Question 12.38

a. $\alpha = 0.01$.
b. 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".
c. A chi-square test is required because a categorical response variable with four levels (race/ethnicity) from two populations (sex) was recorded and the proportions in the response variable are being compared among populations in the null hypothesis.
d. This is an observational study without obvious randomization.
e. 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

f. The table of observed frequencies were given but are shown again 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

g. The χ^2 test statistic is 1650.77 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
```

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

Table 4. Residuals (observed-expected frequencies) for the observations of race by sex categories.

	White	Black	Hispanic	Other
Male	14.39	-13.2	1.79	1.16
Female	-26.05	23.9	-3.24	-2.11

Appendix − R Commands

```
> freq <- c(12855,14946,7019,439,1924,6784,1948,103)
> obs <- matrix(freq,nrow=2,byrow=TRUE)
> rownames(obs) <- c("Male","Female")
> colnames(obs) <- c("White","Black","Hispanic","Other")
> aids.chi <- chisq.test(obs,correct=FALSE)
> aids.chi$expected
> aids.chi$observed
> aids.chi$residuals
```

Notes From Professor

-