The Chi-Square Distribution: Homework

EXERCISE 1

a. Explain why the "goodness of fit" test and the "test for independence" are generally right tailed tests.
b. If you did a left-tailed test, what would you be testing?

(2) - (23) For each word problem, use a solution sheet to solve the hypothesis test problem. Round expected frequency to two decimal places.

EXERCISE 2

A 6-sided die is rolled 120 times. Fill in the expected frequency column. Then, conduct a hypothesis test to determine if the die is fair. The data below are the result of the 120 rolls.

face value	frequency	expected frequency
1	15	
2	29	
3	16	
4	15	
5	30	
6	15	

EXERCISE 3

The marital status distribution of the U.S. male population, age 15 and older, is as shown below. (Source: U.S. Census Bureau, *Current Population Reports*)

Marital Status	Percent	Expected Frequency
never married	31.3	
married	56.1	
widowed	2.5	
divorced/separated	10.1	

Suppose that a random sample of 400 U.S. young adult males, 18 – 24 years old, yielded the following frequency distribution. We are interested in whether this age group of males fits the

distribution of the U.S. adult population. Calculate the frequency one would expect when surveying 400 people. Fill in the above table, rounding to two decimal places.

Marital Status	Frequency
never married	140
married	238
widowed	2
divorced/separated	20

(4) - (5) The percent columns in the chart below contain the race/ethnicity of U.S. Public Schools: High School Class of 2009 and the Advanced Placement Examinee Population for that class. (Source: *http//:www.collegeboard.com*). Suppose the right column contains the result of a survey of 1000 local students from the Class of 2009 who took an AP Exam.

Race/Ethnicity	AP Examinee Population	Overall Student Population	Survey Frequency
Asian, Asian American or Pacific Islander	10.2%	5.4%	113
Black or African American	8.2%	14.5%	94
Hispanic or Latino	15.5%	15.9%	136
American Indian or Alaska Native	0.6%	1.2%	10
White	59.4%	61.6%	604
Not reported/other	6.1%	1.4%	43

EXERCISE 4

Perform a goodness-of-fit test to determine whether the local results follow the distribution of the U.S. Overall Student Population, based on ethnicity.

EXERCISE 5

Perform a goodness-of-fit test to determine whether the local results follow the distribution of the U.S. AP Examinee Population, based on ethnicity.

EXERCISE 6

The City of South Lake Tahoe, CA has an Asian population of 1419 people, out of a total population of 23,609 (Source: *U.S. Census Bureau, Census 2000*). Suppose that a survey of 1419 self-reported Asians in the Manhattan, NY, area yielded the data in the table below. Conduct a goodness of fit test to determine if the self-reported sub-groups of Asians in the Manhattan area fit that of the Lake Tahoe area.

Race	Lake Tahoe Frequency	Manhattan Frequency
Asian Indian	131	174
Chinese	118	557
Filipino	1045	518
Japanese	80	54
Korean	12	29
Vietnamese	9	21
Other	24	66

(7) - (8) UCLA conducted a survey of more than 263,000 college freshmen from 385 colleges in fall 2005. The results of student expected majors by gender were reported in *The Chronicle of Higher Education* (2/2/06). Suppose a survey of 5000 graduating females and 5000 graduating males was done as a follow-up in 2010 to determine what their actual major was. The survey results are in the table below.

Major	Women – expected major	Women – actual major	Men – expected major	Men – actual major
Arts & Humanities	14.0%	670	11.4%	600
Biological Sciences	8.4%	410	6.7%	330
Business	13.1%	685	22.7%	1130
Education	13.0%	650	5.8%	305
Engineering	2.6%	145	15.6%	800
Physical Sciences	2.6%	125	3.6%	175
Professional	18.9%	975	9.3%	460
Social Sciences	13.0%	605	7.6%	370
Technical	0.4%	15	1.8%	90
Other	5.8%	300	8.2%	400
Undecided	8.0%	420	6.6%	340

EXERCISE 7

Conduct a hypothesis test to determine if the actual major of graduating females fits the distribution of their expected majors.

EXERCISE 8

Conduct a hypothesis test to determine if the actual major of graduating males fits the distribution of their expected majors.

EXERCISE 9

A recent debate about where in the United States skiers believe the skiing is best prompted the following survey. Test to see if the best ski area is independent of the level of the skier.

Levels of Skiers in Various U.S. Ski Areas

U.S. Ski Area	beginner	intermediate	advanced
Tahoe	20	30	40
Utah	10	30	60
Colorado	10	40	50

EXERCISE 10

Car manufacturers are interested in whether there is a relationship between the size of car an individual drives and the number of people in the driver's family (that is, whether car size and family size are independent). To test this, suppose that 800 car owners were randomly surveyed with the following results. Conduct a test for independence.

Family Size's Effect on Car Size

Family Size	sub & compact	mid-size	full-size	van & truck
1	20	35	40	35
2	20	50	70	80
3 - 4	20	50	100	90
5+	20	30	70	70

EXERCISE 11

College students may be interested in whether or not their majors have any effect on starting salaries after graduation. Suppose that 300 recent graduates were surveyed as to their majors in college and their starting salaries after graduation. Below are the data. Conduct a test for independence.

Starting Salary by Major

Major	< $30,000	$30,000 - $39,999	$40,000 +
English	5	20	5
Engineering	10	30	60
Nursing	10	15	15
Business	10	20	30
Psychology	20	30	20

EXERCISE 12

Some travel agents claim that honeymoon hot spots vary according to age of the bride and groom. Suppose that 280 East Coast recent brides were interviewed as to where they spent their honeymoons. The information is given below. Conduct a test for independence.

Marrying Age of Bride by Location of Honeymoon

Location	20 - 29	30 - 39	40 - 49	50 and over
Niagara Falls	15	25	25	20
Poconos	15	25	25	10
Europe	10	25	15	5
Virgin Islands	20	25	15	5

EXERCISE 13

A manager of a sports club keeps information concerning the main sport in which members participate and their ages. To test whether there is a relationship between the age of a member and his or her choice of sport, 643 members of the sports club are randomly selected. Conduct a test for independence.

Sport	18 - 25	26 - 30	31 - 40	41 and over
racquetball	42	58	30	46
tennis	58	76	38	65
swimming	72	60	65	33

EXERCISE 14

A major food manufacturer is concerned that the sales for its skinny French fries have been decreasing. As a part of a feasibility study, the company conducts research into the types of fries sold across the country to determine if the type of fries sold is independent of the area of the country. The results of the study are below. Conduct a test for independence.

Type and Quantity of Fries Ordered Regionally

Type of Fries	Northeast	South	Central	West
skinny fries	70	50	20	25
curly fries	100	60	15	30
steak fries	20	40	10	10

EXERCISE 15

According to Dan Lenard, an independent insurance agent in the Buffalo, N.Y. area, the following is a breakdown of the amount of life insurance purchased by males in the following age groups. He is interested in whether the age of the male and the amount of life insurance purchased are independent events. Conduct a test for independence.

Life Insurance Purchase by Age of Consumer

Age of Males	none	$50,000 - $100,000	$100,001 - $150,000	$150,001 - $200,000	$200,000 +
20 - 29	40	15	40	0	5
30 - 39	35	5	20	20	10
40 - 49	20	0	30	0	30
50 +	40	30	15	15	10

EXERCISE 16

Suppose that 600 thirty–year–olds were surveyed to determine whether or not there is a relationship between the level of education an individual has and salary. Conduct a test for independence.

Salary by Level of Education Attained

Annual Salary	not a high school grad.	high school graduate	college graduate	masters or doctorate
< $30,000	15	25	10	5
$30,000 - $40,000	20	40	70	30
$40,000 - $50,000	10	20	40	55
$50,000 - $60,000	5	10	20	60
$60,000 +	0	5	10	150

EXERCISE 17

A plant manager is concerned her equipment may need recalibrating. It seems that the actual weight of the 15 oz. cereal boxes it fills has been fluctuating. The standard deviation should be at most 1/2 oz. In order to determine if the machine needs to be recalibrated, 84 randomly

selected boxes of cereal from the next day's production were weighed. The standard deviation of the 84 boxes was 0.54. Does the machine need to be recalibrated?

EXERCISE 18

Consumers may be interested in whether the cost of a particular calculator varies from store to store. Based on surveying 43 stores, which yielded a sample mean of $84 and a sample standard deviation of $12, test the claim that the standard deviation is greater than $15.

EXERCISE 19

Isabella, an accomplished *Bay to Breakers* runner, claims that the standard deviation for her time to run the 7 ½ mile race is at most 3 minutes. To test her claim, Rupinder looks up 5 of her race times. They are 55 minutes, 61 minutes, 58 minutes, 63 minutes, and 57 minutes.

EXERCISE 20

Airline companies are interested in the consistency of the number of babies on each flight, so that they have adequate safety equipment. They are also interested in the variation of the number of babies. Suppose that an airline executive believes the average number of babies on flights is 6 with a variance of 9 at most. The airline conducts a survey. The results of the 18 flights surveyed give a sample average of 6.4 with a sample standard deviation of 3.9. Conduct a hypothesis test of the airline executive's belief.

EXERCISE 21

According to the U.S. Bureau of the Census, United Nations, in 1994 the number of births per woman in China was 1.8 . This fertility rate has been attributed to the law passed in 1979 restricting births to one per woman. Suppose that a group of students studied whether or not the standard deviation of births per woman was greater than 0.75. They asked 50 women across China the number of births they had. Below are the results. Does the students' survey indicate that the standard deviation is greater than 0.75?

# of births	frequency
0	5
1	30
2	10
3	5

EXERCISE 22

According to an avid aquariest, the average number of fish in a 20–gallon tank is 10, with a standard deviation of 2. His friend, also an aquariest, does not believe that the standard deviation is 2. She counts the number of fish in 15 other 20–gallon tanks. Based on the results

that follow, do you think that the standard deviation is different from 2? Data: 11; 10; 9; 10; 10; 11; 11; 10; 12; 9; 7; 9; 11; 10; 11

EXERCISE 23

The manager of "Frenchies" is concerned that patrons are not consistently receiving the same amount of French fries with each order. The chef claims that the standard deviation for a 10–ounce order of fries is at most 1.5 oz., but the manager thinks that it may be higher. He randomly weighs 49 orders of fries, which yields: mean of 11 oz., standard deviation of 2 oz.

Try these true/false questions.

EXERCISE 24

As the degrees of freedom increase, the graph of the chi-square distribution looks more and more symmetrical.

EXERCISE 25

The standard deviation of the chi-square distribution is twice the mean.

EXERCISE 26

The mean and the median of the chi-square distribution are the same if df = 24.

EXERCISE 27

In a Goodness-of-Fit test, the expected values are the values we would expect if the null hypothesis were true.

EXERCISE 28

In general, if the observed values and expected values of a Goodness-of-Fit test are not close together, then the test statistic can get very large and on a graph will be way out in the right tail.

EXERCISE 29

The degrees of freedom for a Test for Independence are equal to the sample size minus 1.

EXERCISE 30

Use a Goodness-of-Fit test to determine if high school principals believe that students are absent equally during the week or not.

EXERCISE 31

The Test for Independence uses tables of observed and expected data values.

EXERCISE 32

The test to use when determining if the college or university a student chooses to attend is related to his/her socioeconomic status is a Test for Independence.

EXERCISE 33

The test to use to determine if a six-sided die is fair is a Goodness-of-Fit test.

EXERCISE 34

In a Test of Independence, the expected number is equal to the row total multiplied by the column total divided by the total surveyed.

EXERCISE 35

In a Goodness-of Fit test, if the p-value is 0.0113, in general, do not reject the null hypothesis.

EXERCISE 36

For a Chi-Square distribution with degrees of freedom of 17, the probability that a value is greater than 20 is 0.7258.

EXERCISE 37

If df = 2, the chi-square distribution has a shape that reminds us of the exponential.