## Inference for two-way tables

- Row by column table
- Explanatory variable
- Response variable
- cell
- **Expected Counts** = $\frac{(row\ total)(column\ total)}{table\ total}$

- $\chi^2$ **Statistic**- measure how far the observed counts in a two way table are from the expected counts. $\chi^2 = \sum \frac{(O-E)^2}{E}$
- **d.f.** = (rows -1)(columns -1)
- **Assumption**: No more than 20% are less than 5 and all must be greater than or equal to one. In a 2 x 2 table none of the cells should be less than 5.

**Test for Homogeneity**: Comparing two or more populations or treatments.
**$H_o$**: The true proportions are the same for all populations
**$H_a$**: The true proportions are not the same for all populations

**Uses of $\chi^2$ test for Independence**: Investigate association between two categorical variables in a **single** population.
**$H_o$**: Relationships are independent OR no relationship between two categorical variables (Independent) OR There is no association between the variables.
**$H_a$**: the two variables are not independent OR the two variables are dependent OR there is an association between the two variables.

Example (Test for Independence):
An AP Statistics student has compiled the following data from a random sample of 90 seniors at her high school.

**Political Philosophy vs. Gender**

| | Liberal | Moderate | Conservative | Totals |
|---|---|---|---|---|
| Males | 16 | 10 | 6 | 32 |
| Females | 20 | 22 | 16 | 58 |
| Totals | 36 | 32 | 22 | 90 |

Do these data provide evidence of an **association** between political philosophy and gender at the 5% level?

1. Find expected values. row total x column total/table total
   Example Row 1 column 1: Liberal males 32 x 36 / 90 = 12.8

| Categories | Observed | Expected | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|
| Liberal male | 16 | 12.8 | .800 |
| Moderate Male | 10 | 11.4 | .172 |
| Cons. male | 6 | 7.8 | .415 |
| Liberal female | 20 | 23.2 | .441 |
| Moderate female | 22 | 20.6 | .095 |
| Cons. female | 16 | 14.2 | .228 |
| Total | 90 | 90 | 2.151 |

2. **Assumptions**: no expected cell counts < 5

3. **$H_o$**: There is **no association** between political philosophy and gender among the seniors at this high school or political philosophy and gender are **independent** in this population.

4. **$H_a$**: There is **an association** between political philosophy and gender among the seniors at this high school or political philosophy and gender are not independent (meaning they are dependent) in this population.

5. **Test Statistic**: $\chi^2 = 2.15$ d.f. = (2-1)(3-1) = 2
6. $\chi^2 = 2.15 < \chi^2_{.05} = 5.99$ p-value = .3398
7. **Conclusion**: Fail to reject null hypothesis and conclude there appears to be no relationship between political affiliation and gender among the seniors at this high school. (no doubt of independence between the variables)

Example 2 (Test for homogeneity): Students use many criteria when selecting courses in college. "Teacher who is very easy grader" is often one criterion. Three teachers are scheduled to teach statistics next semester at a local college. A sample of previous grade distributions for these three teachers is shown below.

| | **Professor** | | |
|---|---|---|---|
| **Grades** | **#1** | **#2** | **#3** |
| A | 12 | 11 | 27 |
| B | 16 | 29 | 25 |
| C | 35 | 30 | 15 |
| Other | 27 | 40 | 23 |

At the 0.01 significant level, is there sufficient evidence to conclude "The distribution of grades is not the same for all three professors?"

## How to do this on your calculator:

Press the blue 2nd button and $x^{-1}$ (MATRIX) button. Your screen should show NAMES MATH EDIT. Now go to EDIT. Press enter. This is MATRIX [A]—your observed values. The first table above is a 2 x 3 MATRIX (2 rows and 3 columns). So on your calculator type in 2 enter 3 enter. Now enter the data in the matrix as it appears in the table: 16 enter 10 enter 6 enter 20 enter 22 enter 16 enter. Your calculator's matrix should look identical to the above table. Now go to STAT TEST Scroll down to C which is $\chi^2$-TEST. Press enter. Now go to calculate. Your chi-square statistic and your p-value. Go to MATRIX one more time. Go to EDIT and go to [B]. These are your expected counts. You can put those in ( ) next to each of the observed values so you do not have to do it by hand.

Now do the remaining questions on the Class Exercises Sheet handed out yesterday.