Class	Observed (o)	Expected (e)	Deviation $(o - e)$	$(o - e)^2$	$\frac{(o - e)^2}{e}$
Females	85	75	10	100	100/75
Males	65	75	−10	100	100/75

THE CHI-SQUARE TEST

The chi-square statistic is a mathematical comparison of the data obtained with the data expected according to the rules of probability. A chi-square (χ^2) value represents the degree to which the data fit the hypothesis on which the experiment was based. If chi-square is small, the data fit the hypothesis well; a large chi-square value represents a large deviation and throws doubt on the hypothesis. The actual formula for chi-square need not concern us here. In all problems where we wish to compare the frequencies in an observed sample with an expected frequency distribution we calculate a value known as x^2, which for our purposes may be regarded as identical to χ^2.

Let us apply this to a concrete example.

Problem: A genetics student crosses two fruit flies and obtains 150 offspring. He or she sorts the offspring according to sex and finds 85 females and 65 males. Is this close enough to the expected 75 of each sex so that the observed sex distribution can be regarded as due to chance? (An alternative hypothesis might be that males die as embryos more often than females.)

Solution: The experimenter obtained 150 offspring of which 85 were female and 65 male where you would have expected 75 of each sex. To calculate the x^2 value for these results we can conveniently make a table like the one above. The observed and expected numbers of each sex of offspring are entered in the two left columns of the table. Now we subtract the expected from the observed offspring in each class to get the deviation for each class $(o - e)$. In the next column we square each deviation. Finally we divide the squared deviation for each class by the expected number in each class and add up all the results to obtain x^2, which in this case is 2.6 ($\Sigma = x^2 = 2.6$).

To determine whether this x^2 value, which we treat as chi-square, is significant, we look at a table of chi-square values (Table B-2). You will see that there is one item in the table that we have not yet mentioned and that we need before we can read off the probability: **degrees of freedom.** The degrees of freedom in a system are usually defined as the number of classes minus one. In this case, with two classes of offspring (male and female), there is only 1 degree of freedom. One way to look at degrees of freedom is to see that it represents the number of choices you can make in assigning an individual to a class. In this case, you have only one decision to make. If you decide that a particular fruit fly is not male, it must be female. If degrees of freedom are thought of as a number of decisions to be made, you can make one less decision than the number of classes, because after that number of decisions, the last class has been determined automatically. In our fruit fly problem we have an x^2 value of 2.6 and 1 degree of freedom. From the table we see that this value falls between 1.6 in the 0.2 (20%) column and 2.7 in the 0.10 column (10%). This means that we would obtain a chi-square value of 2.6 slightly more than 10% of the time if we performed numerous crosses and counted the sex of the offspring. We say that the chi-square value is **not significant.** The hypothesis that half the offspring of a fruit fly are male and half are female need not be questioned. If we obtained a chi-square value which fell in the 5% or 1% columns, we would suspect that another hypothesis would explain our results better than the one we used. A value in the 5% column would be **significant,** and a value in the 1% column would be **highly significant.**

TABLE B-2 CHI-SQUARE VALUES

Degrees of Freedom (n)	DEVIATION FROM HYPOTHESIS NOT SIGNIFICANT Probability Values (P)								DEVIATION SIGNIFICANT	DEVIATION HIGHLY SIGNIFICANT
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
1	0.004	0.016	0.06	0.15	0.46	1.1	1.6	2.7	3.8	6.6
2	0.1	0.2	0.4	0.7	1.4	2.4	3.2	4.6	6.0	9.2
3	0.4	0.6	1.0	1.4	2.4	3.7	4.6	6.3	7.8	11.3
	Chi-square value consistent with hypothesis								Not consistent	

(Modified from Table IV in R. A. Fisher and F. Yates, *Satistical Tables for Biological, Agricultural, and Medical Research* (6th ed., 1974). Published by Longman Group Ltd., London. Previously published by Oliver and Boyd, Ltd., Edinburgh. Reprinted by permission of the authors and publishers.)

GENETIC DATA SHEET

	1	2	3	4	Total
henotype Class					
umber of individuals ctual count)	a_1	a_2	a_3	a_4	
xpected number	e_1	e_2	e_3	e_4	

$$\chi^2 = \frac{(a_1 - e_1)^2}{e_1} + \frac{(a_2 - e_2)^2}{e_2} + \frac{(a_3 - e_3)^2}{e_3} + \frac{(a_4 - e_4)^2}{e_4}$$

χ^2 TABLE

d.f. \ P =	.95	.90	.80	.70	.50	.30	.10	.05	.01
1	.004	.016	.064	.148	.455	1.07	2.71	3.84	6.64
2	.103	.211	.446	.713	1.38	2.41	4.60	5.99	9.2
3	.352	.584	1.00	1.42	2.37	3.66	6.25	7.82	11.3

d.f. = degrees of freedom = number of classes - 1
P = probability

Name ______________________

Chi-Squared Practice

A student sets up a monohybrid cross to study the genetic inheritance of curly, flightless wings in fruit flys. Thus, a pure strain of curly winged flys is bred against a pure strain of wild, normal winged flys. **240** F_1 offspring were counted, consisting of **101** males and **139** females. Calculate a Chi-squared probability (X^2) to determine if the variation between the actual and expected gender ratios can be explained by random chance alone.

DIRECTIONS:

1. Fill in the Data Table below and follow the example formula to calculate the Chi-squared value. SHOW YOUR WORK !!!
2. Use the Chi-squared Probabililty Table to interpret your calculated X^2 value.
3. EXPLAIN your results: Is the variation in the gender data significant ???

GENETIC DATA SHEET

	1	2	3	4	Total
Phenotype Class					
Number of individuals (actual count)	a_1	a_2	a_3	a_4	
Expected number	e_1	e_2	e_3	e_4	

$$\chi^2 = \frac{(a_1 - e_1)^2}{e_1} + \frac{(a_2 - e_2)^2}{e_2} + \frac{(a_3 - e_3)^2}{e_3} + \frac{(a_4 - e_4)^2}{e_4}$$

χ^2 TABLE

d.f. \ P =	.95	.90	.80	.70	.50	.30	.10	.05	.01
1	.004	.016	.064	.148	.455	1.07	2.71	3.84	6.64
2	.103	.211	.446	.713	1.38	2.41	4.60	5.99	9.2
3	.352	.584	1.00	1.42	2.37	3.66	6.25	7.82	11.3

d.f. = degrees of freedom = number of classes - 1
P = probability
This Table was taken from Table III of Fisher: *Statistical Methods for Research Workers*, published by Oliver & Boyd, Edinburgh, and by permission of the author and publishers.

GENETIC DATA SHEET

	1	2	3	4	Total
ıenotype Class					
ımber of individuals ctual count)	a_1	a_2	a_3	a_4	
‹pected number	e_1	e_2	e_3	e_4	

$$\chi^2 = \frac{(a_1 - e_1)^2}{e_1} + \frac{(a_2 - e_2)^2}{e_2} + \frac{(a_3 - e_3)^2}{e_3} + \frac{(a_4 - e_4)^2}{e_4}$$

χ^2 TABLE

d.f. \ P =	.95	.90	.80	.70	.50	.30	.10	.05	.01
1	.004	.016	.064	.148	.455	1.07	2.71	3.84	6.64
2	.103	.211	.446	.713	1.38	2.41	4.60	5.99	9.2
3	.352	.584	1.00	1.42	2.37	3.66	6.25	7.82	11.3

d.f. = degrees of freedom = number of classes - 1
P = probability

NAME ____________________

Chi-Squared Practice #2

An Un"bean"leavably Fun Activity
or
I Bet You Haven't "Bean" There and Done This Before !!!

Goal: To learn about and practice calculating and interpreting a Chi-squared statistic by analyzing the actual and expected distribution of various mixed beans in a sample.

Practice Activity #1:

1. Carefully pour all of the red and white beans into the large empty ziplock bag and reseal.
2. After mixing and shaking the bag of beans for 1 full minute, collect a sample of beans (ONE large handful)
3. Count the numbers of red and white beans and record in the data table below.
4. Assume there are **400** total red beans and **200** total white beans in the original bag. Follow the example formula to calculate the Chi-squared value. SHOW YOUR WORK !!
5. Use the Chi-squared Table to interpret your calculated X^2 value.

   How many degrees of freedom are there? ________

   Is the variation in the bean distributions significant or can it be explained by some acceptable degree of random chance ?? EXPLAIN ??

GENETIC DATA SHEET

	1	2	3	4	Total
Phenotype Class	Red	White			
Number of individuals (actual count)	a_1 80	a_2 32	a_3	a_4	112
Expected number	e_1	e_2	e_3	e_4	112

$$\chi^2 = \frac{(a_1 - e_1)^2}{e_1} + \frac{(a_2 - e_2)^2}{e_2} + \frac{(a_3 - e_3)^2}{e_3} + \frac{(a_4 - e_4)^2}{e_4}$$

χ^2 TABLE

d.f. \ P =	.95	.90	.80	.70	.50	.30	.10	.05
1	.004	.016	.064	.148	.455	1.07	2.71	3.84
2	.103	.211	.446	.713	1.38	2.41	4.60	5.99
3	.352	.584	1.00	1.42	2.37	3.66	6.25	7.82

d.f. = degrees of freedom = number of classes - 1
P = probability
This Table was taken from Table III of Fisher: *Statistical Methods for Research Workers*, published by Oliver & Boyd, Edinburgh, and by permission of the author and publishers.

Practice Activity #2:

1. Return all the red and white beans to the large ziplock bag.
2. Carefully add all the pinto and black beans to the large ziplock bag with the other beans
3. After mixing and shaking the bag of beans for 1 full minute, collect a sample of beans (ONE large handful)
4. Count the numbers of red, white, pinto, and black beans and record in the data table below.
5. In addition to the **400** red beans and **200** white beans, assume there are also **200** pinto beans and **100** black beans in the original bag.
   Follow the example formula to calculate the Chi-squared value. SHOW YOUR WORK !!
5. Use the Chi-squared Table to interpret your calculated X^2 value.

   How many degrees of freedom are there? ________

   Is the variation in the bean distributions significant or can it be explained by some acceptable degree of random chance ?? EXPLAIN ??

GENETIC DATA SHEET

	1	2	3	4	Total
Phenotype Class	Red	White	Pinto	Black	
Number of individuals (actual count)	a_1 45	a_2 17	a_3 28	a_4 7	97
Expected number	e_1	e_2	e_3	e_4	97

$$\chi^2 = \frac{(a_1 - e_1)^2}{e_1} + \frac{(a_2 - e_2)^2}{e_2} + \frac{(a_3 - e_3)^2}{e_3} + \frac{(a_4 - e_4)^2}{e_4}$$

χ^2 TABLE

d.f. \ P =	.95	.90	.80	.70	.50	.30	.10	.05
1	.004	.016	.064	.148	.455	1.07	2.71	3.84
2	.103	.211	.446	.713	1.38	2.41	4.60	5.99
3	.352	.584	1.00	1.42	2.37	3.66	6.25	7.82

d.f. = degrees of freedom = number of classes - 1
P = probability
This Table was taken from Table III of Fisher: *Statistical Methods for Research Workers*, published by Oliver & Boyd, Edinburgh, and by permission of the author and publishers.