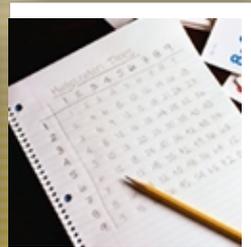
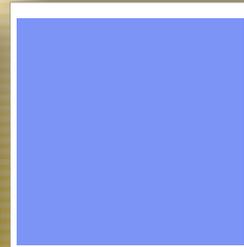
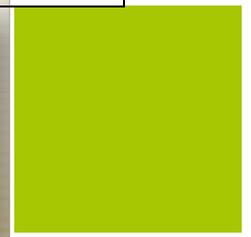
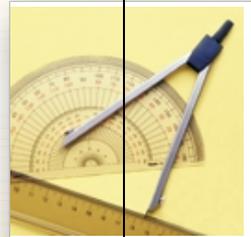


SPSS 回归分析



一、线性回归模型的拟定

影响因变量的自变量只有一种变量的称之为一元，有两个或两个以上的时候称之为多元，假如他们之间有线性关系，就是线性回归。线性关系确实定措施：

对于多元的情况，绘制每个自变量与因变量的散点图，并经过有关性检验拟定它们之间存在线性关系；对于不在线性关系的变量，则找出它们之间的非线性关系，之后再利用非线性关系将有关数据线性化（如 $y=a*\ln x$ ，则将全部数据取自然对数）。从而拟定全部变量都与因变量存在线性关系，为进行多元线性回归做好准备。

二、线性回归模型



回归分析的理论基础.doc

❖ 一元线性回归方程: $y = a + bx$

- a 称为截距
- b 为回归直线的斜率

❖ 多元线性回归方程: $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$

- b_0 为常数项
- b_1 、 b_2 、 \dots 、 b_n 称为 y 相应于 x_1 、 x_2 、 \dots 、 x_n 的回归系数

❖ 一元线性回归模型确实定:一般先做散点图,以便进行简朴地观察,若散点图的趋势大约呈线性关系,能够建立线性方程,若不呈线性分布,可建立其他方程模型,并比较 R^2 来拟定一种最佳方程式(曲线估计)。

二、多元线性回归模型的统计检验

1、多重共线性分析（仅多元回归分析检验）

多重共线性是指解释变量之间存在线性有关关系的现象，该现象的存在会造成：有可能回归方程的**F**统计量高度明显，而每个**t**统计量不明显，严重影响回归效果。测度多重共线性一般有下列方式：

（**spass**操作：分析-回归-线性-在统计量对话框中选共线性分析）

1、容差越大则与方程中其他自变量的共线性越低，应进入方程。具有太小容忍度的变量不应进入方程，**spss**会给出警告。

◇判断措施：容差 <0.1 一般以为该变量与其他变量有多重共线性问题。

2、方差膨胀因子(**VIF**):容差的倒数

3、条件索引。指最大特征根与第*i*个特征根比的平方根。

◇判断措施：一般，当条件索引在**0-10**之间时阐明多重共线性较弱；当条件指数在**10-100**之间阐明多重共线性较强；当条件指数不小于**100**时阐明存在严重的多重共线性。

注：对存在共线性的变量应清除，逐渐删除容差最小的变量，**spass**会自动清除共线性变量，并给出有关成果。（选“进入”措施则不会）**SPSS**

二、线性回归模型的统计检验

2、回归方程的拟合优度

回归直线与各观察点的接近程度称为回归方程的拟合优度，也就是样本观察值汇集在回归线周围的紧密程度。其检验统计量与一元线性回归相同：

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

阐明：**R²**体现了回归方程所能解释的因变量变差的百分比；**1-R²**则体现了因变量总变差中，回归方程所无法解释的百分比。**R²**越接近于**1**，则阐明回归平方和占了因变量总变差平方和的绝大部分百分比，因变量的变差主要由自变量的不同取值造成，回归方程对样本数据点拟合得好。**Spss**中的有关成果在“**模型汇总^b**”中

二、线性回归模型的统计检验

3、回归方程的明显性检验（方差分析F检验）

回归方程的明显性检验是要检验被解释变量与全部的解释变量之间的线性关系是否明显。

◇原假设 $H_0: \beta = 0$ 即:回归系数与0无明显差别

一元回归分析检验统计量:

$$F = \frac{SSR / 1}{SSE / (n - 2)} = \frac{\sum (\hat{y} - \bar{y})^2 / 1}{\sum (y - \hat{y})^2 / (n - 2)} \sim F(1, n - 2)$$

多元回归分析检验统计量:

$$F = \frac{SSR / p}{SSE / (n - p - 1)} = \frac{\sum (\hat{y} - \bar{y})^2 / p}{\sum (y - \hat{y})^2 / (n - p - 1)} \sim F(p, n - p - 1)$$

◇计算F统计量的值和相伴概率p（spss中为sig，默认应不小于0.05），

假如F值较大且 $p \leq \alpha$ ，则拒绝 H_0 ，即:回归系数与0有明显差别，则阐明自变量造成的因变量的线性变动远不小于随机原因对因变量的影响，自变量于因变量之间的线性关系较明显。反之，不能拒绝 H_0 。

二、线性回归模型的统计检验

4、回归系数的明显性检验（t检验）

回归系数的明显性检验是要检验回归方程中被解释变量与每一种解释变量之间的线性关系是否明显。 $H_0: \beta_i = 0$ 即:回归系数与0无明显差别

一元回归分析检验统计量

$$t = \frac{\beta_1}{\sigma} \sim t(n-2)$$
$$\sqrt{\sum (x_i - \bar{x})^2}$$

其中, $\sigma = S_y = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$

多元回归分析检验统计量

$$t_i = \frac{\beta_i}{\sigma} \sim t(n-p-1)$$
$$\sqrt{\sum (x_{ij} - \bar{x}_i)^2}$$

其中, $\sigma = S_y = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-p-1}}$

◇计算t统计量的值和相伴概率p（spss中为sig，默认应不小于0.05）， $p \leq a$ ，拒绝 H_0 ，即:回归系数与0有明显差别，则阐明自变量造成的因变量的线性变动远不小于随机原因对因变量的影响，自变量于因变量之间的线性关系较明显。反之，不能拒绝 H_0 。spss中有关成果在“系数^a”中。

注：对于明显性不高的变量应清除，再回归。

二、多元线性回归模型的统计检验

5、残差分析（在统计量中勾选durbin-watson，在绘制中勾选正态图）

残差是指由回归方程计算得到的预测值与实际样本值之间的差距，定义为：

$$e_i = y_i - \hat{y}_i = y_i - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

对于线性回归分析来讲，假如方程能够很好的反应被解释变量的特征和规律性，那么残差序列中应不包括明显的规律性。

1、检验残差的分布是否为正态，一般用**p-p**图，假如各观察的散点均匀地分布在对角线上，能够判断残差服从正态分布。

2、检验残差的独立性（**DW**检验）。

DW检验用来检验残差的自有关。

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2} \approx 2(1 - \rho)$$

检验统计量为：

◇**判断**：**DW=2**表达无自有关，在**0-2**之间阐明存在正自有关，在**2-4**之间阐明存在负的自有关。一般情况下，**DW**值在**1.5-2.5**之间即可阐明无自有关现象。残差序列不存在自有关，能够以为回归方程基本概括了因变量的变化；不然，以为可能某些与因变量有关的原因没有引入回归方程或回归模型不合适等。

三、回归模型的成果检验

- ❖ 回归分析模型中，除进行多种统计检验外，在数据较多的情况下，可在回归分析中去掉最终一种数据，用以检验。
- ❖ 在得到回归模型后，用其预测最终一种数据，并与真实值进行比较，看其是否在置信区间内，来检验模型的合理性。**Spss**操作如下：

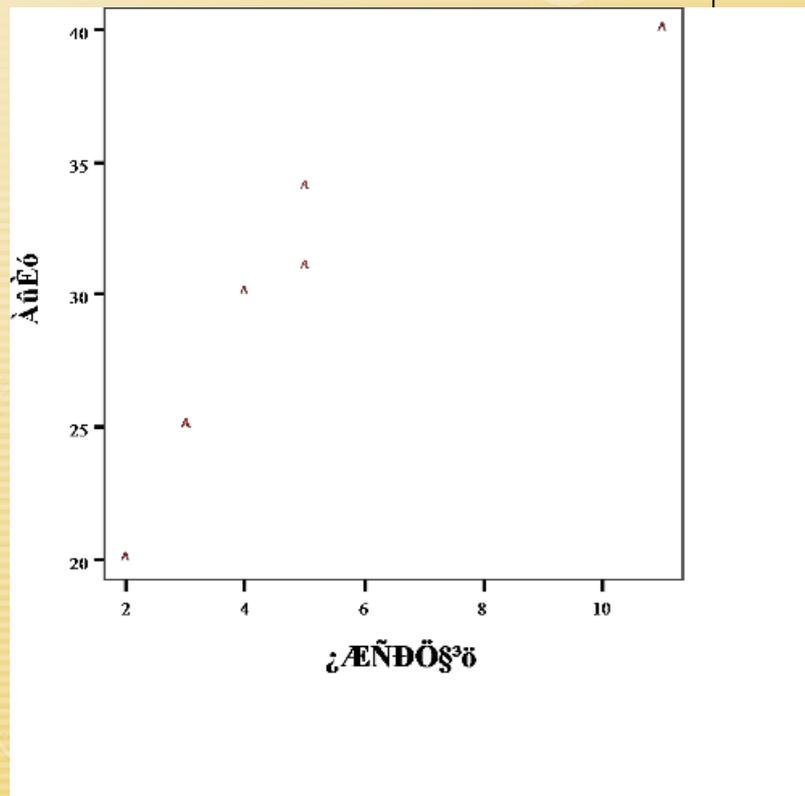
在“回归分析”对话框中，选“保存”按钮，在其中勾选保存预测值和均值置信区间。

一元线性回归举例1

❖ **问题：**某企业近年来科研支出 x 与利润 y 的统计资料如下表（单位：10万元）。将 x 与 y 的数据绘制出散点图，观察 x 与 y 具有线性关系。计算出 y 有关 x 的线性回归模型 $y=ax+b$ ，并检验该模型是否明显以及给出模型的原则误差。

科研支出	利润
11.00	40.00
4.00	30.00
5.00	34.00
3.00	25.00
2.00	20.00

散点图：图形→旧对话框→散点/点状，选择简朴分布，再选入有关变量，最终拟定。如右图，初步拟定它们之间存在线性关系。



一元线性回归举例1

- 选择措施：分析→回归→线形→将y送入自变量框，将x送入因变量框，在统计量对话框中选择回归系数的置信区间和DW残差独立性检验，在绘制对话框中选正态概率图（p-p图），→拟定运营。



统计量对话框阐明：

- 估计：输出有关回归系数的统计量，涉及回归系数、回归系数的原则差、原则化的回归系数、t统计量及其相应的p值等。
- 置信区间：输出每个回归系数的95%的置信度估计区间。
- 协方差矩阵：输出解释变量的有关系数矩阵协方差阵。
- 模型拟合度：输出可决系数、调整的可决系数、回归方程的原则误差、回归方程F检验的方差分析。
- Durbin-watson：残差独立性检验



成果分析:模型汇总表

- ❖ 鉴定系数 $R=0.909$, $R^2=0.826$, 阐明 y 有82.6%是由变量 x 引起的,
 - ❖ 估计误差是指估计值与观察值的平均差别程度, 用 s 表达, s^2 是总体方差的无偏估计量。估计原则误差的值越小, 则估计量与其真实值的近似误差越小。
- 注:** 估计误差只能说越小越好, 并不能精确阐明回归线的代表性怎样, 所以其实际参照意义有限, 一般不作分析(涉及回归系数的原则误差)。

模型汇总 ^b					
模型	R	R 方	调整 R 方	❖ 原则估计的误差	Durbin-Watson
1	.909 ^a	.826	.783	3.24037	1.500

a. 预测变量: (常量), x 。
b. 因变量: y

成果分析：方差分析表

- √ 方差分析表：对回归模型进行明显性检验，回归明显性概率 $\text{sig.} = 0.012 < 0.05$ ，所以回归明显。
- √ 总变差平方和SST为242，回归平方和SSR为200，残差平方和SSE为42。 $R^2 = \text{SSR} / \text{SST}$
- √ df为自由度
 $F(0.012, 1, 4) = 19.048$ 远不小于 $F(0.05, 1, 4) = 7.709$
所以回归方程有关性比较明显。（F值可在Excel中用FINV函数取得）。

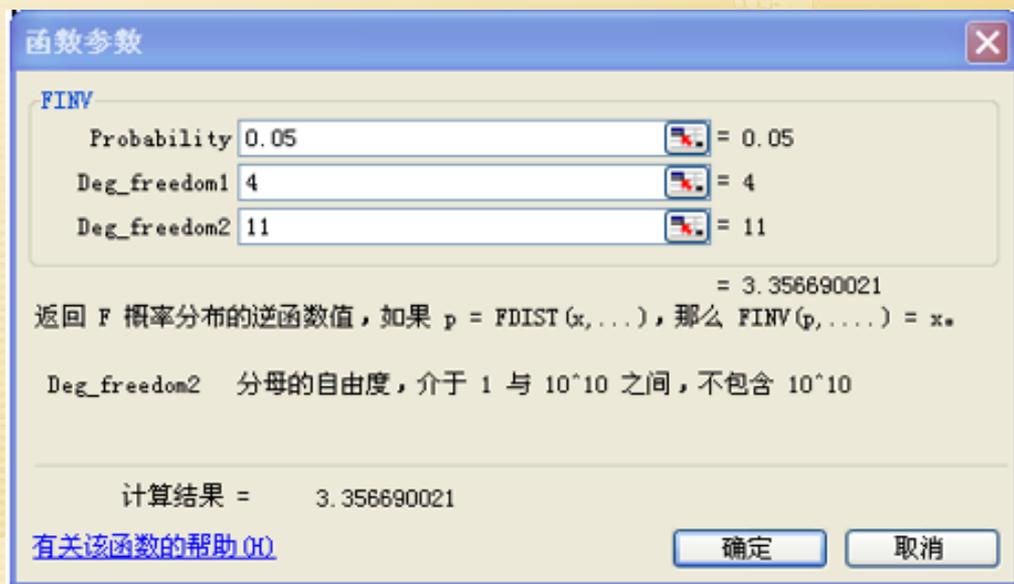
Anova ^b						
模型		平方和	df	均方	F	Sig.
1	回归	200.000	1	200.000	19.048	.012 ^a
	残差	42.000	4	10.500		
	总计	242.000	5			

a. 预测变量：(常量), x。

b. 因变量：y

Excel中获取F值（注：要求为2023版或者用wps）

环节：在Excel函数选项  的下拉列表中选择其他函数，在选择类别中选择“统计”，再选择“FINV”函数，点击拟定，打开函数参数窗口，如下：



输入相应的控制水平、第1自由度和第2自由度，便可得到相应的F值（计算成果）

成果分析：系数表

❖ 回归模型为： $y=20+2x$

模型		非标准化系数		标准系数	t	Sig.	B 的 95.0% 置信区间	
		B	标准 误差	试用版			下限	上限
1	(常量)	20.000	3.085		6.417	.008	9.981	29.619
	x	2.000	.522	.911	3.835	.031	.340	3.660

a. 因变量：y

- √ 常量和解释变量x的t统计量的相伴概率p分别为0.008和0.031，不大于明显性水平0.05，所以，在0.05的明显性水平下常量和解释变量x都经过了t检验。
- √ 常量和解析变量x系数的95%的置信区间分别为：**(9.981, 29.619)**，**(0.340, 3.660)**

成果分析：残差分析

残差统计量 ^a					
	极小值	极大值	均值	❖ 原则 偏差	N
预测值	23.8000	41.8000	29.8000	7.07107	5
残差	-3.80000	❖ 4.20 230	.00000	3.19374	5
❖ 原则 预 测值	-.849	1.697	.000	1.000	5
❖ 原则 残 差	-1.030	1.139	.000	.866	5

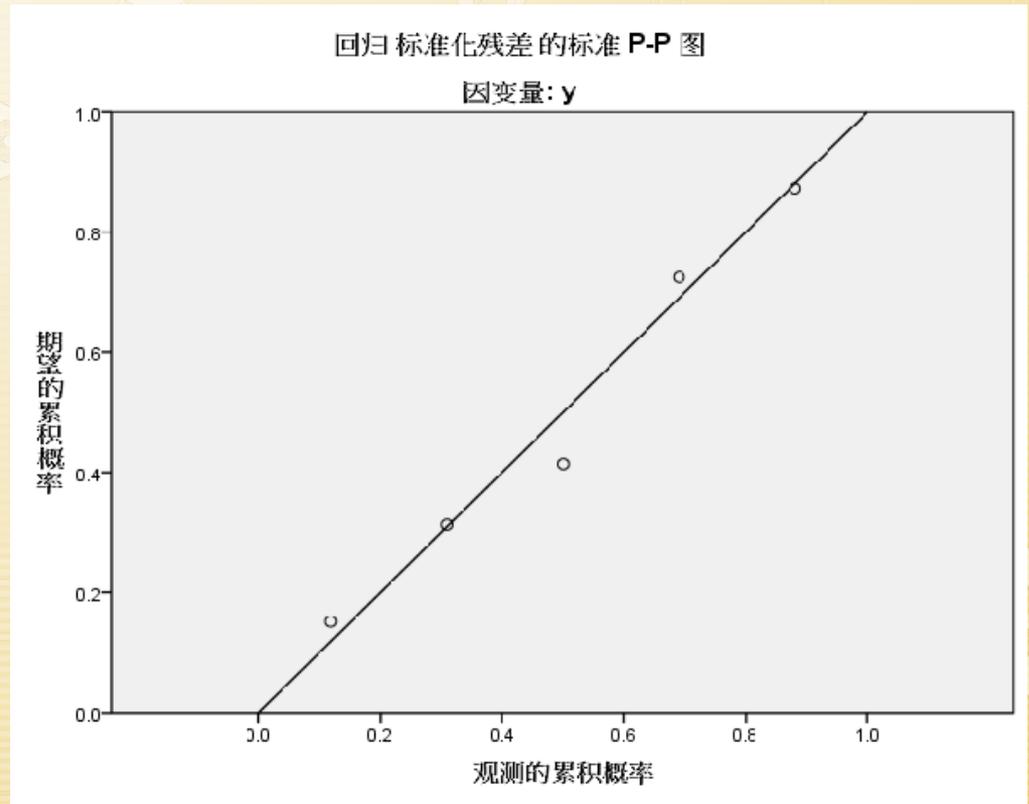
a. 因变量: y

残差p-p图来判断回归模型的优劣。

成果分析：残差分析

1、为了判断随机扰动项是否服从正态分布，观察右图所示的原则化残差的P-P图，能够发觉，各观察的散点均匀地分布在对角线上，据此能够判断残差服从正态分布。**注：可经过残差图观察是否有异常点，如有则应进行有关处理后再进行回归**

2、由模型汇总表可得 $DW=1.5$ 在1.5-2.5之间，阐明残差经过独立性检验，无自有关现象。



★鉴于模型的各项检验都有明显性，残差服从正态分布且相互独立，所以该模型是一种很好的模型。

一元线性回归举例2

❖ 问题:

❖ 既有**1987~2023**年湖南省全社会固定资产投资总额**NINV**和**GDP**两个指标的年度数据，见下表。试研究全社会固定资产投资总额和**GDP**的数量关系，并建立全社会固定资产投资总额和**GDP**之间的线性回归方程。

❖ **Spss**操作参照例题1

年份	固定资产投资	GDP
1987.00	120.38	509.44
1988.00	144.71	614.07
1989.00	114.51	682.80
1990.00	121.24	744.44
1991.00	156.39	833.30
1992.00	234.40	997.70
1993.00	324.58	324.58
1994.00	422.18	422.18
1995.00	523.00	2195.70
1996.00	684.14	2647.16
1997.00	667.39	2993.00
1998.00	796.90	3118.10
1999.00	883.90	3326.80
❖ 2023.00	1012.20	
		3691.88
❖ 2023.00	1174.30	3983.00

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/747145066005006160>