Section VII

Chi-square test for comparing proportions and frequencies

F test for means

proportions: chi-square test

Z test for comparing proportions between two independent groups

$$Z = \frac{P_1 - P_2}{SE_d}$$

$SE_d = \sqrt{P_1(1-P_1)/n_1 + P_2(1-P_2)/n_2}$ *

Definition: $Z^2 = \chi^2(1)$ = chi-square stat with one degree of freedom (df=1).

What if there are many groups / many proportions to compare and we worry about multiplicity? Can do overall (omnibus) χ^2 test. Overall χ^2 test with **k-1** degrees of freedom is for testing null hypothesis that $\pi_1=\pi_2=\pi_3= \ldots \pi_k = \pi$. Analog to overall F test in linear models / ANOVA for comparing many βs or means.

* Technically, under the null hypothesis $\pi_1=\pi_2=\pi$ so $SE_d = \sqrt{p(1-p)/n_1 + p(1-p)/n_2}$ $= \sqrt{p(1-p)\,[1/n_1 + 1/n_2]}$ where $p = (n_1 p_1 + n_2 p_2)/(n_1+n_2)$ is the estimate of π.

Chi-square test (omnibus test) Controlling for multiplicity in hypothesis testing

Ex: Troublesome morning sickness

137/350 = 39% have troublesome morning sickness overall after treatment (100% had it before treatment)

observed frequencies

	no tx	accupress.	dummy	total
yes	67	29	41	137
no	52	90	71	213
total	119	119	112	350
Pct yes	56%	24%	37%	39%

What frequencies would be expected if there is no association between treatment and outcome? (Null hypothesis: $\pi_1=\pi_2=\pi_3=\pi$)

Expected frequencies **if** no assn.

	no tx	accupress.	dummy	total
yes	46.6	46.6	43.8	137
no	72.4	72.4	68.1	213
total	119	119	112	350

39% yes, 61% no in each group

Calculating expected frequency – Example for the "yes, no tx" cell.

Expected freq= 119 (137/350) = 46.6

If there is no association, the observed and expected frequencies should be similar.

Chi-square statistic is a measure of squared differences between observed and expected frequencies.

$$\chi^2 = \sum_{\text{All cells}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

In this example (with six cells)

$$\chi^2 = \frac{(67-46.6)^2}{46.6} + \frac{(52-72.4)^2}{72.4} + \ldots$$

$$\ldots + \frac{(71-68.1)^2}{68.1} = 25.91$$

df =(# rows-1)(# cols –1)= (2-1)(3-1)=2

p value < 0.001 (get from =CHIDIST(χ^2 ,df) in =EXCEL or chi-square table)

If this overall p value is NOT significant, we conclude all proportions are not significantly different from each other at the α level, that is, there is no association.

F statistic for means is the analog of the chi-square statistics for proportions

$$F = \frac{\sum (\bar{Y}_i - \bar{Y})^2 n_i/(k-1)}{S_p^2}$$

Where $\bar{Y}_i$ is the mean of the ith group

n_i is the sample size in the ith group (i=1, 2, 3, ...k)

$\bar{Y}$ = overall mean, k=number of groups

and S_p^2 is the squared pooled standard deviation defined by

$$S_p^2 = \frac{(n_1-1)\, S_1^2 + (n_2-1)\, S_2^2 + \ldots + (n_k-1) S_k^2}{(n_1+n_2+\ldots+n_k) - k}$$

If the overall F based p value is not significant, we conclude none of the means are significantly different from each other.

Rule of thumb for chi-square significance

If the null hypothesis is **true**, the expected (average) value of the $\chi 2$ statistic is equal to its degrees of freedom or $E(\chi 2)=df$. So, **if** the null hypothesis is **true**, $\chi 2/df \approx 1.0$. Therefore a $\chi 2$ value less than its df (or equivalently $\chi 2/df < 1$) is never statistically significant.

Technical note- Fisher's exact test alternative calculation of the chi-square test p value

Conventionally, the p values for the Z and χ^2 statistics are obtained by looking them up on the Gaussian distribution or the corresponding χ^2 distribution, which is derived from the Gaussian distribution. However, when the sample size is small and **the expected frequencies are less than 5 in at least 20% of the cells**, it can be shown that the central limit theorem approximation may not be accurate so the distribution of Z or χ^2 may not be Gaussian. So p values from the Gaussian or χ^2 tables are incorrect. In this case, the exact, correct p value can be computed based on the multinomial distribution (which we have not studied), although this computation is very difficult without a computer program. The algorithm for computing the exact p value was developed by RA Fisher so p values computed this was are said to be computed using “Fishers exact test”. However, the purpose is still to compare frequencies and proportions as with the Z and chi-square tests. In principle, the Fisher procedure could always be used in place of looking up a p value on the chi-square distribution.

Chi-square goodness of fit test

Sometimes we wish to compare observed frequencies to those that would be expected (predicted) if the data follows a known distribution, such as the Poisson distribution or the Gaussian (Normal) distribution or some other theory or model. Such a comparison of observed data to expected results under the model is called a goodness-of-fit test.

Example: Comparison to a Poisson.

We observe n=100 persons and record the number of colds each had in a three month period

Number colds	number persons (o)	expected number (e)
0	39	39.9
1	37	36.7
2	17	16.9
3	7	5.2
4+	0	1.5

Under the Poisson distribution with mean=92/100, the mean in the data, we computed the expected (e) number of persons with 0, 1, 2, 3 and 4 or more colds. $e=100\ (0.92^y\ e^{-0.92})\ /y!$

The chi-square statistic = $\sum (o-e)^2/e = 2.12$.

While the computation of the chi-square value is similar to the chi-square for comparing proportions, the degrees of freedom (df) are the number of categories minus one or 5-1=4 in this example. The p value here is p = 0.7145. We do not reject the null hypothesis that the data fits a Poisson distribution. In general, to show that the data fits the model, one must have a NON significant p value.

Goodness of fit
Gregor Mendel pea hybrid (1865)

If round (R) is dominant over angular (a)
25% RR, 50% Ra, 25% aa
so 75% round phenotype expected in hybrids

	seed form - observed				seed form- expected		
plant	round	angular	total	pct round	round	angular	total
1	45	12	57	78.9%	42.75	14.25	57
2	27	8	35	77.1%	26.25	8.75	35
3	24	7	31	77.4%	23.25	7.75	31
4	19	10	29	65.5%	21.75	7.25	29
5	32	11	43	74.4%	32.25	10.75	43
6	26	6	32	81.3%	24.00	8.00	32
7	88	24	112	78.6%	84.00	28.00	112
8	22	10	32	68.8%	24.00	8.00	32
9	28	6	34	82.4%	25.50	8.50	34
10	25	7	32	78.1%	24.00	8.00	32
total	**336**	**101**	**437**	**76.9%**	**327.75**	**109.25**	**437**

Chi square = 5.297, df= 10-1=9
chi square/df = 0.59
p value = 0.8077

Data fits the genetic theory.