Homework Problem Set 15: Goodness-of-Fit Tests and Tests of Independence

Instructions: The following exercises will test your understanding of goodness-of-fit tests and tests of independence using the X^2 test statistic. For each test conduct four steps (state hypotheses, compute the test statistic, compute the p-value, and make a decision). For goodness-of-fit tests in this homework assignment the degrees of freedom will always be one less than the number of possible outcomes, whereas for tests of independence the degrees of freedom will always be $(r-1)(c-1)$ where r and c represent the number of rows and columns in the table of counts. Use a significance level of $\alpha = 0.05$ for all tests.

Law of Independent Assortment

Another of Mendel's laws is the *Law of Independent Assortment* which states that alleles for *different* traits assort themselves independently in the process of being passed from parent to offspring. Roughly speaking, alleles from different traits occur together only by chance. Statistically this means that two traits are *independent*. We know now that this is not a universal law as genes that are close to each other on a chromosome will tend to be passed together. This is called *linkage* and is what was investigated in the study by William Bateson, Edith Saunders, and Reginald Punnett discussed in class. However it does (at least approximately) apply to some traits such as those examined in pea plants by Mendel. In one experiment he crossed plants that he suspected of each carrying two alleles for the *shape* of the peas — round (R) or wrinkled (r) — and two alleles for the color of the peas as in the example discussed in class — yellow (Y) or green (g). Thus each parent will pass one of four possible combinations of alleles: RY, rY, Ry, or ry. Each offspring will therefore receive one of sixteen possible combinations of alleles as shown in the table below. It is also assumed that the alleles for round and yellow exhibit complete dominance so the offspring will be round if it receives at least one R allele, otherwise it will be wrinkled, and it will be yellow if it receives at least one Y allele, otherwise it will be green.

	Trait	
Outcome	Shape	Color
RRYY	round	yellow
rRYY	round	yellow
RRyY	round	yellow
rRyY	round	yellow
rRYY	round	yellow
rrYY	wrinkled	yellow
rRyY	round	yellow
rryY	wrinkled	yellow
RRyY	round	yellow
rRyY	round	yellow
RRyy	round	green
rRyy	round	green
rRyY	round	yellow
rryY	wrinkled	yellow
rRyy	round	green
rryy	wrinkled	green

By the Law of Segregation each of the sixteen possible outcomes shown above is equally likely with probability

1/16. But assume like Mendel we can only observe the shape and color of the peas of an offspring. These outcomes and their respectively probabilities are given in the table below.

		Count	
Traits	Probability	Observed	Expected
round and yellow	9/16	315	
wrinkled and yellow	3/16	101	
round and green	3/16	108	
wrinkled and green	1/16	32	

The probabilities come from the first table (e.g., 9 of 16 outcomes result in an offspring with peas that are round and yellow).[1] The table also shows the observed counts from one of Mendel's experiments to determine if the Law of Independent Assortment applies to the traits of shape and color in pea plants. The expected counts are not given, so you will need to compute them for what follows. Conduct a *goodness-of-fit test* to determine whether or not the probabilities given above fit the data.

Detection of Non-Randomness

Goodness-of-fit tests have been used to detect situations where a device that is supposed to be random is not working correctly (or is being intentionally manipulated). Examples include mechanical and computer devices for research, gambling, and encryption. Here is a simple example of one such application. Timmy Fiebelkorn's friends suspect him of cheating when playing the tabletop role-playing game *Wizards & Wyverns*. Part of the game involves creating characters with three attributes: strength, dexterity, and intelligence. The level of each attribute is on an integer scale from 0 (low) to 4 (high). The level determined by flipping a coin four times, counting the number of times the coin comes up heads. The probability distribution of the number of times a fair coin comes up heads (h) is a *binomial* distribution, where the probabilities can be computed using the equation

$$P(h) = \frac{4!}{h!(4-h)!} 0.5^h 0.5^{4-h}.$$

So the probability of each possible number of heads and thus each attribute level can be easily computed. These are given in the table below.

Level	Probability
0	0.0625
1	0.2500
2	0.3750
3	0.2500
4	0.0625

Timmy's friends examined 100 of the *Wizards & Wyverns* characters he created, each with three attributes, for a total of 300 attributes. Of these 300 attributes, 15 were at Level 0, 59 were at Level 1, 112 were at Level 2, 81 were at Level 3, and 33 were at Level 4. These are the *observed counts* based on a sample size of 300. Conduct a *goodness-of-fit test* to determine if there is evidence that Timmy is cheating by using the null hypothesis that the probabilities given above are correct (note that these probabilities were derived under the assumption that Timmy was *not* cheating).

Milena Plays Pounce

Recall the example from lecture of my daughter playing the game *Pounce*. This was one of the examples I used when I introduced the concept of a significance test. Suppose she plays 50 trials of Pounce. On each

[1] These probabilities can also be derived by assuming that the dominant and recessive traits have probabilities of 3/4 and 1/4, respectively, and multiplying probabilities together assuming independence. For example, $P(\text{wrinkled and yellow}) = P(\text{wrinkled})P(\text{yellow}) = 3/4 \times 1/4 = 3/16$.

trial she is presented with three words, one of the three words is spoken, and then she needs to choose the correct word. Suppose she is correct 25 times, and thus incorrect 25 times. If she guesses the correct word her probability of being correct is 1/3, and her probability of being incorrect is 2/3. If p is the probability of a correct response then we can conduct a significance test of the hypotheses $H_0: p = 1/3$ versus $H_a: p > 1/3$. The test statistic is

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}},$$

where $\hat{p} = 25/50$, $p = 1/3$, and $n = 50$. This yields a test statistic of $z = 2.5$ and a p-value of approximately 0.006. Another approach would be to use a goodness-of-fit test. Let $p_c = 1/3$ and $p_w = 2/3$ be the probabilities of a correct and a wrong response from Milena *assuming she is guessing*. Conduct a *goodness-of-fit test* assuming these probabilities for the null hypothesis. You should find that the value of the X^2 test statistic is equal to the square of the z test statistic (i.e., $z^2 = 6.25$), and that the p-value is (approximately, due to rounding) twice as large as that given above (i.e., 0.012, this is because the goodness-of-fit test is implicitly two-sided whereas the p-value given above was for a one-sided test).

Sex Role Stereotypes in Personnel Decisions

Rosen and Jerdee (1974) conducted a randomized experiment to investigate sex role stereotypes in personnel decisions.[2] The subjects were 48 male bank supervisors who were attending a management institute. Each manager was given a hypothetical personnel file and asked if the applicant described in the file should be promoted. The contents of the files were *identical* except in half of the files the applicant was listed as a man, and in the other half the applicant was listed as a woman. For those managers who were given the file of a male applicant, 21 recommended promotion, and for those managers who were given the file of a female applicant, 14 recommended promotion. The data from this study can be summarized as a table of observed counts.

	Promotion		
Applicant	yes	no	Total
male	21	3	24
female	14	10	24
Total	35	13	48

In an earlier homework you conducted a significance test of the null hypotheses $H_0: p_m - p_f = 0$ where p_m and p_f denote the probability that it will be decided to promote a male or female applicant, respectively. The test statistic was

$$z = \frac{\hat{p}_m - \hat{p}_f}{\sqrt{\hat{p}(1-\hat{p})(1/n_m + 1/n_f)}},$$

where $\hat{p}_m = 21/24$, $\hat{p}_f = 14/24$, $\hat{p} = 35/48$, $n_m = 24$, and $n_f = 24$. The value of the test statistic was $z \approx 2.274$ and the p-value (assuming a two-sided test) is then approximately 0.023. By definition if $p_m = p_f$ (as assumed by the null hypothesis above) then the gender of the applicant and the decision are *independent*. Thus a test of this null hypothesis can also be done using the X^2 test statistic for a test of independence. Conduct a *test of independence* for applicant gender and the promotion decision using the data given above. You should find that your X^2 test statistic is equal to the square of the z test statistic given earlier (approximately, due to rounding error) and that the p-value is the same.[3]

[2] Rosen, B. & Jerdee, J. (1974). Influence of sex role stereotypes on personnel decisions. *Journal of Applied Psychology*, *59*, 9–14.

[3] This relationship between the two approaches only applies two tests of independence with counts in a table with two rows and two columns. If there is more than two rows or more than two columns (as in the following problems) then only the X^2 test statistic can be used.

Incubation Temperature and Turtle Sex

The table below shows the observed counts from a study investigating the effect of incubation temperature (in degrees Celsius) on the sex of turtles.[4]

	Sex		
Temp	male	female	Total
27.2	2	25	27
27.7	17	7	24
28.3	26	4	30
28.4	19	8	27
29.9	27	1	28
Total	91	45	136

The next table shows the approximate *expected counts* under the assumption that incubation temperature and sex are *independent* (the expected counts have been rounded to the hundredths place).

	Sex		
Temp	male	female	Total
27.2	18.07	8.93	27
27.7	16.06	7.94	24
28.3	20.07	9.93	30
28.4	18.07	8.93	27
29.9	18.74	9.26	28
Total	91	45	136

Conduct a *test of independence* of incubation temperature and sex. Note that the expected counts are given, but you might find it useful to check that you know how to compute them based on the observed counts.

Snoring as a Risk Factor for Coronary Heart Disease

A research study published in the *British Medical Journal* examined the relationship between snoring and health.[5] The table below shows the observed counts when a sample of 2484 individuals were classified by whether or not they had coronary heart disease (CHD) and frequency of snoring.

	Snoring Frequency			
CHD	Low	Medium	High	Total
yes	23	35	51	109
no	1356	603	416	2375
Total	1379	638	467	2484

Conduct a *test of independence* of snoring frequency and coronary heart disease.

Law of Independent Assortment (Solution)

The null hypothesis that the probabilities given in the table (i.e., the 9:3:3:1 ratio) are correct, and the alternative hypothesis is that they are incorrect. The value of the test statistic is $X^2 \approx 0.47$. The p-value (based on 3 degrees of freedom) is approximately 0.925. Thus we would not reject the null hypothesis. The

[4] I have not been able to track down the original source of these data. I assume that they are real but I am not sure.

[5] Norton, P. G. & Dunn, E. V. (1985). Snoring as a risk factor for disease: An epidemiological survey. *British Medical Journal*, *291*, 630–632.

Law of Independent Assortment appears to fit these traits. In modern terms there does not appear to be linkage between shape and color.

Detection of Non-Randomness (Solution)

The null hypothesis is that the probabilities of each attribute level are as given in the first table, while the alternative hypothesis is that these probabilities are incorrect. The value of the test statistic is $X^2 \approx 15.476$, and the p-value (using a degrees of freedom of 4) is approximately 0.004. This would lead us to reject the null hypothesis and conclude that the probabilities assuming that Timmy is not cheating do not fit the observed counts. There is some evidence that he is (at least sometimes) cheating when determining the attributes for his characters.

Milena Plays Pounce (Solution)

The null hypothesis is that the probabilities $p_c = 1/3$ and $p_w = 2/3$ are correct. The value of the test statistic is $X^2 = 6.25$, which yields a p-value (using one degree of freedom) of approximately 0.012. This would lead us to reject the null hypothesis, and thus conclude that Milena was not guessing.

Sex Role Stereotypes in Personnel Decisions (Solution)

The null hypothesis is that the gender of the applicant and the promotion decision are independent, while the alternative hypothesis is that they are not independent. The value of the test statistic is $X^2 \approx 5.169$, which yields a p-value (based on one degree of freedom) of approximately 0.023. As noted in the problem these are related to the results obtained using the z test statistic. The decision is to reject the null hypothesis and conclude that the gender of the applicant and the promotion decision are not independent (i.e., they are associated).

Incubation Temperature and Turtle Sex (Solution)

The null hypothesis is that incubation temperature and turtle sex are independent, and the alternative hypothesis is that they are not independent. The value of the test statistic using the given expected values is $X^2 \approx 59.823$ (if the expected counts are computed more precisely the test statistic comes out to about $X^2 \approx 59.799$). The p-value (based on 4 degrees of freedom) is very small so the decision is to reject the null hypothesis that incubation temperature and turtle sex are not independent (i.e., there is an association between these two variables).

Snoring as a Risk Factor for Coronary Heart Disease (Solution)

The null hypothesis is that coronary heart disease (yes or no) and snoring frequency (low, medium, or high) are independent. The alternative hypothesis is that they are not independent. The value of the test statistic is $X^2 \approx 73.66$, and the p-value (using a degrees of freedom of 2) is very small. This leads us to reject the null hypothesis that coronary heart disease and snoring frequency are independent and conclude that there is evidence of an association.