Class 20: Chi-Square for Larger Tables (Text: Sections 9.1, 9.2. Section 9.3 is optional)

The Chi-Square Test

Null hypothesis: There is no association between variables.
Alternative hypothesis: There **is** an association between the variables, but we do not specify what.

We have a two-way table of **observed** counts and create a tables of **expected** counts:

$$\text{Expected cell count} = \frac{\text{Row total} \times \text{Column total}}{n}$$

Then we calculate the **Chi-Square** statistic with $df = (\#\text{ rows} - 1)(\#\text{ cols} - 1)$

$$\chi^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

Notes on Using the Chi-Square Test

- Large values of χ^2 give significance.
- Small values of χ^2 give no significance.

- Rejecting null hypothesis does not tell us which way the interaction happens.
- Since χ^2 is always positive; there is only an upper tail. We never multiply the probabilities by 2.
- **Conditions**: We can use the Chi-Square test when
  - For tables larger than 2×2: Average of expected counts ≥ 5; all expected counts ≥ 1
  - For 2×2 tables: All expected counts ≥ 5

The example in the previous class could be done either chi-square test of by the z test. Now we do an example with a table that is larger than 2×2, which can only be done by chi-square.

Ex: Is the following drug effective?
In an experiment, 600 people were divided randomly into three groups and given a placebo, a single dose, or a double dose of a drug. The **observed** results were:

	Placebo	Single	Double	
Improve	50	57	78	185
Didn't	180	133	102	415
	230	190	180	600

Step 1:

Step 2:
We calculate the table of **expected** counts:

	Placebo	Single	Double	
Improve				
Didn't				

Aside: On the test, you may have to write some of the terms of χ^2 and you will be expected to know its degrees of freedom and find its P-value, but you will not have to calculate the whole χ^2.
Step 3:

Step 4:

Ex: Since the Chi-Squared test does not tell us how the interaction happens if we reject the null hypothesis, how do we know if the drug is effective? (Maybe it makes people worse?)

Ex: Is a double dose of the drug significantly better than single dose? (Use Chi Squared.)

Ex: As more women work, the effect of childcare on infants has been investigated. One study[1] looked at the relationship between infant-mother attachment and the time spent in childcare, with the following results:

	Low (0-3 hours/week)	*Moderate (4-19 hours/week)*	*High (20-54 hours/week)*
Anxious	24	35	5
Secure	11	10	8

(a) Does the data provide evidence of a difference in attachment patterns with the amount of time spent in childcare? Give the hypotheses, the test statistic, and the *p*-value with your conclusion. Use both the 5% and the 1% levels.

(b) Combine the moderate and high childcare groups together, and test again. What is your conclusion?

[1] By J. Jacobson and D. Willie, reported in *Statistical Record of Women Worldwide* (Galen Research, 1991) and *Introduction to Practice of Statistics*, 13-th ed, Mendenhall, Beaver and Beaver.

Chi-Square Goodness of Fit Test (Optional: Section 9.3):
Example from UA Biology Major, 2006: Distribution of Creosote Bushes
The question the student answered was whether the number of creosote bushes varied with setting.

*The set-up here is slightly different from the previous example as the data has only **one** observed row. This alters how we calculate the degrees of freedom, but nothing else. **This is called a Chi-Square Goodness of Fit Test.***

Areas of the same size were marked out in four regions. To do this, the pace length was found by walking a known 100 ft and dividing this 100 ft by the number of paces needed to complete the 100 ft.

$$\text{Pace length} = \frac{100 \text{ ft}}{45.5 \text{ paces}} = 2.20 \text{ ft/pace}$$

This pace length was then used to measure an area of 95 ft x 95 ft in each region. The numbers of creosote bushes were counted in each one.

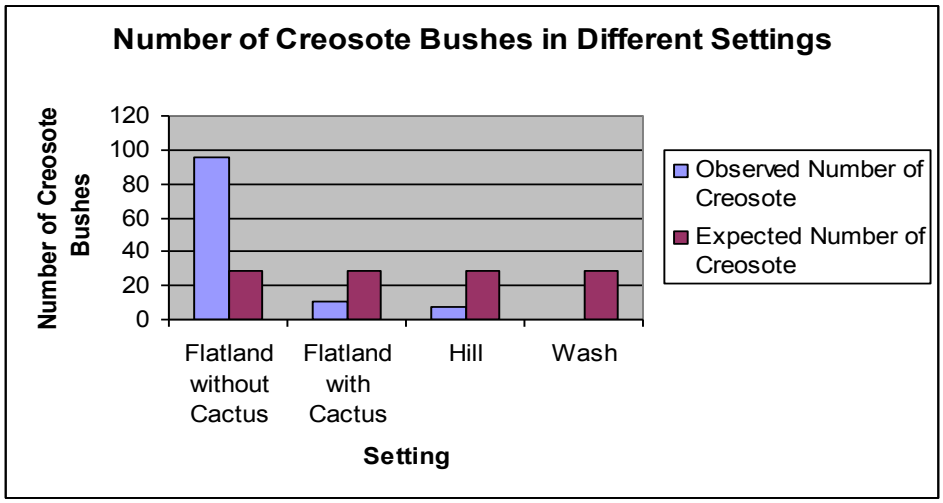
Step 1:

Step 2:

Figure 1: Number of Creosote Bushes Observed and Expected in Each Setting

	Flatland	Flatland without Cactus	Hill	Wash
Number of Creosote Bushes Observed				
Number of Creosote Bushes Expected				

Figure 2: Number of Observed and Expected Creosote Bushes in Different Settings



For a Goodness of Fit Test:

$$\text{Degrees of Freedom} = (\text{Number of Variables} - 1)$$

Step 3:

Step 4: