Outline

1. Homogeneity.
   (a) parameters $\pi_{i,j}$  \(\left( \sum_{j=1}^{J} \pi_{i,j} = 1 \right)\)
   (b) null hypothesis: $H_0$: for every $j$,  $\pi_{1,j} = \pi_{2,j} = \cdots = \pi_{I,j}$
   (c) “expected counts” under null hypothesis: $n_{i,j} = \frac{n_i n_j}{n_{..}}$
   (d) Test statistic: $X^2 = \sum \frac{(n_{i,j} - \hat{n}_{i,j})^2}{\hat{n}_{i,j}}$
   (e) Distribution of test statistic under null hypothesis: If all $\hat{n}_{i,j}$ are large, $X^2$ has an approximately chi-square distribution. $X^2$ has the chi-square distribution with $(I-1)(J-1)$ degrees of freedom.

2. The chi-square distribution.

3. Independence
   (a) parameters $\pi_{i,j}$ \(\left( \sum_{j=1}^{J} \pi_{i,j} = 1 \right) \pi_i \pi_j$
   (b) null hypothesis: $H_0$: for every $i,j$,  $\pi_{i,j} = \pi_i \pi_j$
   (c) Same expected counts, test statistic and distribution.

4. New situation: a categorical explanatory variable and a quantitative response variable. Restriction: the explanatory variable has just two categories.
   (a) Same three data collection scenarios.
   (b) Parameters; $\mu_1$, $\mu_2$.
   (c) Null hypothesis: $H_0 : \mu_1 = \mu_2$.

No further homework

Useful R

```
> m=matrix(c(20,28,11,41), nrow=2,ncol=2, byrow=T)
> m
 [,1] [,2]
[1,]  20  28
[2,]  11  41
> rownames(m)=c('Told','Not Told')
> colnames(m)=c('Diagnosed','Not Diagnosed')
> m
    Diagnosed Not Diagnosed
Told  20     28
Not Told 11     41
```
> chisq.test(m)

    Pearson's Chi-squared test with Yates' continuity correction

data:  m
X-squared = 3.9979, df = 1, p-value = 0.04556