Chi-square Goodness-of-fit test

For this lab, we are going to see how STATA performs the Chi-square goodness-of-fit test.

About the data

```
. use http://www.ats.ucla.edu/stat/stata/notes/hsb2
```

The data used here is called **hsb2,** high school and beyond. This data file contains 200 observations from a sample of high school students with demographic information about the students, such as their gender (**female**), socio-economic status (**ses**) and ethnic background (**race**). It also contains a number of scores on standardized tests, including tests of reading (**read**), writing (**write**), mathematics (**math**) and social studies (**socst**).

Today's Lab

First of all, look at the dataset:
. describe

Let's look at "race", which is the variable we are interested in.
Here, please note that race is stored using real (floating point) numbers not strings.

. sort race
. list race

so we can see that race==1 denotes Hispanic, race==2 denotes Asian, race==3 denotes African-American, and race==4 denotes white.

Q1: What is the proportion of students of each race?

Now ***suppose*** the true proportion of the races is 10% for Hispanic and 10% for Asian, and 10% for African-American, and 70% for whites. **We are going to test whether there is statistical evidence that there is significant difference between the given true proportions and the observed sample proportions**. In STATA the Chi-Square (χ^2) goodness of fit test is invoked by **csgof**:

. csgof race, expperc(10 10 10 70)

We obtain a p-value is 0.16. How do you interpret these findings and what are your conclusions, based on this result?

Q2: What are your conclusions about the test reuslts? What should the null & the alternative hypotheses be? What is the degree of freedom?

Next, let's look at **ses** (socio-economic status) variable. In the sample, are the proportions in each social-economic group the same.

Q3: What should be the expected proportions in this case?

We now perform a Chi-Square (χ^2) goodness of fit test to investigate equal proportions:

. csgof ses, expperc(33.3333,33.3333,33.3334)

Q4: What do you think about our belief now? What should H_0 and H_1 be? What is the degree of freedom?

Q5: Investigate if there are differences between the reading and writing literacy proportions!