Chi-Square Test: Notes

<u>When to use a Chi-square test</u>

Researchers often need to decide if the results they observe in an experiment are close enough to predicted theoretical results so that the tested hypothesis can be supported or rejected. For example, do a series of coin flips match what you'd expect to get by chance, or is their evidence the coin is unfair? Does the number of women interviewed for a job position match the proportion of women in the applicant pool, or is there evidence of bias? Does the number of white-eyed fruit fly offspring match the number expected if the white-eyed trait is recessive, or are white-eyes inherited in some other way?

Chi-square tests come in two types:

Chi-square test for independence

Chi-square goodness of fit test: used to test if the observed data match theoretical or expected results. **We will focus on this test**. *Example*: Do the phenotypes you observe in a fruit fly cross match the pattern expected if the trait is dominant?

A Chi-square test is used when:

1. Your response variable is ___________________________
2. Your response falls into different ______________________
3. You have a hypothesis for the responses you _____________
4. You want to know if the difference between the responses you __________ and the responses you ______________ is significant or not.

If the chi-square test shows your data is not significantly different from what you expected, you **support** the hypothesis. If your data is significantly different you **reject** the hypothesis as an explanation for your data. Remember, no statistical test can ever **prove** a hypothesis, only fail to reject it.

<u>Steps to the Chi-square test</u>

1. Identify the hypothesis you are testing
2. Calculate the expected number of responses in each category if this hypothesis explains your data. Place these numbers in the "expected" column of your chi-square table (see below).
3. Place your data in the "observed" column of your chi-square table (see below).
4. Calculate your chi-square value

   x^2 = Chi-square value   **Σ = Sum of**

   $$x^2 = \Sigma \frac{(\text{Observed Value} - \text{Expected Value})^2}{\text{Expected Value}}$$

5. Determine your **degrees of freedom**. The meaning of your chi-square value depends on the degrees of freedom.
   The **degrees of freedom is one less than the possible total number of different expected results.** When using a chi-square table this is the # of rows - 1.

Example Chi-Square Table

(helps you calculate X^2) Your X^2 value

	Observed	Expected	Obs-Exp	$(Obs-Exp)^2$	$\frac{(Obs-Exp)^2}{Exp}$
Category 1					
Category 2					
…..					
				X^2 total	
				Degrees of Freedom	

6. Use a chi-square probability table to determine your **probability (p) value**, or the likelihood your data fits your hypothesis. If your p-value is **> 0.05**, your data is not significantly different from your expected values, if your p-value is **<0.05** your data is significantly different from your expected values.

Probability (p) Values for Chi-square Test

df	0.99	0.95	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.01
1	0.0001	0.003	0.015	0.064	0.148	0.275	0.455	0.708	1.07	1.64	2.71	3.84	6.63
2	0.020	0.103	0.211	0.446	0.713	1.02	1.39	1.83	2.41	3.22	4.61	5.99	9.21
3	0.115	0.352	0.584	1.00	1.42	1.87	2.37	2.95	3.67	4.64	6.25	7.81	11.3
4	0.297	0.711	1.06	1.65	2.19	2.75	3.36	4.04	4.88	5.99	7.78	9.49	13.3
5	0.554	1.15	1.61	2.34	3.00	3.66	4.35	5.13	6.06	7.29	9.24	11.1	15.1
6	0.872	1.64	2.20	3.07	3.83	4.57	5.35	6.21	7.23	8.56	10.6	12.6	16.8
7	1.24	2.17	2.83	3.82	4.67	5.49	6.35	7.28	8.38	9.80	12.0	14.1	18.5
8	1.65	2.73	3.49	4.59	5.53	6.42	7.34	8.35	9.52	11.0	13.4	15.5	20.1

Tips: Find you **df** value in the left column. Look across the row until you find where your X^2 value falls. Your corresponding probability range can be found in the top row.

Example interpretation: If the p-value for the problem fell between 0.6 and 0.5, there is a 50-60% chance the difference between the observed and expected values are due random chance alone, therefore there is no significant difference between obs and exp values and there is no evidence to reject the hypothesis.

Example Problem #1

A search committee posted a position for a biology professor. They received 220 applications, 25% of which came from women. The committee came up with a short list 25 candidates, 5 women and 20 men, for the job. You want to know if there is evidence for the search committee being biased against women.

1. Identify the hypothesis you are testing: ____
   ____

2. Calculate the expected number of candidates in each category based on the hypothesis.

   Women:

   Men:

	Observed	Expected	Obs-Exp	$(Obs\text{-}Exp)^2$	$\frac{(Obs\text{-}Exp)^2}{Exp}$
Women					
Men					
Total				X^2 total	
				Degree of Freedom	

3. What is the probability range for your chi-squared value? ________________________

4. Based on this probability, do we support or reject the hypothesis above? ____________

5. Write a statement that interprets this statistical result in the context of the problem above.
   ____
   ____
   ____
   ____

Example Problem #2

A research hypothesizes ebony colored bodies and vestigial wings are recessive traits in *Drosophila*. Under this hypothesis she expected to observe 9 Wild Wild : 3 Wild Vestigial : 3 Ebony Wild : 1 Ebony Vestigial offspring if she mates two heterozygous wild wild flies.

She observes 53 Wild Wild : 16 Wild Vestigial : 25 Ebony Wild : 8 Ebony Vestigial. Do these results support her inheritance hypothesis?

6. Identify the hypothesis you are testing: ____
____

7. Calculate the expected number of phenotypes in each category based on the hypothesis.

   Wild Wild:

   Wild Vestigial:

   Ebony Wild:

   Ebony Vestigial:

	Observed	Expected	Obs-Exp	$(Obs\text{-}Exp)^2$	$\frac{(Obs\text{-}Exp)^2}{Exp}$
Wild Wild	53				
Wild Vestigial	16				
Ebony Wild	25				
Ebony Vestigial	8				
Total				X^2 total	
				Degree of Freedom	

8. What is the probability range for your chi-squared value? ____________________________

9. Based on this probability, do we support or reject the hypothesis above? ______________

10. Write a statement that interprets this statistical result in the context of the problem above.
____
____
____
____