Lecture 10. Tests of homogeneity, and connections to confidence intervals

Tests of homogeneity

Example 10.1

150 patients were randomly allocated to three groups of 50 patients each. Two groups were given a new drug at different dosage levels, and the third group received a placebo. The responses were as shown in the table below.

	Improved	No difference	Worse	
Placebo	18	17	15	50
Half dose	20	10	20	50
Full dose	25	13	12	50
	63	40	47	150

Here the row totals are fixed in advance, in contrast to Example 9.4 where the row totals are random variables.

For the above table, we may be interested in testing H_0: the probability of "improved" is the same for each of the three treatment groups, and so are the probabilities of "no difference" and "worse," ie H_0 says that we have homogeneity down the rows. □

- In general, we have independent observations from r multinomial distributions each of which has c categories,
- ie we observe an $r \times c$ table (n_{ij}), $i = 1, \dots, r$, $j = 1, \dots, c$, where $(N_{i1}, \dots, N_{ic}) \sim \text{Multinomial}(n_{i+}; p_{i1}, \dots, p_{ic})$ independently for $i = 1, \dots, r$.
- We test $H_0 : p_{1j} = p_{2j} = \dots = p_{rj} = p_j$ say, $j = 1, \dots, c$ where $p_+ = 1$, and $H_1 : p_{ij}$ are unrestricted (but with $p_{ij} \geq 0$ and $p_{i+} = 1$, $i = 1, \dots, r$).
- **Under $\mathbf{H_1}$**: $\text{like}((p_{ij})) = \prod_{i=1}^r \frac{n_{i+}!}{n_{i1}! \dots n_{ic}!} p_{i1}^{n_{i1}} \dots p_{ic}^{n_{ic}}$, and $\text{loglike} = \text{constant} + \sum_{i=1}^r \sum_{j=1}^c n_{ij} \log p_{ij}$.

  Using Lagrangian methods (with constraints $p_{i+} = 1$, $i = 1, \dots, r$) we find $\hat{p}_{ij} = n_{ij}/n_{i+}$.
- **Under $\mathbf{H_0}$**:
  $\text{loglike} = \text{constant} + \sum_{i=1}^r \sum_{j=1}^c n_{ij} \log p_j = \text{constant} + \sum_{j=1}^c n_{+j} \log p_j$.

  Lagrangian techniques here (with constraint $\sum p_j = 1$) give $\hat{p}_j = n_{+j}/n_{++}$.

- Hence

$$2 \log \Lambda = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \log \left(\frac{\hat{p}_{ij}}{\hat{p}_j} \right) = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \log \left(\frac{n_{ij}}{n_{i+} n_{+j}/n_{++}} \right),$$

  ie the same as in Example 9.5.
- We have $|\Theta_1| = r(c-1)$ (because there are $c-1$ free parameters for each of r distributions).
- Also $|\Theta_0| = c - 1$ (because H_0 has c parameters $p_1, \dots, p_c$ with constraint $p_+ = 1$).
- So df $= |\Theta_1| - |\Theta_0| = r(c-1) - (c-1) = (r-1)(c-1)$, and under H_0, $2 \log \Lambda$ is approximately $\chi^2_{(r-1)(c-1)}$ (ie same as in Example 9.5).
- We reject H_0 if $2 \log \Lambda > \chi^2_{(r-1)(c-1)}(\alpha)$ for an approximate size α test.
- Let $o_{ij} = n_{ij}$, $e_{ij} = n_{i+} n_{+j}/n_{++}$ $\delta_{ij} = o_{ij} - e_{ij}$, and use the same approximating steps as for Pearson's Chi-squared to see that $2 \log \Lambda \approx \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$.

Example 10.2

Example 10.1 continued

o_{ij}	Improved	No difference	Worse	
Placebo	18	17	15	50
Half dose	20	10	20	50
Full dose	25	13	12	50
	63	40	47	150

e_{ij}	Improved	No difference	Worse	
Placebo	21	13.3	15.7	50
Half dose	21	13.3	15.7	50
Full dose	21	13.3	15.7	50

We find $2 \log \Lambda = 5.129$, and we refer this to χ_4^2.

From tables, $\chi_4^2(0.05) = 9.488$, so our observed value is not significant at 5% level, and the data are consistent with H_0.

We conclude that there is no evidence for a difference between the drug at the given doses and the placebo.

For interest, $\sum\sum(o_{ij} - e_{ij})^2/e_{ij} = 5.173$, leading to the same conclusion. □

Confidence intervals and hypothesis tests

- Confidence intervals or sets can be obtained by inverting hypothesis tests, and vice versa.
- Define the **acceptance region** A of a test to be the complement of the critical region C.
- NB By 'acceptance', we really mean 'non-rejection'
- Suppose $X_1, \dots, X_n$ have joint pdf $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$, $\theta \in \Theta$.

Theorem 10.3

(i) *Suppose that for every $\theta_0 \in \Theta$ there is a size α test of $H_0 : \theta = \theta_0$. Denote the acceptance region by $A(\theta_0)$. Then the set $I(\mathbf{X}) = \{\theta : \mathbf{X} \in A(\theta)\}$ is a $100(1-\alpha)$% confidence set for θ.*

(ii) *Suppose $I(\mathbf{X})$ is a $100(1-\alpha)$% confidence set for θ. Then $A(\theta_0) = \{\mathbf{X} : \theta_0 \in I(\mathbf{X})\}$ is an acceptance region for a size α test of $H_0 : \theta = \theta_0$.*

Proof:

- First note that $\theta_0 \in I(\mathbf{X}) \Leftrightarrow \mathbf{X} \in A(\theta_0)$.
- For (i), since the test is size α, we have $\mathbb{P}(\text{accept } H_0 \mid H_0 \text{ is true}) = \mathbb{P}(\mathbf{X} \in A(\theta_0) \mid \theta = \theta_0) = 1 - \alpha$
- And so $\mathbb{P}(I(\mathbf{X}) \ni \theta_0 \mid \theta = \theta_0) = \mathbb{P}(\mathbf{X} \in A(\theta_0) \mid \theta = \theta_0) = 1 - \alpha$.
- For (ii), since $I(\mathbf{X})$ is a $100(1-\alpha)$% confidence set, we have $\mathbb{P}(I(\mathbf{X}) \ni \theta_0 \mid \theta = \theta_0) = 1 - \alpha$.
- So $\mathbb{P}(\mathbf{X} \in A(\theta_0) \mid \theta = \theta_0) = \mathbb{P}(I(\mathbf{X}) \ni \theta_0 \mid \theta = \theta_0) = 1 - \alpha$. □

In words,

(i) says that a $100(1-\alpha)$% confidence set for θ consists of all those values of θ_0 for which $H_0 : \theta = \theta_0$ is not rejected at level α on the basis of $\mathbf{X}$,

(ii) says that given a confidence set, we define the test by rejecting θ_0 if it is not in the confidence set.

Example 10.4

Suppose $X_1, \dots, X_n$ are iid $N(\mu, 1)$ random variables and we want a 95% confidence set for μ.

- One way is to use the above theorem and find the confidence set that belongs to the hypothesis test that we found in Example 10.1.
- Using Example 8.3 (with $\sigma_0^2 = 1$), we find a test of size 0.05 of $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ that rejects H_0 when $|\sqrt{n}(\bar{x} - \mu_0)| > 1.96$ (1.96 is the upper 2.5% point of $N(0,1)$).
- Then $I(\mathbf{X}) = \{\mu : \mathbf{X} \in A(\mu)\} = \{\mu : |\sqrt{n}(\bar{X} - \mu)| < 1.96\}$ so a 95% confidence set for μ is $(\bar{X} - 1.96/\sqrt{n}, \bar{X} + 1.96/\sqrt{n})$.
- This is the same confidence interval we found in Example 5.2. □

Simpson's paradox*

For five subjects in 1996, the admission statistics for Cambridge were as follows:

	Women			**Men**		
	Applied	Accepted	%	Applied	Accepted	%
Total	1184	274	23 %	2470	584	24%

This looks like the acceptance rate is higher for men. But by subject...

	Women			**Men**		
	Applied	Accepted	%	Applied	Accepted	%
Computer Science	26	7	27%	228	58	25%
Economics	240	63	26%	512	112	22%
Engineering	164	52	32%	972	252	26%
Medicine	416	99	24%	578	140	24%
Veterinary medicine	338	53	16%	180	22	12%
Total	1184	274	23 %	2470	584	24%

In all subjects, the acceptance rate was higher for women!

Explanation: women tend to apply for subjects with the lowest acceptance rates.

This shows the danger of pooling (or collapsing) contingency tables.