AIDS Cases

1. $\alpha = 0.01$.
2. H_O:"The distribution of people in the race categories is the same for male and females" versus H_A:"The distribution of people in the race categories is NOT the same for male and females".
3. A chi-square test is required because a categorical response variable with four levels (race/ethnicity) from two groups or populations (sex) was recorded and the proportions in the response variable are being compared among groups in the null hypothesis.
4. This is an observational study without obvious randomization.
5. The test statistic computed below should reasonably follow a χ^2-distribution because all cells in the expected table (Table 1) have values greater than five.

Table 1. Expected frequency table for distribution of individuals into race and sex categories.

```
          White    Black Hispanic Other
Male    11323.7 16649.5   6870.5 415.3
Female   3455.3  5080.5   2096.5 126.7
```

6. The table of observed frequencies were given but are also in Table 2.

Table 2. Observed frequency table for distribution of individuals into race and sex categories.

```
        White Black Hispanic Other
Male    12855 14946     7019   439
Female   1924  6784     1948   103
```

7. The χ^2 test statistic is 1650.8 with 3 df (Table 3).

Table 3. Results from the chi-square test for examining the distribution of races by male and female.

```
X-squared = 1650.767, df = 3, p-value < 2.2e-16
```

8. The p-value is $p < 0.00005$.
9. The H_O is rejected because the $p - value < \alpha$.
10. It appears that the distribution of the individuals into the races/ethnicities differs between the sexes. An analysis of the row percentages table (Table 4) shows that there are many more white male and black females than would be expected.

Table 4. Row percentage table for the observations of race by sex categories.

```
        White Black Hispanic Other   Sum
Male     36.5  42.4     19.9   1.2 100.0
Female   17.9  63.1     18.1   1.0 100.1
```

R Appendix.

```r
library(NCStats)
obs <- matrix(c(12855,14946,7019,439,1924,6784,1948,103),nrow=2,byrow=TRUE)
rownames(obs) <- c("Male","Female")
colnames(obs) <- c("White","Black","Hispanic","Other")
( aids.chi <- chisq.test(obs,correct=FALSE) )
round(aids.chi$expected,1)
percTable(obs,margin=1,digits=1)
```