate to their daily diets.

The reduction in blood pressure (before minus after) for each of the participants in the two groups is shown in the dotplots below.

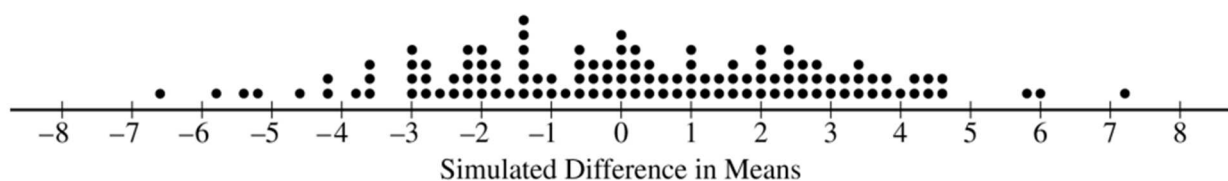


- (a) Determine and compare the medians of the reduction in blood pressure for the two groups.

The researchers found the mean reduction in blood pressure for those who consumed dark chocolate is $\bar{x}_{dark} = 6.08$ mmHg and the mean reduction in blood pressure for those who consumed white chocolate is $\bar{x}_{white} = 0.42$ mmHg.

- (b) One researcher indicated that because the difference in sample means of 5.66 mmHg is greater than 0 there is convincing statistical evidence to conclude that the population mean blood pressure reduction for those who consume dark chocolate is greater than for those who consume white chocolate. Why might the researcher's conclusion, based only on the difference in sample means of 5.66 mmHg, not necessarily be true?

A simulation was conducted to investigate whether there is a greater reduction of blood pressure for those who consume dark chocolate than for those who consume white chocolate. The simulation was conducted under the assumption that no difference exists. The results of 120 trials of the simulation are shown in the following dotplot.



- (c) Use the results of the simulation to determine whether the results from the 25 participants in the study provide convincing statistical evidence, at a 5 percent level of significance, that adding dark chocolate to a daily diet will result in a greater reduction in blood pressure, on average, than adding white chocolate to a daily diet. Justify your answer.

2012021

The manager of a large company that sells pet supplies online wants to increase sales by encouraging repeat purchases. The manager believes that if past customers are offered \$10 off their next purchase, more than 40 percent of them will place an order. To investigate the belief, 90 customers who placed an order in the past year are selected at random. Each of the selected customers is sent an e-mail with a coupon for \$10 off the next purchase if the order is placed within 30 days. Of those who receive the coupon, 38 place an order.

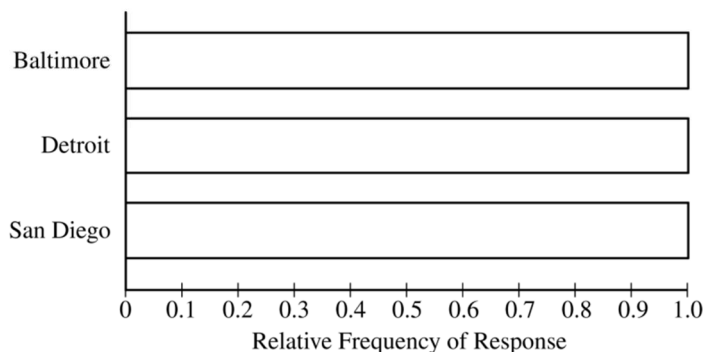
- (a) Is there convincing statistical evidence, at the significance level of $\alpha = 0.05$, that the manager's belief is correct? Complete the appropriate inference procedure to support your answer.
- (b) Based on your conclusion from part (a), which of the two errors, Type I or Type II, could have been made? Interpret the consequence of the error in context.

2021

A research center conducted a national survey about teenage behavior. Teens were asked whether they had consumed a soft drink in the past week. The following table shows the counts for three independent random samples from major cities.

	Baltimore	Detroit	San Diego	Total
Yes	727	1,232	1,482	3,441
No	177	431	798	1,406
Total	904	1,663	2,280	4,847

- (a) Suppose one teen is randomly selected from each city's sample. A researcher claims that the likelihood of selecting a teen from Baltimore who consumed a soft drink in the past week is less than the likelihood of selecting a teen from either one of the other cities who consumed a soft drink in the past week because Baltimore has the least number of teens who consumed a soft drink. Is the researcher's claim correct? Explain your answer.
- (b) Consider the values in the table.
- (i) Construct a segmented bar chart of relative frequencies based on the information in the table.



- (ii) Which city had the smallest proportion of teens who consumed a soft drink in the previous week? Determine the value of the proportion.
- (c) Consider the inference procedure that is appropriate for investigating whether there is a difference among the three cities in the proportion of all teens who consumed a soft drink in the past week.
- (i) Identify the appropriate inference procedure.
- (ii) Identify the hypotheses of the test.

4. The Behavioral Risk Factor Surveillance System is an ongoing health survey system that tracks health conditions and risk behaviors in the United States. In one of their studies, a random sample of 8,866 adults answered the question “Do you consume five or more servings of fruits and vegetables per day?” The data are summarized by response and by age-group in the frequency table below.

Age-Group (years)	Yes	No	Total
18–34	231	741	972
35–54	669	2,242	2,911
55 or older	1,291	3,692	4,983
Total	2,191	6,675	8,866

Do the data provide convincing statistical evidence that there is an association between age-group and whether or not a person consumes five or more servings of fruits and vegetables per day for adults in the United States?

5. Psychologists interested in the relationship between meditation and health conducted a study with a random sample of 28 men who live in a large retirement community. Of the men in the sample, 11 reported that they participate in daily meditation and 17 reported that they do not participate in daily meditation.

The researchers wanted to perform a hypothesis test of

$$H_0 : p_m - p_c = 0$$

$$H_a : p_m - p_c < 0,$$

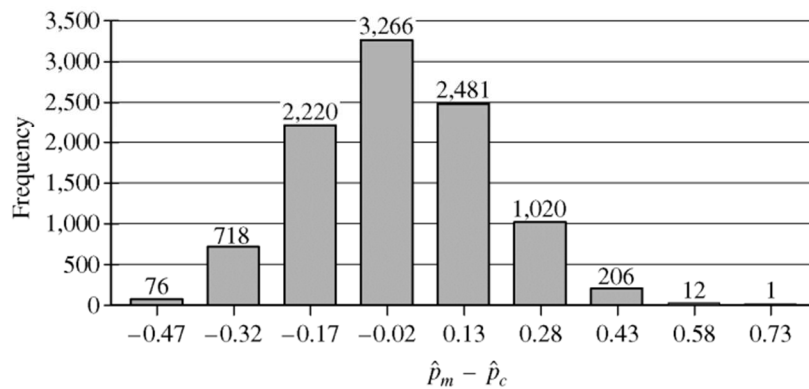
where p_m is the proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation and p_c is the proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily meditation.

- (a) If the study were to provide significant evidence against H_0 in favor of H_a , would it be reasonable for the psychologists to conclude that daily meditation causes a reduction in blood pressure for men in the retirement community? Explain why or why not.

The psychologists found that of the 11 men in the study who participate in daily meditation, 0 had high blood pressure. Of the 17 men who do not participate in daily meditation, 8 had high blood pressure.

- (b) Let \hat{p}_m represent the proportion of men with high blood pressure among those in a random sample of 11 who meditate daily, and let \hat{p}_c represent the proportion of men with high blood pressure among those in a random sample of 17 who do not meditate daily. Why is it not reasonable to use a normal approximation for the sampling distribution of $\hat{p}_m - \hat{p}_c$?

Although a normal approximation cannot be used, it is possible to simulate the distribution of $\hat{p}_m - \hat{p}_c$. Under the assumption that the null hypothesis is true, 10,000 values of $\hat{p}_m - \hat{p}_c$ were simulated. The histogram below shows the results of the simulation.



- (c) Based on the results of the simulation, what can be concluded about the relationship between blood pressure and meditation among men in the retirement community?