KEY

WLHS / AP Bio / Monson

Name____________________
Date____________________ Per________

PACKET: Statistics Practice Problems – 2 – Chi-Squared Test!

This packet provides you with practice working with the Chi-Squared (X^2) Test for different examples / data sets. For extra background or review on these topics, Mr. Anderson (Bozeman Science) has a podcast that will help you (link below).

LINK to Mr. Anderson's Tutorial:
Chi-Squared Test: **http://www.bozemanscience.com/chi-squared-test**

PART 1: Introduction / Review of the Chi-Squared Test

The X^2 value is calculated from the following formula:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

1) Identify what each of the following parts of the formula mean by explaining it in words:

A) X^2 = this is the chi-squared value we are calculating

B) Σ = Sum; add together

C) O_i = Observed data value

D) E_i = expected data value

2) We use a X^2 test to decide whether to ***reject*** or ***fail to reject*** our **null hypothesis**. Explain what a null hypothesis is: → states that there is no statistically sig. difference between the observed + expected values. (i.e. everything is normal + nothing is out of the ordinary... e.g. in a coin flip exp., the coin is fair)

3) Once we have calculated the X^2 value, we then compare it to a critical value which comes from looking at a X^2 table under the correct p-value column. We always use a p-value of 0.05. The critical value we use depends on the **degrees of freedom** in the experiment.

A) How do we determine the **degrees of freedom** for the experiment? it is the # of possible outcomes minus 1

B) If you have an experiment with 4 potential outcomes, how many degrees of freedom are there? 3
(4 - 1 = 3)

C) Using a X^2 table (next page), what is the critical value for an experiment with 1 degree of freedom? ~~2.71~~ 3.84 With four degrees of freedom? ~~7.78~~ 9.49

Table 5.3. Chi-square value

Degrees of Freedom	Probability						Non-significant		Significant	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	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21

PART 2: Chi-Squared Test Practice

To perform a X^2 test, we need the following pieces of information:

1 – the null hypothesis
2 – your observed data
3 – the expected valued for your data
4 – the degrees of freedom in your data
5 – the critical value to compare with your X2 value

Example / Data Set #1: A Genetic Experiment

Our first data set considers a pretend gene in a hypothetical bird population. You are studying a rare bird gene called the "Turducken" gene. All birds in the population can be either TT, Tt, or tt for this gene. ***In a homozygous recessive bird (tt genotype), the bird turns into a Turducken!*** The Punnett square for the Turducken gene between two heterozygous parents is shown here.

	T	t
T	TT	Tt
t	Tt	tt

In this population of birds that you are studying, two heterozygous regular bird parents (Tt) have 200 offspring and 97 of them are Turduckens. You plan to use a X^2 test to determine if you should consider this out of the ordinary.

1) How many outcomes are possible in this situation? (how many types of offspring are possible) 2

2) How many degrees of freedom do you have? 1

3) What is the number of Turduckens in the offspring? This is your observed value of Turduckens,

$O_{Turduckens}$: 97

4) What is the number of regular birds in the offspring? This is your observed value of regular birds,

$O_{RegularBirds}$: 103

5) What is the null hypothesis (the "nothing is out of the ordinary" hypothesis) in this situation? the F1 generation offspring #s are not statistically diff. than what we expected based on the Punnett Square

6) Based on the Punnett Square for 2 heterozygous parents, what **proportion (% or fraction)** of offspring would you expect to be Turduckens in this new generation? tt = 1/4 or 25%

7) Use your answer from the previous question to calculate **how many** Turduckens you would expect in this new generation of 200 since we know both parents have genotype Tt. This is your **expected value** for Turduckens, $E_{Turduckens}$. 50 Turduckens expected

8) Based on the Punnett Square for 2 heterozygous parents, what **proportion (% of fraction)** of offspring would you expect to be regular birds in this new generation? 3/4 or 75%. **How many** birds in this new generation do you expect to be regular birds? This is your expected value for the regular birds $E_{RegularBirds}$. 150 regular birds expected

9) Calculate the X^2 value using your previous answers according to the X^2 formula:

$$\begin{aligned} x^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(O_{Turducken} - E_{Turducken})^2}{E_{Turducken}} - \frac{(O_{RegularBirds} - E_{RegularBirds})^2}{E_{RegularBirds}} \\ &= \frac{(97-50)^2}{50} + \frac{(103-150)^2}{150} = 44.18 + 14.7 = 58.9 \end{aligned}$$

10) Using the degrees of freedom you calculated in #2, find the critical value you will compare to a X^2 table and write it here. 3.84 Now, compare the X^2 value to this critical value.

the X^2 value is 58.9. This is much higher than the critical value of ~~2.71~~ 3.84

11) Based on your comparison in the previous question, will you reject or fail to reject your null hypothesis? What do you conclude about the number of Turduckens in this new generation?

We reject the null hypothesis; the # of observed Turduckens is statistically different from the # expected... some other factor is at work!

Example / Data Set #2: Mendel's Pea Experiment

In our 2nd example, we consider Mendel's Pea Experiment where he crossed two heterozygote parents YyRr x YyRr with each other. In this experiment, the Y and y genes determine whether a pea is yellow (yellow is dominant, Y) or green (only yy genotypes are green). The R and r genes are those that make a pea smooth (smooth is dominant, R) or wrinkled (only rr genotypes are wrinkled). The Punnett Square for this crossing is shown here.

	YR	Yr	yR	yr
YR	YYRR	YYRr	YyRR	YyRr
Yr	YYRr	YYrr	YyRr	Yyrr
yR	YyRR	YyRr	yyRR	yyRr
yr	YyRr	Yyrr	yyRr	yyrr

In his experiment, Mendel OBSERVED the following number of offspring of each phenotype:

Yellow & Smooth	Green & Smooth	Yellow & Wrinkled	Green & Wrinkled
315	108	101	32

= 556

Again, we can use a X^2 test to help us decide if this is in line with what we expect.

1) How many outcomes (phenotypes) are possible? 4 How many degrees of freedom? 3

2) What is the null hypothesis (the "nothing is out of the ordinary" hypothesis) in this situation?

The offspring phenotypes observed are not statistically different from the phenotypes expected in this cross.

3) Based on the proportions in the Punnett Square, calculate the expected number of each phenotype expected in the offspring.

9/16, 3/16, 3/16, 1/16 ×556

$E_{YellowSmooth}$ = 313

$E_{GreenSmooth}$ = 104

$E_{YellowWrinkled}$ = 104

$E_{GreenWrinkled}$ = 35

4) Complete the following table, and then calculate the X^2 value.

Phenotype:	Yellow, Smooth	Green, Smooth	Yellow, Wrinkled	Green, Wrinkled
$O_i - E_i$	315 - 313 = 2	108 - 104 = 4	101 - 104 = -3	32 - 35 = -3
$(O_i - E_i)^2$	4	16	9	9
$\frac{(O_i - E_i)^2}{E_i}$	4/313 = 0.013 +	16/104 = 0.15 +	9/104 = 0.087 +	9/35 = 0.26 =

X^2 value = 0.51

5) Based on your X^2 value, do you reject or fail to reject the null hypothesis? Explain.

For 3 degrees of freedom, the critical value is ~~6.25~~ 7.82. The X^2 value for the pea plant cross results is 0.51, much lower than the critical value.

Therefore, we fail to reject the null hypothesis.

In other words, the observed results are not statistically different from the expected phenotypic proportions!

Example / Data Set #3: Computer time and headache incidence

Your school has installed new lights in the computer labs. Some have expressed concern that the lights may be associated with an increase in headaches. So, as an experiment, your biology class has offered to look at some of the data and offer an analysis. You have the following information: ← use for expected proportions!

Student Type:	No Headache:	Headache:	Totals:
Class in new computer lab	257	37	294
No class in new computer lab	541	45	586
Totals:	798	92	880

1) What is the null hypothesis in this situation? The headache results for students w/ classes in the computer lab are not statistically different than the headache results for students with no classes in the computer lab.

2) What are the possible outcomes in this situation? ***(HINT: think of the "No class in computer lab" group as the control group, and the "Class in new computer lab" as the experimental group).***

Possible outcomes for students with classes in the new computer lab are: headache or no headache (2)

3) Based on your answer to the previous question, the number of degrees of freedom is: 1

4) Calculate the **proportion (% / fraction)** of students reporting a headache and the proportion of students not reporting a headache from **the control group** (i.e. the group of students who had no class in the new computer lab).

Proportion (%) of students without a headache = 92.3% (.923)

Proportion (%) of students with a headache = 7.7% (.077)

5) Use your results from the previous question to calculate the **expected** values of students with and without headaches in the experimental group (i.e. they DO have a class in the new computer lab).

~~(.923)(294)~~ (.077)(294)

$E_{Headache}$ = 23

(.923)(294)

$E_{NoHeadache}$ = 271

6) Complete the following table.

Condition	Observed (O)	Expected (E)	O − E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
Headache	37	23	14	196	8.5
No Headache	257	271	−14	196	.72

sum = 9.22

7) Based on your degrees of freedom your **critical value** in this case is: ~~2.71~~ 3.84

8) What is your X^2 value? 9.22

9) Do you reject or fail to reject the null hypothesis? Explain.

Reject. The X^2 value is well above the "highly significant" (.01) value for 1 degree of freedom. The headache results for students w/a class in the computer lab are statistically significantly different from the "control" group. (no class in comp. lab)