Section 11.2a – Inference for Two-Way Tables – Chi Test of <u>Homogeneity</u>

Page 697

- Introduction: Read first two paragraphs on page 697

- **Example:** Does Background Music Influence What Customers Buy? – Comparing Conditional Distributions – Page 697
  - Market researchers suspect that background music may affect the mood and buying behavior of customers. One study in a European restaurant compared three randomly assigned treatments: no music, French accordion music, and Italian string music. Under each condition, the researchers recorded the number of customers who ordered French, Italian, and other entrees. Here is a table that summarizes the data:

	Music			
Entrees	**None**	**French**	**Italian**	**Total**
French	30	39	30	99
Italian	11	1	19	31
Other	43	35	35	113
Total	84	75	84	243

  - Calculate the conditional Distribution (in proportions) for the type of wine sold.
    - No Music:

    - French Music:

    - Italian Music:

  - Make an appropriate graph for comparing the conditional distributions.

  - What conclusions can we make from the graphs?

- Read: First two paragraphs on page 699

- Chi-Square Goodness of Fit vs. Chi-Square Homogeneity

- **Expected Counts and the Chi-Square Statistic** (2-way table data) – Page 701

OBSERVED Counts				
	Music			
Entrees	**None**	**French**	**Italian**	**Total**
French	30	39	30	99
Italian	11	1	19	31
Other	43	35	35	113
Total	84	75	84	243

EXPECTED Counts				
	Music			
Entrees	**None**	**French**	**Italian**	**Total**
French				99
Italian				31
Other				113
Total	84	75	84	243

- **Finding Expected Counts:**
  - When H_0 is true, the expected count in any cell of a two-way table is:
    - $Expected\ Count =$

- **When to use Chi-Square Test for Homogeneity:**
  - Examining data which come from independent SRSs from DIFFERENT POPULATIONS to see if the distribution of a categorical variable is the same for several populations, or to see if all the proportions are basically equal.

- **Claim:** (In Context)
  - H_0: There is no difference in the distribution of _____ at this store when
    _____.
  - H_A: There is a difference in the distribution of _____ at this store when
    _____.

- **Conditions/State the Name of the Test:**
  - Random sample from each population?
  - Large Sample? All expected counts are at least 5. Need to show the table!
  - Independent: $Population \geq 10n$

- **Calculations:**
  - $\chi^2 = \Sigma \frac{(observed - expected)^2}{expected}$
  - $Degree\ of\ Freedom = (r-1)(c-1)$

- **Conclusions:**
  - My p-value of _____ is smaller/larger than an alpha level of _____.
  - I reject/fail to reject the null.
  - Conclusion in context (make sure to include something about the distribution)
    - Reject the Null – the proportions are significantly different, or there is a significant difference in the distributions.
    - Fail to Reject the Null – the proportions are not significantly different – saying they are equal is the kiss of death. ☹

- **Finish Music Example**:

Conditions:

Calculations:

Conclusion:

- **CALCULATOR HELP!**
  - Enter the observed in Matrix A and the calculator will find and store the expected in Matrix B!!! $2nd\ X^{-1} \rightarrow Edit$ Dimensions: $Row\ x\ Column$
  - But you are expected to know the math the calculator is using to find the expected values. You will be tested on this!!
  - $Stat \rightarrow Tests$: $\chi^2 - Test$

- **Reading Computer Output** – Page 710

```
Chi-Square Test: None, French, Italian

Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

           None      French     Italian      Total
1            30          39          30          99
          34.22       30.56       34.22
          0.521       2.334       0.521

2            11           1          19          31
          10.72        9.57       10.72
          0.008       7.672       6.404

3            43          35          35         113
          39.06       34.88       39.06
          0.397       0.000       0.422

Total        84          75          84         243

Chi-Sq = 18.279, DF = 4, P-Value = 0.001
```

Figure 11.7 Minitab output for the two-way table in the restaurant study. The output gives the observed counts, the expected counts, and the individual components of the chi-square statistic.

- **Example:** Are Cell-Only Telephone Users Different? – The chi-square test for homogeneity
  - o Random digits dialing telephone surveys used to exclude cell phone numbers. If the opinions of people who have only cell phones differ from those of people who have landline service, the poll results may not represent the entire adult population. The Pew Research Center interviewed separate random samples of cell-only and land-line telephone users who were less than 30 years old. Here's what the Pew survey found about how these people describe their political party affiliation:

OBSERVED Count			
	Cell-Only Sample	**Landline Sample**	**Total**
Democrat or lean Democrat	49	47	
Refuse to Lean Either Way	15	27	
Republican or Lean Republican	32	30	
Total			

- Does this data provide convincing evidence at the $\alpha = 0.05$ level that the distribution of party affiliation differs in the under-30 cell-only and landline user populations?

Claims:

Conditions:

Calculations:

Conclusion:

A 2-prop z test is equivalent to a Chi-Square test of homogeneity if you have 2 populations you are sampling from. You will investigate this idea on the last problem in your homework!

Section 11.2a – Inference for Two-Way Tables
Chi-Square Test for Independence
Page 713

- **When to Use Chi-Square Test of Independence/Association:**
    - Examine data (two-way table) which come from an SRS taken from a SINGLE POPULATION were each individual is classified according to two variables.
    - You are testing whether there is an association between the two variables or whether they are independent.
    - Watch for the words association or independent!!

- **Claim:**
    - H_0: There is no relationship/association between two variables (the 2 measured variables are independent)
        - Meaning: If all the proportions are basically equal then there is no association, the variables do not influence the proportions.
    - H_A: There is a relationship/association between two variables (the 2 measured variables are no independent.
        - Meaning: If the proportions are significantly different or not equal then there is an association, the variables are having influence on the proportions.

- **Conditions/State the Name of the Test:**
    - Random sample from each population?
    - Large Sample? All expected counts are at least 5. Need to show the table!
    - Independent: $Population \geq 10n$

- **Calculations:**
    - $\chi^2 = \sum \frac{(observed - expected)^2}{expected}$
    - $Degree\ of\ Freedom = (r-1)(c-1)$

- **Conclusions:**
    - My p-value of _____ is smaller/larger than an alpha level of _____.
    - I reject/fail to reject the null.
    - Conclusion in context (make sure to include something about the <u>distribution</u>)
        - Reject the Null – there is an association between _____ and _____.
        - Fail to Reject the Null – there is no association between _____ and _____.

- **Example:** Angry People and Heart Disease
  - o A study followed a random sample of 8474 people with normal blood pressure for about four years. All the individuals were free of heart disease at the beginning of the study. Each person took the Spielberger Trait Anger Scale test, which measures how prone a person is to sudden anger. Researchers also recorded whether each individual developed coronary heart disease (CHD). This includes people who had heart attacks and those who needed medical treatment for heart disease.

	Low Anger	Moderate Anger	High Anger	Total
CHD	53	110	27	190
No CHD	3057	4621	606	8284
Total	3110	4731	633	8474

  - o Is this an observational study or an experiment? Justify your answer.

  - o Does the data provide convincing evidence of an association between anger level and heart disease in the population of interest?

Claims:

Conditions: (remember to state test and show observed/expected table)

Calculations: (remember to show degree of freedom)

Conclusion:

Summary/Review: How Do You Tell the Difference?

- Instead of focusing on the question asked, it's much easier to look at how the data were produced.

    o **Chi-Square for Homogeneity**
        - If the data come from two or more independent random samples or treatment groups in a randomized experiment.
            - Hypothesis Mentions: Equal or unequal proportions.

    o **Chi-Square for Independence**
        - If the data come from a single random sample, with the individuals classified according to two categorical variables.
            - Hypothesis Mentions: Association, Relationship or Independence.

    o **Chi-Square Goodness of Fit**
        - Not a two-way table – One Row of Data
            - Therefore – No Matrix!
            - Hypothesis Mentions: Comparing to some known distribution or claimed distribution.

- **Worksheet** – Chi-Square: How Do I Tell the Difference?