Probability and Statistics

Random Chance

A tossed penny can land either heads up or tails up. These are mutually exclusive events, i.e. if the coin lands heads up, it cannot also land tails up on the same toss. It is impossible to determine the forces operating on a coin as it falls to the table and lands heads up or tails up. Thus, these events are governed by random chance. If a coin is flipped many times, it will fall heads up about as many times as it falls heads down ($H = T = 0.5$). As there is no other possible way for the coin to fall the probability of **one** of these mutually exclusive events occurring is equal to the sum of their individual probabilities:

$$H + T = 1$$

Given the assumed 50:50 ratio, it is possible to predict the number of times that the coin will fall heads up or down and to determine the deviation of the observed values from the expected (O - E). Flip a penny 40 times and complete Table 1 at the bottom of this tutorial.

Independent Events Occurring Simultaneously -- The Binomial Expansion

When two or more independent events can happen, the probability of them happening simultaneously is equal to the product of their individual probabilities:

$$p(H + T) = p(H) \times p(T)$$

As with the single coin, the various combinations for two coins are mutually exclusive. Thus, if they land HH, they cannot land in any other combination on the same throw. The probability of one of the mutually exclusive events occurring is equal to the sum of their individual probabilities:

$$p(HH) + p(HT) + p(TH) + p(TT) = 1$$

In other words, if you toss two coins, there is a 100% probability that they will land either 2 heads, 2 tails, or heads + tails. In two tosses of a single coin, or one toss of two coins, with p = heads = 0.5 and q = tails = 0,5, the probabilities of the four possibilities given above are:

first coin	second coin	probabilities		
heads (p)	heads (p)	p^2	0.5 x 0.5	0.25
heads (p)	tails (q)	pq	0.5 x 0.5	0.25
tails (q)	heads(p)	pq	0.5 x 0.5	0.25
tails (q)	tails (q)	q^2	0.5 x 0.5	0.25
				1.0

In the probabilities above, we have made a distinction between whether the coins fall head then tail or tail then head. If we are interested only generally in the probability of getting a head and a tail as opposed to two heads or two tails, then there are two alternative ways of satisfying the requirement. Thus the probabilty of getting a head and a tail combination is the sum of their independent probabilities:

$$p(\text{head and tail}) = p(HT) + p(TH).$$

Thus we can express the probabilities as:

$$p^2 + 2pq + q^2 = 1$$

Also note that :

$$\sqrt{p2 + 2pq + q2} \quad = \quad p + q$$

$$(p + q)^2 = p^2 + 2pq + q^2$$

This is essentially an expression of the binomial expansion. The binomial expansion is represented as

$$(p + q)^n$$

where p and q represent the probabilities of alternative, mutually exclusive events (probability of heads vs tails, boy vs girl, wild-type vs mutant allele, etc), and n represents the size of the group or the number of trials. Thus:

$$1 \text{ coin} \quad = \quad (p + q)$$

$$2 \text{ coins} \quad = \quad (p + q)^2$$

$$3 \text{ coins} \quad = \quad (p + q)^3$$

$$4 \text{ coins} \quad = \quad (p + q)^4$$

$$5 \text{ coin s} \quad = \quad (p + q)^5$$

In the case of 3 coins, each (p + q) represents one of the coins and the probability of it landing heads vs tails. If we expand the 3 coin case, then the binomial expansion becomes:

$$p^3 + 3p^2q + 3\ pq^2 + q^3$$

Each expression in the expansion is the probability of one of the possible mutually exclusive outcome:

p^3	=	probability of three heads
$3p^2q$	=	probability of two heads and 1 tail
$3\ pq^2$	=	probability of 1 head and 2 tails
q^3	=	probability of 3 tails

The 3's in two of the expressions indicate how many possible ways one can obtain two heads and 1 tail.. As the number of independent events increases, the binomial expansion can become extremely complex. Individual outcomes can be calculated by the equation:

$$P = \frac{n!}{x!(n-X)!} p^x q^{(n-x)}$$

where n = the total number in the group and x = the number one one class. p = the probability of x occurring and q = probability of the other event occurring. For example, if 10 babies were born in a hospital one evening and you wanted to calculate the probability of 6 girls and 4 boys occurring, the n = 10 babies, x = 6 girls, and (n - x) = 4 boys. p = probability of a girl (0.5) and q = probability of a boy (0.5)

Toss n coins 60 times and record the results the tables in the lab report where:

# coins	**Table**
2	2
3	3
4	4

Use the binomial expansion to determine the expected values.

The Chi-Square Test

X^2 is a test to determine the "goodness of fit" of experimentally derived data to a theoretical expected ratio. In the previous exercises we measured the deviations (observed - expected). Now we must make a judgement as to whether the deviations from expected values are small enough to occur by chance alone, or are large enough to disprove our working hypothesis. From the measured deviations, we can calculate the value X^2, which is a measure of the total deviation in the entire experiment. The formula for X^2 is:

$$X^2 = \Sigma \frac{(\text{observed - expected})^2}{\text{expected}}$$

X^2 is usually calculaed in table form. Once the X^2 value for an experiment has been calculated, it must be evaluated by comparison with a table of X^2 values. In order to use the X^2 test properly, one must understand exactly what is being evaluated.

The X^2 test is always phrased in terms of the "**Null Hypothesis**" (H_O), which states: "**there is no significant difference between observed and expected results**." It makes no difference what the original hypothesis was. The X^2 test is used only to determine whether or not the observed results are statistically the same as the expected. Acceptance or rejection of H_O then gives us a handle by which to evaluate the theory being tested by the experiment. For example, suppose two genes are believed to be unlinked. If this is indeed the case, a dihybrid cross should result in a 9 3 3 1 ratio among the F2 progeny. The cross is performed and the expected results are calculated based on that ratio, based on the assumption of non-linkage. The actual results are then campared to the expected by the X^2 test. If the test leads us to reject the H_O, that the observed and expected results are different, then we must conclude that the hypothesis that enabled us calculate the expected values is, in fact, incorrect. If, on the other hand, if we accept the H_O, then we can conclude that the results are **consistent** with our hypothesis that the two genes are unlinked. It doesn't necessarily **prove** our hypothesis, but it gives us greater confidence that we are on the right track.

To evaluate the H_O, the experimental X^2 value must be compared with a **critical X^2 value** from the table below. To find the appropriate critical X^2 value, we must know the **degrees of freedom** for the experiment, and the **confidence level** or **probability**.

Degrees of Freedom: In each experiment, there is a fixed quantity of individuals divided up among the several classes of data. If there are 1600 individuals divided over only two classes, "tall" and "short", for example, once the number of tall individuals are counted, the number of short individuals is fixed. If there are four classes, AB = 900, Ab = 300, and aB = 300, then the class ab must be 100. It makes no difference which classes we consider to be free and which is fixed. The important thing is that among four classes of data, three are free and one is fixed. Degrees of freedom is easily calculated because:

degrees of freedom = # classes of data - 1.

In the dihybrid cross example, 9 3 3 1 means that there are 4 classes of data, so degrees of freedom would be

4 - 1 = 3

Thus the critical X^2 value would be found in the row for 3 degrees of freedom.

Probability: If one repeats an experiment numerous times, the results will not be identical, but rather will vary around a mean. Likewise, it would be too much to expect that the expected results of a new experiment be exactly identical to the expected results. The X^2 value is a measure of the overall discrepancy in the experiment among all the classes of data. We must then determine how close the results have to be to, or how large the X^2 value can be and still accept the H_O, and at what point we should decide to reject the H_O. This is the probability or confidence level. We would expect that for valid hypotheses the discrepancy between observed and expected would be small. If the discrepancy is small enough, than we accept it. Statisticians usually use the 0.05 probability level to make this discrimination. At the 0.05 probability level, we would say that only 5% of the time would we see large discrepancies between essentially identical observed and expected results. 95% of the time, the discrepancy is smaller. Thus at the 0.05 probability level, there is a 5% chance that we would discard a valid hypothesis. We thus choose the critical X^2 value from the table under the **P = 0.05** column. For the case of 3 degrees of freedom, the X^2 value is 7,815. Thus in an experiment producing a X^2 value less than or equal to 7.815, we would conclude that the discrepancy is low enough to accept the H_o. Greater than 7.815, we would reject the H_o.

It is possible to use other probability levels. For any degree of freedom, the critical X^2 value decreases with increasing probability. Why settle for 5% probability when we can have 95%? or 99%. If you observe a X^2 value at the 95% level, there is very little difference between observed and expected results. But if you **accept** X^2 values only at 95%, and there is actually more variability in the experiment than you would like to accept and you are likely to reject the H_o when you should accept it. At P = 95%, 95% of the time there will be as much or more discrepancy between observed and expected, yet the H_o should be accepted. By increasing the probability level, you increase the chance of discarding correct hypotheses. By decreasing the probability level, you increase the chance of accepting incorrect hypotheses.

There is yet another way to look at a X^2 table. You could take the experimentally determined X^2 value and find the closest critical X^2 value in the table and then determine the closest probability or confidence level. Suppose, for example, that you caluculated a X^2 value of 2.53 at 3 degrees of freedom. 2.53 falls between 2.366 and 4.642, and falls, therefore, between the 0.5 and 0.2 probability levels. You could conclude, then, that you would expect to see that much variation 20 - 50 % of the time. Because 2.53 is less than 7.815, the critical X^2 value, this is an acceptable result.

Table of Chi Square Values

Degrees of Freedom	P=.99	P=0.95	P=0.8	P=0.5	P=0.2	P=0.05	P=0.01
1	0.000157	0.00393	0.0642	0.455	1.642	3.841	6.635
2	0.020	0.103	0.446	1.386	3.219	5.991	9.210
3	0.115	0.352	1.005	2.366	4.642	7.815	11.345
4	0.297	0.711	1.649	3.357	5.989	9.488	13.277
5	0.554	1.145	2.343	4.351	7.289	11.071	15.086
6	0.872	1.635	3.070	5.348	8.558	12.592	16.812
7	1.239	2.167	3.822	6.346	9.803	14.067	18.475
8	1.646	2.733	4.594	7.344	11.030	15.507	20.090
9	2.088	3.325	5.380	8.343	12.242	16.919	21.666
10	2.558	3.940	6.179	9.342	13.442	18.307	23.209
15	5.229	7.261	10.307	14.339	19.311	24.996	30.578
20	8.260	10.851	14.578	19.337	25.038	31.410	37.566
25	11.524	14.611	18.940	24.337	30.675	37.652	44.314
30	14.953	18.493	23.364	29.336	36.250	43.773	50.892

Familiarize yourself with the X^2 test by the following exercises:

1. Go back to tables 1-4 and complete the X^2 sections.

Table 1

Results	Observed	Expected	Obs - Exp	(Obs - Exp)2	÷ Exp
Heads					
Tails					
Totals	40	40		X^2=	

Null Hypothesis = ______________________________

Degrees of Freedom = __________

Probability Level = __________

Critical X2 = __________

☐ Accept H_o

☐ Reject H_o

Table 2

Results	Observed	Expected	Obs - Exp	(Obs - Exp)2	÷ Exp
2 Heads HH					
Heads + Tails HT					
2 Tails TT					
Totals	60	60		X^2=	

Null Hypothesis = ______________________________

Degrees of Freedom = __________

Probability Level = __________

Critical X2 = __________

☐ Accept H_o

☐ Reject H_o

Table 3

Results	Observed	Expected	Obs - Exp	(Obs - Exp)2	÷ Exp
3 Heads HHH					
2 Heads + 1 Tail HHT					
1 Head + 2 Tails HTT					
3 Tails TTT					
Totals	60	60		X^2=	

Null Hypothesis = ______________________________

Degrees of Freedom = __________

Probability Level = __________ ☐ Accept H_o

Critical X2 = __________ ☐ Reject H_o

Table 4

Results	Observed	Expected	Obs - Exp	(Obs - Exp)2	÷ Exp
4 Heads HHHH					
3 Heads + 1 Tail HHHT					
2 Heads + 2 Tails HHTT					
1 Head + 3 Tails HTTT					
4 Tails TTTT					
Totals	60	60		X^2=	

Null Hypothesis = ______________________________

Degrees of Freedom = __________

Probability Level = __________ ☐ Accept H_o

Critical X2 = __________ ☐ Reject H_o

PROBLEMS

1. You have some trick coins that land heads 60% of the time and tails 40%. Use the binomal expansion to calculate the probabilities of HH, HT, and TT. If you flip 2 coins 175 times, what are the numbers of each?

	HH	HT	TT
probability			
175 times			
230 times			

2. A trick penny lands heads 35% of the time and a trick nickel lands heads 55% of the time. Use the binomial expansion to find the probabilities of the various combinations of heads and tails. What would be the numbers if you flipped them 275 times? 465 times?

<table>
<tr><td></td><td>P
H</td><td>N
H</td><td>P
H</td><td>N
T</td><td>P
T</td><td>N
H</td><td>P
T</td><td>N
H</td></tr>
<tr><td>probability</td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>275 times</td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
<tr><td>465 times</td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td><td colspan="2"></td></tr>
</table>

The remaining questions have to do with the frequencies of marbles in the following Jars

Jar 1	**Jar 2**	**Jar 3**	**Jar 4**
80 blue 20 red	70 green 30 yellow	50 black 50white	50 orange 50 brown

3. Use the binomial expansion to calculate the probabilities and combinations by drawing from the jars as indicated:

Jars 3 + 4	combinations								
	probabilities								
Jars 1 + 3	combinations								
	probabilities								
Jars 1, 2, 3	combinations								
	probabilities								

4. Calculate the probabilities of each event below:

a. pulling 10 blue and 10 red from jar 1
b. pulling 25 green and 18 yellow from jar 2
c. pulling 6 black and 15 white from jar 3
d. pulling 10 orange and 10 brown from jar 4

5. You draw from jars 2 and 4 200 times and get the results below. Use X^2 to determine whether or not this result is expected.

Results	**Observed**	**Expected**	**Obs - Exp**	**$(Obs - Exp)^2$**	**÷ Exp**
green + orange	65				
green + brown	75				
yellow + orange	28				
yellow + brown	32				
Totals	200	200		X^2=	

Null Hypothesis =

Degrees of Freedom =

Probability Level =

Critical X2 =

☐ Accept H_o

☐ Reject H_o

6. You draw from jars 1 and 2 450 times and get the results below. Use X^2 to determine whether or not this result is expected.

Results	**Observed**	**Expected**	**Obs - Exp**	**$(Obs - Exp)^2$**	**÷ Exp**
blue +green	230				
red + yellow	38				
blue + yellow	115				
red + green	67				
Totals	450	450		X^2=	

Null Hypothesis =

Degrees of Freedom =

Probability Level =

Critical X2 =

☐ Accept H_o

☐ Reject H_o

7. You draw from jars 1 and 3 600 times and get the results below. Use X^2 to determine whether or not this result is expected.

Results	**Observed**	**Expected**	**Obs - Exp**	**$(Obs - Exp)^2$**	**÷ Exp**
blue + black	65				
red + black	75				
blue + white	28				
red + white	32				
Totals	600	600		X^2=	

Null Hypothesis = ______________________________

Degrees of Freedom = ______________

Probability Level = ______________

Critical X2 = ______________

☐ Accept H_o

☐ Reject H_o