Chapter 7

Recap Sheet

- ## Chi - square Goodness - of - Fit test

Goal : You have one sample and a candidate distribution (e.g. Poisson, Binomial, Normal with given / estimated params). You want to test :

H_0 : the population follows the specified distribution
vs H_1 : it does not

Data preparation :

1. Take a random sample of size n
2. Split the support into k disjoint categories / classes (bins)
3. For each class i compute :
   - **Observed count** O_i (from the sample)
   - **Expected count** $E_i = n p_i$, where p_i is the model probability for class i under H_0.
   - **Rule of thumb** : make sure all $E_i \geq 5$ (some texts allow ≥ 3) ; if not, merge adjacent classes

Test statistic :

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Under H_0 (with large n), χ_0^2 is approximately chi - square with

$$df = k - 1 - m$$

where m is the number of distribution parameters estimated from the data (e.g., m = 1 if you estimated a Poisson mean; m = 2 if you estimated a Normal μ, σ ; m = 0 if all parameters are known).

Decision (level α) : reject H_0 if $\chi_0^2 > \chi^2_{1-\alpha, df}$

P - value : $P(\chi^2_{df} \geq \chi_0^2)$

Ex : Sample : n = 60 boards, counts by number of defects

0	1	2	3
32	15	9	4

1. Estimate parameter (mean of Poisson) by sample mean

$$\hat{\lambda} = \frac{0 \times 32 + 1 \times 15 + 2 \times 9 + 3 \times 4}{60} = 0.75 \quad (m = 1)$$

2. Model probabilities with $\lambda = 0.75$:

$$p_0 = P(X=0) = \frac{e^{-0.75} \times 0.75^0}{0!} = 0.472$$

$$p_1 = P(X=1) = \frac{e^{-0.75} \times 0.75^1}{1!} = 0.354$$

$$p_2 = P(X=2) = \frac{e^{-0.75} \times 0.75^2}{2!} = 0.133$$

$$p_{\geq 3} = 1 - (p_0 + p_1 + p_2) = 0.041$$

3. Expected counts $E_i = n p_i$:

$$E_0 = 28.32 , \ E_1 = 21.24 , \ E_2 = 7.98 , \ E_{\geq 3} = 2.46$$

The last cell's expectation is small, so merge 2 and ≥ 3 :
Observed $O_{\geq 2} = 9 + 4 = 13$; Expected $E_{\geq 2} = 7.98 + 2.46 = 10.44$
Now $k = 3$ classes : $\{0\}, \{1\}, \{\geq 2\}$

4. Test statistic

$$\chi_0^2 = \frac{(32 - 28.32)^2}{28.32} + \frac{(15 - 21.24)^2}{21.24} + \frac{(13 - 10.44)^2}{10.44} = 2.94$$

5. Degrees of freedom : $df = k - 1 - m = 3 - 1 - 1 = 1$

6. Decision / p-value :
Critical values ($df = 1$) : $\chi^2_{0.95,1} = 3.84$, $\chi^2_{0.90,1} = 2.71$
Since 2.94 lies between them, $p \in (0.05 , 0.10)$; exact $p \approx 0.086$
Because $p > 0.05$, fail to reject H_0 → the Poisson model is plausible.

- **Chi-square test of independance in an $r \times c$ contingency table**

Goal : From one sample cross-classified by two categorical variables, decide whether the variables are independant

H_0 : Row variable and column variable are independant
vs H_1 : They are associated (not independant)

Data layout : An $r \times c$ table of observed counts O_{ij} (row i, column j) ; $n = \sum_{i=1} \sum_{j=1} O_{ij}$

Row totals $R_i = \sum_j O_{ij}$; column totals $C_j = \sum_i O_{ij}$

Expected counts under independance : If independant,
$P(\text{row } i , \text{col } j) = P(\text{row } i) P(\text{col } j) \approx \frac{R_i}{n} \times \frac{C_i}{n}$

Thus the expected count in cell ij is :
$$E_{ij} = \frac{R_i C_j}{n}$$
(works for any table size)

Rule of thumb : All E_{ij} should be ≥ 5 (some allow "most ≥ 5, none < 1")

Test statistic and df :
$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Under H_0 (large n), $\chi_0^2 \sim \chi^2_\nu$ with $\nu = (r-1)(c-1)$

Decision at level α : reject H_0 if $\chi_0^2 > \chi^2_{1-\alpha,\nu}$
P-value $= P(\chi^2_\nu \geq \chi_0^2)$

Ex : Table (observed counts)

	Plan 1	Plan 2	Plan 3	Row totals
Salaried	160	140	40	340
Hourly	40	60	60	160
Column totals	200	200	100	500

1. Expected counts $E_{ij} = \frac{R_i C_j}{n}$

Row proportions : $\hat{u}_1 = \frac{340}{500} = 0.68$, $\hat{u}_2 = \frac{160}{500} = 0.32$

Column proportions : $\hat{v}_1 = \frac{200}{500} = 0.40$, $\hat{v}_2 = 0.40$, $\hat{v}_3 = 0\ 20$

So,

- $E_{11} = 500 \times 0.68 \times 0.40 = 136$, $E_{12} = 136$, $E_{13} = 68$
- $E_{21} = 500 \times 0.32 \times 0.40 = 64$, $E_{22} = 64$, $E_{23} = 32$

2. Test statistic :

$$\chi_0^2 = \frac{(160-136)^2}{136} + \frac{(140-136)^2}{136} + \dots + \frac{(60-64)^2}{64} + \frac{(60-32)^2}{32} = 49.63$$

3. $\nu = (r-1)(c-1) = (2-1)(3-1) = 2$

Critical value at $\alpha = 0.05$: $\chi^2_{0.95,2} = 5.99$

Since $49.63 > 5.99$, reject H_0.

Preference for health - insurance plans is associated with job classification (not independant).