第 3 章
分类数据分析

想一想

- 我国改革开放 40 多年来，女性是否和男性一样获得了公平的教育机会？不同性别的人群接受教育的程度是否仍然存在差异？
- 改革开放 40 多年来，我国经济飞速发展，汽车步入百姓家，对于消费者，性别、年龄及月收入等个体特征是如何影响其购车决策的？
- 如果想了解大气环境的污染状况，你认为应该如何利用观测到的污染级别数据进行分析？
- 汽车的大量使用难免会增加交通事故的发生概率，交通事故的发生是否和不同的时段有关联？如果有关联，你该如何判断这种关联的强弱？
- 突发环境事件影响社会公众利益，我国突发的环境事件有什么发生规律？你认为应该如何对其发生规律进行研究？

实际统计分析应用中会出现大量的离散分类计数资料，此类资料可以使用列联表进行分析，本章介绍如何根据分类数据的观测频数和期望频数，利用卡方分布，使用卡方统计量进行分类数据分析。使用卡方统计量进行的分类数据分析在很多著述中也称卡方检验。

3.1 基本概念

3.1.1 分类数据

如第 2 章所述，统计数据的类型主要有分类数据、顺序数据和数值型数据。分类数据是对研究对象进行分类的结果，虽然以数值表示分类的结果，但是数值描述的是研究对象的分类特征。分类数据在实际研究中经常遇见，例如：在研究性别是否影响语言学习成效时，“男”和“女”就是对性别进行的分类；在研究家庭状况是否影响青少年行为时，“完整家庭”和“单亲家庭”就是对家庭状况的分类；在研究青少年行为时，“问题少年”和“非问题少年”就是对青少年行为的分类；等等。如果使用“1”和“0”分别表示被调查者为男性、来自完整家庭、问题少年与被调查者为女性，来自单亲家庭、非问题少年，对这类问题数据的汇总结果即为频数。

实际应用中，还可以通过分组的方法把数值型数据转化为分类数据，如研究语言学习效果时，考试成绩是数值型数据，但是如果按照 0～60、＞60～70、＞70～80、＞80～90、＞90～100 进行分组，数值型数据就变换成分类数据，把落入每一个区间的观测数进行汇总，其结果就是频数。

同一种数值型数据，依据研究目的的不同可以转换为不同的分类数据。如果要研究某次考试的通过率，假定 60 分合格，则可以按照“60 分以下”和“60 分及以上”分成 2 类；如果要研究考试成绩的分布是否合理，则通常的做法是按照前文分法进行分组。此外，实际应用中数值型数据和分类数据之间的转换还需要考虑落入每一个组中的观测数据的数量，组内数据观测频数太多或太少均影响统计分析结果。

3.1.2 卡方统计量

若使用 f_0 和 f_e 分别表示观测频数和期望频数，则卡方统计量可以如下定义。

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e} \tag{3-1}$$

依据概率统计的基本原理，如果从一个随机变量 X 所在的总体中随机抽取若干个观测样本，这些观测样本落在 X 的 k 个互不相交的子集中，其观测频数服从一个多项分布，而这个多项分布当 k 趋于无穷时近似服从卡方分布。基于此，对变量 X 的总体分布的检验可以从对各个观测频数的分析入手。观测频数和期望频数越是接近，则 $f_0 - f_e$ 的绝对值越小，计算出来的卡方值就越小，反之，卡方值就越大。卡方检验通过对卡方的计算结果与卡方

分布中的临界值进行比较，可以做出拒绝或者接受的统计决策。卡方检验的应用主要表现在 2 个方面：拟合优度检验和独立性检验。

3.2　卡方统计量基本应用

3.2.1　拟合优度检验

拟合优度检验使用卡方统计量进行，依据总体分布状况，计算出分类变量中各类别的理论期望频数，与经验分布的实际观测频数作比较，判断期望频数与观测频数是否有显著差异，据此对分类变量进行分析。

拟合优度检验的原假设如下。

H_0：样本来自总体的分布与假设的分布（理论分布）无显著差异。

H_1：样本来自总体的分布与假设的分布（理论分布）有显著差异。

如果卡方值较大，则说明期望频数与观测频数分布差距较大，没有证据支持原假设；如果卡方值较小，则说明期望频数与观测频数分布差距较小，不能拒绝原假设。拟合优度检验可以实现观测数据是否服从指定理论分布的推断研究。

卡方统计量的构建需要使用期望频数，对于连续分布的情况，通常用等分概率值计算期望频数，使用等分概率则可以得到相同的期望频数，便于计算。对于离散分布的情况，期望频数可以使用理论分布的概率值乘以总观测频数得到。

3.2.2　独立性检验

使用卡方统计量不但可以实现一个分类变量的检验，还可以研究 2 个分类变量是否存在联系。对 2 个分类变量的分析称为独立性检验，分析的过程可以通过列联表的形式呈现，因此，独立性检验也被称为列联分析。列联分析应用广泛，例如，海难发生时，对性别是否影响生存状况的研究，产品品牌的区域偏好研究及性别的天然禀赋研究等，列联分析均有应用。

1. 列联表

列联表是将 2 个以上的变量进行交叉分类的频数分布表，由于表中每个变量都可以有 2 个或 2 个以上的类别，列联表会有多种形式，如果行类别记为 R，列类别记为 C，则每一个具体的列联表可以记为 $R \times C$ 列联表。虽然列联表中行和列的位置可以任意放置，但是，若变量之间存在因果关系，通常的做法是把表示原因的变量放在行的位置，表示结果的变量放在列的位置，以便于更好地表示原因对结果的影响。例如，在研究家庭状况对青少年行为的影响时，家庭状况是影响青少年行为的重要因素，家庭状况是自变量，青少年行为是因变量，因此，把家庭状况放在行的位置，把青少年行为放在列的位置。

2. 独立性检验过程

独立性检验的原假设如下。

H_0：行变量和列变量之间是独立的（不存在依赖关系）。

H_1：行变量和列变量之间不独立（存在依赖关系）。

通过计算各个单元格中的期望频数就可以构建卡方统计量，完成列联分析。采用下列公式可以计算任何一个单元格中频数的期望值：

$$f_e = \frac{RT}{n} \times \frac{CT}{n} \times n = \frac{RT \times CT}{n} \tag{3-2}$$

其中：f_e 为给定单元中的期望频数值；RT 为给定单元所在行的合计频数值；CT 为给定单元所在列的合计频数值；n 为观测值的总频数，即样本量。

卡方统计量的自由度（degree of freedom，df），$df = (R-1)(C-1)$。

计算出卡方统计量的值后，查表可以获得相应的临界值。如果卡方统计量的值大于临界值，则拒绝零假设 H_0，认为行变量和列变量之间存在依赖关系；否则，卡方统计量的值小于临界值，则不能拒绝零假设 H_0，现有证据不足以支持行变量和列变量之间存在依赖关系，即行变量和列变量之间不存在依赖关系。

3. 相关程度测量

如果两个分类变量之间相互独立，说明两者之间没有关系；反之，则认为两者之间存在联系。如果两个分类变量之间有联系，可以使用相关系数来度量这种联系的密切程度。列联表中的变量通常是类别变量，所表现的是研究对象的不同品质类别，因此，这种分类数据之间的相关称为品质相关，经常用到的品质相关系数主要有 φ 相关系数、C 列联系数和 V 相关系数 3 种。

φ 相关系数是描述 2×2 列联表数据相关程度的最常用的一种相关系数。根据如下公式计算：

$$\varphi = \sqrt{\frac{\chi^2}{n}} \tag{3-3}$$

其中：χ^2 是按照上文所述计算得到的卡方值；n 是列联表的总频数，即样本量。对于 2×2 的列联表而言，计算的 φ 系数可以控制在 0 到 1 之间。如果列联表的行数 R 或者列数 C 大于 2，φ 系数将随着 R 或 C 的变大而增大，φ 系数值没有上限。此时，可以考虑列联系数。

列联系数 C 主要用于测量列联表大于 2×2 的情况，根据如下公式计算：

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \tag{3-4}$$

由上式可以看出，C 系数不可能大于 1，C 系数的特点是其可能的最大值依赖于列联表的行数和列数，且随着 R 和 C 的增大而增大。当两个变量完全相关时：对于 2×2 列联表，C 系数值为 0.707 1；对于 3×3 列联表，C 系数值为 0.816 5；对于 4×4 列联表，C 系数值为 0.87。如果两个列联表中的行数和列数不一致，计算得到的列联系数则不便于比较。尽管如此，由于其计算简便，且对总体分布没有任何要求，列联系数仍有广泛的适用性。

考虑到 φ 系数无上限，C 系数小于 1，不同行和列的列联表不方便比较系数的情况，V 相关系数被提出。V 系数的计算公式如下：

$$V=\sqrt{\frac{\chi^2}{n\times\min[(R-1),(C-1)]}} \tag{3-5}$$

上式中 $\min[(R-1),(C-1)]$ 表示取（$R-1$）和（$C-1$）中较小的值。当两个分类变量相互独立时，$V=0$；当两个变量完全相关时，$V=1$。所以 V 的取值在 0 到 1 之间。若列联表中存在行或者列为 2 的情况，则 $\min[(R-1),(C-1)]=1$，此时 V 值等于 φ 系数值。

对于同样的分类数据，系数 φ、C 和 V 的度量结果不同。对于不同规格的列联表，行数和列数的差异也会影响相关系数测量值。因此，在对不同列联表变量之间的相关程度进行比较时，不同列联表中行与行、列与列的个数要相同，并且采用同一种系数，得到的系数值才具有可比性。

3.3　卡方检验应用条件

根据上文卡方统计量的计算公式（3-1）：$\chi^2=\sum\frac{(f_0-f_e)^2}{f_e}$ 可以看出，如果期望频数 f_e 的值过小，则（f_0-f_e）的值将会不适当地增大，导致卡方统计量值被高估，从而导致不适当地拒绝零假设 H_0，因此，在使用卡方统计量进行卡方检验时需要对频数进行限制。关于小单元的频数通常有两条准则：

准则一，若只有 2 个单元，则每个单元的期望频数必须是 5 或 5 以上；

准则二，若存在 2 个以上的单元且 20%的单元期望频数小于 5，则不能使用卡方检验。

解决方法是将某些类别合并，使得期望频数不小于 5。例如，某型白血病患者病情缓解时间的虚拟实验数据，经过分组整理后分别如表 3-1 和表 3-2 所示。

表 3-1　可以计算卡方值的数据情况

类别	观测频数	期望频数	类别	观测频数	期望频数
A	18	20	E	10	11
B	12	14	F	25	20
C	6	3	频数合计	80	80
D	9	12			

表 3-2　不可以使用卡方检验的数据情况

类别	观测频数	期望频数	偏差	偏差平方	卡方值	合并后卡方值
A	29	31	–2	4	4/31	4/31
B	15	17	–2	4	4/17	4/17
C	17	15	2	4	4/15	4/15
D	10	13	–3	9	9/13	9/13
E	5	1	4	16	16	25/4
F	4	3	1	1	1/3	
合计	80	80	—	—	17.656 6	7.573 3

表 3-1 中存在 6 个单元类，只有 1 个单元类的期望频数小于 5，根据准则二，表 3-1 中的数据可以计算卡方值，进行卡方检验。表 3-2 中同样存在 6 个单元类别，有 2 个单元类的期望频数小于 5，因此，表 3-2 中的数据不能用于卡方检验。如果强行计算卡方，则可以得到卡方值为 17.656 6，使用显著水平为 0.05 的卡方临界值 $\chi^2_{0.05}(5)=11.070\,5$ 进行检验，因为 17.656 6>11.070 5，所以拒绝原假设，认为观测频数与期望频数之间存在显著差异。但是，仔细观察此数据发现，观测频数与期望频数之间拟合比较理想，两者之间并不存在显著差异。因此，推断结论不符合实际情况。如果将表 3-2 中的 E 和 F 两个单元类合并，使得合并后单元类的期望频数不小于 5，即可解决此问题。重新计算结果如表 3-2 中所示，此时卡方值为 $7.573\,3<\chi^2_{0.05}(4)=9.487\,7$，不能拒绝零假设，得到观测频数与期望频数之间不存在显著差异的更为合理的结论。

此外，关于列联表行和列的放置问题，前文已做说明，此处不再赘述。

3.4 卡方检验实现方法与步骤

根据卡方统计量的构建形式，可以使用列表的方法计算卡方统计量的值，如表 3-2 的计算过程。使用统计分析软件则可以更加便捷地完成统计量的计算与卡方检验的分析工作，SPSS 和统计分析系统（statistical analysis system，SAS）是常用的统计分析软件，下文使用 SPSS 软件对卡方检验的分析过程进行详细说明。

第 1 步：在 SPSS 数据编辑窗口中定义分类变量。

第 2 步：选择“数据”下拉菜单，并选择“加权个案”选项进入主对话框。在主对话框中选择“加权个案”，将相应的变量选入“频率变量”。

第 3 步：选择“分析”下拉菜单，再选择“描述统计”下拉菜单，最后选择“交叉表”选项进入主对话框。在主对话框中将相应的变量分别选入“行”和“列”。

第 4 步：在“统计量”下选择“卡方”。在“名义”下选择“相依系数”和“Phi 和 Cramer 变量”。

第 5 步：单击“继续”按钮，单击“确定”按钮。

SPSS 统计分析系统将报告卡方值和相应的 3 个相关系数。

如果熟悉 SAS 系统，建立数据集后直接使用 FREQ 过程，通过 Tables 和 weight 命令也可以直接得到上述统计量与相关系数值。

3.5 应用案例分析

3.5.1 独立性检验案例 1：超大样本情况——教育机会的公平性研究

想要了解改革开放 40 多年来，女性是否与男性一样享有平等的教育机会，以及教育程

序上是否还存在性别差别，那么应该如何使用卡方统计量来进行统计分析呢？

沿用第 2 章例 2-1 数据，我们把受教育程度划分为 3 种类型：小学及以下（包括未上过学和小学）、中学（包括初中、普通高中和中职）、大学及以上（包括大学专科、大学本科和研究生）；性别变量分为男性和女性 2 种类型。根据第 2 章例 2-1 数据形成频数分布表，如表 3-3 所示，使用卡方统计量进行教育机会的公平性研究。

表 3-3　2018 年全国按性别和受教育程度划分的 6 岁及以上人口的频数分布表　　单位：人

观测频数		受教育程度			总计
		小学及以下	中学	大学及以上	
性别	男	144 428	320 324	77 616	542 368
	女	182 005	268 334	71 487	521 826
总计		326 433	588 658	149 103	1 064 194

资料来源：国家统计局网站，www.stats.gov.cn。本表是 2018 年全国人口变动情况抽样调查样本数据，抽样比为 0.820‰。

表 3-3 中列是受教育程度变量，划分为 3 类；行是性别变量，划分为 2 类，因此，表 3-3 是 2×3 列联表。依据独立性检验原理，分析行变量和列变量是否相互独立，即受教育程度是否独立于性别。建立的零假设和备择假设分别如下。

H_0：受教育程度和性别之间不存在依赖关系（独立）。

H_1：受教育程度和性别之间存在依赖关系（不独立）。

根据卡方统计量的构造公式（3-1），需要计算期望频数值。依据单元格频数期望值计算公式（3-2）可以获得相应的频数期望值。计算过程以及四舍五入的结果如下：

$$f_{11}=\frac{542\,368}{1\,064\,194}\times\frac{326\,433}{1\,064\,194}\times1\,064\,194=166\,367\ ;$$

$$f_{12}=\frac{542\,368}{1\,064\,194}\times\frac{588\,658}{1\,064\,194}\times1\,064\,194=300\,010\ ;$$

$$f_{13}=\frac{542\,368}{1\,064\,194}\times\frac{149\,103}{1\,064\,194}\times1\,064\,194=75\,991\ ;$$

$$f_{21}=\frac{521\,826}{1\,064\,194}\times\frac{326\,433}{1\,064\,194}\times1\,064\,194=160\,066\ ;$$

$$f_{22}=\frac{521\,826}{1\,064\,194}\times\frac{588\,658}{1\,064\,194}\times1\,064\,194=288\,648\ ;$$

$$f_{23}=\frac{521\,826}{1\,064\,194}\times\frac{149\,103}{1\,064\,194}\times1\,064\,194=73\,112\ 。$$

依据公式（3-1）计算的卡方统计量的值为：

$$\chi^2=\frac{(144\,428-166\,367)^2}{166\,367}+\cdots+\frac{(71\,487-73\,112)^2}{73\,112}=8\,776$$

χ^2 的自由度 $=(R-1)(C-1)=(2-1)(3-1)=2$；令显著性水平 $\alpha=0.05$，查表知道：$\chi^2_{0.05}(2)=5.9915$。由于 $\chi^2>\chi^2_{0.05}(2)$，因此拒绝零假设，接受备择假设，认为受教育程度和性别有关联。

使用 SPSS 软件进行独立性检验的操作步骤分别如下。

第一步：在 SPSS 数据编辑窗口中定义受教育程度变量，取值为 1、2 和 3，分别表示小学及以下、中学和大学及以上；定义性别变量取值为 0 和 1，分别表示女性和男性。建立数据集如表 3-4 所示。

表 3-4　SPSS 系统中的数据文件

性别	受教育程度	频数	性别	受教育程度	频数
0	1	182 005	1	1	144 428
0	2	268 334	1	2	320 324
0	3	71 487	1	3	77 616

第二步：依据上文实现方法和操作说明，在对话框中完成相应的菜单选项，执行以后可以得到如表 3-5 系统输出的统计分析报告。

表 3-5　性别*受教育程度交叉制表

			受教育程度			合计
			1	2	3	
性别	0	计数	182 005	268 334	71 487	521 826
		期望的计数	160 066.0	288 647.6	73 112.4	521 826.0
	1	计数	144 428	320 324	77 616	542 368
		期望的计数	166 367.0	300 010.4	75 990.6	542 368.0
合计		计数	326 433	588 658	149 103	1 064 194
		期望的计数	326 433.0	588 658.0	149 103.0	1 064 194.0

系统输出的卡方检验结果如表 3-6 所示。

表 3-6　卡方检验

	值	自由度	渐进 Sig.（双侧）
皮尔森卡方	8 776.058[a]	2	0
似然比	8 788.427	2	0
有效案例中的 N	1 064 194		

注：a. 0 单元格（0%）的期望计数少于 5。最小期望计数为 73 112.44。

根据卡方检验结果，可以判断出性别和受教育程度是有关联的，即性别是影响受教育程度的。两者之间的相关程度如表 3-7 系统输出结果显示。

表 3-7　对称度量

		值	近似值 Sig.
按标量标定	φ	0.091	0
	克莱姆的 V	0.091	0
	相依系数	0.090	0
有效案例中的 N		1 064 194	

注：a. 不假定零假设。

b. 使用渐进标准误差假定零假设。

c. 基于正态近似值。

由表 3-7 可以看出，φ 系数、V 系数和 C 系数分别为 0.091、0.091 和 0.090。根据卡方统计量的值，可以认为性别对受教育程度是存在影响的。但是，两者之间的相关系数约为 0.09，表明性别对受教育程度的影响几乎不存在。为什么会产生这种矛盾的结果呢？这就涉及到，对于超大样本，卡方检验结果的可信性问题。随着样本增大，卡方值通常会相应增大，任何细微的差异都可能导致统计上的显著性，同时，根据 φ 系数、V 系数和 C 系数的构造公式，样本量位于分母位置，样本量越大，相关系数必然越小。

通常在使用卡方统计量进行列联分析时要注意卡方分布的期望值准则，保证小单元格的频数要求，但是这未必能解决此处提出的超大样本的卡方检验适用性问题。仍以第 2 章例 2-1 数据为样本举例说明如下。

把受教育程度分为 8 组，分别用 1、2、3、4、5、6、7、8 表示未上过学、小学、初中、普通高中、中职、大学专科、大学本科和研究生 8 种类型；性别分为男和女 2 组，分别用 1 和 0 表示，形成如表 3-8 数据表。

表 3-8　性别和受教育程度频数表

性别	受教育程度	频数	性别	受教育程度	频数
0	1	41 001	1	1	16 481
0	2	141 004	1	2	127 947
0	3	185 260	1	3	216 604
0	4	61 388	1	4	77 363
0	5	21 686	1	5	26 357
0	6	37 416	1	6	41 065
0	7	30 995	1	7	33 263
0	8	3 076	1	8	3 288

SPSS 系统报告的分析结果分别如表 3-9、表 3-10、表 3-11 所示。

表 3-9　性别*受教育程度交叉制表

			受教育程度								合计
			1	2	3	4	5	6	7	8	
性别	0	观测	41 001	141 004	185 260	61 388	21 686	37 416	30 995	3 076	521 856
		期望	28 186	131 880	197 053	68 036	23 558	38 483	31 509	3 121	521 856
	1	观测	16 481	127 947	216 604	77 363	26 357	41 065	33 263	3 288	542 368
		期望	29 296	137 071	204 811	70 715	24 485	39 998	32 749	3 243	542 368
合计		观测	57 482	268 951	401 864	138 751	48 043	78 481	64 258	6 364	1 064 194
		期望	57 482	268 951	401 864	138 751	48 043	78 481	64 258	6 364	1 064 194

表 3-10　卡方检验

	值	自由度	渐进 Sig.（双侧）
皮尔森卡方	15 697.581	7	0
似然比	16 042.155	7	0
有效案例中的 N	1 064 194		

注：0 单元格（0%）的期望计数少于 5。最小期望计数为 3 120.58。

表 3-11 对称度量

		值	近似值 Sig.
按标量标定	φ	0.121	0
	克莱姆的 V	0.121	0
	相依系数	0.121	0
有效案例中的 N		1 064 194	

注：a. 不假定零假设。

b. 使用渐进标准误差假定零假设。

c. 基于正态近似值。

根据卡方检验可以判断出性别和受教育程度是有关联的，即性别是影响受教育程度的，但是，列联系数则反映了不同的结果，这表明增加分类变量类别细分数据的方法并不能解决超大样本的卡方检验问题。

那么减少分类数据的类型是否有助于缓解该问题，实例说明如下。把受教育程度变量划分为 2 类：大学专科及以下，用 1 表示；本科及以上，用 2 表示。性别变量不变，形成如表 3-12 所示频数分布表。

表 3-12 性别和受教育程度频数表

性别	受教育程度	频数	性别	受教育程度	频数
0	1	487 755	1	1	505 817
0	2	34 071	1	2	36 551

再次执行 SPSS 程序，结果报告如表 3-13、表 3-14、表 3-15 所示。

表 3-13 性别*受教育程度交叉制表

			教育		合计
			1	2	
性别	0	计数	487 755	34 071	521 826
		期望	487 197	34 629	521 826
	1	计数	505 817	36 551	542 368
		期望	506 375	35 993	542 368
合计		计数	993 572	70 622	1 064 194
		期望	993 572	70 622	1 064 194

表 3-14 卡方检验

	值	自由度	渐进 Sig.（双侧）
皮尔森卡方	18.923	1	0
似然比	18.927	1	0
有效案例中的 N	1 064 194		

注：a. 0 单元格（0%）的期望计数少于 5。最小期望计数为 34 629.40。

b. 仅对 2×2 表计算。

表 3-15　对称度量

		值	近似值 Sig.
按标量标定	φ	0.004	0
	克莱姆的 V	0.004	0
	相依系数	0.004	0
有效案例中的 N		1 064 194	

注：a. 不假定零假设。

b. 使用渐进标准误差假定零假设。

c. 基于正态近似值。

同样可以发现，显著的卡方统计量值与不显著的相关系数并存，矛盾并没有因数据粗化而解决。综上所述，在超大样本的情况下，不但要综合卡方统计量和列联系数的值，还需要根据研究问题的具体情况判断分析结果是否符合实际意义。

3.5.2　独立性检验案例 2：大样本情况——我国区域经济发展的均衡性研究

改革开放40 多年来，我国经济飞速发展，为了了解我国中、东、西部区域经济的发展速度和发展水平是否均衡，我们收集了《中国统计年鉴 2019》中 2018 年各省份人均 GDP 数据（单位：万元），同时，标注了各省份所属于的区位，如表 3-16 所示。

表 3-16　各地区人均 GDP（2018）　　单位：万元

地区	区位	人均 GDP	地区	区位	人均 GDP
北京	东部	14.02	安徽	中部	4.77
天津	东部	12.07	福建	东部	9.12
河北	东部	4.78	江西	中部	4.74
山西	中部	4.53	山东	东部	7.63
内蒙古	西部	6.83	河南	中部	5.02
辽宁	东部	5.80	湖北	中部	6.66
吉林	东部	5.56	湖南	中部	5.29
黑龙江	东部	4.33	广东	东部	8.64
上海	东部	13.50	广西	西部	4.15
江苏	东部	11.52	海南	东部	5.20
浙江	东部	9.86	重庆	西部	6.59
四川	西部	4.89	陕西	西部	6.35
贵州	西部	4.12	甘肃	西部	3.13
云南	西部	3.71	青海	西部	4.77
西藏	西部	4.34	宁夏	西部	5.41
新疆	西部	4.95			

资料来源：国家统计局网站，www.stats.gov.cn。

一般情况下，一组数据所分的组数应不少于 5 组且不多于 15 组，我们综合考虑到数据的多少和特点及卡方分布的期望值准则，把数据划分为 5 组，得到各省份人均 GDP 的频数分布表，如表 3-17 所示。

表 3-17　各地区人均 GDP 频数分布表（2018 年）

组别	按人均 GDP 分组/万元	频数/个	频率/%
1	4.5 以下	6	19.35
2	4.5~6.0	13	41.94
3	> 6.0~7.5	4	12.9
4	> 7.5~12.0	5	16.13
5	12.0 以上	3	9.68
合计	—	31	100

我们把区域分为东部、中部和西部 3 类，分别用 1、2 和 3 表示。经济发展水平分为 5 组，分别用 1、2、3、4 和 5 表示。形成如表 3-18 所示的交叉频数分布表。

表 3-18　3×5 交叉频数分布表

		经济发展水平					合计
		1	2	3	4	5	
区位	东 1	1	4	0	5	3	13
	中 2	0	5	1	0	0	6
	西 3	5	4	3	0	0	12
合计		6	13	4	5	3	31

建立的零假设和备择假设分别如下。

H_0：区位与经济发展水平之间不存在依赖关系（独立）。

H_1：区位与经济发展水平之间存在依赖关系（不独立）。

使用 SPSS 软件进行独立性检验，系统报告结果如表 3-19、表 3-20、表 3-21 所示。

表 3-19　区位*发展交叉制表

			经济发展水平					合计
			1	2	3	4	5	
区位	1	计数	1	4	0	5	3	13
		期望	2.5	5.5	1.7	2.1	1.3	13
	2	计数	0	5	1	0	0	6
		期望	1.2	2.5	0.8	1.0	0.6	6.0
	3	计数	5	4	3	0	0	12
		期望	2.3	5.0	1.5	1.9	1.2	12
合计			6	13	4	5	3	31
			6	13	4	5	3	31

表 3-20　卡方检验

	值	自由度	渐进 Sig.（双侧）
皮尔森卡方	22.393[a]	8	0.004
似然比	26.761	8	0.001
有效案例中的 N	31		

注：a. 13 单元格（86.7%）的期望计数少于 5。最小期望计数为 0.58。

表 3-21　对称度量

		值	近似值 Sig.
按标量标定	φ	0.850	0.004
	克莱姆的 V	0.601	0.004
	相依系数	0.648	0.004
有效案例中的 N		31	

从表 3-20 可以看出，卡方值为 22.393，相伴概率为 0.004，故应该拒绝原假设，认为经济发展水平受区位影响。从表 3-21 可知 φ 系数为 0.850，但是本例 3×5 列联表的行数和列数均大于 2，φ 系数将随着行或列的变大而增加，且 φ 系数值无上限，这时依据 φ 系数判断两个变量之间的相关程度就不够准确。对于 C 系数和 V 系数而言，0.648 和 0.601 均相对较高，因此，可以认为区位对经济发展的影响是比较显著的。

从本例中可以看出，对于大样本而言，当卡方检验结果显著，表明两个分类变量之间存在显著相关性时，品质相关系数表达了一致的内容，此时独立性检验效果比较理想。除了进行列联分析之外，使用卡方统计量还可以进行拟合优度的检验，详见下例。

3.5.3　拟合优度检验案例 3：基本原理——机动车辆保险理赔次数研究

改革开放 40 多年来，汽车已成为人们日常生活中不可或缺的一部分，各类保险公司推出的机动车辆保险产品数量不断增加，保险公司之间的产品竞争也日趋激烈。因此，了解机动车辆保险的损失频率对于设计更具竞争力的保险产品至关重要。表 3-22 给出了某非寿险公司 10 000 份机动车辆保险的损失数据，使用拟合优度检验判断理赔次数是否符合泊松分布。

表 3-22　某非寿险公司 100 000 份机动车辆保险损失频率和累积频率

理赔次数	观测到的保单数	频率/%	累计频率/%
0	88 585	88.585	88.585
1	105 77	10.577	99.162
2	779	0.779	99.941
3	54	0.054	99.995
4	4	0.004	99.999
5	1	0.001	100.00
合计	100 000	100.000	—

资料来源：韩天雄. 非寿险精算[M]. 北京：中国财政经济出版社，2010，36.

概率图（probability probability plot，P-P）和分位图（quantile quantile plot，Q-Q）虽然可以直观显示数据拟合的优劣程度，但是，有时候需要通过严密的数学论证才能证明结论的可信性，统计假设检验即为可以信赖的一种论证方法。在假设检验中，通常需要先设定原假设和备择假设。

H_0：数据来源于某个给定的总体。

H_1：数据并非来源于给定的总体。

卡方拟合优度检验常用于离散分布的情况，如果是连续分布则需要通过分组的方式把数据分成多个区间进行考虑。当观测值数量充分大时，统计量会收敛于自由度为 $R-1$ 的卡方分布，其中 k 为分组或分类数。如果计算得到的统计量的值大于临界值 $\chi_\alpha^2(k-1)$，则拒绝原假设，表明原假设中的分布不能拟合样本数据。否则，无法拒绝原假设。通常显著性水平取 0.05。

在卡方拟合优度检验中，要求样本容量足够大且期望频数不太小。为提高模型估计精度，通常要求期望频数不小于 5，总样本量不小于 50。如果不能满足要求，需要将频数较小的组合并。研究表明，当每个组的期望频数大致相等时，检验的效果是最优的。

使用卡方统计量进行拟合优度检验可以判断理赔次数是否服从指定的分布，基本流程大致如下。首先确定泊松分布的参数，本例使用极大似然估计确定分布参数。

泊松分布的概率分布函数为

$$P\{N=k\}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\lambda>0;k=0,1,2,\cdots \tag{3-6}$$

似然函数为

$$L(k;\lambda)=\prod_{i=1}^{n}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} \tag{3-7}$$

对数似然函数为

$$\ln L(k;\lambda)=-n\lambda+\ln\lambda\cdot\sum_{i=1}^{n}k_i-\sum_{i=1}^{n}\ln k_i! \tag{3-8}$$

对式（3-8）求偏导数并令其为 0，有

$$\frac{\partial\ln L}{\partial\lambda}=-n+\frac{\sum_{i=1}^{n}k_i}{\lambda}=0$$

解得 $\hat{\lambda}=\bar{x}$，根据表中数据可以算出样本均值为

$$\bar{x}=\frac{88\,585}{100\,000}\times0+\frac{10\,577}{100\,000}\times1+\frac{779}{100\,000}\times2+\frac{54}{100\,000}\times3+\frac{4}{100\,000}\times4+\frac{1}{100\,000}\times5=0.123\,18$$

因此，泊松分布的参数为 0.123 18，泊松分布为：$P\{N=k\}=\frac{0.123\,18^k}{k!}\times\mathrm{e}^{-0.123\,18}$。

建立原假设和备择假设如下。

H_0：理赔次数服从泊松分布。

H_1：理赔次数不服从泊松分布。

如果 H_0 成立，那么我们可以计算出当理赔次数分别为 0、1、2、3、4、5 的时候，预期能够观测到的保单数，也即前文中的期望频数，计算公式为

$$E_i = N \times P(N = k_i) \tag{3-9}$$

本例概率及期望频数的计算过程如下。

$$P\{N=0\} = \frac{\lambda^0}{0!}e^{-\lambda} = e^{-0.123\,18} = 0.884\,10;$$

$$E_0 = N \times P(N=0) = 100\,000 \times 0.884\,10 = 88\,410$$

$$P\{N=1\} = \frac{\lambda^1}{1!}e^{-\lambda} = 0.123\,18e^{-0.123\,18} = 0.108\,90;$$

$$E_1 = N \times P(N=1) = 100\,000 \times 0.108\,90 = 10\,890$$

$$P\{N=2\} = \frac{\lambda^2}{2!}e^{-\lambda} = 0.123\,18^2 e^{-0.123\,18} = 0.006\,71;$$

$$E_2 = N \times P(N=2) = 100\,000 \times 0.006\,71 = 671$$

$$P\{N=3\} = \frac{\lambda^3}{3!}e^{-\lambda} = 0.123\,18^3 e^{-0.123\,18} = 0.000\,275;$$

$$E_3 = N \times P(N=3) = 100\,000 \times 0.000\,275 = 275$$

$$P\{N=4\} = \frac{\lambda^4}{4!}e^{-\lambda} = 0.123\,18^4 e^{-0.123\,18} = 0.000\,008\,481;$$

$$E_4 = N \times P(N=4) = 100\,000 \times 0.000\,008\,481 = 0.848\,1$$

$$P\{N=5\} = \frac{\lambda^5}{5!}e^{-\lambda} = 0.123\,18^5 e^{-0.123\,18} = 0.000\,000\,209;$$

$$E_5 = N \times P(N=5) = 100\,000 \times 0.000\,000\,209 = 0.020\,9$$

根据卡方统计量的构造公式，可以得到卡方拟合优度检验的结果。自由度为：$\mathrm{df} = k - 1$，其中 k 为分类或者分组的个数。上述数据列于表 3-23。

表 3-23　拟合优度检验

理赔次数	概率	观测频数	期望频数	卡方
0	0.884 10	88 585	88 410	0.346 4
1	0.108 90	10 577	10 890	8.996 2
2	0.006 71	779	671	17.383 0
3	0.000 275	54	27.5	25.536 4
4	0.000 008 481	4	0.848 1	11.713 5
5	0.000 000 209	1	0.020 9	45.881 5
合计	1	100 000	100 000	109.8

自由度为 5，查卡方分布表，在 0.05 的显著性水平下，卡方临界值为 9.487 7，远远小于 109.8，据此可以拒绝零假设，认为理赔次数并不服从泊松分布。

需要注意，虽然总样本量远远大于 50，但是当理赔次数为 4 和 5 时，期望频数较小，不满足大于 5 的条件，因此，有可能造成对卡方的高估，不合理地增加了拒绝 H_0 的可能性。为解决这一问题，把理赔次数为 3、4 和 5 的情况进行合并。合并以后的计算结果如

表 3-24 所示。

表 3-24 理赔次数 3、4 和 5 合并后的拟合优度检验

理赔次数 k	概率	观测频数	期望频数	卡方
0	0.884 10	88 585	88 410	0.346 4
1	0.108 90	10 577	10 890	8.996 2
2	0.006 71	779	671	17.383 0
3 及以上	0.000 283 69	59	28.369	33.073
合计	1	100 000	100 000	59.8

可以看出新的卡方值为 59.80，远远小于没有合并时的 109.86，估计精度是有所改进的，自由度为 3，在显著性水平为 0.05 的条件下，卡方临界值为 7.814 7，59.80 远远大于 7.814 7，因此，拒绝零假设。

虽然满足了期望频数不小于 5 的条件，但是，观测频数之间仍然存在较大差异，理赔次数为 1 的频数为 10 577，而理赔次数为 2 的频数为 779，理赔次数为 3 及以上的频数为 59。期望频数之间的差异也是如此。为了进一步提高检验效果，再次把 2 及以上各组合并，再次计算，结果如表 3-25 所示。

表 3-25 理赔次数 2、3、4 和 5 合并后的拟合优度检验

理赔次数 k	概率	观测频数	期望频数	卡方
0	0.884 10	88 585	88 410	0.346 4
1	0.108 90	10 577	10 890	8.996 2
2 及以上	0.006 993 69	838	699.369	27.480
合计	1	100 000	100 000	36.8

此时新的卡方值为 36.82，进一步缩小，自由度为 2，显著性水平为 0.05 时卡方临界值为 5.991 5，36.82 远大于 5.991 5，因此拒绝零假设，认为理赔次数不服从泊松分布。

尽管例中 3 个卡方值均可以拒绝理赔次数服从泊松分布的零假设，但是，适当合并期望频数较小的组得到的结果具有更高的置信性，具体分析结论依赖于实际研究情境。

3.5.4 拟合优度分析案例 4：离散型分布情况——环境污染级别研究

改革开放40 多年来，经济快速发展的同时伴随着能源消耗的不断增加，环境污染问题日益引起人们的警觉，表 3-26 的数据是某环境测量站在一年中观测到的大气污染情况。在“绿水青山就是金山银山”的绿色发展理念下，我们想知道污染级别是否服从泊松分布，要求对原假设“污染级别服从泊松分布”进行卡方拟合优度检验。

表 3-26 污染级别情况表

污染级别	0	1	2	3	合计/天
观测天数	50	122	101	92	365

根据案例 3 分析，首先估计泊松分布参数，使用极大似然估计或矩估计方法，容易知道泊松分布的参数为 1.644。然后，计算污染级别分别为 0、1、2 和 3 的概率，计算期望频数，最后，构造卡方统计量，具体的计算结果列于表 3-27。

表 3-27　离散分布的拟合优度检验

污染级别	概率	观测频数	期望频数	卡方
0	0.193 2	50	70.518	5.969 9
1	0.317 6	122	115.924	0.318 5
2	0.261 1	101	95.301 5	0.340 8
3	0.228 1	92	83.256 5	0.918 2
合计	1	365	365	7.5

查卡方分布表，在 0.05 的显著性水平下，自由度为 3 的卡方临界值为 7.814 7。计算的卡方值为 7.5，小于临界值 7.814 7，据此无法拒绝零假设，认为理赔次数服从泊松分布。

3.5.5　拟合优度分析案例 5：连续型分布情况——病情缓解时间研究

维护人民健康是建设健康中国的重要保证，改革开放 40 多年以来，各级政府部门不断加大健康服务投入，努力提高医疗卫生的服务质量和水平。医院某科室为了解白血病患者的病情缓解情况，对 21 例急性白血病患者进行了观察，获得如下样本数据：1、1、2、2、3；4、4、5、5、6；8、8、9、10、10；12、14、16、20、24；34。原假设 H_0：病情缓解时间的分布密度为 $f(t)=0.12\mathrm{e}^{-0.12t},(t>0)$。使用卡方拟合优度检验可以对零假设进行统计判断。

本例分布密度函数的参数是已知的，不需要进行参数估计，但是观测频数需要进一步整理才可以进行假设检验。

首先，我们需要将样本数据进行分组。根据期望频数的计算公式（3-9），如果每个组被选中的概率是一样的，那么每个组的期望频数也会相同。因此，在将连续性变量进行分组时，为了简化计算过程，我们通常会选择让观测值落入各个区间或组内的概率相等，也就是等分这些观测值。本例有 21 个观测数据，分为四组较为合适，因此确定每组概率为 0.25，据此可以确定组限，假设 x 表示组限，结果如下：

$$p_1=0.25\Rightarrow\int_0^x 0.12\mathrm{e}^{-0.12t}\mathrm{d}t=0.25\Rightarrow 1-\mathrm{e}^{-0.12x}=0.25$$
$$\Rightarrow x=2.39$$

因此，第一组为[0，2.39]，第二组组限为：

$$p_2=0.25\Rightarrow\int_{2.4}^x 0.12\mathrm{e}^{-0.12t}\mathrm{d}t=0.25$$
$$\Rightarrow x=5.78$$

因此，第二组为[2.4，5.78]。依此类推，第三组和第四组分别为[5.8，11.59]及[11.6，$+\infty$]。通过等分概率，可以保证 21 个观测数据落入这 4 个区间的理论期望频数，有：

$$E_i=N\times P(N=k_i)=21\times 0.25=5.25$$

然后，统计观测频数，在所有 21 个观测值中：只有 1、1、2、2 共计 4 个观测数据落入了第一个区间[0，2.39]，因此，落入该区间的观测频数为 4；有 3、4、4、5、5 共计 5 个观测数据落入第二个区间[2.4，5.78]，所以该区间的观测频数为 5；同样，落入第三和第四区间的观测频数分别为 6 和 6。

最后，构造卡方统计量，具体的计算结果列于表 3-28。

表 3-28　连续分布的拟合优度检验

区间	概率	观测频数	期望频数	卡方
[0，2.39]	0.25	4	5.25	0.298
[2.4，5.78]	0.25	5	5.25	0.012
[5.8，11.59]	0.25	6	5.25	0.107
[11.6，+∞]	0.25	6	5.25	0.107
合计	1	21	21	0.52

查卡方分布表，在 0.05 的显著性水平下，自由度为 3 的卡方临界值为 7.814 7。计算的卡方值为 0.52 小于临界值 7.814 7，据此无法拒绝零假设，认为病情的缓解时间服从密度为“$f(t)=0.12\mathrm{e}^{-0.12t},(t>0)$”的分布。

3.5.6　拟合优度检验的 SPSS 实现

使用 SPSS 软件可以实现拟合优度检验，以案例 4 和案例 5 为例说明如下。

第一步：建立 SPSS 数据集。

针对案例 4 数据，在 SPSS 变量视图中定义：污染级别和观测天数；在数据视图中输入污染级别和观测天数的数值，分别为 0、1、2、3 和 50、122、101、92，完成数据集的构建，结果如图 3-1 所示。

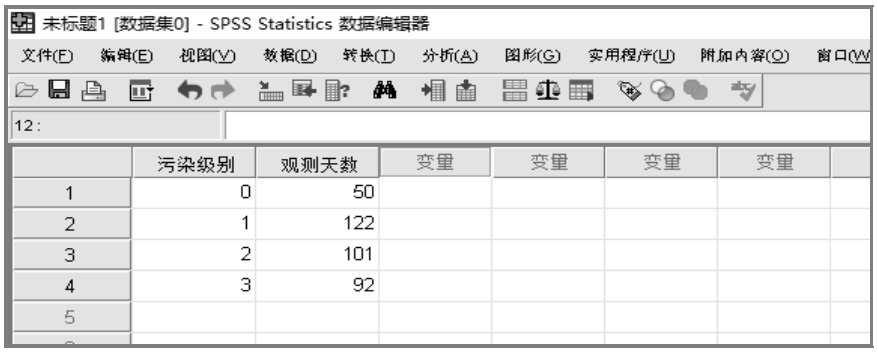

图 3-1　建立数据集

第二步：确定频率变量。

单击主菜单“数据”和下拉子菜单“加权个案”，选中“加权个案”选项，把变量“观测天数”移入频率变量对话框。结果如图 3-2 所示。单击“确定”按钮。

图 3-2　确定频率变量

第三步：确定检验变量和期望频数。

单击主菜单“分析”和下拉子菜单“非参数检验”，单击“卡方”按钮，把变量“污染级别”移入检验变量列表对话框。在“期望值”对话框中有 2 个选项，其中“值”选项一般适用于离散分布的拟合优度检验问题，而选项“所有类别相等”适用于连续分布的拟合问题，说明如下。

1. 离散分布情况

案例 4 中已经计算了期望频数，污染级别 0、1、2 和 3 对应的期望频数分别为 70.518、115.924、95.301 5 和 83.256 5。此 4 个期望值并不相同，因此，本例需要选择“值”按钮，把此 4 个期望值分别依次输入系统，每次输入一个期望频数之后，单击“添加”按钮。结果如图 3-3 所示。

图 3-3　确定检验变量

单击“确定”按钮，SPSS 系统输出拟合优度检验的最终输出报告，分别如表 3-29、表 3-30 所示。

表 3-29 污染级别

	观察数/天	期望数	残差
0	50	70.5	–20.5
1	122	115.9	6.1
2	101	95.3	5.7
3	92	83.3	8.7
总数	365		

表 3-30 检验统计量

	污染级别
卡方	7.547[a]
自由度	3
渐进显著性	0.056

注：a. 0 个单元（0%）具有小于 5 的期望频率。单元最小期望频率为 70.5。

由输出报告可知，渐进显著性概率值为 0.056，大于显著性水平 0.05，因此无法拒绝零假设，认为理赔次数服从泊松分布。

2. 连续分布情况

在连续分布情况下，需要在“期望值”对话框中选择“所有类别相等”选项，就可以完成连续分布的拟合优度检验问题。案例 5 数据的 SPSS 输出结果如表 3-31、表 3-32 所示。

表 3-31 病情缓解时间

	观察频数	期望频数	残差
[0，2.39]	4	5.25	–1.25
[2.4，5.78]	5	5.25	–0.25
[5.8，11.59]	6	5.25	0.75
[11.6，+∞]	7	5.25	0.75
总数	21	21	

表 3-32 检验统计量

	分组
卡方	0.524[a]
自由度	3
渐进显著性	0.914

注：a. 0 个单元（0%）具有小于 5 的期望频率。单元最小期望频率为 5.3。

3.6 小结

使用卡方统计量可以完成对分类数据的独立性检验和拟合优度检验。对于两个分类变

量之间独立性的检验，要注意在不同的样本量情况下卡方统计量的值与品质相关系数之间的关系。对于拟合优度检验，需要注意不同问题对期望频数的要求。

◆ **实践训练**

1. 依据国家统计局分类口径将收入分为 5 个等级，表 3-33 是 2018 年城镇居民与农村居民人均可支配收入汇总结果。

表 3-33　可支配收入与收入组别数据

收入组别	人均可支配收入	
	城镇居民	农村居民
低收入户（20%）	14 386.9	3 666.2
中间偏下户（20%）	24 856.5	8 508.5
中间收入户（20%）	35 196.1	12 530.2
中间偏上户（20%）	49 173.5	18 051.5
高收入户（20%）	84 907.1	34 042.6

资料来源：国家统计局网站，www.stats.gov.cn。

要求：使用卡方检验判断城乡二元制度是否影响居民的人均可支配收入。

2. 表 3-34 为我国改革开放以来国内生产总值（单位：亿元）数据。

表 3-34　GDP 数据

年份	GDP/亿元	年份	GDP/亿元
1978	3 678.7	2001	110 863.1
1980	4 587.6	2002	121 717.4
1985	9 098.9	2003	137 422.0
1986	10 376.2	2004	161 840.2
1987	12 174.6	2005	187 318.9
1988	15 180.4	2006	219 438.5
1989	17 179.7	2007	270 092.3
1990	18 872.9	2008	319 244.6
1991	22 005.6	2009	348 517.7
1992	27 194.5	2010	412 119.3
1993	35 673.2	2011	487 940.2
1994	48 637.5	2012	538 580.0
1995	61 339.9	2013	592 963.2
1996	71 813.6	2014	641 280.6
1997	79 715.0	2015	685 992.9
1998	85 195.5	2016	740 060.8
1999	90 564.4	2017	820 754.3
2000	100 280.1	2018	900 309.5

资料来源：国家统计局网站，www.stats.gov.cn。

要求：使用卡方检验分析不同时期我国经济增长质量的变化规律。

3. 表 3-35 列出了一年内某地区观测到的交通事故发生数，零假设 H_0：事故发生数数据来自泊松分布。要求使用卡方统计量判断 H_0 是否成立。

表 3-35 交通事故发生数

事故发生数	0	1	2	3	4	5	总计
天数	211	108	31	9	5	1	365

4. 某项关于总统选举的民调，部分调查数据整理后如表 3-36 所示，请选择适当的方法判断性别是否对候选人当选有影响。

表 3-36 总统选举民调表（部分）

性别	支持候选人	性别	支持候选人	性别	支持候选人
男	杜威	男	杜鲁门	男	杜鲁门
女	杜鲁门	女	杜威	男	杜鲁门
男	杜鲁门	男	杜威	男	杜威
男	杜鲁门	女	杜威	女	杜鲁门
男	杜威	男	杜鲁门	女	杜鲁门
女	杜威	女	杜鲁门	女	杜威
女	杜鲁门	女	杜鲁门	女	杜鲁门
男	杜鲁门	女	杜鲁门	男	杜威
女	杜威	男	杜鲁门	男	杜鲁门
男	杜威	女	杜威	男	杜威
女	杜威	女	杜鲁门	女	杜鲁门
女	杜威	女	杜威	女	杜鲁门
男	杜鲁门	女	杜鲁门	女	杜鲁门
女	杜威	男	杜威	男	杜鲁门
男	杜威	男	杜鲁门	女	杜威
女	杜鲁门	男	杜鲁门	男	杜威
女	杜威	男	杜鲁门	女	杜威
男	杜威	女	杜威	男	杜威
女	杜威	男	杜威	女	杜威
男	杜鲁门	女	杜威	男	杜威
女	杜威	女	杜威	女	杜威
女	杜威	男	杜鲁门	男	杜威
男	杜威	女	杜鲁门	女	杜鲁门
女	杜威	女	杜威	男	杜鲁门
女	杜鲁门	男	杜鲁门	女	杜威
男	杜威	女	杜威	女	杜鲁门
女	杜威	女	杜鲁门	女	杜威
男	杜威	女	杜威	男	杜鲁门
女	杜鲁门	男	杜鲁门	男	杜威
男	杜威	女	杜威	男	杜威

续表

性别	支持候选人	性别	支持候选人	性别	支持候选人
女	杜威	女	杜鲁门	女	杜鲁门
女	杜威	女	杜威	女	杜威
男	杜鲁门	男	杜鲁门	女	杜威
女	杜威	女	杜威	男	杜鲁门
女	杜威	男	杜鲁门	男	杜威
男	杜鲁门	男	杜威	男	杜威
女	杜威	女	杜鲁门	女	杜威
男	杜威	女	杜威	女	杜鲁门
女	杜威	女	杜威	女	杜威
男	杜鲁门	男	杜鲁门	女	杜鲁门
男	杜威	女	杜威	男	杜威
女	杜威	女	杜鲁门	女	杜威
女	杜鲁门	男	杜威	女	杜鲁门
女	杜威	女	杜威	男	杜威
男	杜鲁门	女	杜威	女	杜鲁门
男	杜威	女	杜鲁门	男	杜威
女	杜威	男	杜威	女	杜鲁门
女	杜威	女	杜威	女	杜威
男	杜鲁门	女	杜鲁门	男	杜威
男	杜鲁门	男	杜威	女	杜鲁门
女	杜威	女	杜威	女	杜威
女	杜威	男	杜鲁门	女	杜威
男	杜鲁门	男	杜威	男	杜鲁门
女	杜威	女	杜鲁门	男	杜威
女	杜鲁门	男	杜威	女	杜威
男	杜鲁门	女	杜威	男	杜鲁门
男	杜鲁门	男	杜鲁门	女	杜威
女	杜威	男	杜威	女	杜威
女	杜威	女	杜威	女	杜威
男	杜威	女	杜威	女	杜鲁门