



安徽理工大学

ANHUI UNIVERSITY OF SCIENCE & TECHNOLOGY

研究生课程

# 试验设计与分析

主讲 闵凡飞 教授

材料科学与工程学院

2020年5月

## 第五章 试验数据分析常用软件简介

# 序言

## 1 统计分析软件的一般特点

- 1) 集成所有成熟的统计分析方法;
- 2) 易于操控数据、文件、程序和运行;
- 3) 结果的释读清晰、正确; 可方便生成各种统计图表; 便于应用;
- 4) 有完备的联机帮助功能;
- 5) 方便与其它软件数据交换。

## 2 软件名称

**Statistical Package for Social Science**

**Statistical Product and Service Solutions**

## 3 软件发展

**60年代：美国斯坦福大学三位研究生研制**

**70年代：SPSS总部成立于芝加哥，推出  
SPSS中小型机版—SPSSX**

**80年代：微机版（V1-3）SPSS/PC+**

**90年代：Windows版（v5-11）**

**2003年：Windows版（v12）**

**2007年：V16, 2008年：V17**

**2012年：V21**

# 1 SPSS基础

## 1.1 基本窗口

### 1) 数据编辑窗口 (data editor)

运行SPSS，自动打开，运行中至少打开其一；

功能：对SPSS的数据进行定义、录入、修改、管理等基本操作的窗口；

包含数据视图工作表 (data view) 和变数视图工作表 (variable view) 两部分。

组成：窗口主菜单 (Data, Transform, Analyze, Graphs)、工具栏、数据编辑区、状态显示区；

SPSS中各统计分析功能都是针对该窗口中的数据进行的；窗口中的数据可以.sav存于磁盘上。

## 2) 输出窗口 (SPSS viewer)

功能：统计分析报表及图形输出窗口；

组成：窗口主菜单、工具栏、结果显示区、状态显示区；在进行第一次分析时自动打开，也可手工打开；可以手工打开若干个可相互切换的viewer窗口；输出窗口可以打开多个或全部关闭；

状态栏上的“+”表示当前输出窗口；

窗口内容可以.spo存于磁盘上。

## SPSS的主菜单

|    |           |      |
|----|-----------|------|
| 1  | File      | 文件操作 |
| 2  | Edit      | 编辑   |
| 3  | View      | 视图   |
| 4  | Data      | 数据   |
| 5  | Transform | 数据转换 |
| 6  | Analyze   | 统计分析 |
| 7  | Graph     | 图表   |
| 8  | Utilities | 实用程序 |
| 9  | Windows   | 窗口   |
| 10 | Help      | 帮助   |

## 2 SPSS数据编辑

### 2.1 建立SPSS数据文件

#### 1) 数据文件的特点

SPSS数据文件是一种有结构的数据文件。数据文件的每一行代表一个个体(case)。每一列代表一个变数(variable)。

在数据视图工作表(data view)中可以输入和编辑数据，但不能输入数学表达式和函数。

SPSS数据文件更常用读（导）入其它格式的数据文件，如.xls, .txt等。



在变数视图工作表（**variable view**）中可对变数进行定义。

变数名(**variable name**)，是变数操控的标志；

变数类型(**type**)，分数值型、字符型、日期型等。

显示宽度(**width**)，数值型默认宽度8.2，如：12345678、12345.67、-1234.56。字符型字符长度默认8位。

注意：显示宽度不影响数据实值。

变数标签(**Variable label**), 对变数名进行标记, 以便在结果输出时方便阅读。

变量标签(**Value label**), 用于注释变量值, 方便分类变量的输入和显示。

列格式(**Column Format**), 显示数据工作表列的宽度。

缺失值(**Missing Values**)限定数据的范围。

变数尺度(**Measurement**), 分**Scale** (数值型)、**Ordinal** (序数型)、**Nominal** (名称型), 在分类汇总等运算中 useful。

## 2) 数据的录入和保存

打开数据视图工作表，可输入、编辑数据，可进行定位、查找、替换等操作。

插入个体：在主菜单的edit栏，有insert cases 项。删除时用edit 中的cut 即可。

插入变数：在主菜单的edit栏，有insert variables 项。删除时用edit 中的cut 即可。

数据单元的移动和复制在同一（或不同）数据视图工作表中进行，剪切、复制、粘贴，也可与其它数据文件所共用（但须注意变数定义格式的一致性）。

保存：存贮成 filename.sav 格式。

### 3) 与其他软件数据共享

在SPSS13可以直接读入许多常用格式的数据文件（\*.sav, systat, EXCEL, Lotus, SYLK, Dbase, Access, txt, dat, csv）等。

使用数据库查询打开。

利用数据库**ODBC**接口读取数据文件。

使用文本导入向导读入文本文件。

更常用复制、粘贴。

## 2.2 SPSS数据的初步整理

### 1) 排序

#### 个体排序 (sort cases)

将所有个案按照用户指定的某一个或多个变数的变量值的升序或降序重新排列。

变数排序(sort variables)。

查找重复个体 (identify duplicated) 。

查找异常值(identify unusual cases)。

## 2) 选取

选取个体 (**goto case**) , 选取变数 (**goto variable**) 。

## 3) 个体选取(**select cases**)

个案选取有多种方式, 选取的个体可另存数据文件。

## 4) 插入

插入个体 (**insert cases**) , 插入变数 (**insert variable**) 。

## 5) 删除

使用 **cut**, 也可直接用**del**键删除。

## 6) 查找、替换

**Find, replace.**

## 7) 数据转置(transpose)

将数据行列互换。

## 8) 重组(restructure)

将数据按某种要求重新组织。

## 9) 文件合并(merge file)

将两个SPSS数据文件合并到一个数据文件中。  
纵向合并（**add cases**）、横向合并（**add variables**）。

## 10) 数据加工 (aggregate)

计算变量、产生计数变量、分类汇总、数据分组；

## 11) 数据文件的其他处理功能

指定加权变数、变数集的定义和使用。



## 2.3 SPSS函数

所有软件都有函数计算，SPSS也不例外，SPSS能够实现一定的函数计算。

在主菜单**transform**栏，有**compute variable...**项，可进行变数的函数计算。

根据函数功能和处理的变量类型，SPSS函数大致可以分成八大类：算术函数、统计函数、分位数函数、逻辑函数、字符串函数、缺失值函数、日期函数和其他函数。

## 算术函数 (1)

| 函数名        | 功能          | 举 例                               |
|------------|-------------|-----------------------------------|
| Abs(算术表达式) | 求绝对值        | Abs(sr-850):分别计算每个个案变量sr与850差的绝对值 |
| Sqrt(正数)   | 求平方根        | sqrt(4):函数值=2                     |
| Sin(弧度)    | 求正弦值        | Sin(30*3.14/180):函数值=0.50         |
| Cos(弧度)    | 求余弦值        | Cos(60*3.14/180):函数值=0.50         |
| Exp(算术表达式) | 求e的若干次幂     | Exp(5):函数值=148.41                 |
| Ln(算术表达式)  | 求以e为底的自然对数值 | Ln(sr):分别计算每个个案变量sr的自然对数值         |

## 算术函数 (2)

| 函数名          | 功 能         | 举 例                      |
|--------------|-------------|--------------------------|
| <b>Lg10</b>  | 求以10为底的对数值  | <b>Lg10(5):函数值=0.7</b>   |
| <b>Rnd</b>   | 求四舍五入后的整数   | <b>Rnd(2.66):函数值=3.0</b> |
| <b>Trunc</b> | 求截去小数部分后的整数 | <b>Trunc(4.7):函数值=4</b>  |
| <b>Mod</b>   | 求除以常数后的余数   | <b>Mod(20,3):函数值=2</b>   |

SPSS函数还有：

CDF & Noncentral CDF(累积概率函数),

Conversion（转换），

current date/time, date arithmetic,

date extraction, date creation,

inverse DF(将概率转化换为正态离差等),

PDF & noncentral PDF(概率密度函数),

Search（求极值、平均数等），

Random numbers（产生各类随机数），

Statistics（计算方差、协方差等），

Significance(卡方、F分布的临界值计算),

.....等。

## 3 描述性统计分析

### 3.1 描述性统计(descriptive statistics)

#### 1) 频数分布分析

频数分布分析主要通过频数分布表、条图和直方图，以及集中趋势和离散趋势的各种统计量，描述数据的分布特征。

计算频率分布表，绘制统计图形。

上述方法适用于定序、定类、定距类型数据。

步骤：【Analyze】，【Descriptive Statistics】，【Frequencies】

## 2) 描述性统计分析 (Descriptives)

主要计数变量的个数，计算最大、最小、极差、平均数、标准差、方差等。

## 3) 扩展性统计分析 (explore)

除了具有以上描述性统计有关的统计数的计算外，还有百分位、方柱图、茎叶图（箱形图）等的计算。

#### 4) 列联表 (Crosstabs)

这是一种两向分组的次数分布表。

#### 5) 比例 (ratio)

计算两个或多个变数的比值 ( $V1/V2$ ) 。

#### 6) P-P Plot, Q-Q plot

察看变数百分位数 (或实值) 与正态分布累积概率之间的关联情况。

## 3.2 平均数比较(compare means)

平均数比较主要涉及假设检验的内容，这将在后续章节再论。而有关变数描述性统计的内容，也能在means等过程实现。

- 1) 平均数(means)
- 2) 单样本  $t$  检验(one-sample  $t$  test)
- 3) 成组比较(independent samples  $t$  test)
- 4) 成对比较(paired samples  $t$  test)
- 5) 单向分组方差分析(one-way ANOVA)



## 4 常用统计图

统计图是用点、线、面等各种几何图形表达统计数据的一种形式，其特点是简明生动，形象具体、通俗易懂且又一目了然。

SPSS具有较强的制图功能，可以绘制多种统计图形，这些图形可以由各种统计分析过程产生，也可以直接由菜单“**Graphs**”图形菜单产生。

SPSS软件系统提供了许多产生统计图形的方法。常用的统计图有条形图（**bar chart**）、饼图（**pie chart**）、线图（**linear chart**）和直方图（**histogram**）等。

## 4.1 条形图

条形图一般用于分类资料，主要适用于彼此独立的资料互相比较。例：某水稻杂种第二代植株米粒性状的分离情况如下表，请绘制性状分离条形图。

| 属性分组 | 次数 |
|------|----|
| 红米非糯 | 96 |
| 红米糯稻 | 37 |
| 白米非糯 | 31 |
| 白米糯稻 | 15 |

## 【Graphs】 【Bar...】

### 对话框选项说明

【Simple】——简单条形图

【Clustered】——分组（类）条形图

【Stacked】——堆积图（分段条形图）

【Summaries for groups cases】——观察值分类描述模式，即对变量中的观察值进行分组后绘图。

【Summaries of separate variables】——变量描述模式，即对每个变量生成一个条形图。

【Values of individual cases】——观察值描述模式，即对应分类轴变量中的每一观察值生成一个条形图。

# 4.2 复式条形图

下表为几种动物性食品的营养成分  
fig1，试绘制复式条形图。

| 品名  | 蛋白质  | 脂肪  | 糖类  | 无机盐 | 水分   | 其他   |
|-----|------|-----|-----|-----|------|------|
| 牛奶  | 3.3  | 4   | 5   | 0.7 | 87   |      |
| 牛肉  | 19.2 | 9.2 |     | 1   | 62.1 | 8.5  |
| 鸡蛋  | 11.9 | 9.3 | 1.2 | 0.9 | 65.5 | 11.2 |
| 咸带鱼 | 15.5 | 3.7 | 1.8 | 10  | 29   | 40   |

分析过程:

**【Graphs】** , **【Legacy Dialog】** ;

**【Clustered】** 一分组（类）条形图;

选择观察值描述模式 **【Values of Individual Cases】** ;

单击 **【Define】** ;得到**Define Clustered Bar**对话框;

将脂肪等数值变数选入 **【Bars Represent:】** 框;

将品名选入 **【Category Labels】** 框;

**【OK】** 。

## 4.3 散点图

散点图就是将两个变数的 $n$ 对观察值分别以坐标点的形式标记于同一直角坐标的平面上。

可由 [graphs], [legacy dialogs], [scatter/dot]

也可由 [graphs], [chart builder], [scatter/dot] 过程实现。例：研究某种昆虫孵化历期平均温度（℃）与历时天数（d）之间关系(fig2)，试制作线图。

|             |      |      |      |      |      |      |      |      |
|-------------|------|------|------|------|------|------|------|------|
| 平均温度<br>(℃) | 11.8 | 14.7 | 15.6 | 16.8 | 17.1 | 18.8 | 19.5 | 20.4 |
| 历时天数<br>(d) | 30.1 | 17.3 | 16.7 | 13.6 | 11.9 | 10.7 | 8.3  | 6.7  |

## 4.4 线图

线图是反映连续性双变数资料关系的主要图形。

这可由[graphs], [chart builder], [line], 或:

[graphs], [legacy dialogs], [line]过程实现。

## 4.5 复式线图

画多条线图。

过程与上相似: [graphs], [legacy dialogs], [line]

## 4.6 直方图（柱形图、矩形图）

直方图适用于表示连续性资料（计量资料）的次数分布。

### 【Graphs】 【Histogram】

SPSS系统根据原始数值的最大值和最小值自动对变量分组，画出直方图。

如果要修改组数等，可以在图形编辑窗口中改变。或在打开【histogram】对话框后，再开启set parameters对话框。调节 ‘number of interval’ 对图形作出适当的改变。



## 5 平均数的假设检验

### 5.1 统计假设检验概述

统计推断（statistical inference），是根据统计数的分布和概率理论，由样本结果（统计数）来推论总体特征（参数）。

统计推断的基本内容：有假设检验(hypothesis test)和参数估计(parameter estimation)两个内容。

统计假设检验是指根据某种实际需要，对未知总体提出一些假设（这些假设通常构成完全事件系），计算在假设条件下出现实际样本的概率，依率作出接受或否定某种假设（显著性）的判断。

## 5.2 单样本的 $t$ 测验

平均数的假设测验主要通过[compare means]实现。以检验实得样本平均数与指定的检验值之间是否存在显著差异。

步骤：

**【Analyze】，【Compare means】，**

**【One sample t test】**

例1：测定某稻田的地表光强4次，得结果为3.4， 2.8， 3.5， 4.1 (千勒克斯)，试测验该结果与根据Beer-Lambert定律推出的理论值  $\mu_0=3.0$  是否有显著差异。

例2. 随机抽查了一个小麦新品系9个样点的千粒重分别是37.2， 37.8， 38.4， 39.3， 51.3， 46.4， 38.9， 37.5， 49.5。问与43克的标准值有无显著差异？

## 5.3 两独立样本的 $t$ 测验

根据两独立样本数据，对是否有显著差异进行测验。

步骤：【Analyze】，【Compare Means】，【Independent-sample  $t$  test】。

待分析的数据需置于同一列（同一变数），但需用一个分组变数将其区分。分析时首先检验两样本方差同质性，方差不同质时用unequal行的  $t$  测验结果。方差同质，看equal行的  $t$  测验结果（不能自动判断，或给出优先结果，这是SPSS众多缺陷中的一个）。

例1：测定前作喷洒过某种有机砷杀虫剂的麦田植株样本4次，得株体中的砷残留量为7.5，9.7，6.8，6.4(毫克)；测定对照(前作未用过有机砷杀雄剂)的植株样本3次，得株体中砷含量为4.2，7.0，4.6。试测验喷洒有机砷杀雄剂是否使后作株体的砷含量显著增高。

例2：为比较甲、乙两地生产的棉花所纺纱线的强力，各抽取7个和8个样本进行测量，得数据如下(单位：公斤)

甲地：1.55 1.47 1.52 1.60 1.43 1.53 1.54

乙地：1.42 2.49 2.46 1.34 1.38 1.54 1.38 1.51

问两种棉花所纺纱线的强力有无显著差异？

简明分析步骤：

**【Analyze】 【Compare Means】 【Independent-Samples T Test】**

**Test Variable[s]框： 砷残留量**

**Grouping Variable框： 样本**

**Define groups...**

**Group1： 键入1**

**Group2： 键入2**

**Continue**

**OK**

## 5.4 成对比较 $t$ 测验

对配对样本是否有显著差异进行测验。

如两种处理施行在基础条件相对一致的个体（或群体）上，其效应的差异是否达到显著。这在许多研究中经常采用。

两样本数据必须两两配对，即：样本个数相同，个案顺序相同。

步骤：

**【Analyze】，【Compare Means】，**

**【Paired-sample  $t$  test】**

例：为测定A、B两种病毒对烟草的致病力，取8株烟草，每一株皆半叶接种A病毒，另半叶接种B病毒，以叶面出现枯斑数的多少作为致病力强弱的指标，得结果于下表。试测验两种病毒致病力的差异显著性。

| 株号 | 病毒A ( $Y_{1j}$ ) | 病毒B ( $Y_{2j}$ ) |
|----|------------------|------------------|
| 1  | 9                | 10               |
| 2  | 17               | 11               |
| 3  | 31               | 18               |
| 4  | 18               | 14               |
| 5  | 7                | 6                |
| 6  | 8                | 7                |
| 7  | 20               | 17               |
| 8  | 10               | 5                |



简明分析步骤：

**【Analyze】，【Compare Means】  
，【Paired-Samples T Test】**

**Paired Variables框：病毒a—病毒b**

**OK**

## 6 方差分析

### 6.1 方差分析基本概念

方差分析将多组数据作为一个整体，将总变异分解成各个变异来源的平方和和自由度，估计各个变异来源的方差，利用 $F$ 测验鉴别组间差异的显著性。这是最广为应用的数据统计分析方法。

在SPSS中，用于方差分析的过程主要是means和 univariate 两个过程，前者主要为单向分组资料，后者为两向分组（或多因素）资料。

## 6.2 单因素的方差分析

多个处理(样本)平均数间的差异是否显著?

例1：一水稻施肥试验，设5个处理：A、氨水，B、废氨水，C、碳酸氢铵，D、尿素，E、不施肥。每处理4盆，完全随机设计。其稻谷产量见下表，试测验各处理平均数的差异显著性。

| 处理 | 观察值 |    |    |    |
|----|-----|----|----|----|
| A  | 24  | 30 | 28 | 26 |
| B  | 27  | 24 | 21 | 26 |
| C  | 31  | 28 | 25 | 30 |
| D  | 32  | 33 | 33 | 28 |
| E  | 21  | 22 | 16 | 21 |

简明分析步骤:

**【Analyze】，【Compare Means】，**

**【One-Way ANOVA】**

或**【Analyze】，【Compare Means】，【means】**

**Dependent List框：产量**

**Factor框：组别**

**Options**

**√Descriptive**

**Continue**

**OK**

例2. 下表为某职业病防治院对31名石棉矿工中的石棉肺患者、可疑患者和非患者进行了用力肺活量（L）测定的数据，问三组石棉矿工的用力肺活量有无差别。

|       |     |     |     |     |     |     |     |     |     |     |     |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 石棉肺患者 | 1.8 | 1.4 | 1.5 | 2.1 | 1.9 | 1.7 | 1.8 | 1.9 | 1.8 | 1.8 | 2   |
| 可疑患者  | 2.3 | 2.1 | 2.1 | 2.1 | 2.6 | 2.5 | 2.3 | 2.4 | 2.4 |     |     |
| 非患者   | 2.9 | 3.2 | 2.7 | 2.8 | 2.7 | 3   | 3.4 | 3   | 3.4 | 3.3 | 3.5 |

## 6.3 方差分析中的多重比较

如果 $F$ 测验差异，并不能说明处理间均存在显著差异。多重比较将进行两两处理间的均值比较检验。常用方法有：

**LSD**，实际上就是  $t$  检验的变形，只是在变异和自由度的计算上利用了全试验误差信息。

**DMRT, Duncan**氏新复极差测验法。

**Q, Tukey**氏固定极差测验法。

**DLSD, Dunnett**氏最小显著差数测验法，等。

实现手段：**【Post Hoc...】**，再行选择。

## 6.4 多因素方差分析

试验因素有固定因素 (**Fixed Factor**) 与随机因素 (**Random Factor**) 之分。前者指试验因素的 $k$ 个处理 (水平) 是经过特意选择的。后者指试验因素的 $k$ 个处理 (水平) 是从该因素总体中随机抽出的 (处理) 样本。

对于固定模型资料, 重在效应的比较分析, 对于随机模型资料, 重在处理 (水平) 间变异度的评价。

例1. 将一种生长激素配成 $M_1, M_2, M_3, M_4, M_5$ , 5种浓度, 并用 $H_1, H_2, H_3$ , 3种时间浸渍某大豆品种的种子, 45天后得各处理每一植株的平均干物重(克)于下表, 试作方差分析。

| $M_i$ | $H_j$ |       |       |
|-------|-------|-------|-------|
|       | $H_1$ | $H_2$ | $H_3$ |
| $M_1$ | 13    | 14    | 14    |
| $M_2$ | 12    | 12    | 13    |
| $M_3$ | 3     | 3     | 3     |
| $M_4$ | 10    | 9     | 10    |
| $M_5$ | 2     | 5     | 4     |



**【Analyze】 , 【General Linear Model】 ,  
【Univariate】**

**Dependent Variable框： 干物质重**

**Fixed Factor[s]框： 激素浓度、浸种时间  
Model...**

**⊙ Custom**

**Model框： 激素浓度、浸种时间**

**Continue**

**Post Hoc...**

**√LSD**

**Continue**

**OK**

例2. 施用A1、A2、A3 3种肥料于B1、B2、B3 3种土壤，以小麦为指示作物，每处理组合种3盆，得产量结果(g)于下表。试作方差分析。

| 肥料种类<br>(A)    | 土壤种类(B)             |                     |                     |
|----------------|---------------------|---------------------|---------------------|
|                | B <sub>1</sub> (油砂) | B <sub>2</sub> (二合) | B <sub>2</sub> (白僵) |
| A <sub>1</sub> | 21.4                | 19.6                | 17.6                |
|                | 21.2                | 18.8                | 16.6                |
|                | 20.1                | 16.4                | 17.5                |
| A <sub>2</sub> | 12.0                | 13.0                | 13.3                |
|                | 14.2                | 13.7                | 14.0                |
|                | 12.1                | 12.0                | 13.9                |
| A <sub>3</sub> | 12.8                | 14.2                | 12.0                |
|                | 13.8                | 13.6                | 14.6                |
|                | 13.7                | 13.3                | 14.0                |

简明分析步骤:

**【Analyze】 , 【General Linear Model】 ,  
【Univariate】**

**Dependent Variable**框: 产量

**Fixed Factor[s]**框: 肥料种类、土壤种类。

**Post Hoc...**

**√LSD**

**Continue**

**OK**

## 6.5 随机区组试验方差分析

在相对均匀一致的局部随机安排一套处理，形成一个区组，这种类型的试验称为随机区组试验。

例：将水稻的3个不同细胞质源的不育系( $A_1, A_2, A_3$ )和5个恢复系( $B_1, B_2, B_3, B_4, B_5$ )杂交，配成15个 $F_1$ 。采用随机区组设计，重复2次，小区计产面积6米<sup>2</sup>，得产量结果见下表，进行分析。

| 处 理            |                | 区组 I | 区组 II |
|----------------|----------------|------|-------|
| A <sub>1</sub> | B <sub>1</sub> | 4.3  | 4.1   |
|                | B <sub>2</sub> | 4.9  | 4.8   |
|                | B <sub>3</sub> | 3.9  | 3.6   |
|                | B <sub>4</sub> | 4.8  | 4.0   |
|                | B <sub>5</sub> | 4.7  | 4.5   |
| A <sub>2</sub> | B <sub>1</sub> | 5.2  | 4.7   |
|                | B <sub>2</sub> | 5.0  | 5.2   |
|                | B <sub>3</sub> | 3.8  | 3.4   |
|                | B <sub>4</sub> | 4.9  | 4.8   |
|                | B <sub>5</sub> | 5.0  | 5.8   |
| A <sub>3</sub> | B <sub>1</sub> | 4.6  | 4.7   |
|                | B <sub>2</sub> | 4.4  | 4.2   |
|                | B <sub>3</sub> | 3.5  | 3.4   |
|                | B <sub>4</sub> | 3.4  | 3.6   |
|                | B <sub>5</sub> | 3.7  | 4.2   |

**【Analyze】，【General Linear Model】，  
【Univariate】**

**Dependent Variable框：产量**

**Fixed Factor[s]框：不育系、恢复系、区组  
Model...**

⊙ **Custom**

**Model框：区组、不育系、恢复系、  
不育系×恢复系。**

**Continue**

**Post Hoc...**

**√LSD**

**Continue**

**OK**

## 6.6 协方差分析

协方差分析是将回归分析和方差分析相结合的一种统计分析方法，其主要的功用是对试验误差进行统计控制。将无法或很难控制的因素作为协变数，利用线性回归排除协变数的影响，使目标变数的分析更精确。

协方差分析的主要功用：

测验多个回归系数 $b_i$ 的差异显著性；

矫正处理平均数并测验其差异显著性；

作出不同变异来源的相关分析。

例：为研究A、B、C三种肥料对于苹果的增产效果，选了24株同龄的苹果树，第一年记下各树的产量（ $X$ ，公斤），第二年将每种肥料随机施于8株苹果上，再记下其产量（ $Y$ ，公斤）。得结果于下表，试作分析。

| 肥料 | 观察值 ( $X$ , $Y$ ) |    |    |    |    |    |    |    |    |
|----|-------------------|----|----|----|----|----|----|----|----|
| A  | $X$               | 47 | 58 | 53 | 46 | 49 | 56 | 54 | 44 |
|    | $Y$               | 54 | 66 | 63 | 51 | 56 | 66 | 61 | 50 |
| B  | $X$               | 52 | 53 | 64 | 58 | 59 | 61 | 63 | 66 |
|    | $Y$               | 54 | 53 | 67 | 62 | 62 | 63 | 64 | 69 |
| C  | $X$               | 44 | 48 | 46 | 50 | 59 | 57 | 58 | 53 |
|    | $Y$               | 52 | 58 | 54 | 61 | 70 | 64 | 68 | 66 |



**【Analyze】，【General Linear Model】，  
【Univariate】**

**Dependent Variable框：施肥产量**

**Fixed Factor[s]框：分组变量**

**Covariate[s]框：初始产量**

**Options**

**√Descriptive Statistics**

**Display means for框：分组变量**

**√Compare Main effects**

**Continue**

**OK**

## 7 回归和相关分析

### 7.1 一元线性回归分析

回归分析用于描述变数之间的数量关系，确定一个或几个自变数对一个依变数的影响程度。

一元线性回归方程： $Y=a+bX$

回归方程的显著性测验，测验自变数与依变数之间的线性关系是否显著，测验方法有 $t$ 测验和 $F$ 测验两种方法。

一元回归时， $F$ 测验与 $t$ 测验等价，即： $F=t^2$ ，两种测验可以相互替代。

例：许多害虫的发生都和气象条件有一定的关系。山东临沂测定1964~1973年(共年)间月下旬的温雨系数(雨量mm/平均温度 $^{\circ}\text{C}$ )和大豆第二代造桥虫发生量(每百株大豆上的虫数)的关系如下表，试建立回归方程。

步骤：【Analyze】，【Regression】，【Linear】

| 温雨系数(X) | 虫口密度(Y) | 温雨系数(X) | 虫口密度(Y) |
|---------|---------|---------|---------|
| 1.58    | 180     | 2.41    | 175     |
| 9.98    | 28      | 11.01   | 40      |
| 9.42    | 25      | 1.85    | 160     |
| 1.25    | 117     | 6.04    | 120     |
| 0.30    | 165     | 5.92    | 80      |

## 7.2 相关分析

相关分析旨在测度变数间关系的性质和密切程度。以相关系数( $r$ )体现两个变数间的线性关系程度。

$r$ :  $[-1,+1]$ ;  $r=1$ : 完全正相关;  $r=-1$ : 完全负相关;  $r=0$ : 无线性相关。

说明: 相关系数只是较好地度量了两个变数间的线性相关程度, 不能描述非线性关系。

步骤: **【Analyze】**, **【Correlate】**, **【Bivariate】**

## 7.3 多元线性回归分析

多元回归方程： $Y=b_0+b_1x_1+b_2x_2+...+b_kx_k$

$b_1$ 、 $b_2$ 、 $b_k$ 为偏回归系数。

$b_1$ 表示在其他自变量保持不变的情况下，自变量 $x_1$ 变动一个单位所引起的依变数 $Y$ 的平均变动。

多元线性回归分析的主要内容：

回归方程的测验，自变数的筛选。

自变数筛选法:

**Enter:** 所选择的自变数将全部进入建立的回归方程中，该项为默认方式。

**Remove:** 将进入方程中的自变数剔除。

**Forward:** 向前筛选法，是自变数不断进入回归方程的过程。

**Backward:** 向后筛选法，是自变数逐步剔除出回归方程的过程。

**Stepwise:** 逐步筛选法，是“向前法”和“向后法”的结合。

例. 测定“丰产3号”小麦的每株穗数 ( $X_1$ )、每穗结实小穗数 ( $X_2$ , 主茎)、百粒重 ( $X_3$ , 克)、株高 ( $X_4$ , 厘米, 主茎) 和每株籽粒产量 ( $Y$ , 克) 的关系, 得结果如下表, 试选择  $Y$  依  $X$  的最优线性回归方程。

步骤: **【Analyze】**, **【Regression】**, **【Linear】**  
[enter]: 再行选择。

## 7.4 曲线拟合

### 1) 多项式拟合

在一元回归分析中，因变量与自变量(时间或其它自变量)之间的关系不呈线性关系，但可用多项式模型分析。

常用多项式模型：

$$Y=a+bX(\text{线性拟合linear})$$

$$Y=a+b_1X+b_2X^2(\text{二次曲线quadratic})$$

$$Y=a+b_1X+b_2X^2+b_3X^3(\text{三次曲线cubic})$$

步骤：

【Analyze】，【Regression】，【Curve Estimation】



例1：某地大气中氰化物测定结果如下表所示，试拟合曲线。

|        |       |       |     |       |      |      |      |      |
|--------|-------|-------|-----|-------|------|------|------|------|
| 污染距离X  | 50    | 100   | 150 | 200   | 250  | 300  | 400  | 500  |
| 氰化物浓度Y | 0.687 | 0.398 | 0.2 | 0.121 | 0.09 | 0.05 | 0.02 | 0.01 |

## 2) 非线性拟合

建立依变量与自变量之间的（非多项式）非线性关系。

非线性模型确定后，估计模型中的参数是在给定初值的基础上反复迭代实现。

步骤 **【Analyze】** **【Regression】** **【Nonlinear】**

注意在 **【Model Expression】** 框中输入一个合适的方程，在该方程中，包括变量、参数和常数等。

必须为每个参数设置初始值。

例2：测定水稻品种IR72籽粒开花后不同天数下的平均单粒重( $Y$ , mg)，得结果于下表。试用Logistic方程描述籽粒增重依开花天数的关系。

|                   |     |      |      |      |       |       |       |       |       |
|-------------------|-----|------|------|------|-------|-------|-------|-------|-------|
| 开花后<br>天数(d)      | 0   | 3    | 6    | 9    | 12    | 15    | 18    | 21    | 24    |
| 平均单<br>粒重<br>(mg) | 0.3 | 0.72 | 3.31 | 9.71 | 13.09 | 16.85 | 17.79 | 18.23 | 18.43 |