



南京大學
NANJING UNIVERSITY

第8章 分类变量的统计描述与参数估计

学习目标

- 了解用于分类变量统计描述的指标有哪些
- 掌握用于对多个分类变量进行联合描述的工具：
交叉表（列联表）
- 掌握如何对多选题进行统计描述

主要内容

- 8.1 指标体系概述
 - 8.1.1 单个分类变量的统计描述
 - 8.1.2 多个分类变量的联合描述
 - 8.1.3 多选题的统计描述
- 8.2 案例1
- 8.3 案例2

8.1 指标体系概述

8.1.1 单个分类变量的统计描述

- 频数分布情况的描述
 - 频数、百分比、累积频数、累积百分比
- 集中趋势的描述
 - 众数：出现次数最多的那个数

8.1.1 单个分类变量的统计描述

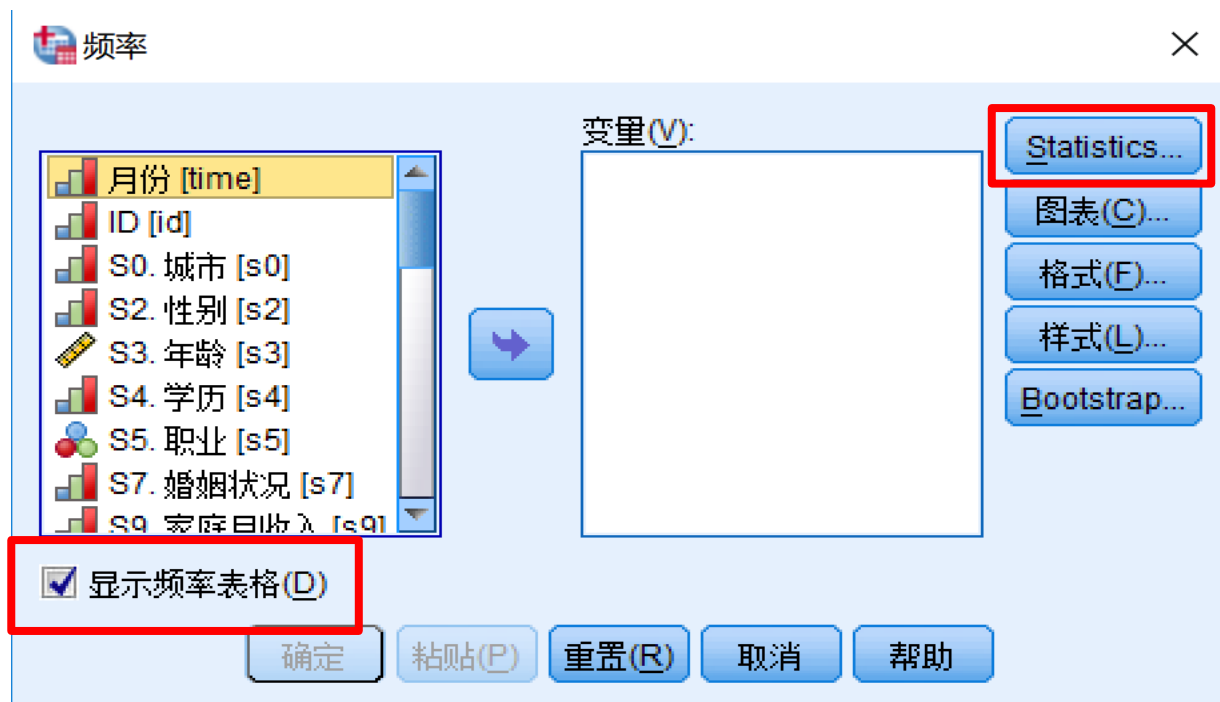
- 相对数

- **比**：任何两个相关联的变量A与B之比。例男生人数与女生人数之比
- **构成比**：指事物内部某一组成部分的观察单位数与该事物各组成部分观察单位总数之比，常用来说明事物内部各组成部分所占的比重或分布情况。
- **率**：说明一段时间（单位时间）内某个事件发生的频率或强度。

$$\text{某事件的发生率} = \frac{\text{观察期内发生某事件的对象数}}{\text{该时期开始时的观察对象数}}$$

8.1.1 单个分类变量的统计描述

- “分析” → “描述统计” → “频率”
 - 频数表：频数、百分比、累积百分比



8.1.1 单个分类变量的统计描述

- “分析” → “描述统计” → “频率”

— 众数

频率: 统计

百分位值

☐ 四分位数(Q)

☐ 分割点(U): 10 相等组

☐ 百分位数(P):

添加(A)

更改(C)

删除(R)

集中趋势

☐ 平均值(M)

☐ 中位数(D)

☒ 众数(O)

☐ 合计(S)

☐ 值为组的中点(L)

离散

☐ 标准偏差(T) ☐ 最小值(I)

☐ 方差(V) ☐ 最大值(X)

☐ 范围(N) ☐ 平均值的标准误差(E)

分布

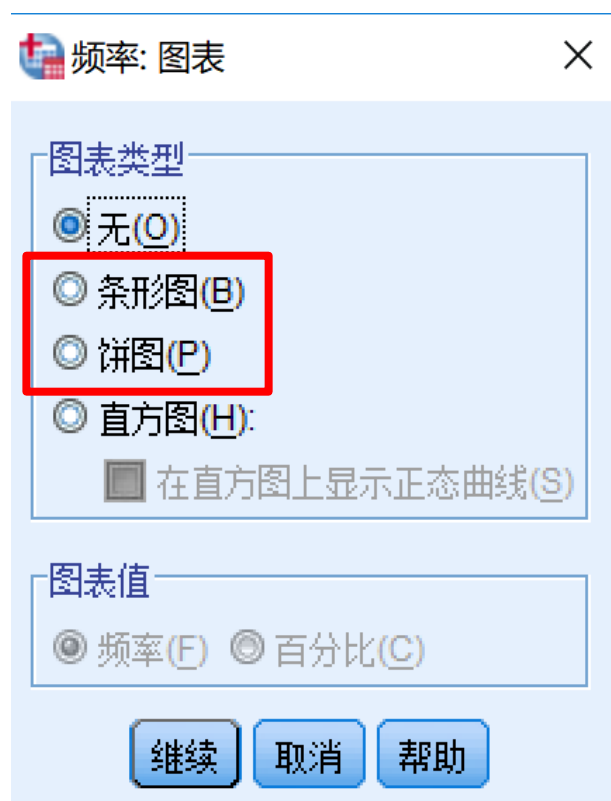
☐ 偏度(W)

☐ 峰度(K)

继续 取消 帮助

8.1.1 单个分类变量的统计描述

- “分析” → “描述统计” → “频率”
— 统计图（条图、饼图）



8.1.2 多个分类变量的联合描述

- 对两个甚至多个分类变量的频数分布进行联合观察。
 - 原始频数
 - 行百分比：该单元格的频数占所在行样本的构成比
 - 列百分比：该单元格的频数占所在列样本的构成比
 - 总百分比：该单元格的频数占总样本的构成比

8.1.2 多个分类变量的联合描述

- “分析” → “描述统计” → “交叉表格”



8.1.2 多个分类变量的联合描述

- “分析” → “描述统计” → “交叉表格”

交叉表格: 单元格显示 ×

计数(T) <ul style="list-style-type: none"><input checked="" type="checkbox"/> 观察值(O)<input type="checkbox"/> 期望值(E)<input type="checkbox"/> 隐藏较小计数(H) 小于 <input type="text" value="5"/>	z-检验 <ul style="list-style-type: none"><input type="checkbox"/> 比较列的比例(P)<input type="checkbox"/> 调整 p 值 (Bonferroni 方法)
百分比 <ul style="list-style-type: none"><input checked="" type="checkbox"/> 行(R)<input checked="" type="checkbox"/> 列(C)<input checked="" type="checkbox"/> 总计(T)	残差 <ul style="list-style-type: none"><input type="checkbox"/> 未标准化(U)<input type="checkbox"/> 标准化(S)<input type="checkbox"/> 调节的标准化(A)
非整数权重 <ul style="list-style-type: none"><input checked="" type="radio"/> 四舍五入单元格计数(N) <input type="radio"/> 四舍五入个案权重(W)<input type="radio"/> 截断单元格计数(L) <input type="radio"/> 截断个案权重(H)<input type="radio"/> 无调节(M)	

继续 取消 帮助

8.1.3 多选题的统计描述

- **应答人数 (Count)**：选择了某选项的人数，或称为原始频数，通常会说A选项的应答人数是…
- **应答人数百分比 (Percent of Cases)**：选择该项的人数占总人数的比例。反映选项在人群中的受欢迎程度。
- **应答人次 (Responses)**：选择选项的人次。针对单个选项，应答人次等同于应答人数，但对于整道多选题来说，应答人次往往大于应答人数。
- **应答次数百分比 (Percent of Responses)**：在做出的所有选择中，选择该项的次数占总次数（总反应数）的比例。

8.1.3 多选题的统计描述

- “分析” → “多重响应” → “定义变量集”
 - 定义多选题变量集

定义多响应集 ×

设置定义

集合中的变量(V):

多响应集(S):

添加(A)
更改(C)
删除(R)

将变量编码为

☒ 二分法(D) 计数值(O):

☐ 类别(G) 范围(E): 到(T)

名称(N):

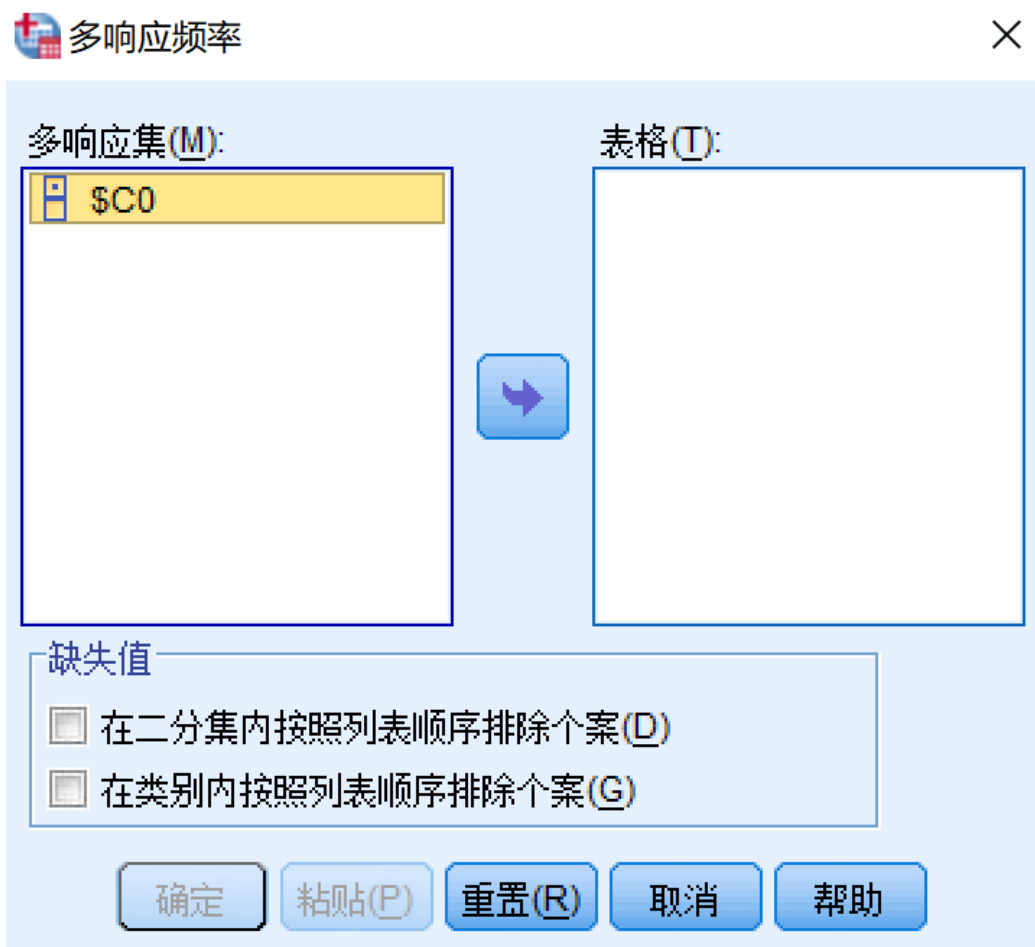
标签(L):

注: 在此处定义的集只能在多重响应频率和交叉表格程序中使用。对于在其它位置使用的集, 请使用数据菜单上的定义多响应集。

关闭 帮助

8.1.3 多选题的统计描述

- “分析” → “多重响应” → “频率”
 - 对多选题进行频数分析



8.1.3 多选题的统计描述

- “分析” → “多重响应” → “交叉表”
 - 对多选题进行列联表分析





南京大學
NANJING UNIVERSITY

8.2 案例1


使用频率过程进行描述

- 数据文件：CCSS_Sample.sav
- 要求：对“学历”进行统计描述
- 具体过程：
 - “分析” → “描述统计” → “频率” → 将变量S4选入“变量”列表
 - 勾选左下角“显示频率表格”框
 - 点击“统计”按钮，选择“众数”
 - 点击“图表”按钮，选择“条形图”

使用频率过程进行描述



使用频率过程进行描述

 频率: 统计

百分位值

☐ 四分位数(Q)

分割点(U): 10 相等组

☐ 百分位数(P):

添加(A)

更改(C)

删除(R)

集中趋势

☐ 平均值(M)

☐ 中位数(D)

☒ 众数(O)

☐ 合计(S)

☐ 值为组的中点(L)

离散

☐ 标准偏差(T)

☐ 最小值(I)

☐ 方差(V)

☐ 最大值(X)

☐ 范围(N)

☐ 平均值的标准误差(E)

分布

☐ 偏度(W)

☐ 峰度(K)

继续

取消

帮助

 频率: 图表

图表类型

☐ 无(O)

☒ 条形图(B)

☐ 饼图(P)

☐ 直方图(H):

☐ 在直方图上显示正态曲线(S)

图表值

☒ 频率(F)

☐ 百分比(C)

继续

取消

帮助

使用频率过程进行描述

Statistics

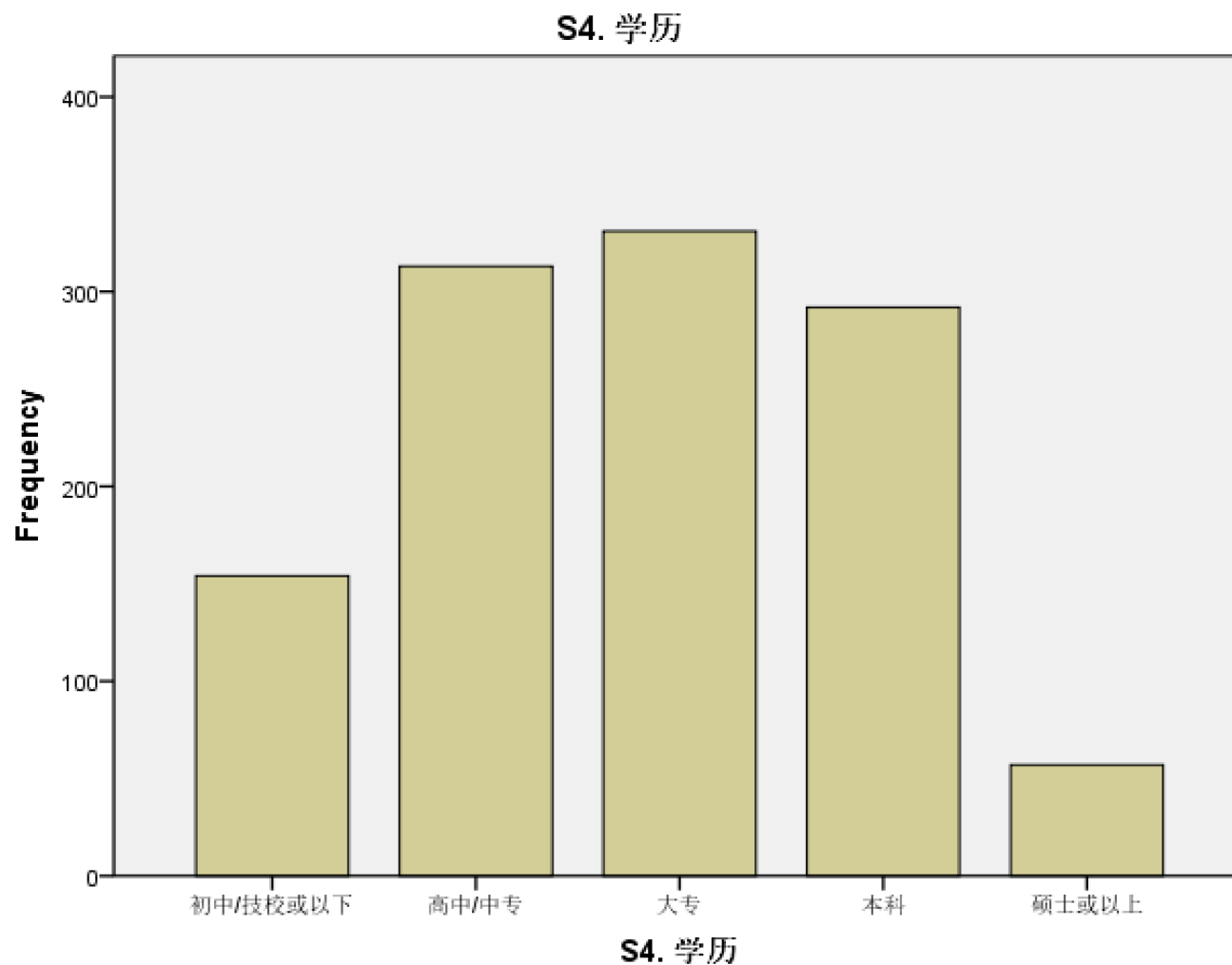
S4. 学历

N	Valid	1147
	Missing	0
Mode		3

S4. 学历

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	初中/技校或以下	154	13.4	13.4	13.4
	高中/中专	313	27.3	27.3	40.7
	大专	331	28.9	28.9	69.6
	本科	292	25.5	25.5	95.0
	硕士或以上	57	5.0	5.0	100.0
Total		1147	100.0	100.0	

使用频率过程进行描述



使用交叉表过程进行描述

- 数据文件：CCSS_Sample.sav
- 要求：对“性别”和“学历”进行联合统计描述
- 具体过程：
 - “分析” → “描述统计” → “交叉表” → 将变量S2选入“行”列表，将变量S4选入“列”列表
 - 点击“单元格”按钮，选择行、列、总百分比

使用交叉表过程进行描述

交叉表格



月份 [time]

ID [id]

S0. 城市 [s0]

S3. 年龄 [s3]

S5. 职业 [s5]

S7. 婚姻状况 [s7]

S9. 家庭月收入 [s9]

C0. 请问您的家庭...

C0. 请问您的家庭...

C0. 请问您的家庭...

O1. 是否拥有家用...

A3. 首先, 请问与...

A3a. 您为什么这...

A3a. 您为什么这...

行(O):

S2. 性别 [s2]

列(C):

S4. 学历 [s4]

层1的1

上一页(V)

下一页(N)

☐ 在表层中显示层变量(L)

☐ 显示集群条形图(B)

☐ 取消表格(T)

精确(X)...

Statistics...

单元格(E)...

格式(F)...

样式(L)...

Bootstrap...

确定

粘贴(P)

重置(R)

取消

帮助

使用交叉表过程进行描述

交叉表格: 单元格显示 ×

计数(T)

☒ 观察值(O)
☐ 期望值(E)
☐ 隐藏较小计数(H)
小于

z-检验

☐ 比较列的比例(P)
☐ 调整 p 值 (Bonferroni 方法)

百分比

☒ 行(R)
☒ 列(C)
☒ 总计(T)

残差

☐ 未标准化(U)
☐ 标准化(S)
☐ 调节的标准化(A)

非整数权重

☒ 四舍五入单元格计数(N) ☐ 四舍五入个案权重(W)
☐ 截断单元格计数(L) ☐ 截断个案权重(H)
☐ 无调节(M)

继续

取消

帮助

使用交叉表过程进行描述

S2. 性别 * S4. 学历 Crosstabulation

			S4. 学历					Total
			初中/技校或以 下	高中/中专	大专	本科	硕士或以上	
S2. 性别	男	Count	74	167	191	169	36	637
		% within S2. 性别	11.6%	26.2%	30.0%	26.5%	5.7%	100.0%
		% within S4. 学历	48.1%	53.4%	57.7%	57.9%	63.2%	55.5%
		% of Total	6.5%	14.6%	16.7%	14.7%	3.1%	55.5%
	女	Count	80	146	140	123	21	510
		% within S2. 性别	15.7%	28.6%	27.5%	24.1%	4.1%	100.0%
		% within S4. 学历	51.9%	46.6%	42.3%	42.1%	36.8%	44.5%
		% of Total	7.0%	12.7%	12.2%	10.7%	1.8%	44.5%
Total	Count		154	313	331	292	57	1147
	% within S2. 性别		13.4%	27.3%	28.9%	25.5%	5.0%	100.0%
	% within S4. 学历		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total		13.4%	27.3%	28.9%	25.5%	5.0%	100.0%



南京大學
NANJING UNIVERSITY

8.3 案例2

对多选题进行描述

- 数据文件：CCSS_Sample
- 要求：对多选题C0进行统计描述；按婚姻状况来考察贷款情况。
- 具体过程：
 - 明确多选题对应的变量：C0_1, C0_2, C0_3
 - 设置多选题变量集
 - 分析→多重响应→定义变量集
 - 多选题名称：C0
 - 多选题标签：家庭每月还贷情况

对多选题进行描述

- 分析→多重响应→频率
- 分析→多重响应→交叉表
 - 将s7选入“行”列表，将\$C0选入“列”列表
 - 点击s7后再点击“定义范围”按钮，确定参与分析的s7中的类别
 - 点击“选项”按钮选择“行百分比”

对多选题进行描述

定义多响应集

×

设置定义

- S5. 职业 [s5]
- S7. 婚姻状况 [s7]
- S9. 家庭月收入 [s9]
- O1. 是否拥有家用...
- A3. 首先, 请问与...
- A3a. 您为什么这...
- A3a. 您为什么这...
- A4. 那么与现在相...
- A4. 那么与现在相...



集合中的变量(V):

- C0. 请问您的家庭目...
- C0. 请问您的家庭目...
- C0. 请问您的家庭目...

添加(A)

更改(C)

删除(R)

多响应集(S):

将变量编码为

☒ 二分法(D) 计数值(O): 1

☐ 类别(G) 范围(E): 到(T)

名称(N):

C0

标签(L):


家庭每月还贷情况

注: 在此处定义的集只能在多重响应频率和交叉表格程序中使用。对于在其它位置使用的集, 请使用数据菜单上的定义多响应集。


关闭

帮助


对多选题进行描述

 多响应频率 ×

多响应集(M):

 \$C0

表格(T):



缺失值

☐ 在二分集内按照列表顺序排除个案(D)
☐ 在类别内按照列表顺序排除个案(G)

确定

粘贴(P)

重置(R)

取消

帮助

对多选题进行描述

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
\$C0 ^a	163	14.2%	984	85.8%	1147	100.0%

a. Dichotomy group tabulated at value 1.

\$C0 Frequencies

		Responses		Percent of Cases
		N	Percent	
家庭每月还贷情况 ^a	C0. 请问您的家庭目前有下列还贷支出吗：房贷	118	59.3%	72.4%
	C0. 请问您的家庭目前有下列还贷支出吗：车贷	33	16.6%	20.2%
	C0. 请问您的家庭目前有下列还贷支出吗：其他一般消费还贷	48	24.1%	29.4%
Total		199	100.0%	122.1%

a. Dichotomy group tabulated at value 1.

对多选题进行描述

多响应交叉表格

行(W): s7(??)

列(N): \$C0

层(L):

多响应集(M):

定义范围(G)...

多响应交叉表格: 定义变量范围

最小(N): 1


最大(X): 3

继续 取消 帮助

确定 粘贴(P) 重置(R) 取消

有序(O)	↓
有序(O)	↓
变量	↓
有序(O)	↓
名义(N)	↓
有序(O)	↓
有序(O)	↓
有序(O)	↓
有序(O)	↓
有序(O)	↓

对多选题进行描述

 多响应交叉表格: 选项 ×

单元格百分比

☒ 行(W) ☐ 列(C) ☐ 总计(T)

☐ 跨响应集匹配变量(M)

百分比基于

☒ 个案(S) ☐ 响应(R)

缺失值

☐ 在二分集内按照列表顺序排除个案(E)
☐ 在类别内按照列表顺序排除个案(X)

继续

取消

帮助

对多选题进行描述

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
s7*\$C0	163	14.2%	984	85.8%	1147	100.0%

s7*\$C0 Crosstabulation

			家庭每月还贷情况 ^a			Total
			C0. 请问您的家庭目前有下列还贷支出吗：房贷	C0. 请问您的家庭目前有下列还贷支出吗：车贷	C0. 请问您的家庭目前有下列还贷支出吗：其他一般消费还贷	
S7. 婚姻状况	已婚	Count	91	23	30	120
		% within s7	75.8%	19.2%	25.0%	
	未婚	Count	27	10	17	42
		% within s7	64.3%	23.8%	40.5%	
	离异/分居/丧偶	Count	0	0	1	1
		% within s7	0.0%	0.0%	100.0%	
Total		Count	118	33	48	163

Percentages and totals are based on respondents.

a. Dichotomy group tabulated at value 1.

THE END