



# 社会网络分析

## 第13讲 社会网络的实证研究

邵鹏 博士

副教授 硕士生导师

西安工程大学管理学院

shaopengde@sohu.com

inebm.cn



# 提纲

- 1 理论模型构建（模型与假设）
- 2 实证分析结构与工具（研究方法）
- 3 数据录入与描述性分析（数据收集）
- 4 方差分析
- 5 信度与效度分析（量表与样本检验）
- 6 相关与回归分析
- 7 中介与调节效应分析



# 1. 理论模型构建



# 构建模型的理论

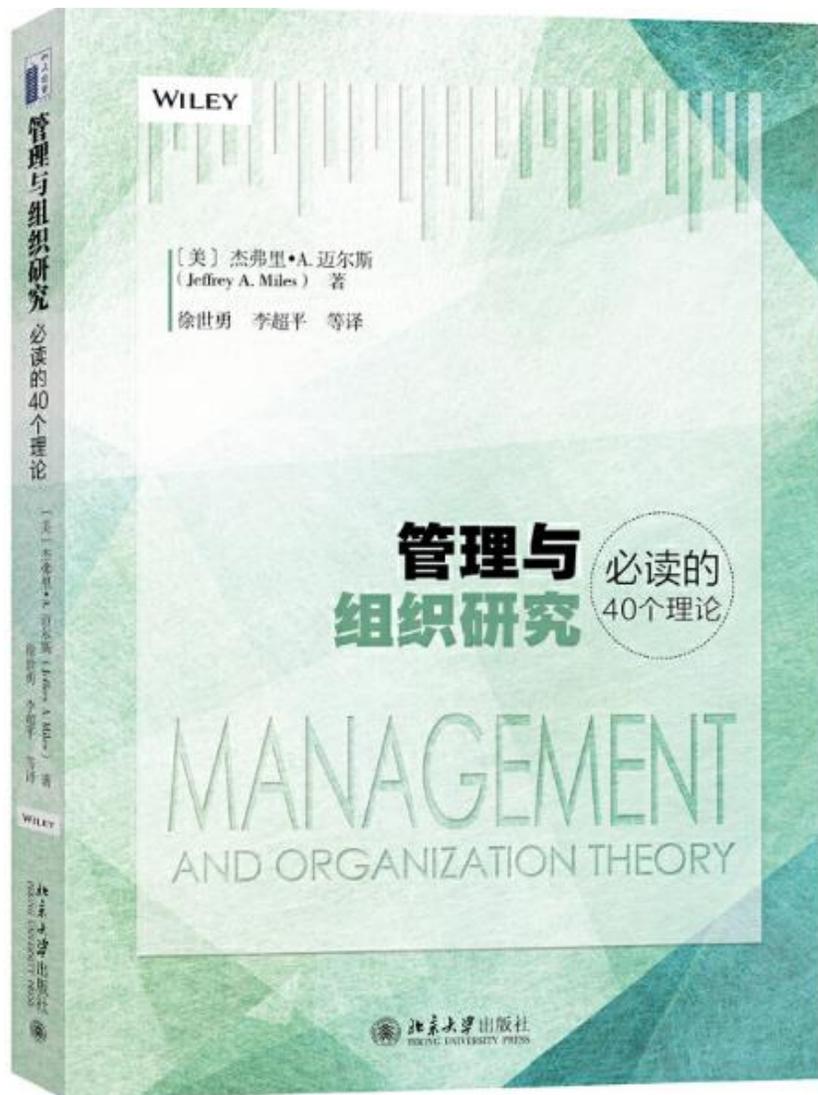
啤酒与尿不湿



相关关系与因果关系的区别？

为什么构建模型要有理论基础？

# 理论



- |               |                |                 |
|---------------|----------------|-----------------|
| 1 吸收能力理论 15   | 16 镜像理论 123    | 31 社会认知理论 228   |
| 2 行动者-网络理论 22 | 17 制度理论 129    | 32 社会比较理论 235   |
| 3 代理理论 30     | 18 知识基础理论 136  | 33 社会交换理论 242   |
| 4 议程设置理论 37   | 19 媒介丰富性理论 143 | 34 社会促进理论 249   |
| 5 依恋理论 45     | 20 心智模型理论 150  | 35 社会身份认同理论 256 |
| 6 归因理论 52     | 21 组织生态理论 157  | 36 社会网络理论 263   |
| 7 平衡理论 59     | 22 组织公正理论 164  | 37 利益相关者理论 271  |
| 8 控制理论 67     | 23 计划行为理论 171  | 38 结构权变理论 278   |
| 9 创新扩散理论 74   | 24 前景理论 178    | 39 结构化理论 285    |
| 10 动态能力理论 81  | 25 心理契约理论 185  | 40 交易成本理论 292   |
| 11 有效市场理论 88  | 26 资源基础理论 192  |                 |
| 12 伦理理论 95    | 27 角色理论 199    |                 |
| 13 场理论 102    | 28 自我决定理论 206  |                 |
| 14 博弈论 109    | 29 意义构建理论 214  |                 |
| 15 目标设定理论 116 | 30 社会资本理论 221  |                 |

# 理论



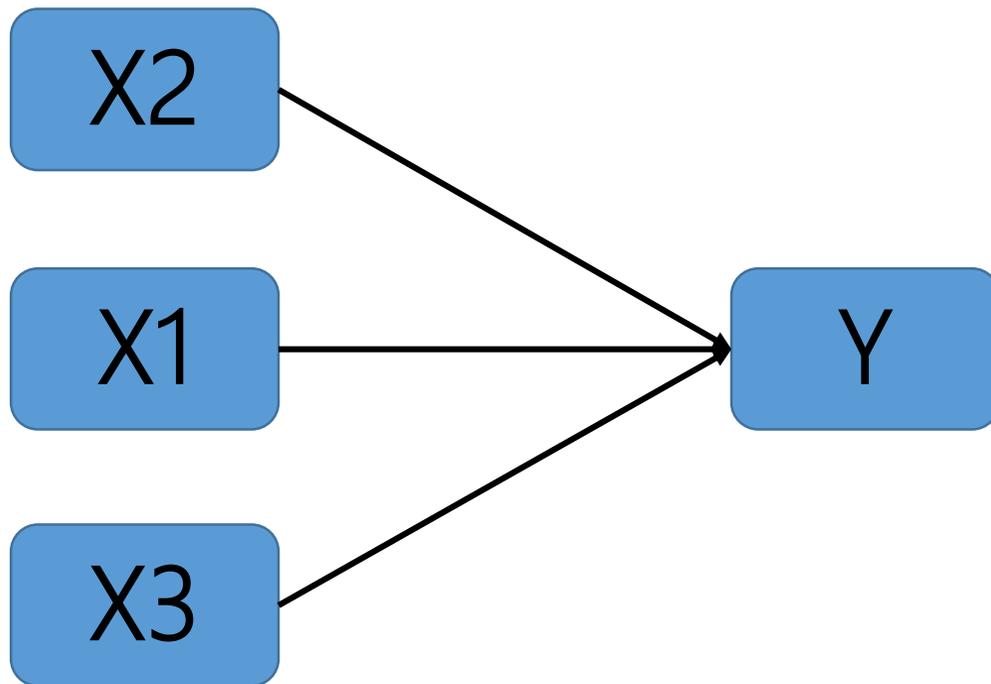
- 1情感事件理论
- 2积极情绪的拓展-建构理论
- 3协同演化理论
- 4认知评价理论
- 5认知失调理论
- 6道德补偿理论
- 7组织复杂性理论
- 8组织构型理论
- 9资源保存理论
- 10解释水平理论
- 11权变理论
- 12竞合理论
- 13决策双系统理论
- 14自我损耗理论
- 15情绪传染理论
- 16 情绪即社会信息理论
- 17伦理型领导理论
- 18事件系统理论
- 19公平启发理论
- 20情感信息理论
- 21目标定向理论
- 22内隐领导理论
- 23互动仪式链理论
- 24工作要求-资源模型
- 25领导-成员交换理论
- 26松散耦合理论
- 27中等阶层一致性理论
- 28少数派影响理论
- 29道德许可理论
- 30归属需求理论
- 31新制度理论
- 32组织学习理论
- 33个人-环境匹配理论
- 34种群生态理论
- 35调节定向理论
- 36相对剥夺理论
- 37资源配置理论
- 38资源依赖理论
- 39角色调和理论
- 40自我分类理论
- 41自我控制理论
- 42自我验证理论
- 43服务型领导理论
- 44信号理论
- 45社会传染理论
- 46社会契约理论
- 47社会支配倾向理论
- 48社会信息加工理论
- 49社会相互依赖理论
- 50社会学习理论
- 51社会表征理论
- 52现代管家理论
- 53战略选择理论
- 54结构适应理论
- 55竞赛理论
- 56特质激发理论
- 57变革型领导理论
- 58不确定管理理论
- 59不确定性-认同理论
- 60高层梯队理论





# 因果模型

X1、X2、X3是自变量，对因变量Y产生影响。变量间的影响关系（ $X \rightarrow Y$ ）是直接的因果链关系。

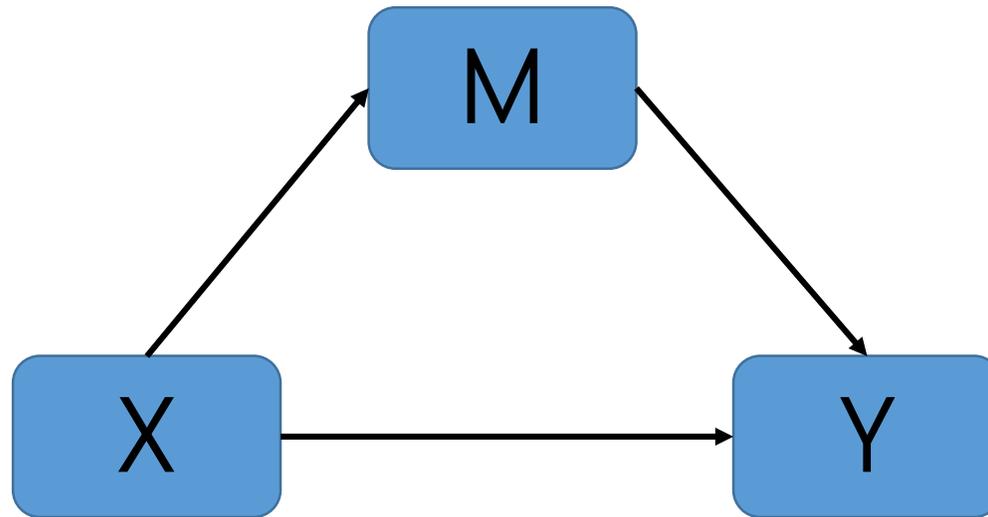




# 有中介变量的模型

中介效应是指变量间的影响关系 ( $X \rightarrow Y$ ) 不是直接的因果链关系而是通过一个或一个以上变量(M)的间接影响产生的。

M为中介变量，X通过M对Y产生的的间接影响称为中介效应。

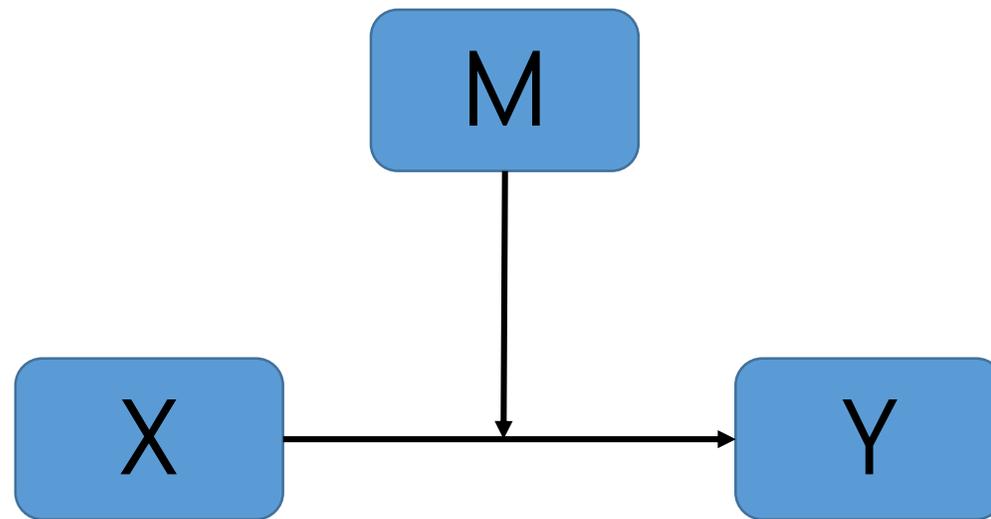




# 有调节变量的模型

调节效应是交互效应的一种，是有因果指向的交互效应，而单纯的交互效应可以互为因果关系；

调节变量一般不受自变量和因变量影响，但是可以影响自变量和因变量；  
常见的调节变量有性别、年龄、收入水平、文化程度、社会地位等。





# 例1：高铁网络与城市发展

## 高铁网络对城市生态效率的影响——基于中国277个地级市的空间计量研究

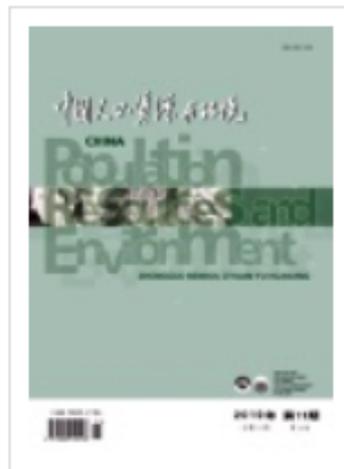
📖 记笔记

罗能生 田梦迪✉ 杨钧 李建明 王玉泽

湖南大学经济与贸易学院 新乡学院管理学院

📄 导出/参考文献 ◀ 分享 ▶ 🔔 创建引文跟踪 ☆ 收藏 🖨 打印

摘要：高铁开通产生了很大的社会经济效应与环境效应,综合研究高铁对这两种效应即生态效率的影响具有重要的理论意义和现实意义。本文采用2006—2016年中国277个地级市数据,将社会网络分析(SNA)和空间杜宾模型(SDM)相结合,检验我国高铁网络发展对城市生态效率的影响效应。研究发现:在全国范围内,高铁网络度数中心度(degree centrality)和中间中心度(betweenness centrality)的优化对城市生态效率产生显著正向效应,但是受城市异质性作用产生的影响效应具有明显差异性,具体表现为高铁网络中心度优化对中心区域产生正效应,对非中心区域产生负效应,即高铁网络在一定程度上加剧了城市生态效率不平衡、不协调发展的负面效应。在此基础上本文对结果进行机制检验,发现城市服务业集聚在高铁网络提高城市生态效率中具有中介作用。与此同时,采用人口地理矩阵和投入导向规模报酬不变的超效率DEA模型重新测算的城市生态效率分别对上述结果进行稳健性检验,结论依旧保持不变。因此,从长远发展来看,各级政府仍应大力发展高铁建设,推动高铁网络不断优化,带动经济效益和环境效应的提升,但是在进行高铁线路规划时,... 更多



中国人口·资源与环境  
China Population, Resources and Environment

2019年11期



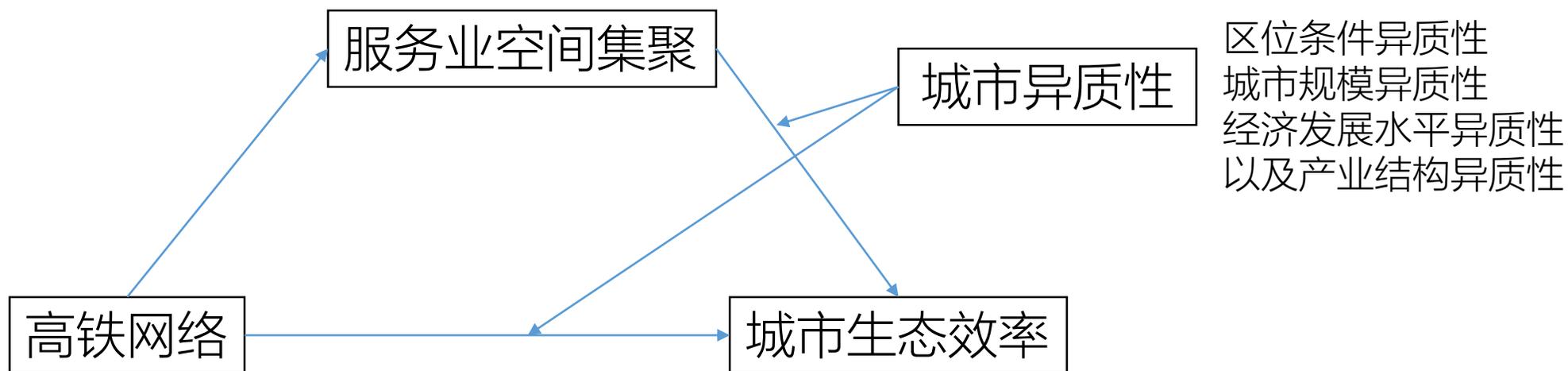
# 假设与模型

H1: 从全国范围来看，高铁网络的完善通过推动经济增长、降低环境污染两条途径，引致城市生态效率提升。

H2: 高铁网络对城市生态效率的作用效应可能受城市异质性影响，呈现出差异性。

H3: 高铁网络通过促进服务业空间集聚进而促进城市生态效率的提升

H4: 高铁网络将服务业空间集聚作为生态效率传导机制受城市异质性影响，呈现出差异性。





# 高铁网络构建

- 根据两个城市之间有无高铁构建复杂网络。铁路网络定义有三种空间，R空间以车次作为节点，L空间以车站作为节点，P空间以线路作为节点。本文采用P空间以线路来衡量节点联系，若两个城市之间开通高铁线路，赋值为“1”，反之赋值为“0”，
- 将城市6月份之前开通的高铁定义为当年开通，之后开通的高铁定义为下一年开通，由此构建二值网络进行测量。

2016年网络度数中心度前9名城市相关指标的数值

变量	南京	北京	广州	武汉	长沙	株洲	上饶	杭州	常州
度数中心度	14.660	14.136	13.613	13.089	13.089	13.089	13.089	13.089	11.518
中间中心度	5.539	6.736	14.861	26.311	6.227	6.227	2.956	9.643	1.430
服务业密度	0.148	0.452	0.174	0.206	0.146	0.086	0.084	0.244	0.094

# 分析



表 7 基于城市发展水平和城市产业结构的异质性分析

解释变量	发达城市		欠发达城市		产业优化度高城市		产业优化度低城市	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
生态效率滞后项	0.134 3 <sup>***</sup>	0.057 7 <sup>*</sup>	0.427 9 <sup>***</sup>	0.426 9 <sup>***</sup>	0.180 8 <sup>***</sup>	0.085 0 <sup>***</sup>	0.316 9 <sup>***</sup>	0.317 6 <sup>***</sup>
	(4.53)	(1.91)	(23.92)	(23.88)	(6.86)	(3.15)	(15.52)	(15.55)
节点城市的度数中心度	0.452 2 <sup>**</sup>		0.031 9		0.446 8 <sup>***</sup>		-0.183 4	
	(2.32)		(-0.36)		(2.9)		(1.44)	
节点城市的中间中心度		0.229 4 <sup>***</sup>		0.037 6		0.245 4 <sup>***</sup>		-0.007 4
		(8.20)		(0.83)		(10.62)		(-0.17)
控制其它变量	YES							
$W_X \times X$	YES							
$\rho$	0.523 3 <sup>***</sup>	0.529 6 <sup>***</sup>	0.124 5	0.115 6	0.663 4 <sup>***</sup>	0.653 3 <sup>***</sup>	0.225 9	0.225 8
	(3.24)	(3.25)	(0.77)	(0.71)	(4.34)	(4.19)	(1.61)	(1.61)
时间效应	YES							
地区效应	YES							
R-squared	0.134 8	0.074 0	0.231 0	0.358 1	0.156 0	0.162 6	0.390 4	0.456 2
N	940	940	1 830	1 830	1 260	1 260	1 510	1 510

## 例2：企业社交工作平台影响员工即兴能力的机理研究

**摘要：**在超竞争环境(Hypercompetitive Environments)下,企业员工必须不断提升其即兴能力来应对内外部的变化。以阿里巴巴的钉钉和腾讯的企业微信为代表的企业社交工作平台作为一类新兴的信息技术,为员工即兴能力的提升提供了新的可能。本文在进行文献综述的基础上,采用探索性案例研究方法,识别了企业社交工作平台的核心技术特征,并基于在线社会网络的视角提出了企业社交工作平台影响员工即兴能力的机理模型,最后进行了问卷实证检验。本文的主要研究结论如下:第一,可及性、社交性及交互性是企业社交工作平台的3个核心技术特征;第二,企业社交工作平台的3个核心技术特征通过员工在线社会网络提升员工的即兴能力。

**关键词：**企业社交工作平台 技术特征 即兴能力 在线社会网络

# 例2：企业社交工作平台影响员工即兴能力的机理研究



图1 本文的研究框架

## 两种方法

案例分析：明确研究问题，提出研究变量

统计分析：提出假设，问卷调查，结构方程模型检验

表1 常见的企业社交工作平台的归类

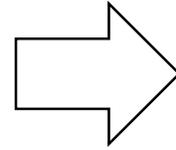
类型	典型软件应用
企业办公系统+社交元素	钉钉、微洽、伙伴、金蝶云之家、今目标、明道企业社会化协作平台、泛微OA、用友企业空间、北森tita、纷享销客、i8小时、Tower、Yammer等
社交平台+办公场景	微信、微博、QQ、Facebook、Twitter等

表2 企业社交工作平台技术特征各变量及典型条目

变量	条目数	典型条目
可及性	27	一位内部审计师：我们一般会把常用的文档共享在里面(QQ群)，大家都可以自由下载
社交性	39	一位项目经理：微洽是有微会议的功能的，很方便就能把相关人员拉进来，能够快速有效地开展讨论
交互性	20	一位IT主管：在这个社区上每个人都能很容易发布自己想发布的帖子，不管是关于自己的工作感悟也好还是对于公司的建议也好

# 例2：企业社交工作平台影响员工即兴能力的机理研究

## 案例访谈



## 案例编码

采用 Miles 和 Huberman ( 1984 ) 的三角测量法，通过多种数据来源对案例进行分析，以保证研究结论的信度。

- 对 X 电信公司的高管及各主要岗位职员共 18 人的深度访谈、X 电信公司公开披露的年度报告以及官方网站发布的相关信息；
- 对 Y 终端公司 ZJ 分公司推广企业社交工作平台的项目负责人进行了深入访谈，通过官方网站等途径查找了关于该公司的相关介绍，通过网络查找了关于该公司使用企业社交工作平台的相关报道及已有研究。

表2 企业社交工作平台技术特征各变量及典型条目

变量	条目数	典型条目
可及性	27	一位内部审计师：我们一般会把常用的文档共享在里面(QQ群)，大家都可以自由下载
社交性	39	一位项目经理：微洽是有微会议的功能的，很方便就能把相关人员拉进来，能够快速有效地开展讨论
交互性	20	一位IT主管：在这个社区上每个人都能很容易发布自己想发布的帖子，不管是关于自己的工作感悟也好还是对于公司的建议也好

# 例2：企业社交工作平台影响员工即兴能力的机理研究

假设提出

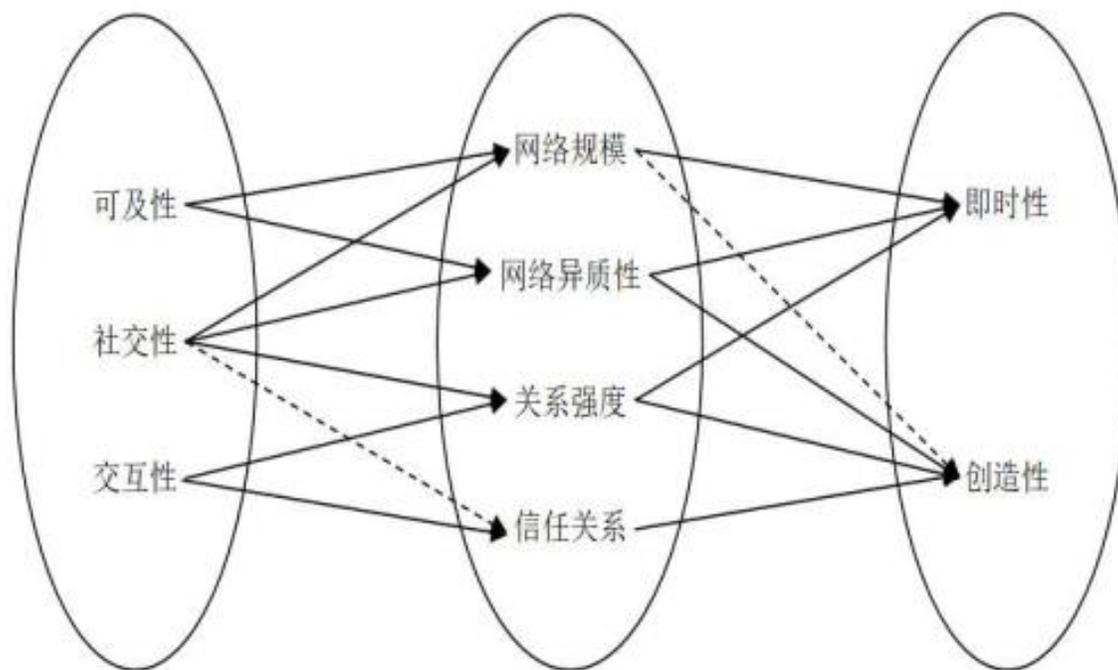


图2 企业社交工作平台影响员工即兴能力机理图  
注：其中实线代表强关系，虚线代表弱关系。



# 数据收集

## 预调研

向 108 名 MBA 学员发放了预调研问卷，得到有效问卷 90 份。根据预调研的数据，我们进一步调整了创造性变量其中一个题项的表述。

## 正式问卷发放

共发放了 286 份，回收 253 份，其中有效问卷 231 份，回收率为 88.5%，有效回收率为 80.8%。主要从 MBA 学生课堂、企业实地发放以及在线发放 3 种途径进行正式问卷的发放。

表 3 正式问卷的发放及回收情况

渠道	发放数量	有效回收数量	有效回收率
MBA 课堂	150	136	90.6%
企业实地发放	70	52	74.3%
在线发放	66	43	65.2%
合计	286	231	80.8%

## 信效度检验—SPSS

表4 各变量的量表来源及信度

变量	量表来源	题项数	Cronbach's $\alpha$	组合信度
可及性	(Nelson et al., 2005)	3	0.84	0.84
社交性	(Animesh et al., 2011)	5	0.83	0.83
交互性		5	0.85	0.85
网络规模	(Hoye et al., 2009)	3	0.78	0.79
网络异质性	(谢雅萍、黄美娇, 2014)	4	0.80	0.81
信任关系		4	0.73	0.78
关系强度		3	0.88	0.90
即时性	(Vera and Crossan, 2005)	3	0.82	0.81
创造性		4	0.81	0.81

# 分析



## 路径分析—AMOS

表5 模型最终的检验结果

			路径系数	标准误差	P值
网络规模	←	可及性	0.562	0.100	***
网络异质性	←	可及性	0.345	0.093	***
网络规模	←	社交性	0.313	0.104	0.003
关系强度	←	社交性	0.698	0.186	***
关系强度	←	交互性	0.291	0.076	***
信任关系	←	交互性	0.593	0.089	***
即时性	←	网络规模	0.118	0.051	0.019
即时性	←	关系强度	0.298	0.044	***
创造性	←	网络异质性	0.129	0.053	0.015
创造性	←	关系强度	0.302	0.082	***
创造性	←	信任关系	0.154	0.069	0.028
即时性	←	自我效能感	0.225	0.062	***
创造性	←	自我效能感	0.316	0.061	***
创造性	←	线下网络	0.374	0.090	***
即时性	←	线下网络	0.243	0.087	0.005
CMIN/DF=2.112 DF=721 P=0.000 RMSEA=0.056					
GFI=0.913 NFI=0.908 CFI=0.921					

注：\*\*\*表示P<0.001。



## 2. 实证分析结构与工具



# 实证论文结构

## 0 引言

## 1 理论基础

## 2 模型与假设

### 2.1 模型构建

### 2.2 假设提出

## 3 研究设计

### 3.1 量表设计

### 3.2 问卷调查

预调查、修正问卷、正式调查

### 3.3 信度、效度分析、验证性因子分析

### 3.4 样本描述

## 4 分析

### 4.1 相关分析

### 4.2 回归分析

### 4.3 中介、调节效应分析

### 4.4 结构方程模型路径分析

## 5 讨论与建议

## 6 结论与展望



# SPSS的发展

SPSS为IBM公司推出的一系列用于统计学分析运算、数据挖掘、预测分析和决策支持任务的软件产品及相关服务的总称，有Windows和Mac OS X等版本。

SPSS是世界上最早的统计分析软件，由美国斯坦福大学的三位研究生Norman H. Nie、C. Hadlai (Tex) Hull 和 Dale H. Bent于1968年研究开发成功，同时成立了SPSS公司，并于1975年成立法人组织、在芝加哥组建了SPSS总部。

2009年7月28日，IBM公司宣布将用12亿美元现金收购统计分析软件提供商SPSS公司。如今SPSS已出至版本22.0，而且更名为IBM SPSS。迄今，SPSS公司已有40余年的成长历史。



最初软件全称为“社会科学统计软件包”(Solutions Statistical Package for the Social Sciences)  
但是随着SPSS产品服务领域的扩大和服务深度的增加正式将英文全称更改为“统计产品与服务解决方案”(Statistical Product and Service Solutions)



# SPSS的特点

01

## 操作简便

界面非常友好，大多数操作可通过鼠标拖曳、点击“菜单”、“按钮”和“对话框”来完成

02

## 编程方便

SPSS的命令语句、子命令及选择项的选择绝大部分由“对话框”的操作完成。因此，用户无需花大量时间记忆大量的命令、过程、选择项。

03

## 功能强大

具有完整的数据输入、编辑、统计分析、报表、图形制作等功能。

04

## 数据接口

能够读取及输出多种格式的文件。比如\*.dbf文件，ASCII数据文件，\*.xls文件等。

05

## 针对性强

SPSS针对初学者、熟练者及精通者都比较适用。只需要掌握简单的操作分析，大多青睐于SPSS。



# SPSS的界面

The image displays two side-by-side windows of the SPSS Data Editor. Both windows are titled 'Untitled1 [DataSet0] - SPSS Data Editor'. The left window is in 'Variable View' mode, showing a grid with columns labeled 'var' and rows numbered 1 to 16. The right window is in 'Variable View' mode, showing a table with columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, and Measure. The rows are numbered 1 to 17. The 'Name' column for row 1 is highlighted in blue. The status bar at the bottom right of the right window indicates 'SPSS Processor is ready'.



# SPSS的界面

The screenshot shows the SPSS Viewer interface. The title bar reads '\*Output1 [Document1] - SPSS Viewer'. The menu bar includes File, Edit, View, Data, Transform, Insert, Format, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The left sidebar shows a tree view with 'Output' expanded to 'Correlations'. The main area displays the following information:

[DataSet1] C:\Users\Suzy\Desktop\SPSS讲座\示例11.sav

**Descriptive Statistics**

	Mean	Std. Deviation	N
性别	1.65	.489	20
成绩	89.9500	4.38268	20

**Correlations**

		性别	成绩
性别	Pearson Correlation	1	.458*
	Sig. (2-tailed)		.042
	N	20	20
成绩	Pearson Correlation	.458*	1
	Sig. (2-tailed)	.042	
	N	20	20

\*. Correlation is significant at the 0.05 level (2-tailed).

SPSS Processor is ready

# AMOS



The screenshot displays the AMOS software interface. The title bar reads "Unnamed project : Group number 1 : Input". The menu bar includes "File", "Edit", "View", "Diagram", "Analyze", "Tools", "Plugins", and "Help". The toolbar on the left contains various icons for creating and editing path diagrams, including rectangles, ovals, arrows, and text boxes. The central workspace is currently empty. The bottom status bar shows "Path diagram" and "Tables" tabs, and a message: "Not estimating any user-defined estimand."



# 学习推荐



SPSS有话说

微信号 SpssSpeaking

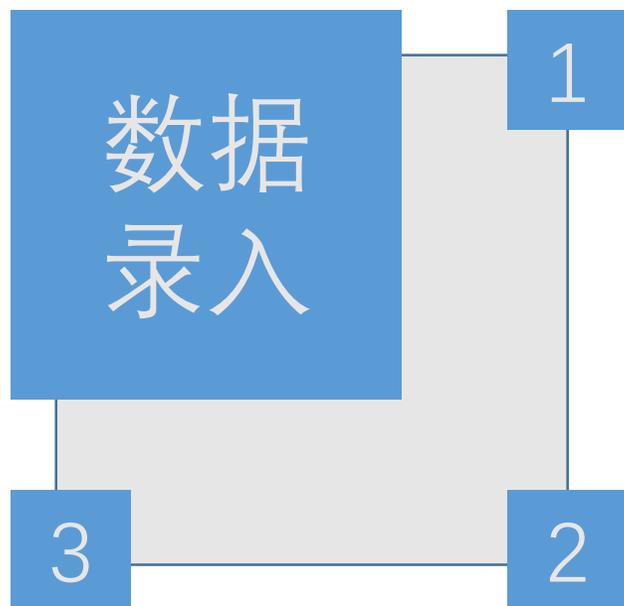
功能介绍 共享spss、amos、mplus等统计分析方法与操作，帮助大家解决统计难、操作难、毕业难的问题。



# 3. 数据录入与描述性分析



# 数据录入的步骤



## 定义变量名

姓名、性别、成绩...  
地区、时间、营业额...  
身高、体重、智力...

## 指定每个变量的各种属性

变量名、变量标签值、变量的储存类型、缺失值、变量的测量尺度

## 录入数据

变量名不能与SPSS保留字相同。

SPSS的保留字有ALL、END、BY、EQ、GE、GT、LE、  
LT、NE、NOT、OR、TO、WITH。



# 数据录入的格式

01

**第一**

“一个观测占一行，一个变量占一列”

不同观测对象的数据不能在同一记录中出现，即同一观测数据应当独占一行。

02

**第二**

每一个测量指标或影响因素只能占据一列的位置，即同一指标的观测值都应当录入到同一个变量中去。



# 数据录入的格式

- **Name**：变量名。变量名必须以字母、汉字及@开头，总长度不超过8个字符，共容纳4个汉字或8个英文字母，英文字母不区别大小写，最后一个字符不能是句号。
- **Type**：变量类型。变量类型有8种，最常用的是Numeric数值型变量。其它常用的类型有：String字符型，Date日期型，Comma逗号型（隔3位数加一个逗号）等。
- **Width**：变量所占的宽度。
- **Decimals**：小数点后位数。
- **Label**：变量标签。关于变量涵义详细说明。
- **Values**：变量值标签。关于变量各个取值的涵义说明。
- **Missing**：缺失值的处理方式。
- **Columns**：变量在Date View中所显示的列宽（默认列宽为8）。
- **Align**：数据对齐格式（默认为右对齐）。
- **Measure**：数据的测度方式。名义尺度、定序尺度和等间距尺度三种（默认为等间距尺度）。

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	编号	Numeric	8	0		None	None	8	≡ Right	Ordinal
2	姓名	String	8	0		None	None	8	≡ Right	Nominal
3	性别	Numeric	8	0		{1, 男}...	None	8	≡ Right	Nominal
4	成绩	Numeric	8	2		None	None	8	≡ Right	Scale



# 数据的直接录入

示例1，将下面的数据按要求录入到SPSS中

姓名	性别	期末成绩
小王	男	96.5
小张	女	90

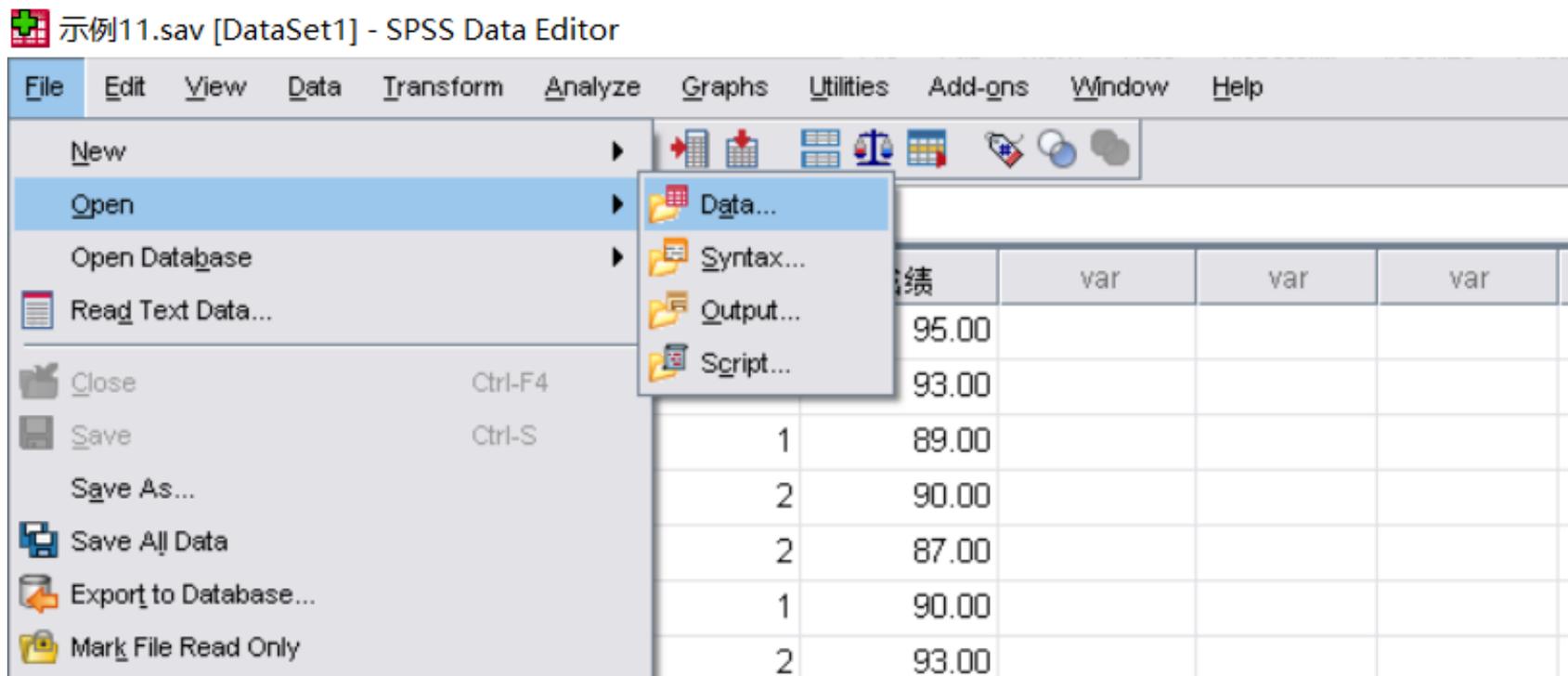
要求：

- ① 姓名：字符型；宽度8；小数点0；列宽8；左对齐；称名变量
- ② 性别：数字型；宽度8；小数点0；1代表男，2代表女；列宽8；左对齐；称名变量
- ③ 期末成绩：数字型；宽度10；小数点1；列宽8；右对齐；等比变量



# 外部数据获取

SPSS中可以直接读入许多常用格式的数据文件，选择菜单File→Open Data或直接单击快捷键工具栏上的 快捷按钮，系统就会弹出Open File对话框，单击“文件类型”列表框，在里面能够看到可以直接打开的数据文件格式。





# 描述性分析

## Frequencies：频数分布

调用此过程可进行频数分布表的分析。频数分布表是描述性统计中最常用的方法之一，此外还可对数据的分布趋势进行初步分析。

## Descriptives：一般性描述

调用此过程可对变量进行描述性统计分析，计算并列出一系列相应的统计指标，且可将原始数据转换成标准Z分值并存入数据库，所谓Z分值是指某原始数值比其均值高或低多少个标准差单位，高的为正值，低的为负值，相等的为零。

## Explore：探索性分析

调用此过程可对变量进行更为深入详尽的描述性统计分析，故称之为探索性统计。它在一般描述性统计指标的基础上，增加有关数据其它特征的文字与图形描述，显得更加细致与全面，有助于用户思考对数据进行进一步分析的方案。

## Crosstabs：交叉列表

调用此过程可进行计数资料和某些等级资料的交叉表分析，在分析中，可对二维至多维交叉表资料进行统计描述和 $\chi^2$ 检验，并计算相应的百分数。



# 描述性分析

示例2.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform **Analyze** Graphs Utilities Add-ons Window Help

1: 编号 1

	编号	姓名
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	
8	8	
9	9	
10	10	

Reports

**Descriptive Statistics**

- 123 Frequencies...
- Descriptives...
- Explore...
- Crosstabs...
- 1/2 Ratio...
- P-P Plots...
- Q-Q Plots...

Tables

Compare Means

General Linear Model

Generalized Linear Models

Mixed Models

Correlate

Regression

Loglinear

Neural Networks

Classify

Data Reduction



# 描述性分析

频数分布

一般性描述

探索性分析

交叉列表



# 频数分布表(Frequencies)

The screenshot shows the SPSS 'Frequencies' dialog box. On the left, a list of variables includes '编号' (ID), '姓名' (Name), and '性别' (Gender). The 'Variable(s):' box on the right contains '成绩' (Score). Below the variable lists is a checked checkbox for 'Display frequency tables'. On the right side, there are three buttons: 'Statistics...', 'Charts...', and 'Format...'. At the bottom are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. Five callout boxes with Chinese text point to these elements: '待分析的变量' (Variables to be analyzed) points to the 'Variable(s):' box; '定义需要计算的统计量' (Define statistics to be calculated) points to the 'Statistics...' button; '定义需要绘制的统计图' (Define charts to be drawn) points to the 'Charts...' button; '定义表格' (Define tables) points to the 'Format...' button; and '是否显示频数表' (Whether to display frequency tables) points to the 'Display frequency tables' checkbox.



# Frequencies : 定义统计量

The screenshot shows the 'Frequencies: Statistics' dialog box with the following sections and callouts:

- Percentile Values** (Callout: 百分位数):
  - Quartiles
  - Cut points for: 10 equal groups
  - Percentile(s):
  - Buttons: Add, Change, Remove
- Central Tendency** (Callout: 集中趋势):
  - Mean
  - Median
  - Mode
  - Sum
- Dispersion** (Callout: 离散趋势):
  - Std. deviation    Minimum
  - Variance    Maximum
  - Range    S.E. mean
- Distribution** (Callout: 分布特征描述):
  - Skewness
  - Kurtosis

Additional options:  Values are group

Buttons: Continue, Cancel, Help



# Frequencies : 定义统计图

The image shows a screenshot of the 'Frequencies: Charts' dialog box in a statistical software package. The dialog box has a title bar with a red cross icon and the text 'Frequencies: Charts'. It contains two main sections: 'Chart Type' and 'Chart Values'. In the 'Chart Type' section, there are radio buttons for 'None', 'Bar charts', 'Pie charts', and 'Histograms:'. Under 'Histograms:', there is a checkbox labeled 'With normal curve'. In the 'Chart Values' section, there are radio buttons for 'Frequencies' and 'Percentages'. At the bottom of the dialog box are three buttons: 'Continue', 'Cancel', and 'Help'. Several Chinese callout boxes with arrows point to specific elements in the dialog box:

- '统计图类型' (Chart Type) points to the 'Chart Type' section.
- '无图形' (No chart) points to the 'None' radio button.
- '条图' (Bar chart) points to the 'Bar charts' radio button.
- '饼图' (Pie chart) points to the 'Pie charts' radio button.
- '直方图' (Histogram) points to the 'Histograms:' radio button.
- '直方图加上正态曲线' (Histogram with normal curve) points to the 'With normal curve' checkbox.
- '以频数绘制条图或饼图' (Draw bar chart or pie chart with frequencies) points to the 'Frequencies' radio button.
- '以构成比绘制条图或饼图' (Draw bar chart or pie chart with percentages) points to the 'Percentages' radio button.



# Frequencies : 定义表格

The screenshot shows the 'Frequencies: Format' dialog box with the following options and callouts:

- 按数值升序** (Ascending values):  Ascending values
- 按数值降序** (Descending values):  Descending values
- 按频数升序** (Ascending counts):  Ascending counts
- 按频数降序** (Descending counts):  Descending counts
- 频数表排列次序** (Multiple Variables):  Compare variables,  Organize output by variables
- Suppress tables with more than n categories
- Maximum number of categories: 10
- Buttons: Continue, Cancel, Help

# Frequencies : 结果解释

Statistics

成绩		
N	Valid	10
	Missing	0
Mean		89.9000
Std. Error of Mean		1.02686
Median		90.0000
Mode		90.00 <sup>a</sup>
Std. Deviation		3.24722
Variance		10.544
Skewness		-.035
Std. Error of Skewness		.687
Kurtosis		-.920
Std. Error of Kurtosis		1.334
Range		10.00
Minimum		85.00
Maximum		95.00
Sum		899.00
Percentiles	25	86.7500
	50	90.0000
	75	93.0000

Mean		均数
Std. Error of Mean		标准误
Median		中位数
Mode		众数
Std. Deviation		标准差
Variance		方差
Skewness		峰度系数
Std. Error of Skewness		峰度系数的标准误
Kurtosis		偏度系数
Std. Error of Kurtosis		偏度系数的标准误
Range		全距
Minimum		最小值
Maximum		最大值
Sum		合计
Percentiles	25	25%位数
	50	50%位数
	75	75%位数



# Frequencies : 结果解释

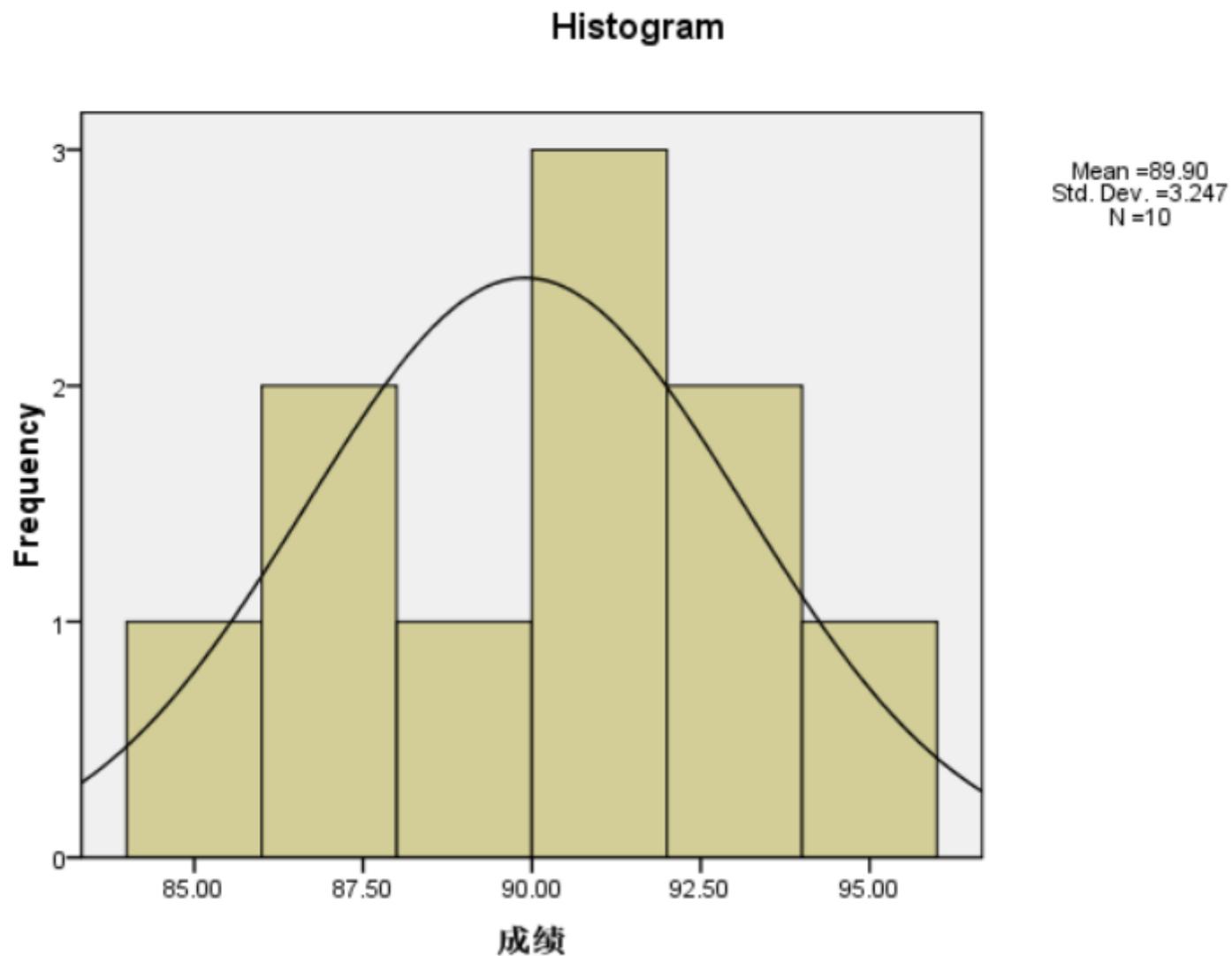
Frequency	Percent	Valid Percent	Cumulative Percent
频数	百分比	有效百分比	累积百分比

## 成绩

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 85	1	10.0	10.0	10.0
86	1	10.0	10.0	20.0
87	1	10.0	10.0	30.0
89	1	10.0	10.0	40.0
90	2	20.0	20.0	60.0
91	1	10.0	10.0	70.0
93	2	20.0	20.0	90.0
95	1	10.0	10.0	100.0
Total	10	100.0	100.0	



# Frequencies : 结果解释





# 一般性描述(Descriptives)





# Descriptives : 定义统计量

Descriptives: Options

Mean  Sum

**Dispersion**

Std. deviation  Minimum  
 Variance  Maximum  
 Range  S.E. mean

**Distribution**

Kurtosis  Skewness

**Display Order**

Variable list  
 Alphabetic  
 Ascending means  
 Descending means

Continue Cancel Help





# Descriptives : 结果解释

编号	姓名	性别	成绩	Z成绩
1	赵	2	95.00	1.57057
2	钱	2	93.00	0.95466
3	孙	1	89.00	-0.27716
4	李	2	90.00	0.03080
5	周	2	87.00	-0.89307
6	吴	1	90.00	0.03080
7	郑	2	93.00	0.95466
8	王	2	91.00	0.33875
9	冯	1	86.00	-1.20103
10	陈	1	85.00	-1.50898

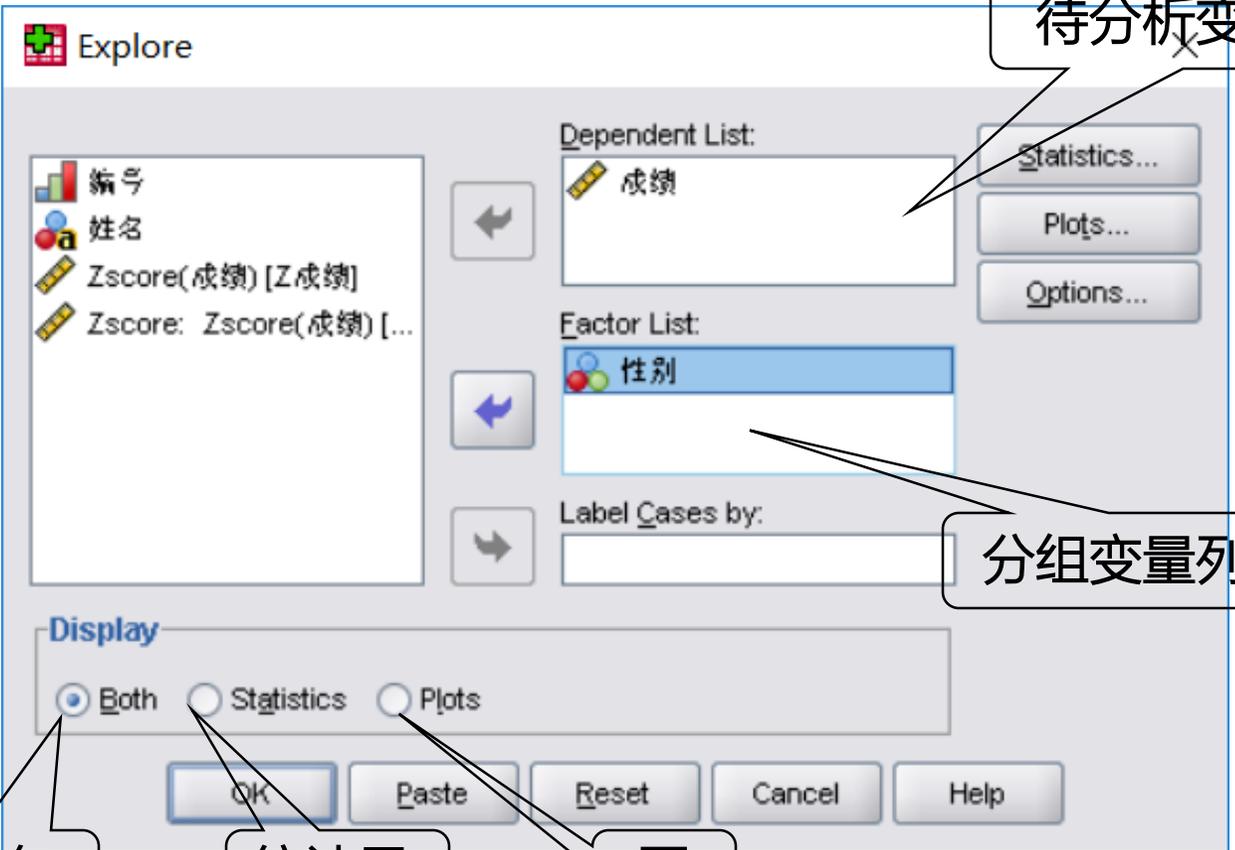
系统以z成绩为变量名将原始数据转换成标准z分值，存放在原数据库中。新变量具有均值为0、标准差为1的特征，亦即变量的标准化过程。

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
	Statistic	Statistic	Statistic	Statistic	Statistic
Zscore(成绩)	10	-1.50898	1.57057	.0000000	1.0000000
Valid N (listwise)	10				



# 探索性分析(Explore)



待分析变量列表

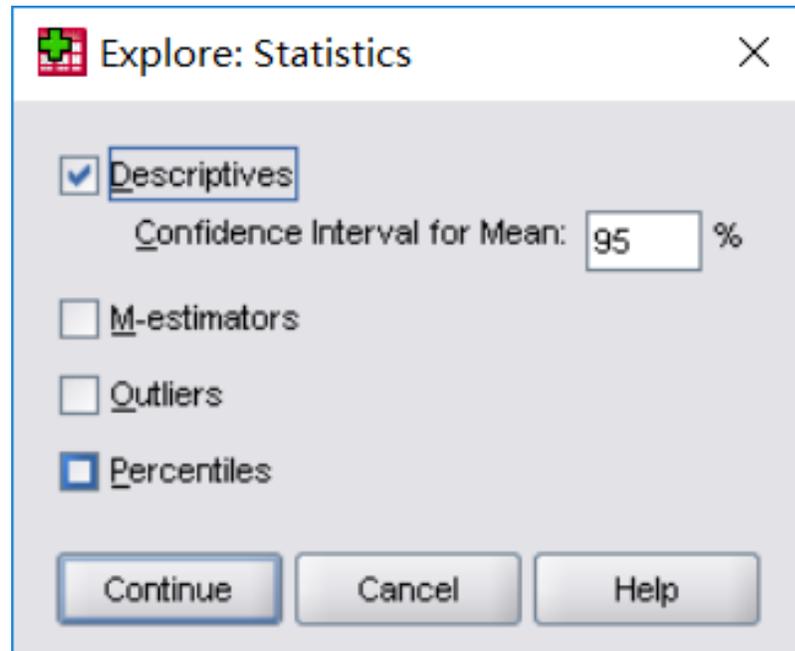
分组变量列表

两者均有

统计量

图

# Explore : 定义统计量



M-estimators : 作中心趋势的粗略最大似然确定, 输出四个不同权重的最大似然确定数。  
Outliers : 输出五个最大值与五个最小值。  
Percentiles : 输出第5%、10%、25%、50%、75%、90%、95%位数。

Descriptives : 输出均数、均数的95%可信区间、去掉5%极端值的均数、中位数、方差、标准差、最小值、最大值、全距、四分位数间距、偏度系数、峰度系数。



# Explore : 定义统计图

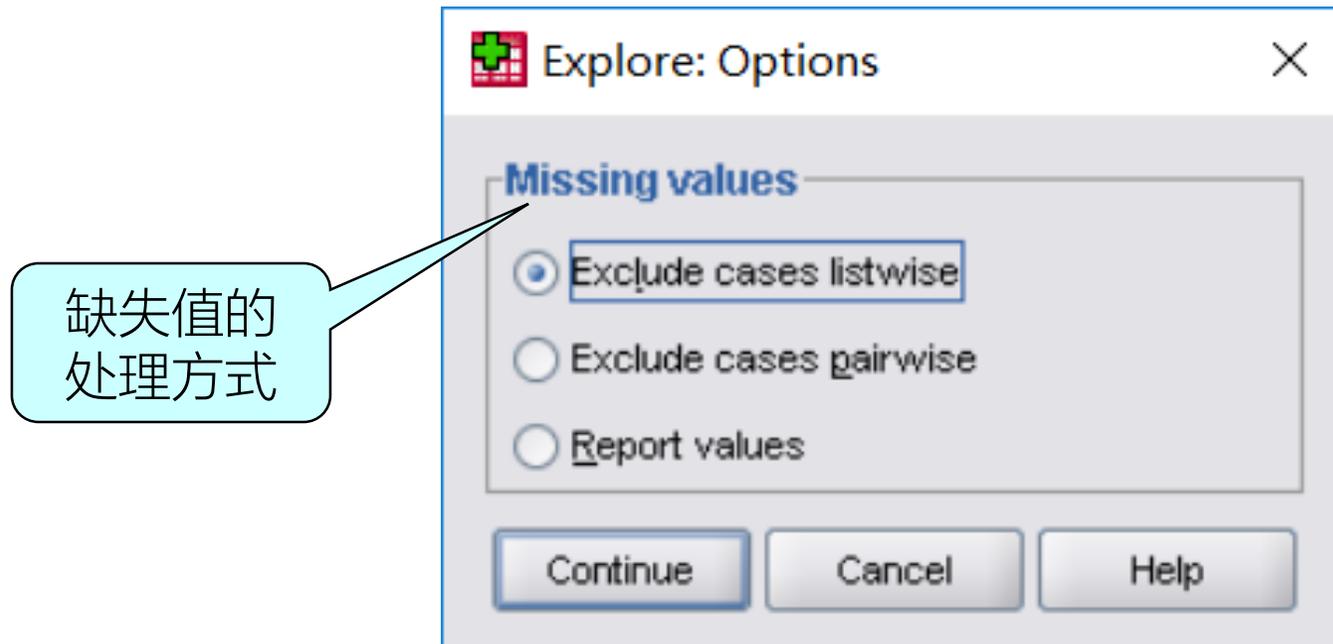
The screenshot shows the 'Explore: Plot' dialog box with the following settings and annotations:

- Boxplots:**  Factor levels together (Annotated: 箱图绘制方式)
- Boxplots:**  Dependents together
- Boxplots:**  None
- Descriptive:**  Stem-and-leaf (Annotated: 茎叶图)
- Descriptive:**  Histogram (Annotated: 直方图)
- Normality plots with tests (Annotated: 正态性检验/正态分布图)
- Spread vs Level with Levene Test:**  None
- Spread vs Level with Levene Test:**  Power estimation
- Spread vs Level with Levene Test:**  Transformed Power: Natural log
- Spread vs Level with Levene Test:**  Untransformed

Buttons: Continue, Cancel, Help



# Explore：缺失值的处理



Exclude case listwise:不分析有任一缺失值的记录

Exclude case pairwise:不分析计算某统计量时有缺失值的记录

Report values:报告缺失值



# Explore：结果解释

Case Processing Summary

		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
成绩	男	4	100.0%	0	.0%	4	100.0%
	女	6	100.0%	0	.0%	6	100.0%

原始数据的基本情况：男性有效数据4，缺失0，合计4；女性有效数据6，缺失0，合计6。



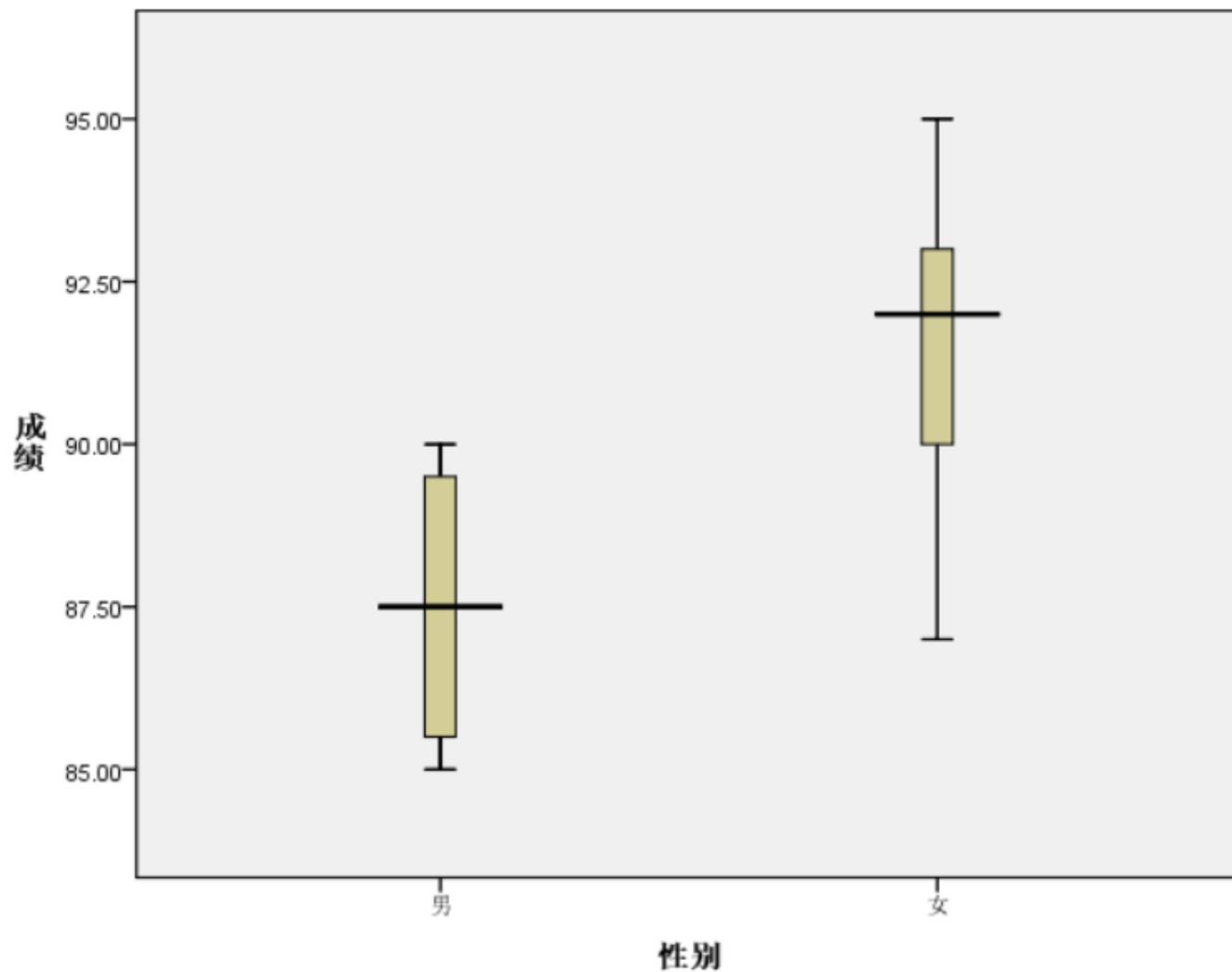
# Explore : 结果解释

Descriptives

性别		Statistic	Std. Error
成绩	男	Mean	87.5000
	95% Confidence Interval for Mean	Lower Bound	83.7121
		Upper Bound	91.2879
	5% Trimmed Mean	87.5000	
	Median	87.5000	
	Variance	5.667	
	Std. Deviation	2.38048	
	Minimum	85.00	
	Maximum	90.00	
	Range	5.00	
	Interquartile Range	4.50	
	Skewness	.000	1.014
	Kurtosis	-4.339	2.619
	女	Mean	91.5000
95% Confidence Interval for Mean		Lower Bound	88.5504
		Upper Bound	94.4496
5% Trimmed Mean		91.5556	
Median		92.0000	
Variance		7.900	
Std. Deviation		2.81069	
Minimum		87.00	
Maximum		95.00	
Range		8.00	
Interquartile Range		4.25	
Skewness		-.608	.845
Kurtosis		.203	1.741

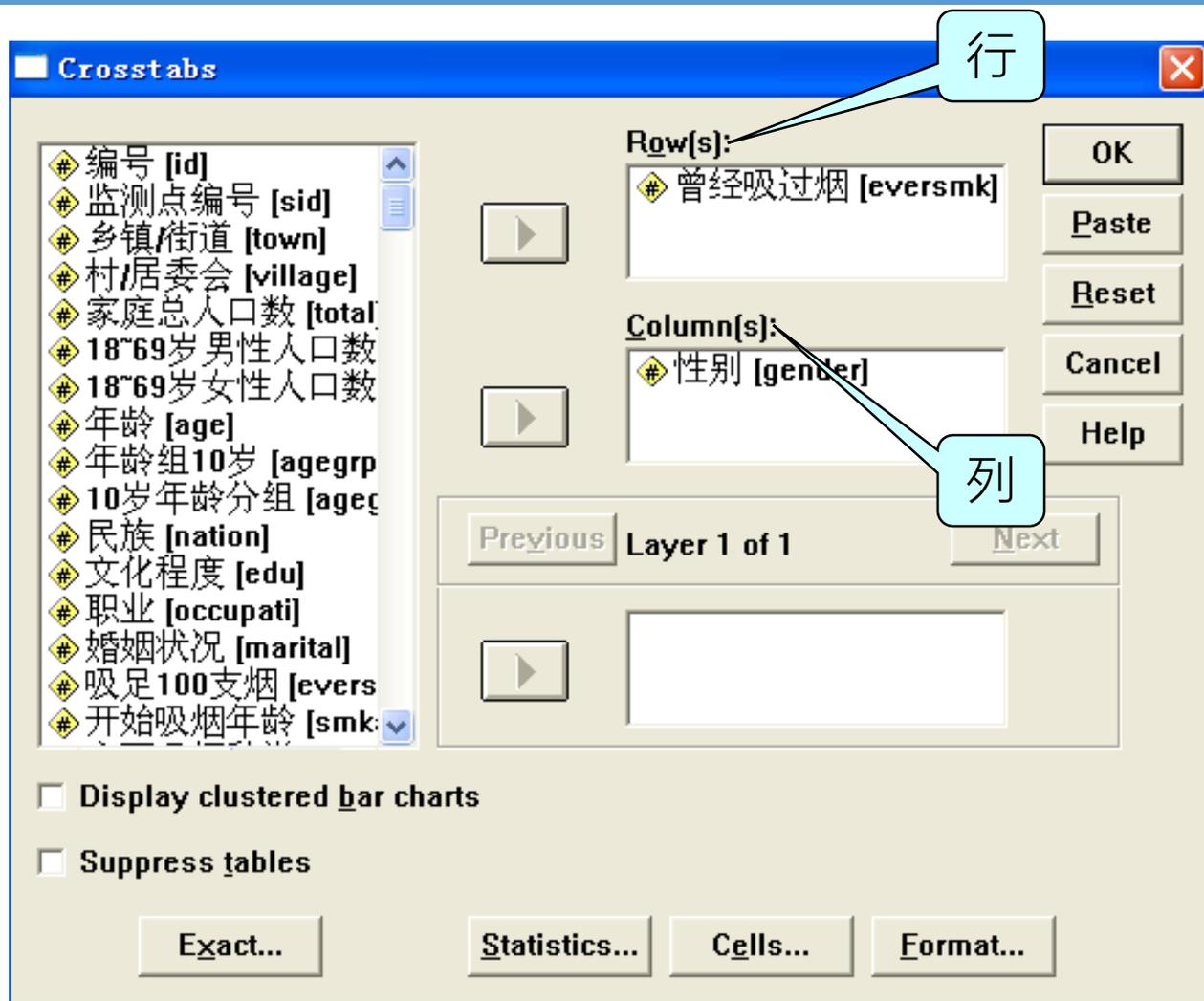


# Explore : 结果解释





# 交叉列表(Crosstabs)





# Crosstabs : 定义统计量

X<sup>2</sup>检验

Crosstabs: Statistics

Chi-square

Correlations

**Nominal**

Contingency coefficient

Phi and Cramer's V

Lambda

Uncertainty coefficient

**Ordinal**

Gamma

Somers' d

Kendall's tau-b

Kendall's tau-c

**Nominal by Interval**

Eta

Kappa

Risk

McNemar

Cochran's and Mantel-Haenszel statistics

Test common odds ratio equals: 1

Continue Cancel Help



# Crosstabs : 定义交叉表内容

The screenshot shows the 'Crosstabs: Cell Display' dialog box with the following options and callouts:

- Counts:**
  - Observed (实际观察数)
  - Expected (理论数)
- Percentages:**
  - Row (行百分数)
  - Column (列百分数)
  - Total (合计百分数)
- Residuals:**
  - Unstandardized (残差)
  - Standardized
  - Adjusted standardized
- Noninteger Weights:**
  - Round cell counts
  - Round case weights
  - Truncate cell counts
  - Truncate case weights
  - No adjustments

Buttons: Continue, Cancel, Help



# Crosstabs : 结果解释

## Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
曾经吸过烟 * 性别	999	99.9%	1	.1%	1000	100.0%



# Crosstabs : 结果解释

曾经吸过烟 \* 性别 Crosstabulation

			性别		Total
			男	女	
曾经吸过烟	从来都不吸	Count	108	513	621
		% within 性别	23.5%	95.2%	62.2%
	吸, 现在仍在吸	Count	283	20	303
		% within 性别	61.5%	3.7%	30.3%
	以前吸过, 现在已经不吸了	Count	69	6	75
		% within 性别	15.0%	1.1%	7.5%
Total	Count	460	539	999	
	% within 性别	100.0%	100.0%	100.0%	



# Crosstabs : 结果解释

## Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	542.476 <sup>a</sup>	2	.000
Likelihood Ratio	615.619	2	.000
Linear-by-Linear Association	456.325	1	.000
N of Valid Cases	999		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 34.53.



# 4. 方差分析

# 方差分析



单因素  
方差分析

双因素  
方差分析



# 单因素方差分析

单因变量的单因素方差分析主要解决多于两个总体样本或变量间均值的比较问题。是一种对多个（大于两个）总体样本的均值是否存在显著差异的检验方法。其目的也是对不同的总体的数据的均值之间的差异是否显著进行检验。

单因素方差分析的应用条件：在不同的水平（因素变量取不同值）下，各总体应当服从方差相等的正态分布。



# 单因素方差分析

示例3，某企业需要一种零件，现有三个不同的地区的企业生产的同种零件可供选择，为了比较这三个零件的强度是否相同，每个地区的企业抽出6件产品进行强度测试，其值如表所示。假设每个企业零件的强度值服从正态分布，试检验这三个地区企业的零件强度是否存在显著差异。

	1	2	3
1	116	110	89
2	98	103	85
3	100	118	99
4	115	106	73
5	83	107	97
6	105	116	102

解：首先建立假设

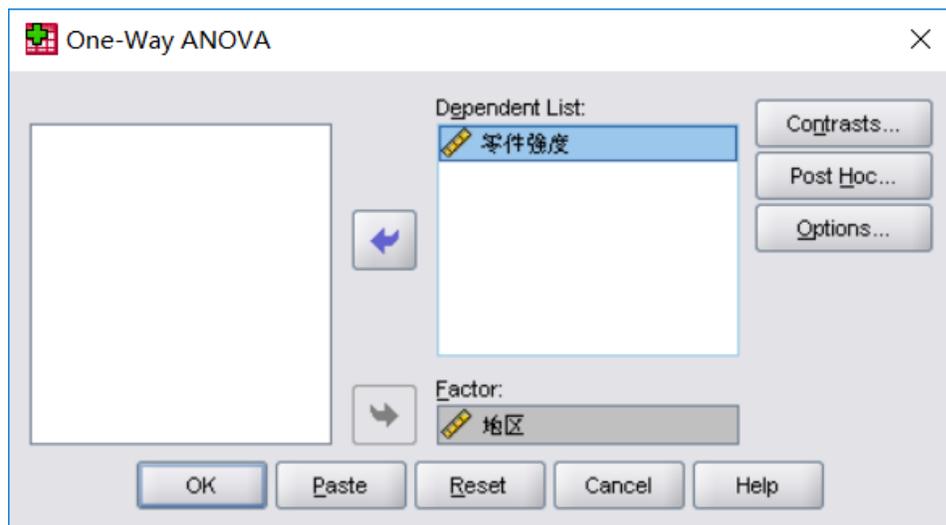
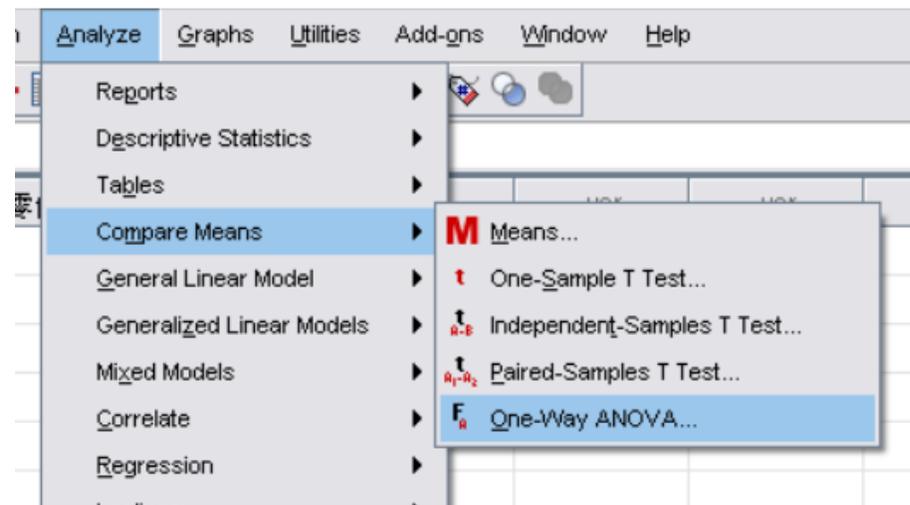
$H_0$ ：三个地区的零件强度无显著差异；

$H_1$ ：三个地区的零件强度有显著差异。



# 单因素方差分析

1、单击Analyze → Compare Means → One-Way ANOVA，打开 One-Way ANOVA对话框。



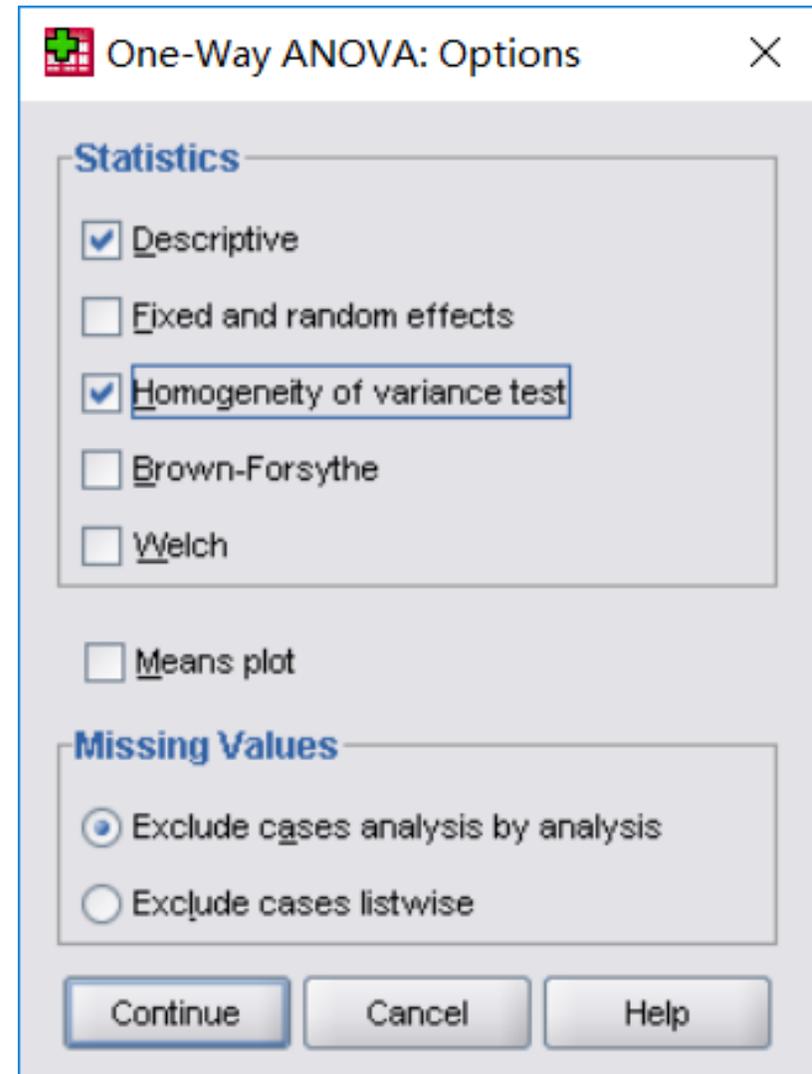
2、从左框中选择因变量“零件强度”进入Dependent list框内，选择因素变量“地区”进入Factor框内。点击OK就可以得到方差分析下表。



# 单因素方差分析

3、单击Option按钮，打开Option对话框如图所示：在Option选项中选择输出项。主要有不同水平下样本方差的齐性检验，缺失值的处理方式及均值的图形。

本例中选择Homogeneity of variance test 进行不同水平间方差齐性的检验以及 Descriptive 基本统计描述。在Missing Value 栏中选择系统默认项。





# 单因素方差分析

完成所有选择后返回主对话框，然后单击OK，就可以得到三个地区零件强度分析表。

Descriptives

零件强度								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	6	102.8333	12.25425	5.00278	89.9733	115.6934	83.00	116.00
2	6	110.0000	5.89915	2.40832	103.8092	116.1908	103.00	118.00
3	6	90.8333	10.81511	4.41525	79.4836	102.1831	73.00	102.00
Total	18	101.2222	12.45016	2.93453	95.0309	107.4135	73.00	118.00

Test of Homogeneity of Variances

零件强度			
Levene Statistic	df1	df2	Sig.
1.203	2	15	.328

方差齐性检验，Sig值大于0.05，符合方差齐性假设



# 单因素方差分析

由于F统计量值的P值明显小于显著性水平0.05，故拒绝假设H0，认为这三个地区的零件强度有显著差异。

ANOVA

零件强度	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1125.444	2	562.722	5.591	.015
Within Groups	1509.667	15	100.644		
Total	2635.111	17			

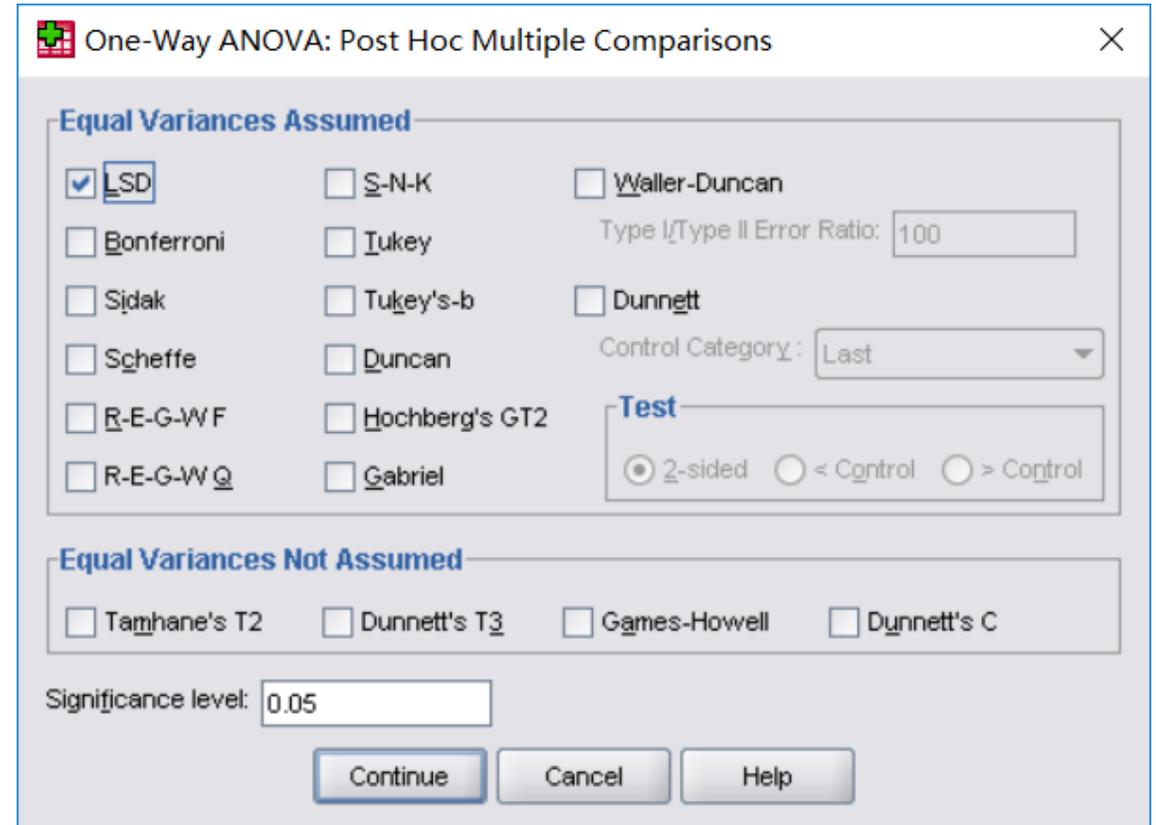


# 单因素方差分析

4、如果需要将水平间两两比较，可以单击Post Hoc 按钮，打开多重比较对话框。

在该对话框中列出了许多多重比较检验，涉及到许多的数理统计方法，在实际中只选用其中常用的方法即可。

对话框下部的Significance level表示显著性水平，默认值是0.05，也可以根据需要重新输入其它值。





# 单因素方差分析

如果满足在水平间方差相等的条件，常用LSD（least-significant difference最小显著性差异法），表示用 t 检验完成各组均值间的配对比较。

当方差不等的情况下，可以选择Tamhane's T2, 用t检验进行各组均值间的配对比较。

Multiple Comparisons

零件强度  
LSD

(I) 地 区	(J) 地 区	Mean Difference (I- J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-7.16667	5.79208	.235	-19.5122	5.1789
	3	12.00000	5.79208	.056	-.3455	24.3455
2	1	7.16667	5.79208	.235	-5.1789	19.5122
	3	19.16667*	5.79208	.005	6.8211	31.5122
3	1	-12.00000	5.79208	.056	-24.3455	.3455
	2	-19.16667*	5.79208	.005	-31.5122	-6.8211

\*. The mean difference is significant at the 0.05 level.

从表可以看出，地区2与地区3之间的差异是非常显著的， $p < 0.05$ 。



# 双因素方差分析

- 单因变量的双因素方差分析是对观察的现象（因变量）受两个因素或变量的影响进行分析，检验不同水平组合之间对因变量的影响是否显著。
- 双因素方差分析应用条件：因变量和协变量必须是数值型变量，且因变量来自或近似来自正态总体。因素变量是分类变量，变量可以是数值型或字符型的。各水平下的总体假设服从正态分布，而且假设各水平下的方差是相等的。
- 双因素方差分析过程可以分析出每一个因素的作用；各因素之间的交互作用；检验各总体间方差是否相等；还能够对因素的各水平间均值差异进行比较等。



# 双因素方差分析

示例4：右表是某商品S在不同地区和不同时期的销售量表。已知数据服从正态分布，则要检验地区因素及时间因素对销售量的影响是否显著。

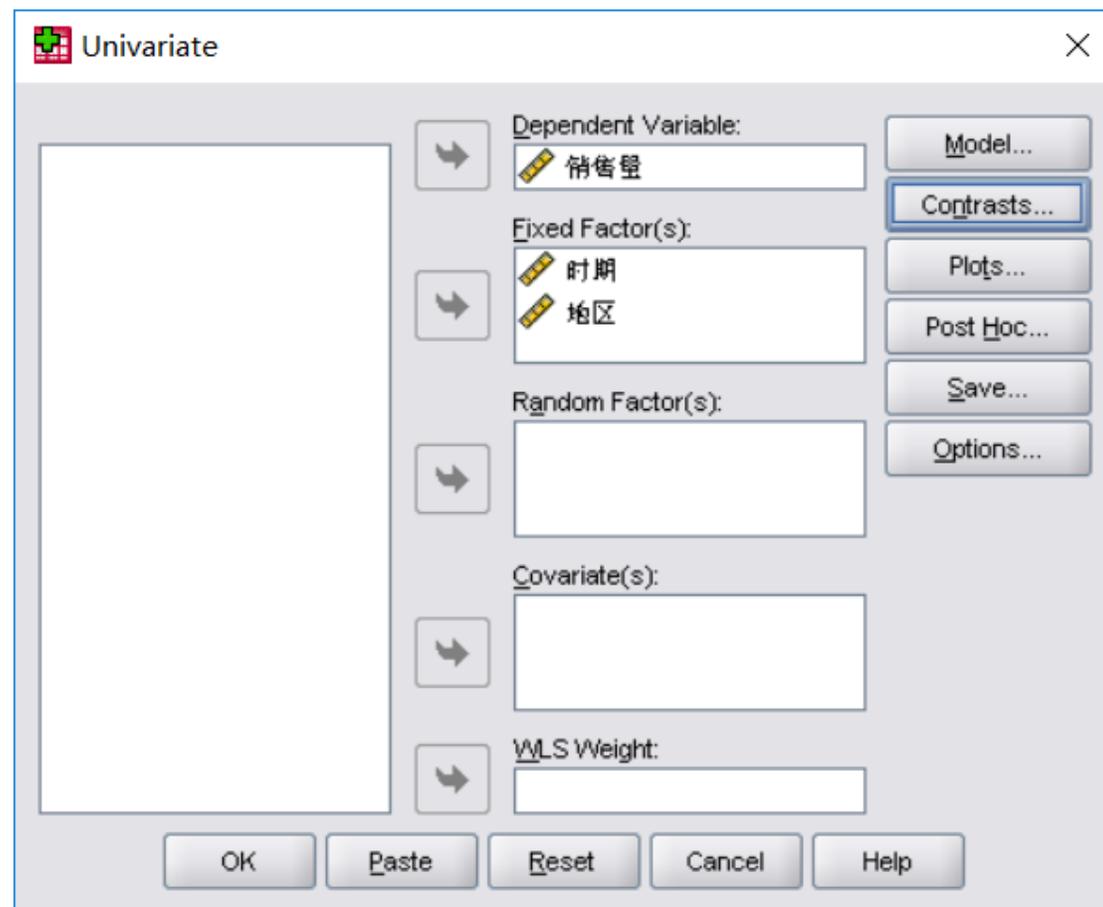
地区 \ 时期	1	2	3	4	5
1	6.5	14.2	13.4	2.4	6.2
2	1.8	7.1	9.4	1.5	4.8
3	3.6	10.8	7.2	1.7	4.9
4	3.7	8.9	8.6	2.3	4.6
5	7.6	12.6	7.5	2.8	5.2



# 双因素方差分析

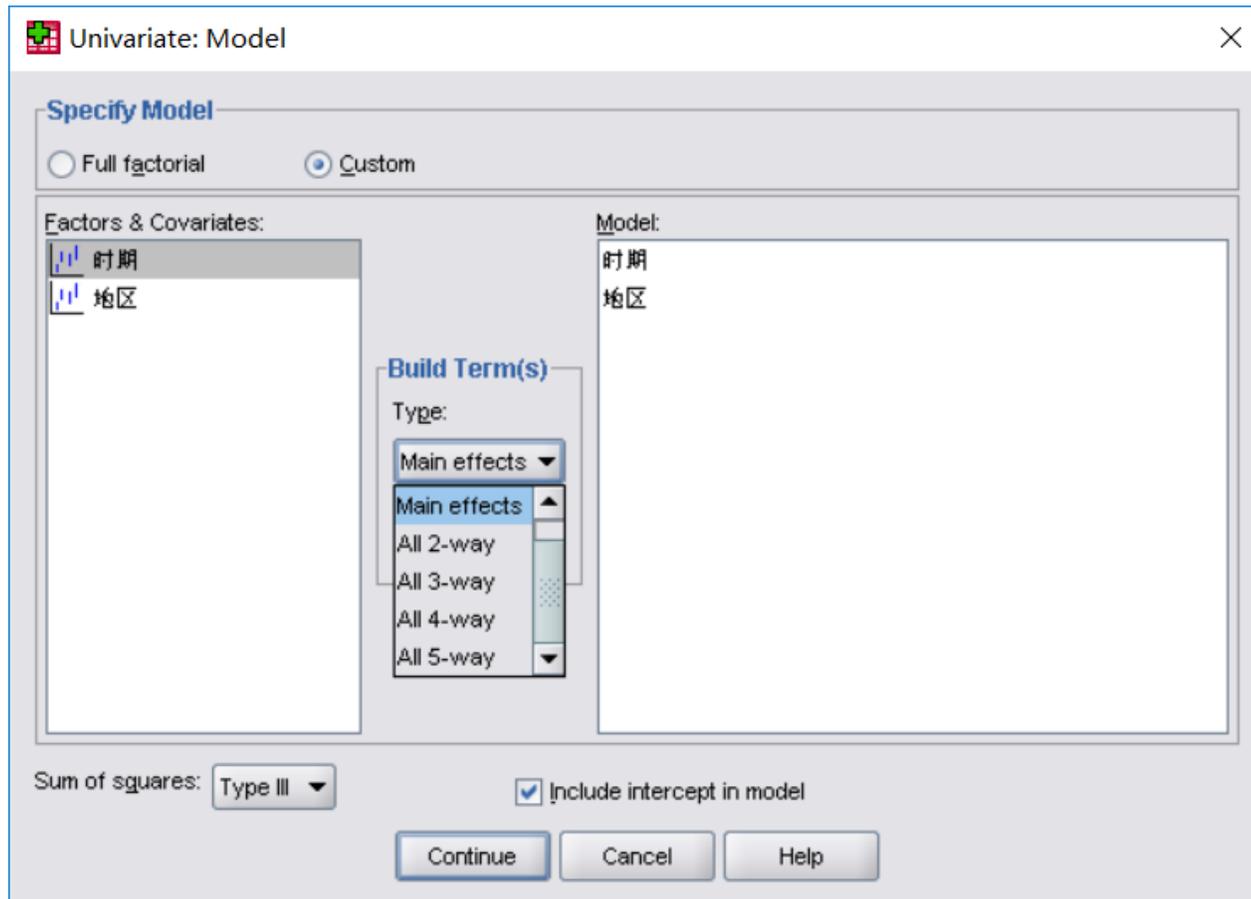
由于销售量受地区和时间两个因素的影响，这是一个双因素方差分析的问题。

- 1、单击Analyze → General linear Model → Univariate，打开Univariate主对话框。
- 2、选择要分析的变量“销售量”进入Dependent Variable 框中，选择因素变量“地区”和“时期”进入Fixed Factor框中。
- 3、单击Model按钮选择分析模型，得到Model对话框。如图所示：在Specify框中，指定模型类型。





# 双因素方差分析



- Custom选项为自定义模型，本例选择此项并激活下面的各项操作。
- 先从左边框中选择因素变量进入Model框中，然后选择效应类型。一般不考虑交互作用时，选择主效应Main，考虑交互作用时，选择交互作用Interaction。可以通过单击Build Term下面的小菜单完成，本例中选择主效应。



# 双因素方差分析

4、从表中数据可以看出，F值对应概率P值都小于显著性水平0.05，这说明地区和时期对销售量的影响都是显著的。

Tests of Between-Subjects Effects

Dependent Variable: 销售量

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	289.717 <sup>a</sup>	8	36.215	14.679	.000
Intercept	1015.060	1	1015.060	411.438	.000
时期	42.498	4	10.625	4.307	.015
地区	247.218	4	61.805	25.052	.000
Error	39.474	16	2.467		
Total	1344.250	25			
Corrected Total	329.190	24			

a. R Squared = .880 (Adjusted R Squared = .820)

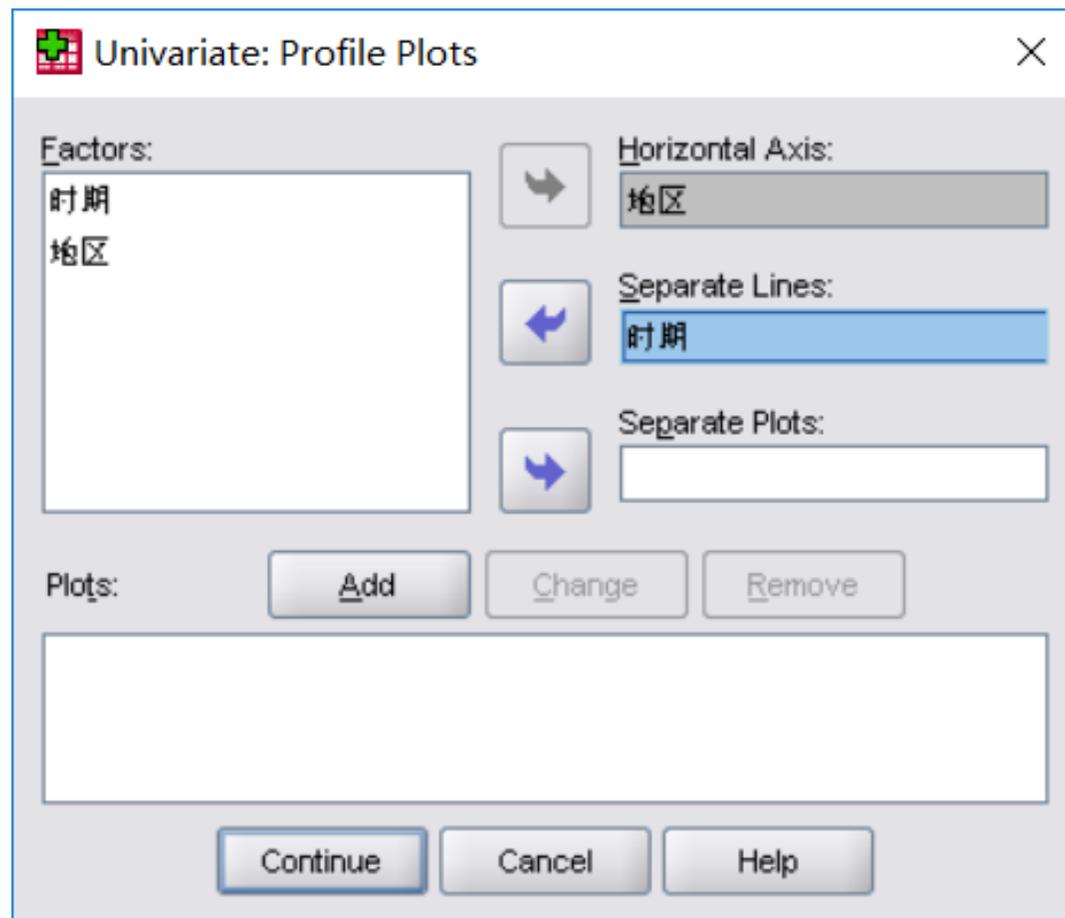


# 双因素方差分析

5、如果需要进行图形展示，可单击Plots按钮，打开图形对话框如图所示。

(1) 在Factor框中选择因素变量进入横坐标Horizontal Axis框内，然后单击add按钮，可以得到该因素不同水平的因变量均值的分布。

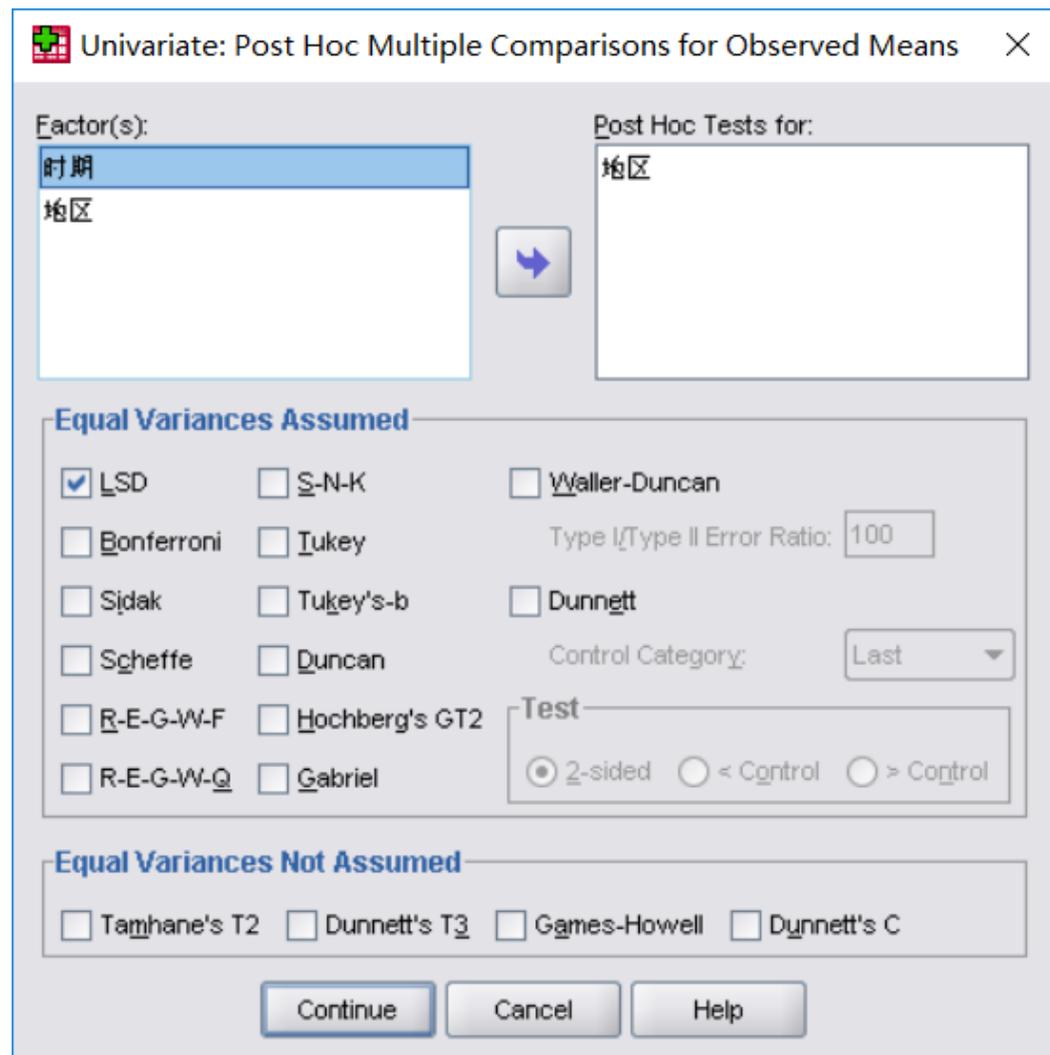
(2) 如果要了解两个因素变量的交互作用，将一个因素变量送入横坐标后，将另一个因素变量送入Separate Lines分线框中，然后单击add按钮。就可以输出反映两个因素变量的交互图。本例中选择“地区”为横坐标。





# 双因素方差分析

6、如需要将因素A各水平间均值进行两两比较，单击Post Hoc按钮，打开Post Hoc Multiple多重比较对话框如图所示。从Factor框中选择因素变量进入Post Hoc Test for框中，然后选择多重比较方法。本例中各组方差相等，选择LSD方法。



# 双因素方差分析

7、单击Options按钮，打开Univariate：Options对话框，从中选择需要输出的显著性水平，默认值为0.05。在进行所有的选择后，单击OK，就可以得到输出结果。由多重比较LSD表中得到不同地区销售量的比较表。

销售量  
LSD

(I) 地 区	(J) 地 区	Mean Difference (I- J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-6.0800'	.99340	.000	-8.1859	-3.9741
	3	-4.5800'	.99340	.000	-6.6859	-2.4741
	4	2.5000'	.99340	.023	.3941	4.6059
	5	-.5000	.99340	.622	-2.6059	1.6059
2	1	6.0800'	.99340	.000	3.9741	8.1859
	3	1.5000	.99340	.151	-.6059	3.6059
	4	8.5800'	.99340	.000	6.4741	10.6859
	5	5.5800'	.99340	.000	3.4741	7.6859
3	1	4.5800'	.99340	.000	2.4741	6.6859
	2	-1.5000	.99340	.151	-3.6059	.6059
	4	7.0800'	.99340	.000	4.9741	9.1859
	5	4.0800'	.99340	.001	1.9741	6.1859
4	1	-2.5000'	.99340	.023	-4.6059	-.3941
	2	-8.5800'	.99340	.000	-10.6859	-6.4741
	3	-7.0800'	.99340	.000	-9.1859	-4.9741
	5	-3.0000'	.99340	.008	-5.1059	-.8941
5	1	.5000	.99340	.622	-1.6059	2.6059
	2	-5.5800'	.99340	.000	-7.6859	-3.4741
	3	-4.0800'	.99340	.001	-6.1859	-1.9741
	4	3.0000'	.99340	.008	.8941	5.1059

Based on observed means.

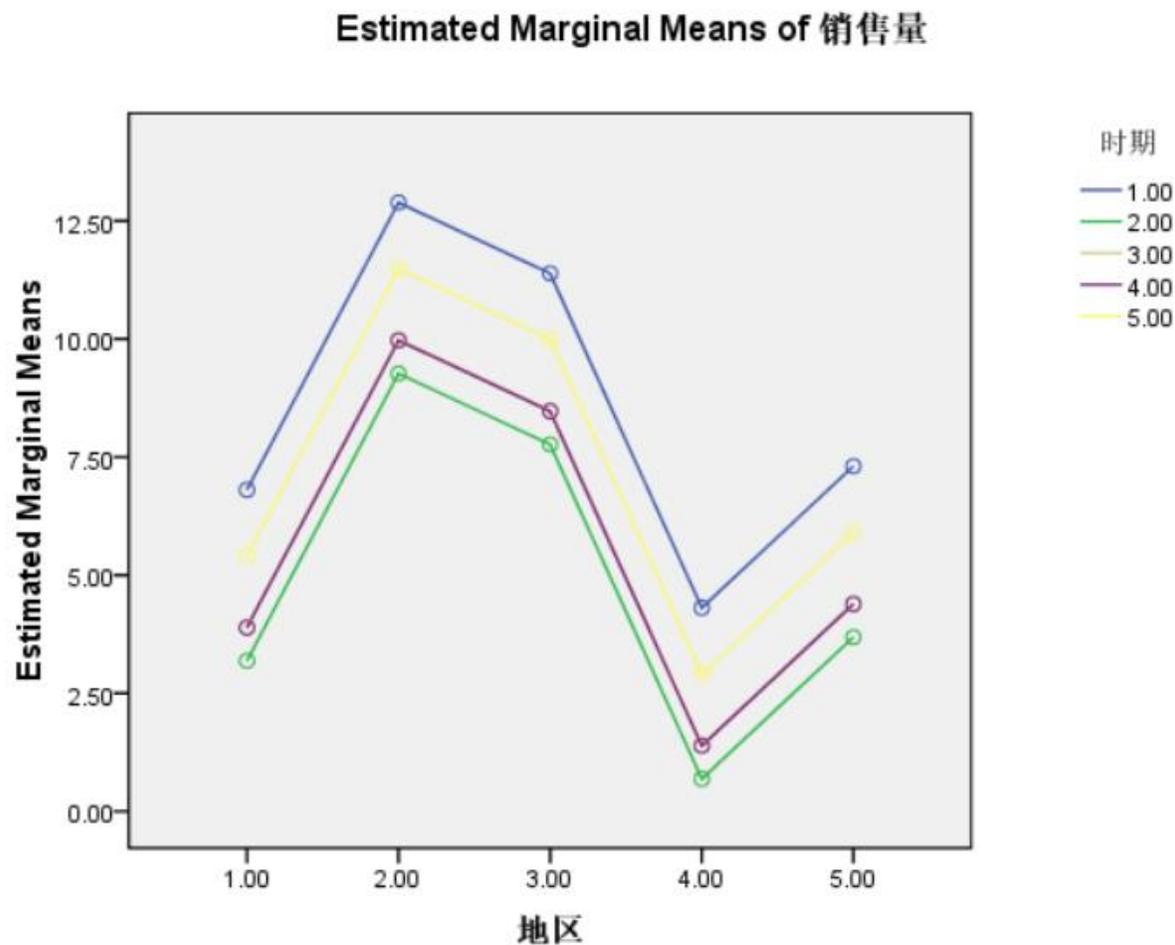
The error term is Mean Square(Error) = 2.467.

\*. The mean difference is significant at the .05 level.



# 双因素方差分析

两个因素变量地区和时期的折线之间无交叉，因此两个因素之间基本上没有交互作用。





# 5. 信度与效度分析



# 信度与效度

问卷调查中引用他人的测量工具(量表或问卷)或者根据研究目的和理论编制或修订符合自己需要的工具。

为了保证问卷具有较高的可靠性和有效性，在正式分析或形成正式问卷之前，会对测量结果进行信效度检验。

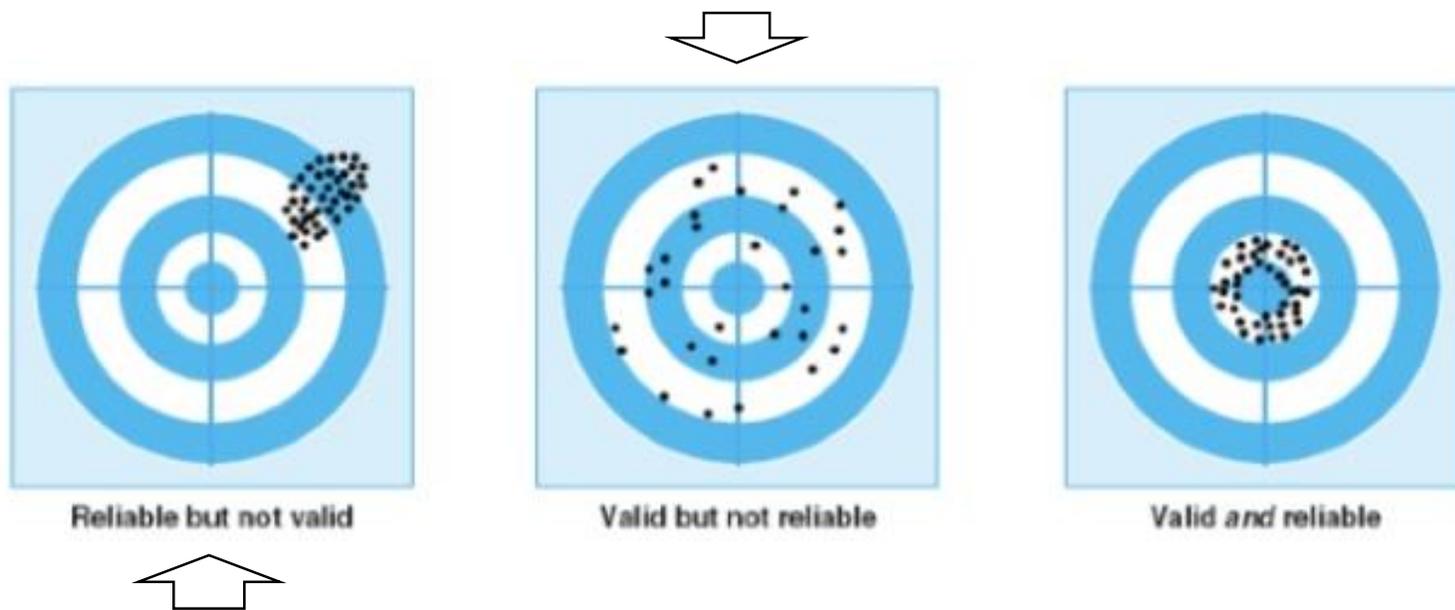
信度 ( Reliability ) 用于测量样本有没有真实作答量表类题项。Reliability关注测量的稳定性、一致性 ( consistency ) 。

效度 ( validity ) 是测量工具确能测出其所要测量特质的程度。Validity侧重测量的“准确性” ( accurate assessment ) 。



# 信度与效度

买了个新的体重称来跟踪自己体重变化，  
第一次161斤，重新又站上去了一次171斤，  
又上去一看又变成161斤



效度 Validity  
关注是不是能正中靶心；  
信度 Reliability  
关注多大程度上能总是命中靶心。

你所检验的是不是你想要检验的？  
你瞄准的靶子是不是你该射击的靶子？  
你行驶的方向是不是你要去的地方？



# 信度类型

- 重测信度：用同样的测量工具,对同一组被调查者间隔一定时间重复测试,两次测试结果的相关程度。
- 复本信度：让同一组被调查者一次填答两份问卷复本,计算两个复本的相关系数。复本信度法要求两个复本除表述方式不同外,在内容、格式、难度和对应题目的提问方向等方面要完全一致,而在实际调查中,很难使调查问卷达到这种要求,因此采用这种方法者较少。
- 分半信度：指在测验后将测验项目分成相等的两组（两半），计算两项项目分之间的相关。分半信度相关越高表示信度高，或内部一致性程度高。它是常用信度检验方法之一，反映了测验项目内部一致性程度，即表示测验测量相同内容或特质的程度。
- 评分者信度：多个评分者给同一批人的答卷评分的一致性程度。
- 内部一致性信度：用于评价问卷的内部一致性。Cronbach  $\alpha$ 系数取值在0到1之间, 系数越高,信度越高,问卷的内部一致性越好。



# 信度Cronbach $\alpha$

最常用的是信度指标是内部一致性信度，其数值也称克隆巴赫信度系数 (Cronbach  $\alpha$  系数值)

- 如果 $\alpha$  在0.8以上,则该测验或量表的信度非常好；
- 如果 $\alpha$ 在0.7以上都是可以接受；
- 如果 $\alpha$ 在0.6以上,则该量表应进行修订,但仍不失其价值；
- 如果 $\alpha$ 低于0.6,量表就需要重新设计题项。



# 信度不达标的检查

- 第一：使用‘描述分析’检验下是否有奇怪的**异常值**，如果有则需要使用‘数据处理->异常值’功能处理后再分析；
- 第二：如果信度系数值依然很低（比如低于0.5），此时可考虑把所有量表题合并在一起进行一次信度分析（题项越多通常信度系数会越高）；
- 第三：如果数据中有**反向题**，需要先使用‘数据处理->数据编码’将反向题处理后再分析；
- 第四：删除不合理的项，留下有意义的项；
- 第五：**加大样本量**，样本量越大通常情况下信度会越高；
- 第六：问卷设计时一个维度尽量4~7个题较好，**题项越多信度会越高**，而且如果不达标还可以删除个别不合理项。



# 效度类型：内容效度

内容效度是指测量项目对整个测量内容范围的代表程度，即考察测验项目的代表性或对所测量内容取样的适当性，也就是考察测验项目能否切实测量到测验所要测量的目的或行为。

- 基于理论文献综述、开放式问卷调查、知名相关变量和个别访谈几方面综合调查的结果，在问卷测试过程中多次请专家进行审查和修订，基本保证了问卷的有效性和题项能够涵盖拖延各个方面的原因，并具有代表性，基本上反应了当代大学生拖延原因的真实情况。
- 通过了理论思考、开放式访谈、成立专业的项目编制小组及专家审核等一系列步骤，保证了量表能够真实的反映青少年情绪调节策略运用的实际情况。
- 量表在编制的过程中，研究者对量表的项目进行过多次审核和修改，因此能够保证量表具有较好的内容效度。



# 效度类型：结构效度

结构效度是指测验实际测量到的与所要测量的理论构想或特质之间的符合程度，它是评判测量工具质量的重要标准之一。

- 验证性因素分析是考察已有的理论模型和现实调查数据拟合程度的一种验证，是考察量表结构效度的一种十分有效的方法。
- 验证性因素分析是通过检验数据拟合程度来考察结构效度，寻求数据与模型的最佳拟合，检验因素指标的效度与理论构想之间的关系，因此其得到的结果更具确定性。



# EFA与CFA

因素分析是基于相关关系对众多数据进行降维（即简化）的数据处理方法，目的在于挖掘出众多数据后的某种结构。

问卷收集完成后，先进行探索性因素分析，再进行验证性因素分析。

SPSS能进行EFA和CFA，而AMOS软件做的是CFA。

探索性因素分析（ Exploratory Factor Analysis ）

探索性因素分析问卷题项是否能归于一个因子（聚集在一起表示这些题项相关，可以被同一个潜在变量解释）是否落在同一个因子中。

验证性因素分析（ Confirmatory Factor Analysis ）

验证性因素分析通过AMOS将潜在变量与对应题项绘制测量模型。模型拟合质量好说明测量模型绘制的潜在变量与题项的关系通过数据验证。



# 探索性因素分析（EFA）

探索性因素分析是假设测量变量（量表题项）能够最终形成几个因子（潜在变量）是未知的，希望通过因素分析，看看量表题项能够形成几个因子，以及每个因子对应哪些题项。

验证性因素分析则先根据实际的研究调查情况，将潜在变量与潜在变量对于的题项关系固定，然后用数据来拟合验证这种关系模型是否成立，如果拟合质量好，说明模型关系得到验证，反之则需要需要进行题项删除或修改。

- 探索性因素分析着重在探索模型结构，验证性因素分析着重在验证模型结构。
- 探索性因素分析常用在问卷的设计初期，帮助分析者建立模型框架，而验证性因子常用在成熟问卷的信效度分析中。

SPSS 分析——降维——因子分析——抽取：基于特征值



# 验证性因素分析 ( CFA )

验证性因素分析整体模型适配度是否达到标准可从以下指标来考查：

- 卡方自由度比 $X^2/df$ 越小，表示模型的拟合度越好。一般而言， $X^2/df < 3$ 表示模型整体拟合度较好； $3 < X^2/df < 5$ 表示模型整体可以接受，但需要改进； $X^2/df > 10$ 表明整体模型非常差。
- 渐进均方根误差 ( RMSEA )，该指标受样本数量影响较小。该指标值越小，则模型拟合度越好。一般认为， $RMSEA > 0.1$ 表示模型拟合度不佳； $0.08-0.1$ 表示模型尚可，具有普通适配； $0.05-0.08$ 表示模型拟合好； $RMSEA < 0.05$ 表示模型拟合度非常好。
- 拟合良好性指标 ( GFI )、常规拟合指数 ( NFI )、比较拟合指标 ( CFI )、调整拟合良好性指标 ( AGFI ) 越接近1则表示模型的拟合度越好，一般认为值应在0.9以上。



# 6. 相关与回归分析



# 相关与回归分析

- 相关分析和回归分析是统计分析方法中最重要内容之一，是多元统计分析方法的基础。相关分析和回归分析主要用于研究和分析变量之间的相关关系，在变量之间寻求合适的函数关系式，特别是线性表达式。
- 对变量之间的相关关系进行分析（Correlate）。其中包括简单相关分析（Bivariate）和偏相关分析（Partial）。
- 建立因变量和自变量之间回归模型（Regression），其中包括线性回归分析（Linear）和曲线估计（Curve Estimation）。
- 数据条件：参与分析的变量数据是数值型变量或有序变量。



# 相关与回归分析

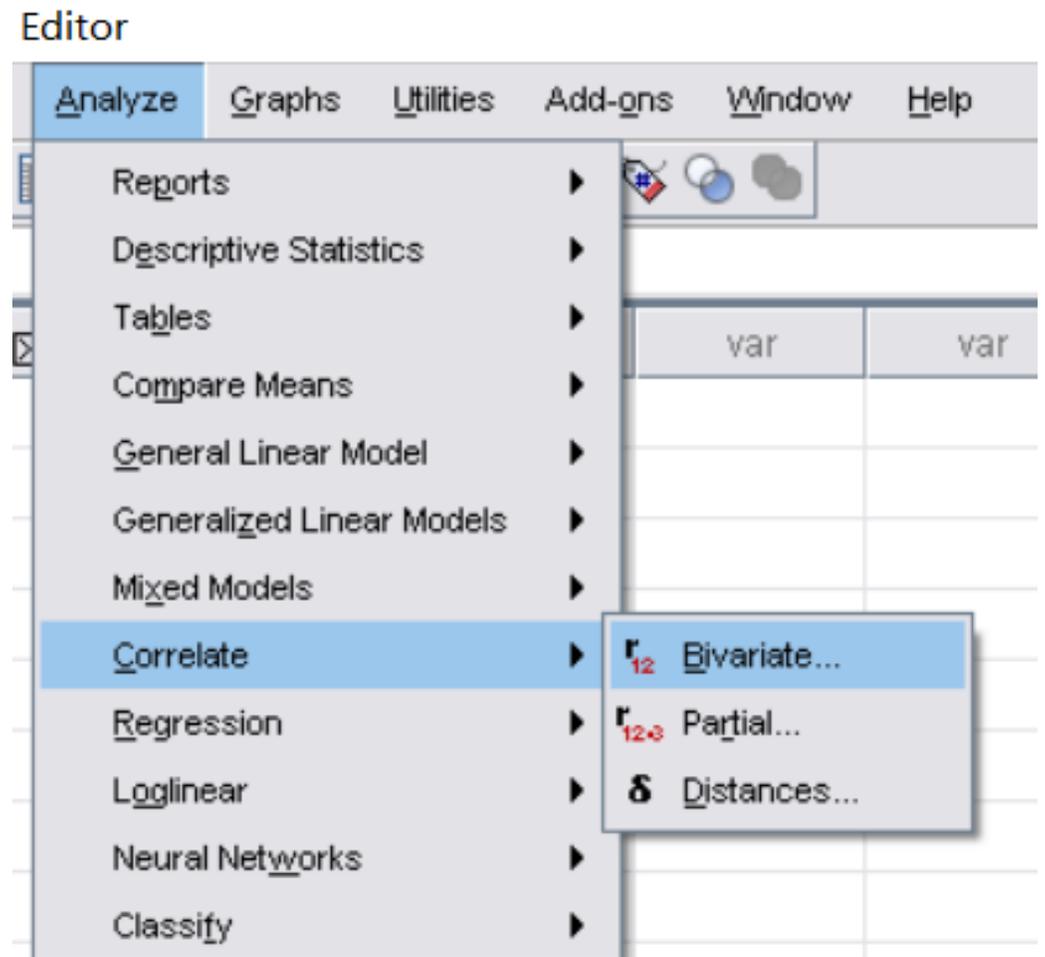
**相关分析**

**回归分析**



# 相关分析

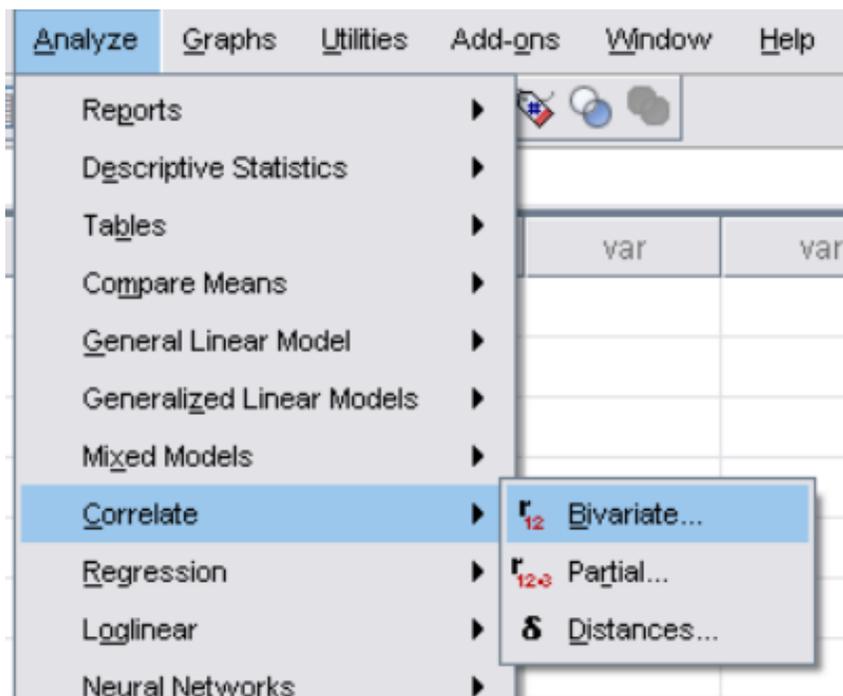
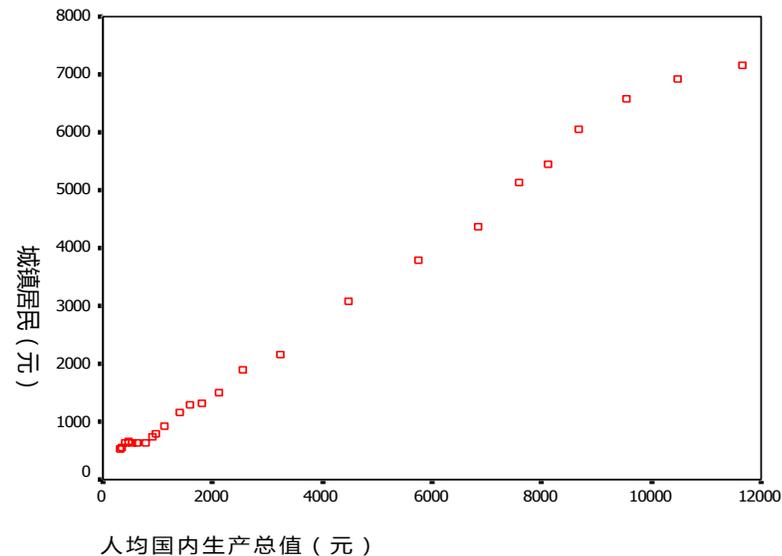
在SPSS中，可以通过Analyze菜单进行相关分析(Correlate)，Correlate菜单如图所示。





# 相关分析

两个变量之间的相关关系称简单相关关系。有两种方法可以反映简单相关关系。一是通过散点图直观地显示变量之间关系，二是通过相关系数准确地反映两变量的关系程度。



(1) 散点图。SPSS软件的绘图命令集中在Graphs菜单。

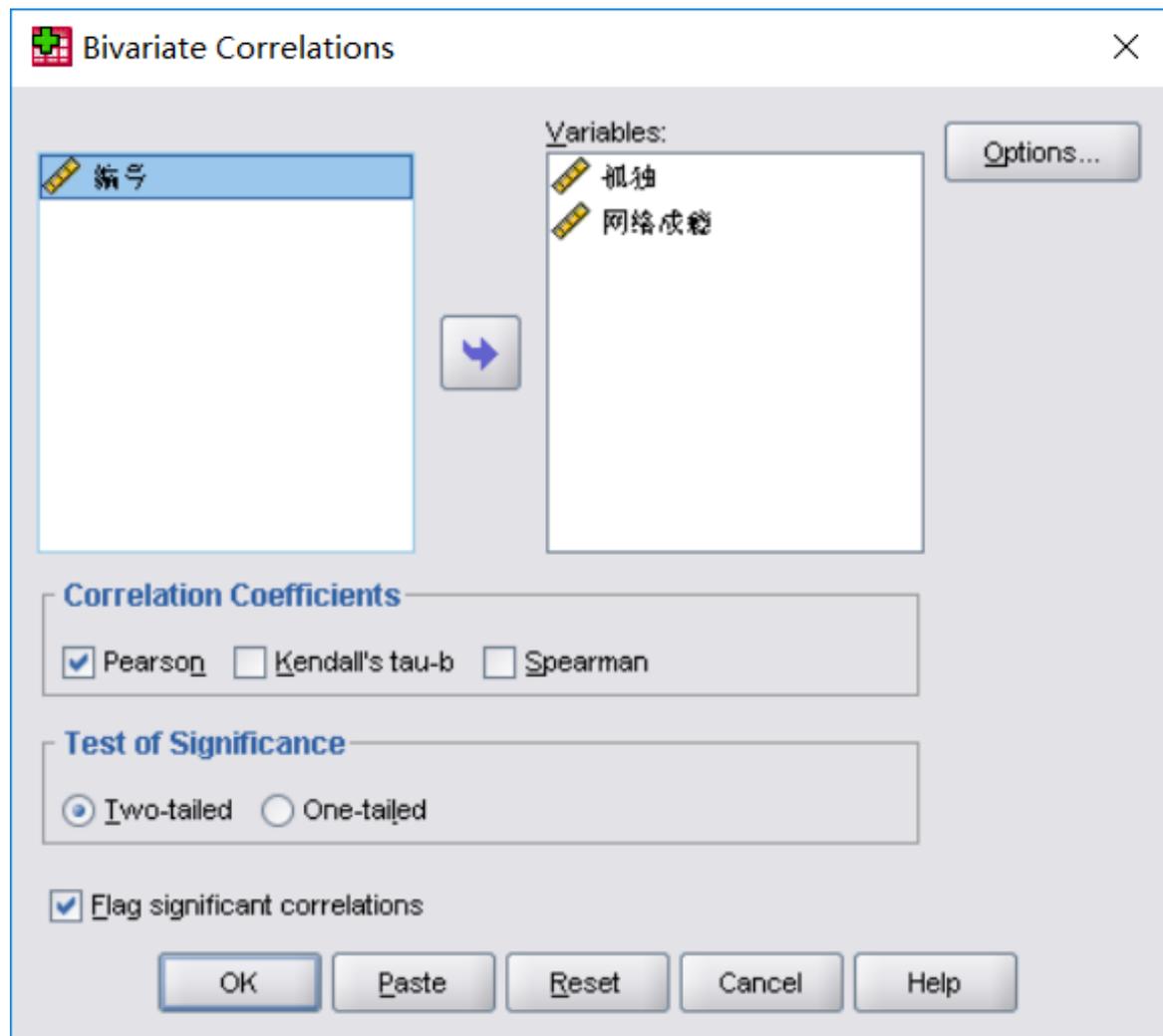
(2) 相关系数：

①打开数据库后，单击Analyze → Correlate → Bivariate 打开Bivariate对话框，见图所示。



# 相关分析

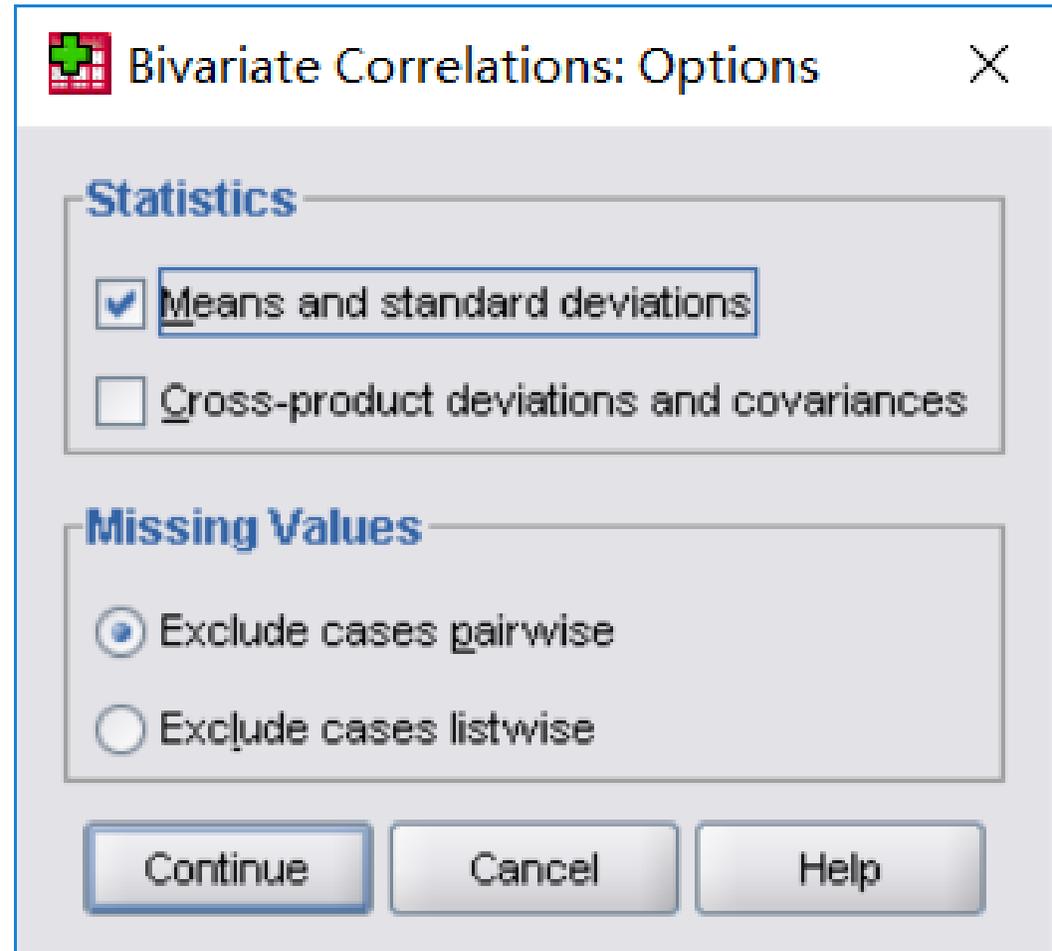
②从左边的变量框中选择需要考察的两个变量进入 Variables 框内，从Correlation Coefficients 栏内选择相关系数的种类，有 Pearson相关系数，Kendall's一致性系数和 Spearman等级相关系数。从检验栏内选择检验方式，有双侧检验和单侧检验两种。





# 相关分析

③ 单击Options按钮，选择输出项和缺失值的处理方式。本例中选择输出基本统计描述。





# 相关分析

单击OK，可以得到相关分析的结果。

**Descriptive Statistics**

	Mean	Std. Deviation	N
孤独	39.48	8.945	169
网络成瘾	12.18	3.687	169

**Correlations**

		孤独	网络成瘾
孤独	Pearson Correlation	1	.273**
	Sig. (2-tailed)		.000
	N	169	169
网络成瘾	Pearson Correlation	.273**	1
	Sig. (2-tailed)	.000	
	N	169	169

\*\* . Correlation is significant at the 0.01 level (2-tailed).



# 线性回归分析

线性回归是统计分析方法中最常用的方法之一。

如果所研究的现象有若干个影响因素，且这些因素对现象的综合影响是线性的，则可以使用线性回归的方法建立现象（因变量）与影响因素（自变量）之间的线性函数关系式。

线性回归的假设理论：

- 正态性假设：即所研究的变量均服从正态分布；
- 等方差假设：即各变量总体的方差是相等的；
- 独立性假设：即各变量之间是相互独立的；
- 残差项无自相关性：即误差项之间互不相关， $\text{Cov}(\varepsilon_i, \varepsilon_j)=0$



# 线性回归分析

## 线性回归模型的检验项目

- 回归系数的检验 ( t检验 )
- 回归方程的检验 ( F检验 )
- 拟合程度判定 ( 可决系数 $R^2$  )
- D.W检验 ( 残差项是否自相关 )
- 共线性检验 ( 多元线性回归 )
- 残差图示分析 ( 判断异方差性和残差序列自相关 )



# 线性回归分析

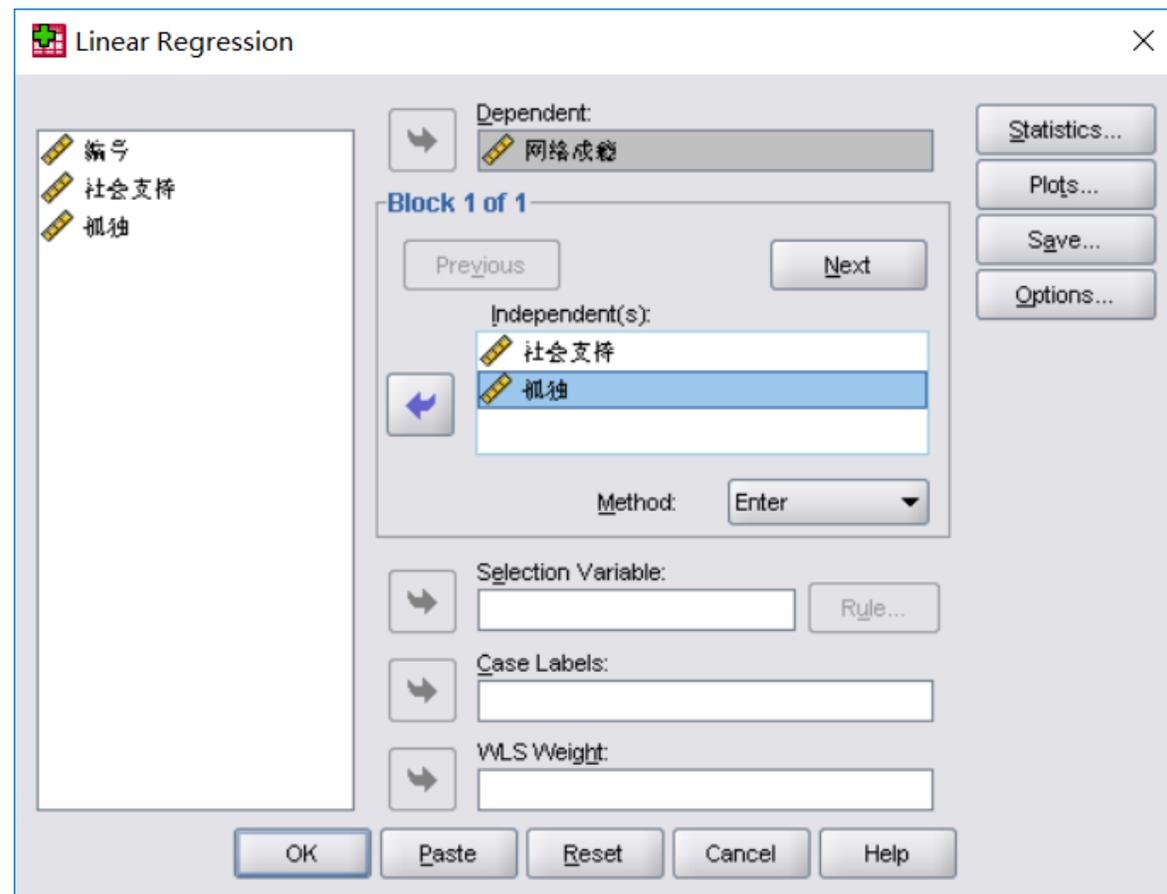
## 3、线性回归分析的具体步骤

SPSS软件中进行线性回归分析的选择项为

Analyze→Regression→Linear。

具体操作步骤如下：

1、打开数据文件，单击Analyze → Regression → Linear打开Linear 对话框

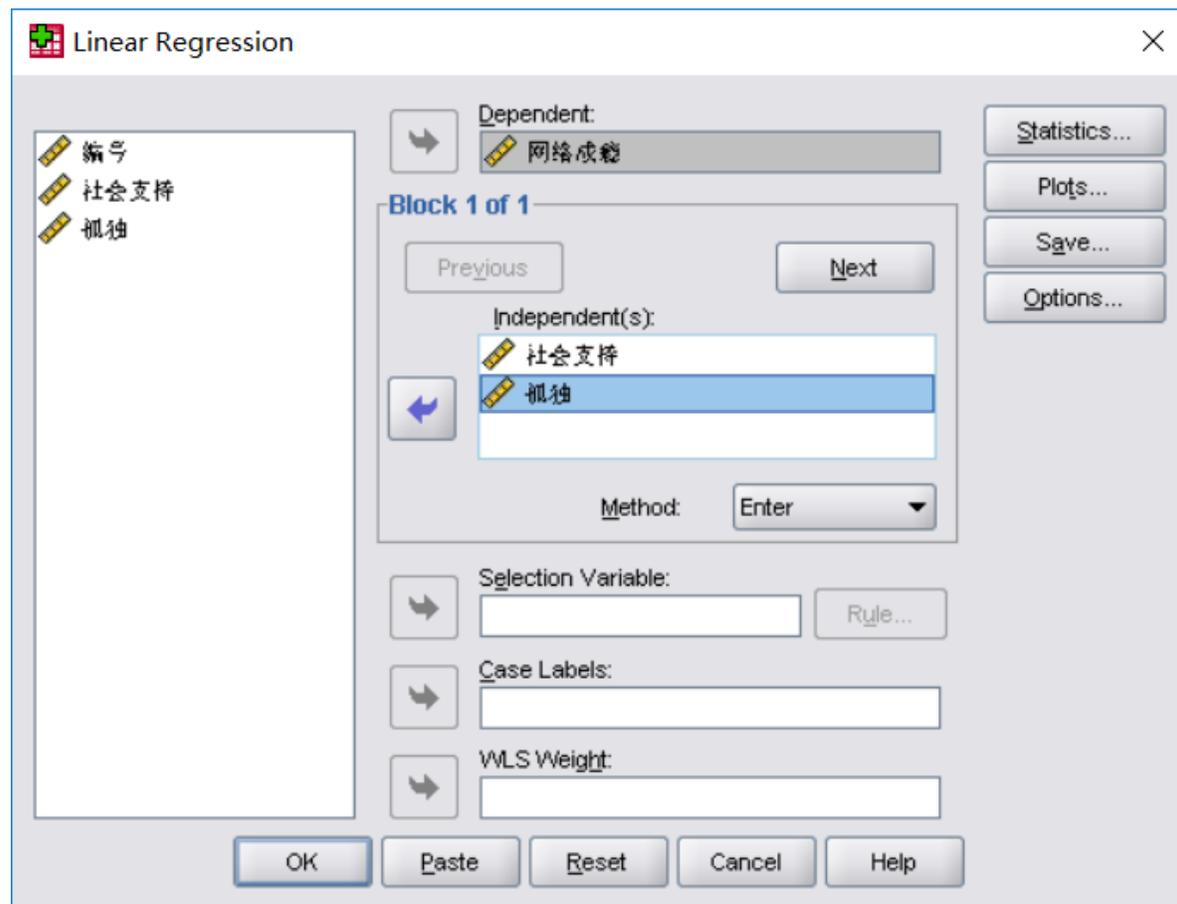




# 线性回归分析

2、从左边框中选择因变量进入Dependent 框内，选择一个或多个自变量进入Independent 框内。从Method 框内下拉式菜单中选择回归分析方法，有五种方法：

- 强行进入法(Enter)
- 消去法(Remove)
- 向前选择法(Forward)
- 向后剔除法(Backward)
- 逐步回归法(Stepwise)

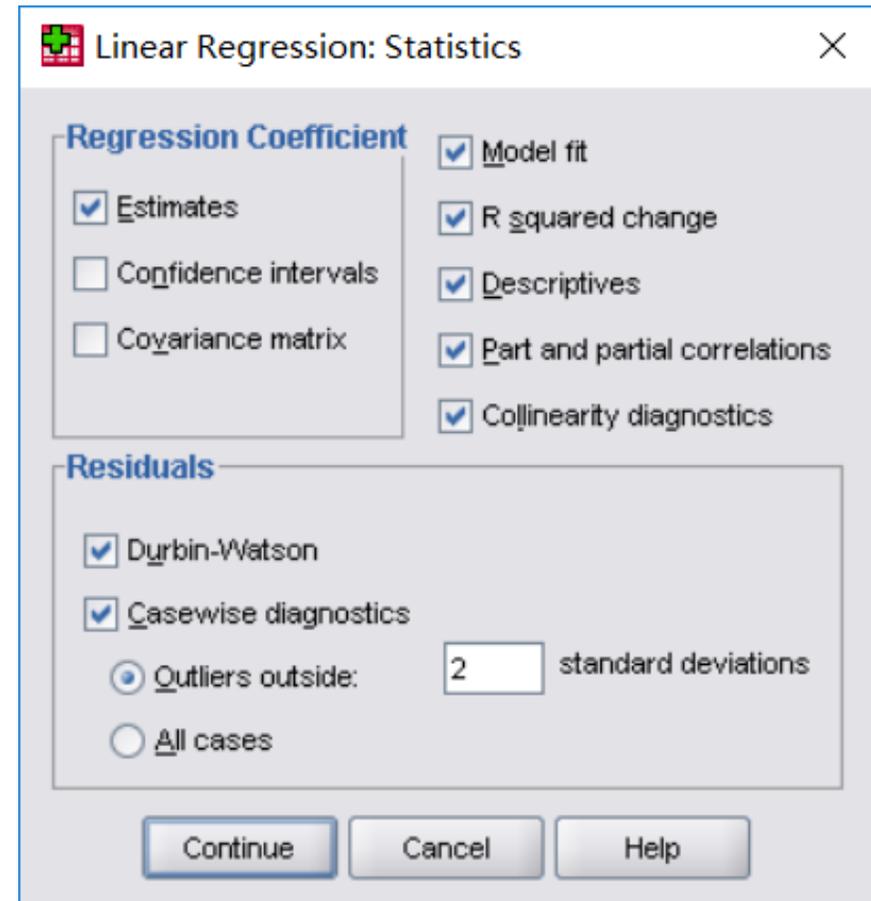




# 线性回归分析

3、单击Statistics，打开Linear Regression：Statistics对话框，可以选择输出的统计量如图所示。

- Regression Coefficients栏，回归系数选项栏。
- Estimates (系统默认): 输出回归系数的相关统计量：包括回归系数，回归系数标准误、标准化回归系数、回归系数检验统计量(t值)及相应的检验统计量概率的P值(sig)。
- Confidence intervals:输出每一个非标准化回归系数95%的置信区间。
- Covariance matrix: 输出协方差矩阵。





# 线性回归分析

**Model fit**是默认项。能够输出复相关系数 $R$ 、 $R^2$ 及 $R^2$ 修正值，估计值的标准误，方差分析表。

**R squared change**: 引入或剔除一个变量时， $R^2$ 的变化。

**Descriptives**: 基本统计描述。

**Part and Partial correlations**：相关系数及偏相关系数。

**Collinearity diagnostics**：共线性诊断。主要对于多元回归模型，分析各自变量的之间的共线性的统计量：包括容忍度和方差膨胀因子、特征值，条件指数等。



# 线性回归分析

4、单击Options按钮，打开Linear Regression：Options对话框，如图所示。可以从中选择模型拟合判断准则Stepping Method Criteria 及缺失值的处理方式。

- Stepping Method Criteria 栏，设置变量引入或剔除模型的判别标准。
- Use probability of F:采用F检验的概率为判别依据。
- Use F value: 采用F值作为检验标准。
- Include constant in equation 回归方程中包括常数项。
- Missing Values 缺失值的处理方式。

Linear Regression: Options

**Stepping Method Criteria**

Use probability of F  
Entry: .05 Removal: .10

Use F value  
Entry: 3.84 Removal: 2.71

Include constant in equation

**Missing Values**

Exclude cases listwise

Exclude cases pairwise

Replace with mean

Continue Cancel Help



# 线性回归分析

5、如果要保存预测值等数据，可单击Save按钮打开Linear Regression : Save对话框。选择需要保存的数据种类作为新变量存在数据编辑窗口。其中有预测值、残差，预测区间等。本例中不做选择。

6、当所有选择完成后，单击OK得到分析结果。

Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.275 <sup>a</sup>	.075	.064	3.566	.075	6.778	2	166	.001	1.985

a. Predictors: (Constant), 孤独, 社会支持

b. Dependent Variable: 网络成瘾

DW量是判断数据是否存在序列相关的关键，如果存在的话就是伪回归。它的值在2附近就表明是不存在序列相关的。



# 线性回归分析

ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	172.395	2	86.197	6.778	.001 <sup>a</sup>
	Residual	2111.030	166	12.717		
	Total	2283.425	168			

a. Predictors: (Constant), 孤独, 社会支持

b. Dependent Variable: 网络成瘾

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	6.983	2.167		3.223	.002					
	社会支持	.070	.164	.033	.427	.670	-.053	.033	.032	.907	1.102
	孤独	.117	.032	.283	3.613	.000	.273	.270	.270	.907	1.102

a. Dependent Variable: 网络成瘾



# 7. 中介与调节效应分析



# 中介效应

中介变量是联系两个变量之间关系的桥梁和纽带。

中介变量主要解决自变量如何影响因变量。中介变量必须与自变量、因变量均存在显著相关。因变量、中介变量必须为连续变量，自变量可以是连续变量也可以是分类变量。

## 完全中介与部分中介

- 中介按照在自变量与因变量之间所起的作用分为完全中介与部分中介。
- **完全中介**是自变量完全通过中介变量起作用；**部分中介**是自变量既可以直接影响因变量，也可以通过中介变量间接影响因变量。

## 简单中介与多重中介

- 简单中介与多重中介的区别在于是否仅有一个中介变量。仅有一个中介变量称为**简单中介**，如果有多个中介变量称为**多重中介**。
- 多重中介进一步按照中介变量是否存在顺序关系分为并行多重中介与链式多重中介。**并行多重中介**表示中介变量之间相互独立，**链式多重中介**表示中介变量之间存在顺序关系。



# 中介效应检验

常用的中介效应检验方法包括逐步回归法(依次检验)、乘积系数Sobel检验、差异系数检验法和Bootstrap法。

- 逐步回归法。
- 乘积系数Sobel检验法和差异系数检验法的准确性效果基本相同，是在逐步回归的基础上控制了第一类错误（“弃真错误”，也就是真的结论当作假的结论舍弃了），因此准确性比逐步回归好很多，但是它们同样存在比较致命的缺陷，这两种检验方法要求两个系数乘积（差值）服从正态分布的假设，一般无法满足。
- Bootstrap法，解决了乘积系数Sobel检验的问题。Sobel检验是基于均值抽样分布的特性对标准误进行估计，因此，一旦系数乘积不服从正态分布，估计的标准误也就不准确了，Bootstrap法通过模拟抽样分布过程，形成了一个（中介效应）抽样分布，从而得到一个更为准确的标准误估计值。



# 中介效应检验

逐步回归法中介关系检验如下：

$$Y=cx+e_1 \quad (1)$$

$$M=ax+e_2 \quad (2)$$

$$Y=c'x+bM+e_3 \quad (3)$$

- 第一步，检验方程(1),如果c显著则继续检验方程(2),如果c不显著(说明X对Y无影响)，则停止中介效应检验；
- 第二步，方程(2) 如果a显著则继续检验方程(3);如果a不显著，则停止检验;
- 第三步，方程(3)若b显著则说明中介效应显著。此时若c'显著则说明是不完全中介效应;若c'不显著则说明是完全中介效应。



# 调节效应分析

如果变量Y与变量X的关系是变量M的函数，称M为调节变量。就是说，Y与X的关系受到第三个变量M的影响。调节变量可以是定性的(如性别、种族、学校类型等)，也可以是定量的(如年龄、受教育年限、刺激次数等)，它影响因变量和自变量之间关系的方向(正或负)和强弱。

要求调节变量最好与自变量、因变量均不存在显著相关，存在相关也是可以的。在数据类型上，自变量和调节变量既可以是连续变量也可以是分类变量，但因变量必须为连续变量。

- 方差分析：自变量和调节变量均是分类变量
- 分组回归：自变量是连续变量，调节变量是分类变量
- 分层回归：自变量是分类变量，调节变量是连续变量
- 分层回归：自变量和调节变量都是连续变量



# 调节效应分析

## 方差分析：自变量和调节变量均是分类变量

自变量和调节变量都是分类变量，实际就是多因素方差分析中的交互作用显著性分析，交互作用显著即说明调节效应显著。

步骤：【分析】 - 【一般线性模型】 - 【单变量】 - 【自变量、调节变量放入固定因子，因变量放入因变量框】。

- 问题：检验性别在年级与拖延行为之间是否存在调节作用？
- 自变量：年级（分类变量）
- 调节变量：性别（分类变量）
- 因变量：拖延行为

性别与年级的交互作用不显著，故调节效应不存在。



主体间效应检验

源	III 类平方和	自由度	均方	F	显著性	偏 Eta 平方
修正模型	3.356 <sup>a</sup>	7	.479	1.090	.368	.011
截距	2980.601	1	2980.601	6775.650	.000	.904
性别	.012	1	.012	.027	.869	.000
年级	3.110	3	1.037	2.356	.071	.010
性别 * 年级	.646	3	.215	.490	.690	.002
误差	315.407	717	.440			
总计	4011.385	725				
修正后总计	318.764	724				

a. R 方 = .011 (调整后 R 方 = .001)



# 调节效应分析

## 分组回归：自变量是连续变量，调节变量是分类变量

步骤：【数据】 - 【拆分文件】 - 【分割文件】 - 【比较组】 - 【放入调节变量】 - 【确定】 - 【分析】 - 【回归】 - 【线性】 - 【放入中心化或标准化后的自变量和因变量】。

问题：检验性别在自我控制与拖延行为之间是否存在调节作用？

自变量：自我控制（连续变量）；调节变量：性别（分类变量）；因变量：拖延行为

注：自变量和因变量均已标准化处理。

性别	模型		平方和	自由度	均方	F	显著性
男	1	回归	10.925	1	10.925	71.007	.000 <sup>b</sup>
		残差	17.386	113	.154		
		总计	28.312	114			
女	1	回归	16.740	1	16.740	111.193	.000 <sup>b</sup>
		残差	31.014	206	.151		
		总计	47.755	207			

性别	模型		未标准化系数		标准化系数	t	显著性
			B	标准误差	Beta		
男	1	(常量)	4.309	.197		21.818	.000
		自我控制	-.522	.062	-.621	-8.427	.000
女	1	(常量)	4.412	.159		27.770	.000
		自我控制	-.527	.050	-.592	-10.545	.000

a. 因变量：拖延行为

b. 预测变量：(常量), 自我控制

男性组中，自我控制的标准回归系数为-0.621，而在女性组中，自我控制的标准回归系数为-0.592，这两个回归系数具有较大的差异。但这只是主观判断，从统计角度出发，需要采用费雪Z转化比较差异显著性，如果显著表示存在调节。



# 调节效应分析

## 分层回归：自变量是分类变量，调节变量是连续变量

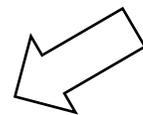
问题：检验性别在自尊与网络成瘾之间是否存在调节作用

(1) 先将自变量转化成虚拟变量：X1、X2、X3；然后将调节变量中心化或标准化处理（M）；求处理之后的调节变量与虚拟变量的乘积：M\*X1、M\*X2、M\*X3

(2) 分层层级回归分析调节效应。第一层：X1、X2、X3、M对Y的回归；第二层：M\*X1、M\*X2、M\*X3对Y的回归。R<sup>2</sup> 改变量是否显著或者XM的回归系数是否显著(显著即调节)。

模型		系数 <sup>a</sup>				显著性
		未标准化系数		标准化系数		
		B	标准误差	Beta	t	
1	(常量)	.031	.031		1.009	.313
	Zscore(性别)	-.058	.031	-.058	-1.864	.063
	Zscore(自尊)	-.566	.031	-.563	-18.245	.000
2	(常量)	.036	.031		1.175	.240
	Zscore(性别)	-.062	.031	-.062	-2.003	.046
	Zscore(自尊)	-.559	.031	-.556	-17.985	.000
	XM	.069	.030	.070	2.277	.023

a. 因变量：Zscore(网络成瘾)



性别与自尊的交互项对网络成瘾不显著，故自尊的调节效应不显著。



# 调节效应分析

## 分层回归：自变量（X）和调节变量（M）都是连续变量

(1) 对两个变量先做中心化或标准化处理：X、M；计算两个变量乘积（交互效应项或调节效应项XM）

(2) 分析层级回归分析调节效应或交互效应。第一层：X、M对Y的回归；第二层：XM对Y的回归。

$$Y=a+bX+cM+e \quad (1)$$

$$Y=a+bX+cM+c'MX+e \quad (2)$$

M为调节变量，MX为调节效应，调节效应是否显著即是分析c'是否显著达到统计学意义上的临界比率水平。

- 检验2个回归方程的复相关系数 $R_1^2$ 和 $R_2^2$ 是否有显著区别，若 $R_1^2$ 和 $R_2^2$ 显著不同，则说明MX交互作用显著，即表明M的调节效应显著；
- 层次回归方程中的c'系数（调节变量偏相关系数），若c'显著，则说明调节效应显著；



# 调节效应分析

## 调节效应和交互效应

- 调节效应和交互效应在统计模型上无本质区别；但调节效应能够指定自变量、调节变量；而交互作用地位是等价的。
- 在交互作用分析中，两个自变量的地位可以是对称的，其中任何一个都可以解释为调节变量；也可以是不对称的，只要其中有一个起到了调节变量的作用，交互效应就存在。
- 在调节效应中，自变量、调节变量是明确的，在一个确定的模型中两者不能互换。



谢谢