MA181 INTRODUCTION TO STATISTICAL MODELLING
GOODNESS-OF-FIT TEST

Example - Mendel's peas Mendel's double intercross data for $(Round, Yellow) \times (Wrinkled, Green)$ peas. The expected frequencies are in the ratios 9:3:3:1 on the assumption that the factors segregate independently.

	RY	WY	RG	WG	Total
Observed frequency	315	101	108	32	556
Expected frequency	$312\frac{3}{4}$	$104\frac{1}{4}$	$104\frac{1}{4}$	$34\frac{3}{4}$	556

$$x^2 = \frac{(2\frac{1}{4})^2}{312\frac{3}{4}} + \frac{(3\frac{1}{4})^2}{104\frac{1}{4}} + \frac{(3\frac{1}{4})^2}{104\frac{1}{4}} + \frac{(2\frac{3}{4})^2}{34\frac{3}{4}} = 0.470$$

Critical regions with $\nu = 3$ are $x^2 > 7.815$ for $\alpha = 0.05$ and $x^2 > 11.34$ for $\alpha = 0.01$. Therefore accept H_0: the two genes segregate independently.

Example - Pharbitis Double intercross data for two genes A and B in Pharbitis. The expected frequencies are again in the ratios 9:3:3:1, on the assumption that A and B segregate independently.

	AB	Ab	aB	ab	Total
Observed frequency	187	35	37	31	290
Expected frequency	$163\frac{3}{8}$	$54\frac{3}{8}$	$54\frac{3}{8}$	$18\frac{1}{8}$	290

$$x^2 = \frac{(23\frac{7}{8})^2}{163\frac{1}{8}} + \frac{(19\frac{3}{8})^2}{54\frac{3}{8}} + \frac{(17\frac{3}{8})^2}{54\frac{3}{8}} + \frac{(12\frac{7}{8})^2}{18\frac{1}{8}} = 25.096$$

Critical regions with $\nu = 3$

$x^2 > 7.815$ for $\alpha = 0.05$

$x^2 > 11.34$ for $\alpha = 0.01$

$x^2 > 16.27$ for $\alpha = 0.001$

Reject H_0 and conclude (very strongly) that the genes are linked.

Estimating parameters

Example - Pharbitis revisited One theory suggests that the probabilities for the four cells can be written as $(2+\theta)/4$, $(1-\theta)/4$, $(1-\theta)/4$ and $\theta/4$ for some parameter θ. The maximum likelihood estimate of θ is $\hat{\theta} = 0.4835$, which leads to the expected frequencies given in the following.

	AB	Ab	aB	ab	Total
Observed Frequency	187	35	37	31	290
Expected frequency	180.054	37.446	37.446	35.054	290

$$x^2 = \frac{(187 - 180.054)^2}{180.054} + \ldots + \frac{(31 - 35.054)^2}{35.054} = 0.902$$

Critical regions with 2 degrees of freedom are

$x^2 > 5.991$ for $\alpha = 0.05$

$x^2 > 9/210$ for $\alpha = 0.01$

Accept H_0: model given as above.

Example Peas in pods The table below gives, in its second column, the frequency distribution of the number Y of peas found in the pod of a four-seeded line of pea. A total of 269 pods were inspected.

$\hat{\pi} = 0.5530$

Number of peas in pod	0	1	2	3	4	Total
Observed frequency	16	45	100	82	26	269
Expected frequency	10.74	53.15	98.62	81.33	25.15	268.99

$$x^2 = \frac{(16 - 10.74)^2}{10.74} + \ldots + \frac{(26 - 25.15)^2}{25.15} = 3.88.$$

Critical regions with three degrees of freedom as on page 1. Do not reject H_0: model given by binomial distribution.

Small expected frequencies No expected frequency should be smaller than one and no more than 20% should be less than five. Otherwise it is necessary to pool cells.

Example - Poisson distribution The number Y, of α-particles emitted by a film of Polonium in 2608 intervals of $\frac{1}{8}$ minute was given on the Poisson distribution handout. The end of the table is as follows:

	Frequency of Intervals	
y	Observed	Poisson, E_y
10	10	11.3
11	4	4.0
12	0	1.3
13	1	0.4
14	1	0.1
≥ 15	0	0.0

The last four cells may be pooled to give the following complete table.

$\hat{\mu} = 3.8715$

y	0	1	2	3	4	5	6	7	8	9
O_y	57	203	383	525	532	408	273	139	45	27
E_y	54.3	210.3	407.1	525.3	508.4	393.7	254.0	140.5	68.0	29.2

10	11	≥ 12	Total
10	4	2	2608
11.3	4.0	1.8	2607.9

$x^2 = 13.0$

Critical regions with 11 degrees of freedom

$x^2 > 19.68$ for $\alpha = 0.05$

$x^2 > 24.72$ for $\alpha = 0.01$

$x^2 > 31.26$ for $\alpha = 0.001$

Accept H_0: model is given by Poisson distribution.