Chi-Square Test for Homogeneity - Key

1. In a study of the television viewing habits of children, a developmental psychologist selects a random sample of 300 first graders - 100 boys and 200 girls. Each child is asked which of the following TV programs they like best: The Lone Ranger, Sesame Street, or The Simpsons. The results are displayed in the table below.

	The Lone Ranger	**Sesame Street**	**The Simpsons**
Boys	50	30	20
Girls	50	80	70

Does this data suggest a difference in viewing preference between boys and girls?

H_0: The distribution of TV viewing preferences for boys and girls is the same.

H_a: The distribution of TV viewing preferences for boys and girls is not the same.

Chi-Square Test for Homogeneity

A random sample of 300 first graders

10% Condition: $10n_b = 10(100) = 1000 < N_b$ and $10n_g = 10(200) = 2000 < N_g$

Large Counts: The expected counts are all above 5.

33.3 36.7 30

66.7 73.3 60

X^2-Test $\Rightarrow$ $X^2 = 19.3182$ and a p-value of 0.0000638425

The p-value of 0.0000638425 is less than the alpha of 0.05, so we reject the null hypothesis in favor of the alternative hypothesis. There is sufficient evidence to suggest that the distribution of TV viewing preferences for boys and girls is not the same.

2. In 2006, the NCAA published a report called "Substance Use: NCAA Study of Substance Use of College Student-Athletes." We use data from this report to investigate the following question: Does steroid use by student athletes differ for the three NCAA divisions?

   The data comes from a random selection of teams in each NCAA division. The sampling plan was somewhat complex, but we can view the data as though it came from a random sample of athletes in each division. The surveys are anonymous to encourage truthful responses. The data is given in the table below.

	Steroids	**No Steroids**
Division I	103	8440
Division II	52	4289
Division III	65	6428

Does the data provide evidence that the use of steroids is different between the three divisions?

H_0: The distribution of steroid use in the three college divisions is the same.
H_a: The distribution of steroid use in the three college division is not the same.

Chi-Square Test for Homogeneity

A random sample of NCAA division teams.

10% Condition: $10n_I = 10\ (8543) = 85430 < N_I$ and $10n_{II} = 10\ (4341) = 43410 < N_{II}$ and $10n_{III} = 10\ (6493) = 64930 < N_{III}$

Large Counts: The expected counts (97, 8446, 49, 4292, 74, 6419) are all greater than 5.

97 8446

49 4292

74 6419

X^2-Test $\Rightarrow X^2 = 1.5704$ and a p-value = 0.4560

The p-value of 0.4560 is greater than alpha of 0.05, so we fail to reject the null hypothesis in favor of the alternative hypothesis. There is insufficient evidence to suggest that the distribution of steroid use in the three college divisions is not the same.

3. The NCAA survey discussed in Problem 2 includes this question: "When, if ever, did you start using anabolic steroids?" The response options are: have never used, before junior high, junior high, high school, freshman year of college, after freshman year of college. We focused on those who admitted use of steroids and compared the distribution of their responses for the years 1997, 2001, and 2005. The data is given in the table below.

	1997	**2001**	**2005**
Junior High or Before	16	15	69
High School	15	42	156
During Freshman Year of College	12	17	65
After Freshman Year of College	18	26	107

Does this data provide evidence that when athletes started using steroids changed between surveys?

H_0: The distributions of when college athletes began steroid use across the three surveys are the same.
H_a: The distributions of when college athletes began steroid use across the three surveys are not the same.

Chi-Square Test for Homogeneity

A random sample of NCAA teams

10% Condition: $10n_{1997} = 10(61) = 610 < N_{1997}$ and $10n_{2001} = 10(100) = 1000 < N_{2001}$ and $10n_{2005} = 10(397) = 3970 < N_{2005}$

Large Counts: The expected counts are all greater than 5.

11 23 10

17 18 38

17 27 71

151 67 107

X^2-Test $\Rightarrow$ $X^2 = 6.8751$ and a p-value = 0.3326

The p-value of 0.3326 is greater than alpha of 0.05, so we fail to reject the null hypothesis in favor of the alternative hypothesis. There is insufficient evidence to suggest that the distribution of when college athletes began steroid use across the three surveys is not the same.

4. A university admissions officer was concerned that males and females were accepted at different rates into the four different schools (business, engineering, liberal arts, and science) at her university. She collected the following data on the acceptance of 1200 males and 800 females who applied to the university.

	Business	**Engineer**	**Lib Arts**	**Science**
Male	334	251	271	344
Female	179	167	201	253

Does this data provide evidence that the distribution of males and females accepted into the schools is equal?

H_0: The distribution of males and females accepted into the four schools is the same.

H_a: The distribution of males and females accepted into the four schools is not the same.

Chi-Square Test for Homogeneity

Not a random sample

10% Condition: $10n_m = 10(1200) = 12000 < N_m$ and $10n_f = 10(800) = 8000 < N_f$

Large Counts: The expected counts are all greater than 5.

308 251 283 357

205 167 189 239

X^2-Test $\Rightarrow$ $X^2 = 8.297$ and a p-value = 0.0403

The p-value of 0.0403 is less than alpha of 0.05, thus we reject the null hypothesis in favor of the alternative hypothesis. There is sufficient evidence to suggest that the distribution of males and females accepted into the four schools is not the same.