# Tests of homogeneity of proportions

Situation: $r$ multinomial samples, each of size $n_i$

ex. sample $n_1$ voters from Idaho
sample $n_2$ voters from Montana
sample $n_3$ voters from Utah
sample $n_4$ voters from Wyoming

determine $Y_{i1}$, $Y_{i2}$, $Y_{i3}$ (# dems, repubs, other) in each sample; interested in whether the proportions of dems, repubs, & others are different among the states

Data can be summarized in a two-way table:

| population | response | | | | |
|---|---|---|---|---|---|
| 1 | $y_{11}$ | $y_{12}$ | $\cdots$ | $y_{1c}$ | $n_{1\cdot}$ |
| 2 | $y_{21}$ | $y_{22}$ | $\cdots$ | $y_{2c}$ | $n_{2\cdot}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | $\vdots$ |
| $r$ | $y_{r1}$ | $y_{r2}$ | $\cdots$ | $y_{rc}$ | $n_{r\cdot}$ |
| | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $\cdots$ | $n_{\cdot c}$ | $n$ |

However, even though there is a two-way table, this situation is *different* from taking one sample of size $n$ and cross-tabulating.

Model:

$$Y_{i1}, Y_{i2}, ..., Y_{ic} \sim \text{multinomial}(n_{i\cdot}, \pi_{i1}, \pi_{i2}, ..., \pi_{ic})$$

that is, the rows arise from multinomial distributions.

Test of homogeneous proportions: tests whether the corresponding (col) proportions in the multinomial distributions are the same:

$$\begin{aligned} \text{H}_0\text{:}\ & \pi_{11} = \pi_{21} = \cdots = \pi_{r1} \\ & \pi_{12} = \pi_{22} = \cdots = \pi_{r2} \\ & \qquad\qquad \vdots \\ & \pi_{1c} = \pi_{2c} = \cdots = \pi_{rc} \end{aligned}$$

In the null hypothesis, there are $c - 1$ parameters.

$$\text{H}_\text{a}\text{:}\ \neq \ \neq \ \cdots$$

In the alternative hypothesis, there are $r \times (c - 1)$ parameters.

Note: this is the multinomial version of “analysis of variance”.

Likelihood: product of multinomial distributions (**product multinomial model**)

ML estimates under $H_0$:

$$\begin{array}{c}\widetilde{\pi}_{11} = \widetilde{\pi}_{21} = \cdots = \widetilde{\pi}_{r1} = \frac{n_{\cdot 1}}{n} \\ \widetilde{\pi}_{12} = \widetilde{\pi}_{22} = \cdots = \widetilde{\pi}_{r2} = \frac{n_{\cdot 2}}{n} \\ \vdots \\ \widetilde{\pi}_{1c} = \widetilde{\pi}_{2c} = \cdots = \widetilde{\pi}_{rc} = \frac{n_{\cdot c}}{n}\end{array}$$

$$\widehat{E}_{ij} = n_{i\cdot}\widetilde{\pi}_{ij} = \frac{n_{i\cdot}n_{\cdot j}}{n}$$

ML estimates under $H_a$:

$$\widehat{\pi}_{ij} = \frac{y_{ij}}{n_{i\cdot}}$$

Curiously, the likelihood ratio statistic for the test of homogeneous multinomials is *identical* to the LR statistic for testing independence in a contingency table:

$$G^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}\log_e\left(\frac{y_{ij}}{\widehat{E}_{ij}}\right)$$

Pearson chi-square statistic is identical too:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(y_{ij}-\widehat{E}_{ij}\right)^2}{\widehat{E}_{ij}}$$

Degrees of freedom:

$$r \times (c-1) - (c-1) = (r-1)(c-1) \quad \text{(same!)}$$

Rejection region:

reject $H_0$ if $G^2 \geq \chi_\alpha^2$

Notes:

- Calculations are the same for both tests (homogeneity & independence)
- Interpretations are different
- Tests are not generally the same for more complex models (additional variables, etc)
- More complex contingency tables (independence) are analyzed with **log-linear models** (SPSS, SAS PROC CATMOD)
- More complex product-multinomial models are analyzed with **logit models** (SAS PROC CATMOD)