11.2 Chi-Square Tests for Homogeneity

Read 696–702

How is section 11.2 different than section 11.1?

What are the two explanations for the differences in the distributions of wine purchases?

How do you state hypotheses for a test of homogeneity?

What is the problem of multiple comparisons? What strategy should we use to deal with it?

How do you calculate the expected counts for a test that compares the distribution of a categorical variable in multiple groups or populations?

What is the formula for the chi-square test statistic? Is it on the formula sheet? What does it measure?

Alternate Example: *Saint-John's-wort and depression*
An article in the *Journal of the American Medical Association* (vol. 287, no. 14, April 10, 2002) reports the results of a study designed to see if the herb Saint-John's-wort is effective in treating moderately severe cases of depression. The study involved 338 subjects who were being treated for major depression. The subjects were randomly assigned to receive one of three treatments—Saint-John's-wort, Zoloft (a prescription drug), or a placebo—for an eight-week period. The table below summarizes the results of the experiment.

	Saint-John's-wort	Zoloft	Placebo	Total
Full response	27	27	37	91
Partial response	16	26	13	55
No response	70	56	66	192
Total	113	109	116	338

(a) Calculate the conditional distribution (in proportions) of the type of response for each treatment.
(b) Make a segmented bar graph to compare the conditional distributions in part (a).
(c) Compare the distributions of response for each treatment.
(d) State the hypotheses we are interested in testing.
(e) Calculate the expected counts and the chi-square statistic.

Read 703–706

How do you calculate the degrees of freedom for a chi-square test for homogeneity?

What are the conditions for a chi-square test for homogeneity?

Alternate Example: *Saint-John's-wort and depression*
(a) Verify that the conditions for this test are satisfied.
(b) Calculate the *P*-value for this test.
(c) Interpret the *P*-value in context.
(d) What is your conclusion?

> *(a) Random: The treatments were randomly assigned.*
> *Large sample size: The expected counts (30.4, 29.3, 31.2, 18.4, 17.7, 18.9, 64.2, 61.9, 65.9) are all at least 5.*
> *Independent: Knowing the response of one patient should not provide any additional information about the response of any other patient.*
>
> *(b) df = (3 – 1)(3 – 1) = 4, P-value =* χ^2`cdf(8.72, 1000, 4)` *= 0.0685.*
> *(c) Assuming that the treatments are equally effective, the probability of observing a difference in the distributions of responses among the three treatment groups as large or larger than the one in the study is about 0.07.*
>
> *(d) Since the P-value is greater than* α *= 0.05, we fail to reject the null hypothesis. We do not have convincing evidence that there is a difference in the distributions of responses for patients with moderately severe cases of depression when taking Saint-John's-wort, Zoloft, or a placebo.*

Can you use your calculators to do a chi-square test of homogeneity?

HW page 694 (19–22), page 724 (27–33 odd)

11.2 Chi-square HOP, continued…

Read 706–709

Has modern technology changed the distribution of birthdays? With more babies being delivered by planned c-section, a statistics class hypothesized that the day-of-the-week distribution for births would be different for people born after 1993 compared to people born before 1980. To investigate, they selected a random sample of people from each both age categories and recorded the day of the week on which they were born. The results are shown in the table. Is there convincing evidence that the distribution of birth days has changed? *(from DeAnna McDonald at UHS)*

	Before 1980	**After 1993**	
Sunday	12	9	**21**
Monday	12	11	**23**
Tuesday	14	11	**25**
Wednesday	10	10	**20**
Thursday	6	17	**23**
Friday	9	9	**18**
Saturday	10	6	**16**
	73	**73**	**146**

How do you conduct a follow-up analysis for a test of homogeneity? When should you do this?

Alternate Example: *Ibuprofen or acetaminophen?*
In a study reported by the *Annals of Emergency Medicine* (March 2009), researchers conducted a randomized, double-blind clinical trial to compare the effects of ibuprofen and acetaminophen plus codeine as a pain reliever for children recovering from arm fractures. There were many response variables recorded, including the presence of any adverse effect, such as nausea, dizziness, and drowsiness. Here are the results:

	Ibuprofen	Acetaminophen plus codeine	Total
Adverse effects	36	57	93
No adverse effects	86	55	141
Total	122	112	234

(a) Calculate the chi-square statistic and *P*-value.

(b) Show that the results of a two-sample *z* test for a difference in proportions are equivalent.

When should you use a chi-square test and when should you use a two-sample *z* test?

- *The chi-square test is always two-sided. That is, it only tests for a difference in the two proportions. If you want to test whether one proportion is larger than the other, use the two-sample z test.*
- *If you want to estimate the difference between two proportions, use a two-sample z interval. There are no confidence intervals that correspond to chi-square tests.*
- *If you are comparing more than two treatments or the response variable has more than two categories, you must use a chi-square test.*
- *You can also use a chi-square goodness-of-fit test in place of a one-sample z test for a proportion if the alternative hypothesis is two-sided. The chi-square test will use two categories (success and failure) and have df = 2 – 1 = 1.*

HW #35 page 725 (35, 37, 39, 43)

11.2 Chi-Square Test for Association/Independence

Read pages 713–718

What does it mean if two variables have an association? What does it mean if two variables are independent?

How is a test of association/independence different than a test of homogeneity?

How do you state hypotheses for a test of association/independence?

How do you calculate expected counts for a test of association/independence?

Remember not to round!

What are the conditions for a test of association/independence?

Alternate Example: *Finger length*
Is your index finger longer than your ring finger? Does this depend on your gender? A random sample of 460 high school students in the U.S. was selected and asked to record if their pointer finger was longer than, shorter than, or the same length as their ring finger on their left hand. The gender of each student was also reported. The data are summarized in the table below.

	Female	Male	Total
Index finger longer	85	73	158
Same	42	44	86
Ring finger longer	100	116	216
Total	227	233	460

(a) Make a graph to investigate the relationship between gender and relative finger length. Describe what you see.
(b) Do the data provide convincing evidence at the $\alpha = 0.05$ level of an association between gender and relative finger length for high school students in the U.S.?
(c) If your conclusion in part (b) was in error, which type of error did you commit? Explain.

$\chi^2 = 2.065, df = 2, P = 0.356$

Don't accept H_0*!*

Read 718–721

An article in the *Arizona Daily Star* (April 9, 2009) included the following table. Suppose that you decide to analyze these data using a chi-square test. However, without any additional information about how the data were collected, it isn't possible to know which chi-square test is appropriate.

Age (years):	18–24	25–34	35–44	45–54	55–64	65+	Total
Use online social networks:	137	126	61	38	15	9	386
Do not use online social networks:	46	95	143	160	130	124	698
Total:	183	221	204	198	145	133	1084

(a) Explain why it is OK to use age as a categorical variable rather than a quantitative variable.
(b) Explain how you know that a goodness-of-fit test is not appropriate for analyzing these data.
(c) Describe how these data could have been collected so that a test for homogeneity is appropriate.
(d) Describe how these data could have been collected so that a test for association/independence is appropriate.

HW page 728 (49, 51, 53–60)