Lecture 12. Chi-square Tests

1. Test for goodness-of-fit.

2. Test for association.

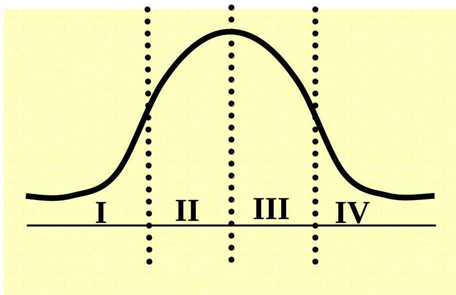
Goodness-of-fit Test（適合度檢定）: An example

假設 $X_i \sim N(0,\ 1)$,   i=1, 2,⋯, 20

取 –0.67, 0, 0.67 三點為 cutoff points，

則落在 I， II， III， IV四個區間內的機率

約略 = 0.25。當 n = 20 時，落在這 4 個



interval 內的**點數**期望上應約為 5（20 × 0.25 = 5）

$(E_1 = E_2 = E_3 = E_4 = 5)$。若實際上，落在 I , II , III , IV 中之**點數**為 ：

$O_1 = 3,\ O_2 = 4,\ O_3 = 11,\ O_4 = 2$。

我們可以比較 $(3,4,11,2)^T$ 與 $(5,5,5,5)^T$ 兩個向量之「差距」。

比法是： $\sum_{i=1}^{4} \left\{ (O_i - E_i)^2 / E_i \right\}$ ← **Test statistic** (**Pearson**)

若{E_i}與{O_i}相差很近，表示無法 reject $N(0,\ 1)$ 之 assumption (or hypothesis)，否則 reject $H_0 : F(\bullet) = N(0,\ 1)$ 。

上面的 test statistic $X^2 \equiv \sum_{i=1}^{4} \left\{ (O_i - E_i)^2 / E_i \right\} \sim \chi_3^2$

General

1. 設 cutoff points 為 $c_1, c_2 \cdots\cdots, c_p$，根據 $F_0(\bullet)$ 計算出 $E_1, \cdots\cdots, E_{p+1}$，又根據 data 算出 $O_1, \cdots\cdots, O_{p+1}$，再去求 $X^2 = \sum_{i=1}^{p+1} \frac{(O_i - E_i)^2}{E_i}$ 若 $X^2 > \chi^2_{p,\ 1-\alpha}$，

則 reject $\begin{cases} H_0 : F(\bullet) = F_0(\bullet) \\ (v.s.\ \ H_a : F(\bullet) \neq F_0(\bullet)) \end{cases}$

2. Extension: If $F_0 = F_0(\theta)$, where **dim(θ)=k** (<p), then the previous E_is must be estimated based on the estimated θ. In this circumstances, X^2 is distributed as chi-square distribution with **d.f.=(p+1)-1-k=p-k**.

3. A worked example:

Data:

Calculation:

SAS code:

列聯表分析裡的卡方”獨立性檢定”(或稱”相關性檢定”)

Contingency Table Analysis: a chi-square test for independence (or test for association)

列聯表分析(contingency table analysis)在統計學裡是一個很重要的分析方式。其所分析的是交叉分類之資料(cross-classification data)。比如，一個人“抽菸與否”與“是否得肺癌”之關係。抽菸（smoke）或不抽菸（non-smoke）以變數 $S = 1$ or $S = 0$ 表示；得肺癌（lung cancer）或未得肺癌，以 $L = 1$ or $L = 0$ 表示。我們可以得到類似 Table 1 之 2x2 表。Table 1 表示 $S = 1$ 的人(抽菸者)有 27 人，其中 8 人得肺癌，19 人未得；而 $S = 0$(非抽菸者)的有 17 人，其中只有 1 人得肺癌。相對於 Table 1，我們以 Table 2 表示每一個空格的機率。

Table 1	$L=1$	$L=0$	Total
$S=1$	**8**	**19**	**27**
$S=0$	**1**	**16**	**17**
Total	**9**	**35**	**44**

Table 2	$L=1$	$L=0$	Total
$S=1$	$\mathbf{P_{11}}$	$\mathbf{P_{10}}$	$\mathbf{P_{1\cdot}}$
$S=0$	$\mathbf{P_{01}}$	$\mathbf{P_{00}}$	$\mathbf{P_{0\cdot}}$
Total	$\mathbf{P_{\cdot 1}}$	$\mathbf{P_{\cdot 0}}$	**1**

若“抽煙與否”與“是否得肺癌”無關(H_0)，則**理論上**(或說**期望上**)，應當 $P(S=s,\ L=l)=P(S=s)\times P(L=l)$，$s=0, 1$，$l=0, 1$。根據初等統計，我們知道兩個事件 A 和 B 如果是獨立的，則 P(AB)=P(A)P(B)，亦即:

$$\left\{\begin{array}{lcl} p_{11} & = & p_{1\bullet}\times p_{\bullet 1} \\ p_{10} & = & p_{1\bullet}\times p_{\bullet 0} \\ p_{01} & = & p_{0\bullet}\times p_{\bullet 1} \\ p_{00} & = & p_{0\bullet}\times p_{\bullet 0} \end{array}\right\} \qquad (1)$$

根據 Table 1 及 Table 2，我們可以用 $\frac{27}{44}$ 估計 $\mathbf{P_{1\cdot}}$，用 $\frac{9}{44}$ 估計 $\mathbf{P_{\cdot 1}}$。**並且，根據(1)中之第一式，$\mathbf{P_{11}}$** 可以用 $\frac{27}{44}*\frac{9}{44}$ 估計(如果 H_0 對)。也就是 L=1&S=1 那一格期望上應該有 44* $\mathbf{P_{11}}$ 個人，或即 $44*\frac{27}{44}*\frac{9}{44}=\frac{27*9}{44}$ 個人。而事實上看到的是 8 個人。以上所述，是 2x2 表的左上角那一格(以紅色表示)。其餘三格的推論方式亦相同。現在以 **E** 表期望上之個數，**O** 表觀察到之個數。則利用統計上的大樣本性質之證明，可以導出:

$$\sum_{i=1}^{4}\frac{(O_i-E_i)^2}{E_i}\sim\chi_1^2$$

根據這個公式，我們可以計算出 Table 1 之列聯表的 χ^2 實現值如下:

$$\frac{(8-\frac{9*27}{44})^2}{\frac{9*27}{44}}+\frac{(19-\frac{35*27}{44})^2}{\frac{35*27}{44}}+\frac{(1-\frac{9*17}{44})^2}{\frac{9*17}{44}}+\frac{(16-\frac{35*17}{44})^2}{\frac{35*17}{44}}=\mathbf{3.62},$$

p-value=0.057。據此，我們可依所定之顯著水準決定是否要 reject H_0。

但是，由於 Table 1 中有所謂的稀疏觀察值(sparse data)，亦即有些格子觀察到的個數很小(<5)。這種情形最好另以 **Fisher's Exact Test** 做檢定。更一般地，對於一個 r ×c 表，我們亦可證明:

$$\sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(r-1)(c-1)}, \textbf{ Q1}$$

SAS Code Example:

```
data lung;
input SMK LungCa count @@;
cards;
1   1     8
1   0    19
0   1     1
0   0    16
;
proc freq order=data;
tables SMK*LungCa/nocol nopercent all;
weight count;
run;
```

Chi-square Table

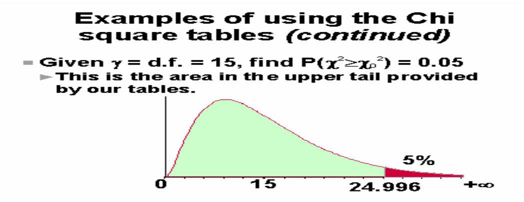


TABLE B.1 Critical Values of the Chi-Square Distribution

ν	α: 0.999	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.001
1	0.000	0.000	0.000	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.841	5.024	6.635	7.879	10.828
2	0.002	0.010	0.020	0.051	0.103	0.211	0.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597	13.816
3	0.024	0.072	0.115	0.216	0.352	0.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	0.091	0.207	0.297	0.484	0.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860	18.467
5	0.210	0.412	0.554	0.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.833	15.086	16.750	20.515
6	0.381	0.676	0.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.458
7	0.599	0.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278	24.322
8	0.857	1.344	1.646	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	1.152	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	1.479	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188	29.588
11	1.834	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757	31.264
12	2.214	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300	32.909
13	2.617	3.565	4.107	5.009	5.892	7.042	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819	34.528
14	3.041	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319	36.123
15	3.483	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	27.488	30.578	32.801	37.697
16	3.942	5.142	5.812	6.908	7.962	9.312	11.912	15.338	19.369	23.542	26.296	28.845	32.000	34.267	39.252
17	4.416	5.697	6.408	7.564	8.672	10.085	12.792	16.338	20.489	24.769	27.587	30.191	33.409	35.718	40.790
18	4.905	6.265	7.015	8.231	9.390	10.865	13.675	17.338	21.605	25.989	28.869	31.526	34.805	37.156	42.312
19	5.407	6.844	7.633	8.907	10.117	11.651	14.562	18.338	22.718	27.204	30.144	32.852	36.191	38.582	43.820
20	5.921	7.434	8.260	9.591	10.851	12.443	15.452	19.337	23.828	28.412	31.410	34.170	37.566	39.997	45.315
21	6.447	8.034	8.897	10.283	11.591	13.240	16.344	20.337	24.935	29.615	32.671	35.479	38.932	41.401	46.797
22	6.983	8.643	9.542	10.982	12.338	14.041	17.240	21.337	26.039	30.813	33.924	36.781	40.289	42.796	48.268
23	7.529	9.260	10.196	11.689	13.091	14.848	18.137	22.337	27.141	32.007	35.172	38.076	41.638	44.181	49.728
24	8.085	9.886	10.856	12.401	13.848	15.659	19.037	23.337	28.241	33.196	36.415	39.364	42.980	45.559	51.179
25	8.649	10.520	11.524	13.120	14.611	16.473	19.939	24.337	29.339	34.382	37.652	40.646	44.314	46.928	52.620
26	9.222	11.160	12.198	13.844	15.379	17.292	20.843	25.336	30.435	35.563	38.885	41.923	45.642	48.290	54.052
27	9.803	11.808	12.879	14.573	16.151	18.114	21.749	26.336	31.528	36.741	40.113	43.195	46.963	49.645	55.476
28	10.391	12.461	13.565	15.308	16.928	18.939	22.657	27.336	32.620	37.916	41.337	44.461	48.278	50.993	56.892
29	10.986	13.121	14.256	16.047	17.708	19.768	23.567	28.336	33.711	39.087	42.557	45.722	49.588	52.336	58.301
30	11.588	13.787	14.953	16.791	18.493	20.599	24.478	29.336	34.800	40.256	43.773	46.979	50.892	53.672	59.703
31	12.196	14.458	15.655	17.539	19.281	21.434	25.390	30.336	35.887	41.422	44.985	48.232	52.191	55.003	61.098
32	12.811	15.134	16.362	18.291	20.072	22.271	26.304	31.336	36.973	42.585	46.194	49.480	53.486	56.328	62.487
33	13.431	15.815	17.074	19.047	20.867	23.110	27.219	32.336	38.058	43.745	47.400	50.725	54.776	57.648	63.870
34	14.057	16.501	17.789	19.806	21.664	23.952	28.136	33.336	39.141	44.903	48.602	51.966	56.061	58.964	65.247
35	14.688	17.192	18.509	20.569	22.465	24.797	29.054	34.336	40.223	46.059	49.802	53.203	57.342	60.275	66.619

(continued)

TABLE B.1 (cont.) Critical Values of the Chi-Square Distribution

ν	α: 0.999	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.001
36	15.324	17.887	19.233	21.336	23.269	25.643	29.973	35.336	41.304	47.212	50.998	54.437	58.619	61.581	67.985
37	15.965	18.586	19.960	22.106	24.075	26.492	30.893	36.336	42.383	48.363	52.192	55.668	59.893	62.883	69.346
38	16.611	19.289	20.691	22.878	24.884	27.343	31.815	37.335	43.462	49.513	53.384	56.896	61.162	64.181	70.703
39	17.262	19.996	21.426	23.654	25.695	28.196	32.737	38.335	44.539	50.660	54.572	58.120	62.428	65.476	72.055
40	17.916	20.707	22.164	24.433	26.509	29.051	33.660	39.335	45.616	51.805	55.758	59.342	63.691	66.766	73.402
41	18.576	21.421	22.906	25.215	27.326	29.907	34.585	40.335	46.692	52.949	56.942	60.561	64.950	68.053	74.745
42	19.239	22.138	23.650	25.999	28.144	30.765	35.510	41.335	47.766	54.090	58.124	61.777	66.206	69.336	76.084
43	19.906	22.859	24.398	26.785	28.965	31.625	36.436	42.335	48.840	55.230	59.304	62.990	67.459	70.616	77.419
44	20.576	23.584	25.148	27.575	29.787	32.487	37.363	43.335	49.913	56.369	60.481	64.201	68.710	71.893	78.750
45	21.251	24.311	25.901	28.366	30.612	33.350	38.291	44.335	50.985	57.505	61.656	65.410	69.957	73.166	80.077
46	21.929	25.041	26.657	29.160	31.439	34.215	39.220	45.335	52.056	58.641	62.830	66.617	71.201	74.437	81.400
47	22.610	25.775	27.416	29.956	32.268	35.081	40.149	46.335	53.127	59.774	64.001	67.821	72.443	75.704	82.720
48	23.295	26.511	28.177	30.755	33.098	35.949	41.079	47.335	54.196	60.907	65.171	69.023	73.683	76.969	84.037
49	23.983	27.249	28.941	31.555	33.930	36.818	42.010	48.335	55.265	62.038	66.339	70.222	74.919	78.231	85.351
50	24.674	27.991	29.707	32.357	34.764	37.689	42.942	49.335	56.334	63.167	67.505	71.420	76.154	79.490	86.661
51	25.368	28.735	30.475	33.162	35.600	38.560	43.874	50.335	57.401	64.295	68.669	72.616	77.386	80.747	87.968
52	26.065	29.481	31.246	33.968	36.437	39.433	44.808	51.335	58.468	65.422	69.832	73.810	78.616	82.001	89.272
53	26.765	30.230	32.019	34.776	37.276	40.308	45.741	52.335	59.534	66.548	70.993	75.002	79.843	83.253	90.573
54	27.468	30.981	32.793	35.586	38.116	41.183	46.676	53.335	60.600	67.673	72.153	76.192	81.069	84.502	91.872
55	28.173	31.735	33.570	36.398	38.958	42.060	47.610	54.335	61.665	68.796	73.311	77.380	82.292	85.749	93.168
56	28.881	32.491	34.350	37.212	39.801	42.937	48.546	55.335	62.729	69.919	74.468	78.567	83.513	86.994	94.461
57	29.592	33.248	35.131	38.027	40.646	43.816	49.482	56.335	63.793	71.040	75.624	79.752	84.733	88.236	95.751
58	30.305	34.008	35.913	38.844	41.492	44.696	50.419	57.335	64.857	72.160	76.778	80.936	85.950	89.477	97.039
59	31.021	34.770	36.698	39.662	42.339	45.577	51.356	58.335	65.919	73.279	77.931	82.117	87.166	90.715	98.324
60	31.738	35.535	37.485	40.482	43.188	46.459	52.294	59.335	66.981	74.397	79.082	83.298	88.379	91.952	99.607
61	32.459	36.301	38.273	41.303	44.038	47.342	53.232	60.335	68.043	75.514	80.232	84.476	89.591	93.186	100.888
62	33.181	37.068	39.063	42.126	44.889	48.226	54.171	61.335	69.104	76.630	81.381	85.654	90.802	94.419	102.166
63	33.906	37.838	39.855	42.950	45.741	49.111	55.110	62.335	70.165	77.745	82.529	86.830	92.010	95.649	103.442
64	34.633	38.610	40.649	43.776	46.595	49.996	56.050	63.335	71.225	78.860	83.675	88.004	93.217	96.878	104.716
65	35.362	39.383	41.444	44.603	47.450	50.883	56.990	64.335	72.285	79.973	84.821	89.177	94.422	98.105	105.988
66	36.093	40.158	42.240	45.431	48.305	51.770	57.931	65.335	73.344	81.085	85.965	90.349	95.626	99.330	107.258
67	36.826	40.935	43.038	46.261	49.162	52.659	58.872	66.335	74.403	82.197	87.108	91.519	96.828	100.554	108.526
68	37.561	41.713	43.838	47.092	50.020	53.548	59.814	67.335	75.461	83.308	88.250	92.689	98.028	101.776	109.791
69	38.298	42.494	44.639	47.924	50.879	54.438	60.756	68.334	76.519	84.418	89.391	93.856	99.228	102.996	111.055
70	39.036	43.275	45.442	48.758	51.739	55.329	61.698	69.334	77.577	85.527	90.531	95.023	100.425	104.215	112.317