

第 14 章 卡方世界——Crosstabs 过程详解

可以毫不夸张的说：国内 80% 的 SPSS 使用者都只利用了 Crosstabs 过程不到 20% 的功能，即交叉表和 $R \times C$ 表的卡方检验。

——张文彤

统计分析都是和变量打交道的，变量则可分为定量变量和分类变量（含有序分类）两大类，对分类变量的描述和推断自然是各种统计软件中非常重要的一部分功能。在 SPSS 中专门用于分类数据的描述/分析的菜单项主要有以下这些：

- ✧ Custom Tables 菜单：功能主要是对数据生成各种描述表格、汇总报表，只能计算各种描述指标，其报表生成能力极为强大，但基本没有统计推断功能。
- ✧ Crosstabs 菜单项：对分类资料提供各种格式的频数分布描述，并进行“简单”的统计推断，包括了各种类型的卡方检验、相关性度量指标、确切概率等。
- ✧ 几个 Logistic 菜单项：拟合两分类/多分类/有序分类的 Logistic 回归模型，该模型现正在得到越来越广泛的应用。
- ✧ Loglinear 菜单：拟合对数线性模型，这是针对分类资料最为严格的建模方法，但使用上也最为复杂。



除了以上这些外，SPSS 菜单项中主要针对分类数据描述/分析的过程还有进行对应分析和最优尺度分析的 Correspondence Analysis 过程和 Optimal Scaling 过程、进行多维尺度分析的 Multidimensional Scaling 过程和 Multidimensional Scaling(PROXSCAL)过程，以及专门用于多选题描述的 Multiple Response 菜单。

可见在以上四大块中，只有 Crosstabs 过程既包括强大的描述功能，又提供了非常有力而实用的统计推断能力，而最为常用的卡方检验更是被它几乎垄断！事实上，针对分类资料的常用统计分析，我们只需要使用 Crosstabs 过程和 Logistic 过程就差不多够用了。正因如此，我们才不惜花费笔墨，专门为它辟出一章来加以讲解。

下面我们来看看 Crosstabs 过程的手段究竟如何：在描述时，对计数资料和有序分类资料可以产生二维至 n 维列联表，并计算相应的百分数指标、期望频数、行/列汇总指标、各种残差指标等；统计推断则包括了成组、配伍、分层的卡方检验、针对不同使用范围的一大批用于度量行、列变量关联度的指标及其检验、流行病学常用关联度指标及其检验等。如果安装了 Exact 模块，还可计算任意维列联表的确切概率（Fisher's Exact Test）值。

下面就让我们来揭开 Crosstabs 过程的神秘面纱，逐步去体会它的实力所在。



Crosstabs 过程不能产生一维频数表（单变量频数表），该功能由 Frequencies 过程实现。

14.1 分类资料数据录入格式简介

分类资料的数据录入格式往往和定量资料不太相同，在定量资料中，由于一般每个观察对象的变量值都不一样，记录格式为一个观察对象一条记录（枚举格式）。而在分类资料中，由于所有的变量值都限于很少的几个类别，如果还按照定量资料的记录格式就非常麻烦，如记录全班 52 位同学的性别，就会写成如图 14.1 中左图所示。

	name	sex
1	张三	男
2	李四	男
3	王二麻子	女

	sex	count
1	男	32
2	女	20

图 14.1 分类资料录入的枚举格式和频数表格式

一共要写 52 条记录！但实际上变量值只有男、女两类，这样录入不是累傻小子吗！为了方便起见，人们往往采用频数表格式来记录，即图 14.1 中右图的格式，这样只需要两条记录就可以了，而传递的总信息量和前面差不太多，只是不能得知具体的个体取值而已。这种格式往往用于没有原始数据库，需要重新录入数据分析时。而在原始数据库中，由于管理和跟踪记录的要求，一般仍是一个观察对象一条记录。SPSS 对这两种录入格式均可识别，对频数表格式只需要用 Weight Cases 过程指定一下频数变量即可。

14.2 Crosstabs 过程详解

14.2.1 引例

例 14.1 某医生用国产呋喃硝胺治疗十二指肠溃疡，以甲氧咪胍作对照组，问两种方法治疗效果有无差别（倪宗瓚，《医学统计学》第二版 P37）？

处理	未愈合	愈合	合计
呋喃硝胺	8	54	62
甲氧咪胍	20	44	64
合计	28	98	126

	drug	result	count
1	呋喃硝胺	未愈合	8.00
2	呋喃硝胺	愈合	54.00
3	甲氧咪胍	未愈合	20.00
4	甲氧咪胍	愈合	44.00

解：由于此处给出的直接是频数表，因此在建立数据集时可以直接输入三个变量——行变量、列变量和记录每个格子中频数的变量，然后用 Weight Cases 过程指定频数变量，最后调用 Crosstabs 过程进行 χ^2 检验。现数据集 2by2.sav 已建好，结构如上图，分析时的操作步骤如下：

Data→Weight Cases

☒ Weight Cases by:

Frequency Variable 框: count

频数变量为 count

Analyze→Descriptive Statistics→Crosstabs

Rows 框: drug

Columns 框: result

Statistics: ☒ Chi-square:

要求进行卡方检验

14.2.2 界面说明

【主对话框】(如图 14.2)

1. Rows 框: 用于选择行*列表中的行变量。
2. Columns 框: 用于选择行*列表中的列变量。
3. Layer 框组: Layer 指的是层, 用于设置分层分析变量。包括 Layer 框和 Layer 按钮组两部分。
 - ✧ Layer 框: 选入分层变量, 比如将性别选入, 则对行/列变量进行描述和分析时会按照性别为男/女的情况分别进行, 但并不给出合计。如果选入多个分层变量, 则分别分层, 如分别对性别和婚姻状况分层分析。
 - ✧ Layer 按钮组: 共包括 和 两个按钮, 用于控制下方的 Layer 框。在 Layer 框选入多个变量时默认是分别进行分析, 如果希望进行嵌套的分层分析, 如按照性别和婚姻状况的不同组合进行分层分析, 则先将性别选入, 然后单击 按钮, 此时 Layer 框跳至第二层, 再将婚姻状况选入 Layer 框, 两个变量即被分配到了不同层。
4. ☐ Display clustered bar charts: 显示重叠条图, 它可以直观的反映出各单元格内频数的多少。
5. ☐ Suppress table: 禁止在结果中输出行*列表。

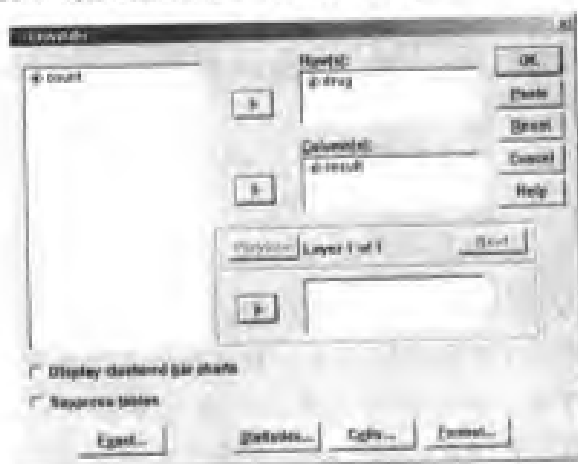


图 14.2 主对话框

【Exact】子对话框 (如图 14.3)

用于设定针对 2*2 以上行*列表是否进行确切概率的计算, 以及具体的计算方法。

! 如果 SPSS 没有安装 Exact 模块, 则主对话框中不会出现 Exact 框。

1. ☒ Asymptotic only: 只计算近似的概率值, 不计算确切概率。
2. ☐ Monte Carlo: 采用蒙特卡罗模拟方法计算确切概率值。蒙特卡罗模拟默认进行 10000 次抽样, 给出确切概率以及 99% 可信区间; 这些默认值均可更改。

3. ☒ Exact: 计算出确切的概率值, 默认计算时间限制在 5 分钟内, 超过此时限则自动停止。该默认值可以更改。

4. 对话框最下方的说明文字用于提示: 如果计算机的计算能力许可, 最好采用确切概率法, 而不是蒙特卡罗模拟方法, 以保证结果的准确。但是对该问题我有不同看法, 在 3*3 及以上的行*列表表中, 确切概率的精确计算是极为漫长的过程。我曾经用 SAS 6.12 在 P133 机上计算过一个 12 格表的确切概率, 整整跑了两个小时后, SAS 告诉我机器内存不足。SPSS 比 SAS 也好不到哪里去, 稍微多两个数据也会叫唤内存不足, 因此一般只需要选用蒙特卡罗模拟算出概率值的 99% 可信区间就行了, 精度完全可以满足需要, 而速度极快 (10000 次模拟一般耗时在 5~10 秒左右)。

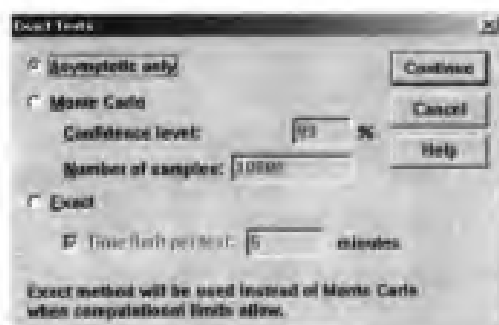


图 14.3 Exact 子对话框

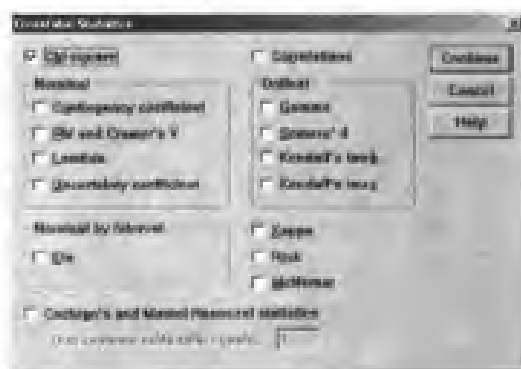


图 14.4 Statistics 子对话框

【Statistics 子对话框】(如图 14.4)

该对话框包含了一大批用于度量行、列变量关联度的指标, 并按照各自的使用范围分成了几大组, 这些指标在科研和医学领域有极大的应用价值 (注意这句话不仅仅是指卡方!), 下面我们就来依次讲解:

1. ☐ Chi-square: 进行卡方检验, 对于四格表资料还会自动给出校正卡方检验和确切概率法的结果。该检验用于判断行、列变量是否独立, 如果数据不满足卡方检验的要求 (不能有单元格的期望数小于 1, 不能有 20% 以上单元格的期望数小于 5), 则系统会在分析结果的最后给出警告, 提示用户采用确切概率法分析。



卡方检验究竟是参数检验还是非参数检验? 如果从原理上来讲, 卡方检验是研究样本分布偏离理论分布的严重程度, 即检验的是分布, 而不是总体参数。因此它属于非参数检验。在分类变量的统计推断中, 真正的参数检验方法是确切概率法, 不过它本身也存在不少问题, 而且当样本量较大时, 卡方检验的效率也比较高, 因此一般都采用卡方检验进行分析。

2. ☐ Correlations: 计算行、列变量的 Pearson 相关系数和 Spearman 等级相关系数, 前者只能用与两变量均为数值变量的情况, 故而在此用处不大; 后者理论上可适用于有序分类变量, 但不如下面 Ordinal 复选框组中的统计量效率高, 所以实际上也没什么用处。

3. Nominal 复选框组: 包含了一组用于反映分类变量相关性的指标, 这些指标在变量属于有序或无序分类时均可使用, 但两变量均为有序分类变量时效率没有下面

Ordinal 复选框组中的统计量高。

- ✧ Contingency coefficient: 计算列联系数, 基于 χ^2 值得出, 其值介于 0~1 之间, 越大表明两变量间相关性越强。
- ✧ Phi and Cramer's V: 这两者也是基于 χ^2 值的, Phi 在四格表 χ^2 检验中介于 0~1 之间, 在 R*C 表 χ^2 检验中介于 0~1 之间; Cramer's V 则介于 0~1 之间。指标的绝对值越大, 则相关性越强。
- ✧ Lambda: 用于反映自变量对应变量的预测效果, 即知道自变量取值时对应变量的预测有多少改进, 其值为 1 时表明知道了自变量就可以完全确定应变量取值, 为 0 时表明自变量对应变量完全无预测作用。共输出行变量为自变量、列变量为自变量、对称时的三个结果, 后者为前两者的对称平均指标。
- ✧ Uncertainty coefficient: 不确定系数, 其值介于 0~1 之间, 和 Lambda 类似, 也用于反映当知道自变量后, 应变量的不确定性下降了多少(比例)。为不确定性大小的度量指标, 共会输出行变量为自变量、列变量为自变量、对称不确定系数三个结果, 后者为前两者的对称平均指标。

4. Ordinal 复选框组: 包含了一组用于反映分类变量一致性的指标, 这些指标只能在两个变量均属于有序分类时使用。它们均是基于 Gamma 统计量衍生出来的。所谓一致性高, 就是指行变量等级高的列变量等级也高, 行变量等级低的列变量等级也低。如果行变量等级高而列变量等级低, 则被称为不一致。

- ✧ Gamma: 介于 -1~1 之间, 当观察值集中于对角线处时, 其取值为 -1 或 1, 表示两者取值绝对一致或绝对不一致; 如两变量完全无关, 则取值为 0。
- ✧ Somers' d: 由下面的 tau-b 改进而来, 只校正了自变量相等的对子。
- ✧ Kendall's tau-b: 由 Gamma 系数改进而来, 对相等的对子进行了校正, 该指标有时也被称为 Kendall's 等级相关系数。
- ✧ Kendall's tau-c: 由 tau-b 改进而来, 在其基础上对表的大小进行了校正。

5. Nominal by Interval 复选框组: 包含了一个变量为数值变量, 而另一个为分类变量时度量两者关联度的指标, 现在的版本中只包括 Eta 复选框, 用于计算 eta 值, 大家在 Means 过程中已经遇到过它了, eta 的平方表示由组间差异所解释的应变量的方差的比例, 即 SS 组间/SS 总。系统一共会给出两个 Eta 值, 分别对应了行变量为应变量(数值变量)和列变量为应变量的情况。

6. ☐ Kappa: 计算 Kappa 值, 即内部一致性系数。这是医学中非常常用的一致性指标, 取值在 0~1 之间, 除根据 P 值判断一致性有无统计学意义外, 根据经验, $Kappa \geq 0.75$ 表明两者一致性较好; $0.75 > Kappa \geq 0.4$ 表明一致性一般, $Kappa < 0.4$ 则表明两者一致性较差。

7. ☐ Risk: 计算 OR 值(比数比)和 RR 值(相对危险度), 学医的人对它们应当再熟悉不过了。




为不是学医的朋友考虑, 这里我还是把两个指标解释一下吧: OR 值的计算公式为(病例中暴露的比例/病例中非暴露的比例)/(对照中暴露的比例/对照中非暴露的比例), 例如 $OR=2$, 则说明病例中暴露于该危险因素者的比例

为对照中的两倍，显然该因素可能和疾病有关（要经过检验）；RR值的计算公式为暴露组的发病密度/非暴露组的发病密度，这里的发病密度可理解为发病概率，例如 $RR=2$ ，则说明暴露组的发病概率为非暴露组的两倍。在推断疾病/危险因素关联上，RR要比OR更加直观和准确。

8. **McNemar**：进行 McNemar 检验，即常用的配对卡方检验。但请注意，这里系统是按二项分布的原理计算出确切概率，相当于进行了精确校正（可以理解为校正数在 0~1 之间波动，更准确嘛），所以结果和手工用校正/不校正公式计算出的都不一样，并且不能给出卡方值，只能给出 P 值。该检验只能针对方形表格进行，即行、列分类数相同（如果分类数都不同，怎么可能是配对数据呢）。

9. **Cochran's and Mantel-Haenszel statistics**：为两个二分类变量进行独立性检验和同质性（齐性）检验，同时可进行分层因素的调整，共包括以下几种统计量：

- ✧ χ^2_{MH} 统计量，即在流行病学中极为常用的分层卡方检验，当层间存在混杂因素时，计算分层卡方可以很好的消除该混杂因素对结果的影响。同时系统还会给出总的调整 OR 值（ OR_{MH} ）。
- ✧ χ^2_{CMH} 统计量，可以认为是 χ^2_{MH} 统计量的进一步发展，它的适用条件较 χ^2_{MH} 严格，但功能更加强大，有兴趣的朋友可参考相关书籍，此处不赘述。
- ✧ 同质性检验，共包括 Breslow-Day 和 Tarone's 两种方法，用于检验各层的风险情况是否一致。

 告诉大家一个小常识，Breslow 教授是缺失值分析理论在地球上的权威，如果统计方法中有他的名字，那该方法多半是从缺失值分析理论中推导而来的。

最后，该复选框下方的 Test common odds ratio equal 框用于设定相应 H_0 假设的 OR 值，默认为 1。

【Cells 子对话框】（如图 14.5）

用于定义列联表单元格中需要显示的指标。

1. Counts 复选框组：是否输出实际观察数(Observed)和理论数(Expected)。
2. Percentages 复选框组：是否输出行百分数(Row)、列百分数(Column)以及合计百分数(Total)。
3. Residuals 复选框组：选择残差的显示方式，可以是实际数与理论数的差值(Unstandardized)、标化后的差值(Standardized，将差值转化为标准正态分布)，或者被标准误差除的单元格残差(Adj. Standardized)。

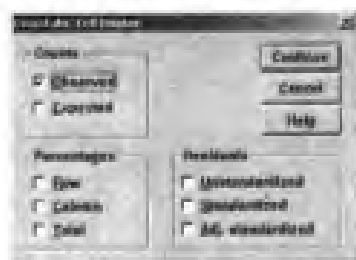


图 14.5 Cells 子对话框



图 14.6 Format 子对话框

【Format】子对话框】(如图 14.6)

用于选择行变量是升序还是降序排列。

14.2.3 结果解释

引例的结果如下:

Crosstabs

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
DRUG * RESULT	126	100.0%	0	.0%	126	100.0%

首先是处理记录缺失值情况报告, 可见 126 例均为有效值。

DRUG * RESULT Crosstabulation

Count		RESULT		
		未愈合	愈合	Total
DRUG	呋喃硝胺	8	54	62
	甲氧咪胍	20	44	64
Total		28	98	126

上面为列出的四格表, 其中加入了变量值标签, 看起来非常清楚, 和给出的原始表格几乎一模一样。

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.133 ^b	1	.013		
Continuity Correction ^a	5.118	1	.024		
Likelihood Ratio	6.304	1	.012		
Fisher's Exact Test				.018	.011
Linear-by-Linear Association	6.084	1	.014		
N of Valid Cases	126				

^a Computed only for a 2x2 table

^b 0 cells (.0%) have expected count less than 5. The minimum expected count is 13.78.

上表给出了一堆检验结果, 从左到右为: 检验统计量值(Value)、自由度(df)、双侧近似概率(Asymp.Sig.2-sided)、双侧精确概率(Exact Sig.2-sided)、单侧精确概率(Exact Sig.1-sided); 从上到下为: Pearson 卡方(即常用的卡方检验)、连续性校正的卡方值(Continuity Correction)、对数似然比方法计算的卡方(Likelihood Ratio)、Fisher's 确切概率法(Fisher's Exact Test)、线性相关的卡方值(Linear by Linear Association)、有效记录数(N of Valid Cases)。另外, Continuity Correction 和 Pearson 卡方值处分别标注有 a 和 b, 表格下方为相应的注解: a. 只为 2*2 表计算, 即只在 2*2 表时系统才计算校正卡方值。b. 0%个格子的期望频数小于 5, 最小的期望频数为 13.78。

好, 得到了如此丰富的结果, 可我们该用哪一个呢? 这个问题往往令许多初学者

头痛不已，实际上按照国内的习惯，我们只需要在未校正卡方、校正卡方和确切概率法三种方法之间选择即可，其余的对我们而言用处不大，可以视而不见。现在来复习一下，教科书上是这样来判断的：

1. 当 $n \geq 40$ 且所有 $T \geq 5$ 时，用普通的 χ^2 检验，若所得 $P \approx \alpha$ ，改用确切概率法；
2. 当 $n \geq 40$ 但有 $1 \leq T < 5$ 时，用校正的 χ^2 检验；
3. 当 $n < 40$ 或有 $T < 1$ 时，不能用 χ^2 检验，改用确切概率法。

我们这里显然符合条件 1，因此无须校正，直接采用第一行的检验结果，即 $\chi^2 = 6.133$ ， $P = 0.013$ 。



给大家讲件真事，也是笑话，如果你笑得出来。有位临床研究生做课题，考虑到统计不太熟，就找另一位研究生帮忙（当然不是学统计的），该朋友做了上面的一堆卡方结果出来，同学看了直发愁，问他应该用哪个，朋友反问：“你是想有差别还是无差别？”他回答：“当然是有差别喽。”答曰：“就用 P 值最小的那个！”



可能这里还是要多说两句，卡方分布本身是一个连续型分布，在分类资料的统计分析中，卡方检验只是一个近似的检验。当例数多时，这个问题不太明显；例数较少时分类资料的不连续分布和卡方分布之间的差异就会达到无法忽视的程度，从而需要连续性校正；如果差异再大下去，校正都不够用，只好放弃卡方检验，采用确切概率法计算了。因此，有观点认为应当尽量采用校正卡方和确切概率法的结果。

实际上，根据模拟研究的结果，在 Pearson 卡方、似然比卡方和 Fisher's 确切概率法的分析结果中，似然比卡方是最为准确的，即使在小样本也是如此。而确切概率法的分布假设在许多时候都不成立，因此建议大家多参考似然比卡方的结果。

14.3 综合分析实例

14.3.1 确切概率的计算

例 14.2 某研究者调查了一批高血压患者的血压控制情况和肥胖度，数据见文件 tables.sav，为枚举格式。汇总情况见下表，试分析两者间有无关系。

		血压控制情况			
		良好	尚可	不良	合计
肥胖程度	不肥胖	15	24	12	51
	轻度肥胖	4	2	7	13
	中/重度肥胖	20	13	11	44
	合计	39	39	30	108

解：数据 tables.sav 为枚举格式，即一条记录代表一个病例。因此分析时无需先用 Weight Cases 过程。从上表可见轻度肥胖的三个格子中频数都非常少，计算可得它

们的理论频数均小于 5，此处必须要用确切概率法计算。由于总例数由 108 例，精确计算可能较费时，这里我们采用蒙特卡罗模拟方法来代替。

Analyze→Descriptive Statistics→Crosstabs

Rows 框: fat

Columns 框: control

Statistics: ☒ Chi-square: Continue

要求进行卡方检验

Exact: ☒ Monte Carlo: Continue

用蒙特卡罗模拟方法计算确切概率

OK

分析结果中的卡方分析表如下:

Chi-Square Tests


	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided)		
				Sig.	99% Confidence Interval	
					Lower Bound	Upper Bound
Pearson Chi-Square	9.199 ^a	4	.056	.055 ^b	.049	.061
Likelihood Ratio	8.813	4	.066	.076 ^b	.070	.083
Fisher's Exact Test	8.463			.069 ^b	.062	.075
Linear-by-Linear Association	.719 ^c	1	.397	.402 ^b	.389	.415
N of Valid Cases	108					

^a 3 cells (33.3%) have expected count less than 5. The minimum expected count is 3.61.

^b Based on 10000 sampled tables with starting seed 2000000.

^c The standardized statistic is -.648.

显然，右侧多出来了蒙特卡罗法计算出的 P 值可信区间，应有双侧和单侧两部分，由于表格太长，此处只列出了双侧部份。可见 Pearson 卡方的近似概率为 0.056，而蒙特卡罗法计算的确切概率为 0.053，99%可信区间为 0.048~0.059，结论为尚不能认为肥胖程度和血压控制情况有关，但由于 P 值在 0.05 附近，两者关系值得进一步关注。

 表格下方的注解处指明蒙特卡罗法的起始随机种子为 2000000，由于它是随机抽样过程，不同的种子得到的结果会略有差别，大家可以使用 Transform→Random Number Seed 过程将随机种子确定为 2000000 来得到和上表完全相同的结果。

14.3.2 配对卡方与一致性检验

例 14.3 两位放射科医生对一批砒肺胸片独自做出了砒肺分级诊断，见下表，请问他们的诊断结果是否基本一致，诊断水平有无差别。

		医生乙诊断结果			
		I 级	II 级	III 级	合计
医生甲 诊断结果	I 级	32	15	0	47
	II 级	1	54	12	67
	III 级	0	7	45	52
	合计	33	76	57	166

解：这是一个典型的分类数据配对资料，根据研究目的，可以采用 Kappa 检验或

配对卡方检验。前者重在研究两个医生的诊断一致性，希望所有的频数都出现在主对角线上，这样一致性就最好；而后者重在研究两者在哪些地方不一致，即两者的诊断结果有怎样的偏好，计算时只利用了不在主对角线上的数据。这里我们将两个分析结果都给出，以回答两个不同方向的问题。数据已录入为 agree.sav，分析步骤如下：

Data → Weight Cases

Weight Cases by:

Frequency Variable 框: count

频数变量为 count

OK

Analyze → Descriptive Statistics → Crosstabs

Rows 框: doct1

Columns 框: doct2

Statistics:

☒ Chi-square: ☒ McNemar

进行成组卡方和配对卡方检验

☒ Kappa

要求进行 Kappa 检验

Continue

OK

主要分析结果如下。首先给出的卡方分析表不仅给出了配对卡方结果，也会给出成组卡方结果，成组卡方用于反映两个医生的诊断有无关系，显然，两者是有关的，即被甲医生诊断为矽肺级别较高的胸片，乙医生的诊断级别也较高。下方的配对卡方 P 值为 0.002，显示两位医生的诊断结果是不同的，观察原始数据，可以发现乙医生较甲医生容易将矽肺期次判断的偏高。

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	170.918 ^a	4	.000	
Likelihood Ratio	174.298	4	.000	
Linear-by-Linear Association	112.236	1	.000	
McNemar Test				.002 ^b
N of Valid Cases	166			

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 9.34

b. Binomial distribution used

Symmetric Measures

	Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement Kappa	.676	.049	12.291	.000
N of Valid Cases	166			

a. Not assuming the null hypothesis

b. Using the asymptotic standard error assuming the null hypothesis.

上表为 Kappa 检验的结果，可见相应的 P 值小于 0.001，因此同样拒绝无效假设，认为两位医生的诊断结果存在一致性，结合其 Kappa 值为 0.676，小于 0.75，可以认为其一致性中等。

14.3.3 分层卡方检验

例 14.4 国外的口服避孕药剂量都相当大，某次病例对照研究调查了口服避孕药与心肌梗死的情况，考虑到年龄是一个可能的混杂因素，将其也纳入调查，得到如下数据，请分析口服 OC 与心肌梗死有无关系。

	年龄<40		年龄≥40	
	服用 OC	未服用 OC	服用 OC	未服用 OC
病例	21	26	18	88
对照	17	59	7	95
合计	38	85	25	183

解：数据已按要求录入为 oc.sav，年龄是否为混杂因素有比较严格的判断步骤，因篇幅所限，这里将其省去，只给出后面的计算结果。题中要求的分层分析可以用 Cochran's and Mantel-Haenszel statistics 来完成，同时可利用 Risk 复选框给出各层各自的 OR 值用于比较，分析步骤如下：

Data→Weight Cases

Weight Cases by:

Frequency Variable 框: count

频数变量为 count

OK

Analyze→Descriptive Statistics→Crosstabs

Rows 框: oc

Columns 框: case

Layer 框: age

按照变量 age 分层

Statistics:

☒ Risk

要求计算 OR、RR 值

☒ Cochran's and Mantel-Haenszel statistics

进行分层卡方检验

Continue

OK

主要的输出结果如下：

OC * CASE * AGE Crosstabulation

Count				
		CASE		Total
AGE		病例	对照	
<40岁	OC 服用OC	21	17	38
	不服用OC	26	59	85
	Total	47	76	123
≥40岁	OC 服用OC	18	7	25
	不服用OC	88	95	183
	Total	106	102	208

上表为按变量 age 取值分层的交叉描述表。

Risk Estimate

AGE		Value	95% Confidence Interval	
			Lower	Upper
<40岁	Odds Ratio for OC (服用OC / 不服用OC)	2.803	1.274	6.167
	For cohort CASE = 病例	1.807	1.176	2.776
	For cohort CASE = 对照	.645	.441	.943
	N of Valid Cases	123		
≥40岁	Odds Ratio for OC (服用OC / 不服用OC)	2.776	1.106	6.965
	For cohort CASE = 病例	1.497	1.124	1.995
	For cohort CASE = 对照	.539	.283	1.027
	N of Valid Cases	208		

上表分层给出了 OR 值、以变量 case 的不同取值为准的 RR 值、以及其相应的可信区间。由于本数据为病例对照资料，所以计算出的 RR 值无实际意义。可见当年龄小于 40 岁时 OR 为 2.803，大于等于 40 岁时为 2.776，且可信区间下限都越过了 1，具有统计学意义。请注意这里可信区间的计算采用的是 Woolf 自然对数转换法，较为保守，区间较 Miettinen 方法宽的多。

Tests for Homogeneity of the Odds Ratio

Statistics		Chi-Squared	df	Asymp. Sig. (2-sided)
Conditional Independence	Cochran's	11.782	1	.001
	Mantel-Haenszel	10.729	1	.001
Homogeneity	Breslow-Day	.000	1	.987
	Tarone's	.000	1	.987

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

上表给出的是分层卡方检验和一致性检验的结果，可见 $\chi^2_{MH} = 10.729$ ， $P = 0.001$ ，表明去除了年龄的混杂作用后，心肌梗死和服用 OC 有关， χ^2_{CMH} 的检验结论与之相同；下方的一致性检验结果 P 值为 0.987，表明不同层间的 OR 值相同，即层间同质。

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			2.791
ln(Estimate)			1.026
Std. Error of ln(Estimate)			.306
Asymp. Sig. (2-sided)			.001
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	1.532
		Upper Bound	5.084
	ln(Common Odds Ratio)	Lower Bound	.427
		Upper Bound	1.626

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

最后这张表格给出的是 OR_{MH} 值（调整了年龄混杂作用后的综合 OR 值）、 OR_{MH} 值的自然对数、可信区间及其相应的 P 值，可见统计检验结论和前面一致，相应的 $OR_{MH} = 2.791$ ，即去除了年龄的混杂效应后，服用 OC 的妇女患心肌梗死的危险度大约为未服 OC 妇女的 2.79 倍。