STAT 371 **Discussion #8**

1. Chi-square (χ^2) test for Multinomial Distribution

1.1 Structure of data

1) K groups
2) There are n_i observations in the i^{th} group
3) Total # of subjects $N = n_1 + n_2 + \cdots\cdots + n_k$

1.2 State of hypothesis

H_0: $Pr(1) = p_1$, $Pr(2) = p_2$, ……, $Pr(k) = p_k$

H_A: At least one of the probabilities specified in H_0 is incorrect

Or H_A: $Pr(1) \neq p_1$, and/or $Pr(2) \neq p_2$, ……, and/or $Pr(k) \neq p_k$

If we let $X_1 = n_1$, $X_2 = n_2$, ……, $X_k = n_k$
Then under the null hypothesis, variables X_1, X_2, ……, X_k follow the multinomial distribution Multinomial(p_1, p_2, ……, p_k, N), where
$p_1 + p_2 + \cdots\cdots + p_k = 1$

1.3 Test statistics

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

where expected value for the i^{th} group is calculated by $N*p_i$.

1.4 Distribution of X^2 under null hypothesis when sample size is large (the expected count in each group is ≥ 5)

$$X^2 \sim \chi^2 \text{ (df=k-1)}$$

1.5 Find the p-value = $Pr(X^2 > \text{observed } X^2)$

1) Use table 9 in the textbook
2) Use R: p-value = 1-pchisq(observed X^2, k-1)

1.6 Use chisq.test() in R

chisq.test(c(n_1, n_2, ······ n_k), p=c(p_1, p_2, ……, p_k))

2. Example: 10.10 from textbook
   Scientists have used Mongolian gerbils when conducting neurological research. A certain breed of these gerbils were crossed and gave progeny of the following colors:

Color	Black	Brown	White
Number of progeny	40	59	42

After these data consistent with the 1:2:1 ratio predicted by a certain genetic model? Use a chi-square test at α=0.05.

1) By hand – discuss in class
2) Use chisq.test(): chisq.test(c(40, 59, 42), p=c(0.25,0.5,0.25))

   Results given by R: Chi-squared test for given probabilities
   data: c(40, 59, 42)
   X-squared = 3.8085, df = 2, p-value = 0.1489

   Notice that the p-value is 0.1489, which is > α. Thus we fail to reject H_0.