十三五 高等院校石油天然气类规划教材

计算机地质制图

李爱荣 ◎ 主编
武富礼 ◎ 主审

石油工业出版社
Petroleum Industry Press

高等院校石油天然气类规划教材

计算机地质制图

李爱荣　主编
武富礼　主审

石油工业出版社

内 容 提 要

本教材优选了几款与石油地质工作密切相关的常用软件进行介绍，包括 Surfer11、Grapher10、GeoMap3.6、Gxplorer、ResForm3.0，并介绍了与地质制图相关的数据分析、数据与图形的转换等知识。

本书是高等院校石油地质类专业的教材，也可供从事相关行业的科技工作者参考。

图书在版编目(CIP)数据

计算机地质制图/李爱荣主编.
北京：石油工业出版社，2016.10
高等院校石油天然气类规划教材
ISBN 978-7-5183-1504-8

Ⅰ.计…
Ⅱ.李…
Ⅲ.地质图—计算机制图—高等学校—教材
Ⅳ.P285.1—39

中国版本图书馆 CIP 数据核字(2016)第 233024 号

出版发行：石油工业出版社
（北京市朝阳区安华里 2 区 1 号楼　100011）
网　址：www.petropub.com
编辑部：(010)64523693
图书营销中心：(010)64523633　(010)64523731
经　销：全国新华书店
排　版：北京苏冀博达科技有限公司
印　刷：北京晨旭印刷厂

2016 年 10 月第 1 版　2016 年 10 月第 1 次印刷
787 毫米×1092 毫米　开本：1/16　印张：17.25
字数：397 千字

定价：35.00 元
（如发现印装质量问题，我社图书营销中心负责调换）

前言

在石油地质工作研究中，各类地质现象可以通过图形来展示，所以地质图件的编制和显示逐步由二维到三维，由定性到半定量、定量以及可视化发展，掌握计算机地质制图也是现代石油地质工作者必备的技能之一。

计算机地质制图是针对石油地质类专业学生的学习和工作需求，选取了目前石油地质研究工作中常用的几款绘图软件，通过对典型图件深入浅出的解析、编制过程的逐步演示，讲解了如何借助通用软件或者行业软件来编制所需的各种地质类图件。通过本教材的学习，读者可以很快地掌握基本地质图件编制过程和方法，大大提高工作效率。

由西安石油大学和西南石油大学联合编写的这本《计算机地质制图》教材，综合了编者近十年的教学、科研实践，以适应教学、科研、生产的需要。全书共分为七章，内容编写分工如下：第二章、第三章和第五章由西安石油大学李爱荣编写，第一章、第四章、第七章由西安石油大学刘桂珍编写，第六章由西南石油大学杨辉廷编写。全书由李爱荣主编，由西安石油大学武富礼教授担任主审。

在本教材的编写过程中，得到了西安石油大学赵靖舟教授、王凤琴教授、时保宏教授的大力支持，以及西南石油大学蔡正旗教授、长江大学李少华教授的关心和指导。西安石文软件公司与北京侏罗纪软件股份有限公司对西安石油大学软件校园版的安装工作给予了全方位的支持和帮助，尤其石文软件公司张章更是自始至终全程协助与指导。西安海卓石油信息技术有限公司向西南石油大学捐赠 ResForm 校园网络版软件，并对安装工作给予了全方位的支持和帮助。西南石油大学丁熊老师和西安石油大学研究生殷悦悦同学在资料的整理过程中做了大量工作。在此一并表示衷心的感谢。

鉴于当下各类软件的升级和完善速度较快，版本不断升级，加上编者水平有限，书中难免有纰漏或者不完善之处，恳请广大读者以最新版本和官方版本的说明为准。

编者

2016 年 7 月

目录

第一章 数据分析

在地质学研究中，为了反映两个或多个变量之间的相互关系，描述相关关系的方向与密切程度，需要采用相关分析；为了反映两个或多个变量之间的依存关系，建立回归方程，需要采用回归分析。

第一节 相关分析

自然界中许多现象之间存在着相互依赖、相互制约的关系。这种关系表现在量上主要有两类。一类是确定性的函数关系，其变量之间有着确定性的关系。在这种情况下给定自变量的数值时，便有一个确定的因变量值与之对应，并且这种关系可以用一个数学表达式反映出来。另一类是不确定的相关关系，其变量之间存在着密切的关系，从一个（或一组）变量的每一确定值，不能求出另一变量的确定值，可是在大量的试验中，这种不确定的关系又具有某种统计规律性。变量之间存在的这种不确定的数量依存关系，称为相关关系。

一、相关关系的种类

客观事物的联系和变化是相当复杂的，其相关关系也表现为各种不同的形式。

（1）按变量之间相互关系的表现形式不同，表现为线性相关和非线性相关。

当 X 值每增加 1 个单位时，Y 值随之而发生大致均等的增加或减少，如果将各对观测值画成散点图，则各个观测点的分布形状近似为直线，这种相关关系称为线性相关。当 X 每增减 1 个单位时，Y 值随之发生不均等的增加或减少。从散点图看，各个观测点的分布形状近似为各种不同的曲线，如抛物线、双曲线、指数和对数等，这种相关关系称为非线性相关。

(2)按变量变化的方向不同,分为正相关和负相关。

当 X 值每增加 1 个单位时,Y 值随之也相应地增加,这种相关关系称为正相关。当 X 增加时,Y 值随之减少,这种相关关系称为负相关。

(3)按变量之间的相关程度,分为完全相关、不完全相关和不相关。

在统计中采用相关系数(r)这一指标来反映相关关系的密切程度。以直线相关为例,如果因变量完全随着自变量而变动,在散点图上可以看出所有的观测点都位于同一条直线上,这是的相关关系就转化为函数关系,称为完全相关,这时$|r|=1$。当因变量完全不随自变量的变动而作相应的变动,即变量之间完全不存在任何依存关系,就称为不相关或零相关,这时$|r|=0$。以上两种是极端情况,大多数相关关系介于其间,即$0<|r|<1$,称为不完全相关。

二、相关关系的描述与度量

为了形象地描述两个变量之间的关系,可以从总体中获取两个变量 X、Y 的一组样本数据$(X_i,Y_i)(i=1,2,\cdots,n)$,将两个变量的样本数据作为坐标点画在坐标平面上。由坐标及这些散点构成的二维数据图称为散点图,它可以近似地反映两个变量相互关系的类型、变动方向和密切程度,如图 1-1 所示。

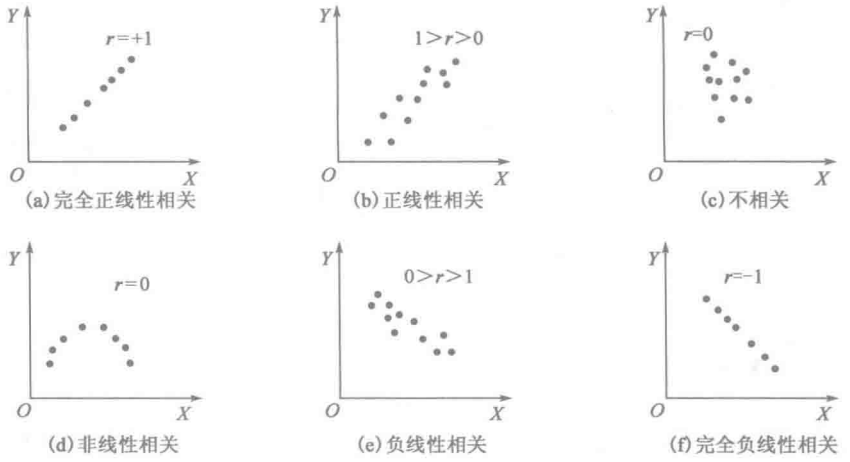


图 1-1 散点图

从图 1-1 可以看出,变量 X 和 Y 的相关关系:完全正线性相关、正线性相关、不相关、非线性相关、负线性相关、完全负线性相关。

三、相关系数的概念及计算

散点图只是对相关关系的初步判断,若要对相关关系进行定量分析,可以计算相关

系数。

根据样本数据计算的对两个变量之间线性相关强度的度量值，称为相关系数。若相关系数是根据总体全部数据计算的，称为总体相关系数，通常用 ρ 表示。总体相关系数的计算公式为：

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\ \sqrt{\mathrm{Var}(Y)}} \tag{1-1}$$

式中　$\mathrm{Cov}(X,Y)$——变量 X 和 Y 的协方差；

$\mathrm{Var}(X)$，$\mathrm{Var}(Y)$——变量 X 和 Y 的方差。

总体相关系数 ρ 是反映两变量之间线性相关程度的一种特定值。对于给定的总体，X 和 Y 的数值是既定的，总体相关系数表现为一个常数。

一般情况下，对总变量 X 和 Y 的全部数据进行观测是不可能的，所以总体相关系数一般是不知道的。通常需要从总体中随机抽取一定数量的样本，通过 X 和 Y 的样本观测值计算样本相关系数来估算总体相关系数。变量 X 和 Y 样本相关系数通常用 r_{XY} 表示，其计算公式为：

$$r_{XY} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \cdot (Y_i - \overline{Y})^2}} \tag{1-2}$$

式中　$\overline{X}$，$\overline{Y}$——X 和 Y 的样本均值。

若令 $L_{YY} = \sum (Y_i - \overline{Y})^2 = \sum Y_i^2 - \frac{1}{n}(\sum Y_i)^2$，则相关系数的公式应该简化为：

$$r_{XY} = \frac{L_{XY}}{\sqrt{L_{XX} L_{YY}}} \tag{1-3}$$

相关系数的特点如下：

(1)相关系数的取值范围在 $-1 \sim +1$ 之间，即 $-1 \leqslant r \leqslant 1$ 。

(2)当 $r>0$ 时，两变量正相关；当 $r<0$ 时，两变量负相关。

(3)当 $0<|r|<1$ 时，两变量存在一定程度的线性相关。$|r|$ 越接近 1，两变量间线性关系越密切；$|r|$ 越接近于 0，两变量的线性相关越弱。

(4)当 $|r|=1$ 时，两变量为完全线性相关，即为函数关系。当 $r=0$ 时，两变量间无线性相关关系。

四、相关系数的显著性检验

通过散点图可以大致判断两类数据是否相关。但是，如果能够求出相关系数时，就能用数字更加准确地说明两类数据的相关性。同一总体的不同样本可以算出不同的相

关系数，到底哪一个更能代表总体的相关程度呢？因此有必要对相关系数进行显著性检验，常用的是相关系数检验法：

设总体的相关系数为 ρ，检验相关系数是否显著，实际上是检验假设 $\rho=0$ 是否成立。抽样的样本容量为 n，给定信度（检验水平）a，根据自由度为 $f=n-2$ 差相关系数得 $r_a(n-2)$，当 $|r|>r_a(n-2)$ 时，拒绝原假设，认为相关系数显著；否则，接受原假设，认为相关系数不显著。

五、利用 Excel 进行相关分析

在 Microsoft Excel 中，可以利用数据分析宏的相关功能进行相关分析并计算出相关系数，具体步骤（以 Office 2010 为例）如下：

（1）在 Excel 中，录入好数据。打开 Excel 菜单“文件”中“选项”对话框，在左侧窗格中单击“加载项”选项，可以看到 Excel 的加载项列表，然后单击“转到”按钮，如图 1-2 所示。

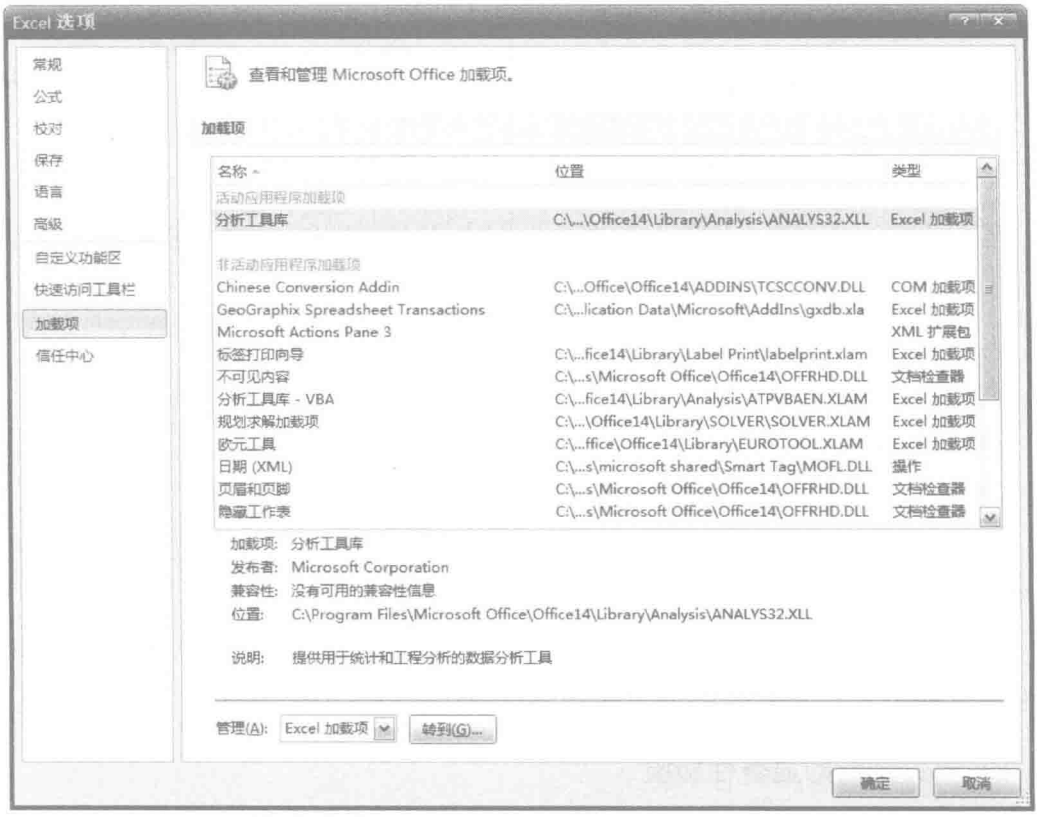


图 1-2　单击“转到”按钮

（2）打开“加载宏”对话框，在“可用加载宏”列表框中勾选“分析工具库”复选框后单

击“确定”按钮，如图 1-3 所示。此时该类加载宏将被添加到 Excel 2010 功能中。

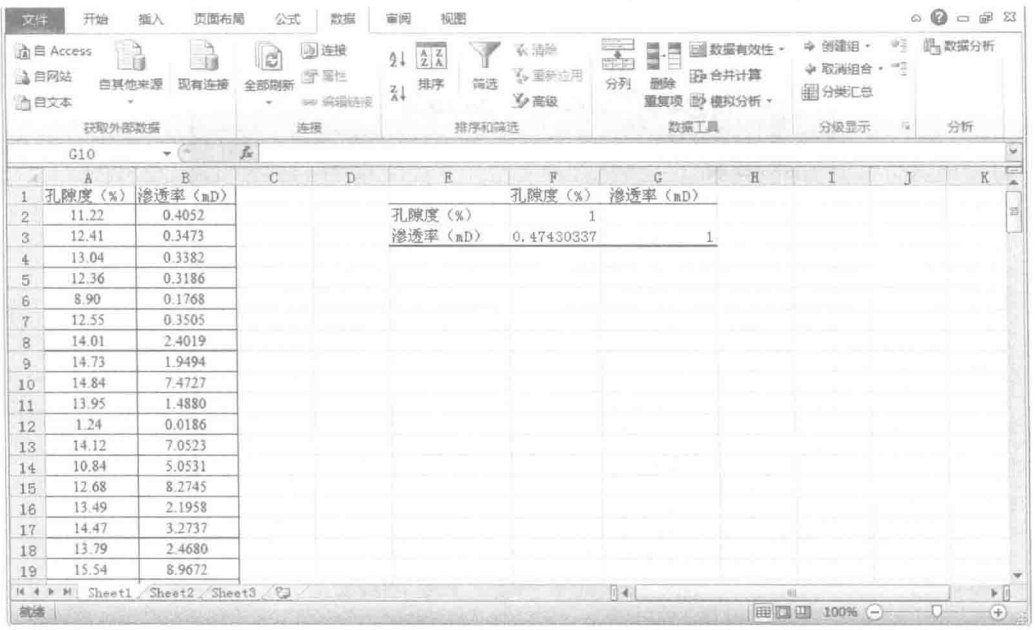
图 1-3　选择“分析工具库”加载宏

(3)选择数据，点击“数据分析”宏中的“相关系数”。

(4)点击“确定”后，在“输入区域”中输入数据所在的单元区域“$A:$B”，选择“标志位于第一行”，输入“输出区域”为$E$1:$I$8 单元格，如图 1-4 所示。

图 1-4　选择输入和输出区域

(5)点击“确定”按钮,就可得到相关系数的分析结果,如图 1-5 所示。本例所得相关系数 $r=0.4743$。



	A	B
1	孔隙度（%）	渗透率（mD）
2	11.22	0.4052
3	12.41	0.3473
4	13.04	0.3382
5	12.36	0.3186
6	8.90	0.1768
7	12.55	0.3505
8	14.01	2.4019
9	14.73	1.9494
10	14.84	7.4727
11	13.95	1.4880
12	1.24	0.0186
13	14.12	7.0523
14	10.84	5.0531
15	12.68	8.2745
16	13.49	2.1958
17	14.47	3.2737
18	13.79	2.4680
19	15.54	8.9672

	F	G
	孔隙度（%）	渗透率（mD）
孔隙度（%）	1	
渗透率（mD）	0.47430337	1

图 1-5　相关系数分析的结果

第二节　回归分析

一、回归分析概述

为了说明变量之间的相关关系,可以用相关系数来加以反映。但是,相关系数仅能说明相关关系的方向和紧密程度,而不能说明变量之间因果的数量关系。回归分析就是对具有相关关系的变量之间数量变化的一般关系进行测定,确定一个相关的数学表达式,以便于进行估计或预测的统计方法。

回归分析主要解决以下几个方面的问题:

(1)对于具有相关关系的两个变量,找出两者之间的回归方程。

(2)对回归方程、参数估计进行显著性检验。

(3)利用回归方程进行分析、评价及预测。

回归分析可分为线性回归和非线性回归。对线性回归与非线性回归的区分有两种理解。一是按回归变量本身是否线性,即是否为一次式来划分。例如,$y=\beta_0+\beta_1 x_1+$

$\beta_2 x_2+\beta_3 x_3+\varepsilon$ 为三元线性回归方程，而 $y=\beta_0+\beta_1 x+\beta_2 x^2+\beta_3 x^3+\varepsilon$ 为一元三次非线性回归方程。二是按回归变量的参数即回归系数是否线性来划分。例如，以上两式都是线性方程，因为它们的回归系数 β_1、β_2、β_3 都是线性的(一次式)，而 $y=\beta_0 x^{\beta_1}+\varepsilon$ 是非线性回归，因 y 不是两参数 β_0、β_1 的线性函数，β_0 与 β_1 是用乘法和指数方法连在一起的。在应用研究中，常见到的是按变量是否为一次性来划分线性与非线性回归方程。

在线性回归分析中，一个因变量与一个自变量的回归称一元线性回归，而一个因变量与多个自变量的回归称多元线性回归。

二、一元线性回归

(一) 回归方程

如果随机变量 y 随自变量 x 的变化而变化，且呈简单线性关系，则 y 依 x 变化的规律可用一元线性回归方程表示。由于随机因素的干扰，y 与 x 线性关系中包含随机误差项 ε，即有：

$$y=\beta_0+\beta_1 x+\varepsilon \tag{1-4}$$

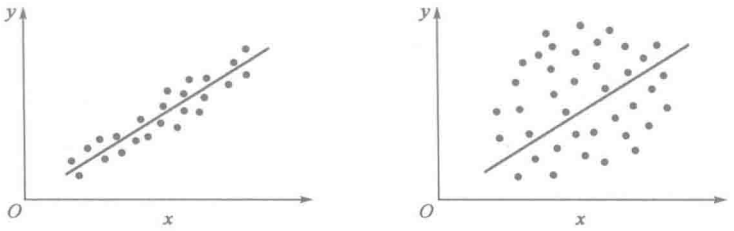
假定 $\varepsilon\sim N(0,\sigma)$，即 $E(\varepsilon)=0$，则对于给定的 x，各次 y 值会有所波动，但平均说来，应有 $E(y)=\beta_0+\beta_1 x$。这就是总体回归直线方程，β_0 为截距，β_1 为回归系数。一般是从总体中抽取部分单位来观察 y 依 x 的变化规律，所以通过样本观测值求出样本回归直线方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x$，用它对总体回归情况进行估计。

(二)估计标准误差

建立了回归方程以后，下一步就是要测定回归估计值的准确性。根据回归方程所计算的多值与实际观察值 y 有一定的离差，这表明用多值来估计 y 存在一定误差，这个误差大小反映了各个散点在回归直线周围的离散程度。二者的差异大，说明散点的离散程度大，回归直线的代表性差，则估计值的准确性小；反之亦然，如图 1-6 所示。所以，测定二者之间的差异大小很重要，它与回归方程的实际应用价值有关，因为要利用回归直线对因变量的未知数值进行推算或预测。

估计标准误差就是用来反映多 $\hat{y}$ 与 y 之间估计误差大小，说明估计值准确程度的统计指标，记为 S_y，意思是各观察值与估计值之间估计误差的平均值：

$$S_y=\sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} \tag{1-5}$$



(a)回归直线的代表性好　　(b)回归直线的代表性差

图 1-6　数据点的分散程度与回归曲线代表性对照

式(1-5)中的 $n-2$ 为自由度，因为 n 个数据点在求得参数估计值 $\hat{\beta}_0$、$\hat{\beta}_1$ 后，受连个正规方程的限制，丧失 2 个自由度，因此用 $n-2$。

为了进一步说明估计标准误差，下面对随机变量 y 的总变差进行分析。

观察值 y 的取值大小是上下波动的，这种波动称为变差，它是由两方面原因引起的：(1)自变量在各次试验中的取值不同；(2)其它因素(包括观察和实验中产生的误差)的影响。对每个观察值来说，变差大小通过与平均数 $\bar{y}$ 的离差表示，而全部 n 次观察值的总变差可由这些离差的平方和 $L_{yy}=\sum(y-\bar{y})^2$ 表示。每个观察值的变差可分解为两部分(图 1-7)，即 $y-\bar{y}=(y-\hat{y})+(\hat{y}-\bar{y})$，所以总变差为：

$$L_{yy}=\sum(y-\bar{y})^2=\sum[(y-\hat{y})+(\hat{y}-\bar{y})]^2$$

$$=\sum(y-\hat{y})^2+\sum(\hat{y}-\bar{y})^2+2\sum(y-\hat{y})(\hat{y}-\bar{y})$$

其中　　$$\sum(y-\hat{y})(\hat{y}-\bar{y})=\sum(y-\hat{\beta}_0-\hat{\beta}_1x)(\hat{\beta}_0+\hat{\beta}_1x-\bar{y})$$

因为　　$$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$$

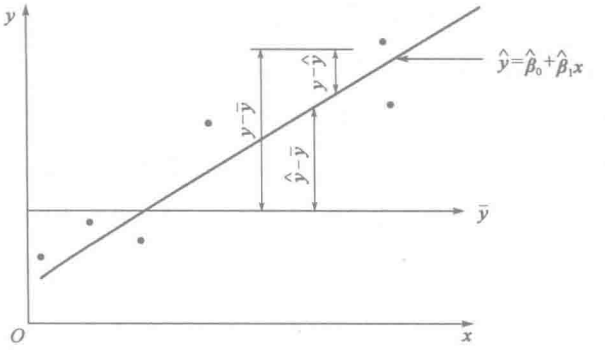


图 1-7　总变差分解图

所以
$$\sum(y-\hat{\beta}_0-\hat{\beta}_1x)(\hat{\beta}_0+\hat{\beta}_1x-\bar{y})$$
$$=\sum(y-\bar{y}+\hat{\beta}_1\bar{x}-\hat{\beta}_1x)(\bar{y}-\hat{\beta}_1\bar{x}+\hat{\beta}_1x-\bar{y})$$
$$=\sum[(y-\bar{y})-\hat{\beta}_1(x-\bar{x})]\hat{\beta}_1(x-\bar{x})$$
$$=\sum[\hat{\beta}_1(x-\bar{x})(y-\bar{y})-\hat{\beta}_1{}^2(x-\bar{x})^2]$$
$$=\hat{\beta}_1\sum[(x-\bar{x})(y-\bar{y})-\hat{\beta}_1(x-\bar{x})^2]$$

因为
$$\hat{\beta}_1=\frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$\hat{\beta}_1\left[\sum(x-\bar{x})(y-\bar{y})-\sum(x-\bar{x})(y-\bar{y})\right]=0$$

即
$$\sum(y-\hat{y})(\hat{y}-\bar{y})=0$$

所以
$$L_{yy}=\sum(y-\hat{y})^2+\sum(\hat{y}-\bar{y})^2 \tag{1-6}$$

式(1-6)说明总变差可分为两部分，其中，$\sum(\hat{y}-\bar{y})^2$ 是估计值 $\hat{y}$ 与平均数 $\bar{y}$ 的离差平方和，根据直线方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x$，可把 $(\hat{y}-\bar{y})$ 看成是 x 的变动所引起的。因此，$\sum(\hat{y}-\bar{y})^2$ 反映了在 y 的总变动中 x 与 y 的直线回归关系而引起的 y 的变化部分，称为回归变差，记作 SSR，即 $SSR=\sum(\hat{y}-\bar{y})^2$。

$\sum(\hat{y}-\bar{y})^2$ 是每个观测点距回归直线的离差平方和，是除 x 对 y 的直线关系以外的其它一些随机因素的影响，使观察值 y 总是围绕回归线上下波动，由此而产生的变差，称为剩余变差，记作 SSE，即 $SSE=\sum(y-\hat{y})^2$。

类似地，总变差 L_{yy} 记作 SST。

总变差、回归变差、剩余变差的关系式可写为：

$$SST=\sum(y-\bar{y})^2=\sum(y-\hat{y})^2+\sum(\hat{y}-\bar{y})^2=SSE+SSR \tag{1-7}$$

回归变差与剩余变差的计算公式为：

$$SSR=\sum(\hat{y}-\bar{y})^2=\sum(\hat{\beta}_0+\hat{\beta}_1x-\hat{\beta}_0-\hat{\beta}_1\bar{x})^2=\hat{\beta}_1^2\sum(x-\bar{x})^2$$

$$=\hat{\beta}_1 \cdot \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \cdot \sum(x-\bar{x})^2$$

$$=\hat{\beta}_1 \cdot \sum(x-\bar{x})(y-\bar{y}) = \hat{\beta}_1 L_{xy} \tag{1-8}$$

可见，有了回归系数 $\hat{\beta}_1$，回归变差就可以通过式(1-8)求得。至于剩余变差，可按下式求得：

$$SSE = \sum(y-\hat{y})^2 = \sum(y-\bar{y})^2 - \sum(\hat{y}-\bar{y})^2 = L_{yy} - \hat{\beta}_1 L_{xy}$$

则

$$S_y = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} = \sqrt{\frac{SSE}{n-2}} \tag{1-9}$$

(三) 回归方程的显著性检验

估计标准误差的大小可以反映回归直线的精确度，即 x 与 y 之间的线性相关程度。但判断估计标准误差的大小要有一个基值，即当估计标准误差为多少时可以认为回归方程的相关显著，回归直线具有代表性。

对回归方程进行显著性检验应用方差分析的原理和方法，即在满足条件的情况下进行 F 检验，检验假设 $H_0:\beta_1=0$ 是否成立。在 H_0 成立时，构造检验统计量：

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F(1,n-2) \tag{1-10}$$

显然，F 值越大，x 和 y 的线性相关关系越显著。根据给定的显著性水平 α 和自由度 1、$n-2$，查附表 1，得临界值 $F_\alpha(1,n-2)$，将根据样本数据计算得到的 F 值与临界值比较，如果 $F \geqslant F_\alpha(1,n-2)$，$F$ 值落入拒绝域，则否定原假设 $H_0:\beta_1=0$，即认为 x 和 y 间存在显著的线性相关关系；否则，接受 H_0，即没有理由认为 x 和 y 之间存在显著的线性相关关系。

下面介绍回归系数的显著性检验。

回归方程的显著性检验是通过变差分解，利用回归变差 SSR 和剩余变差 SSE 构建 F 统计量对 $H_0:\beta_1=0$ 是否成立进行检验；如果直接检验自变量 x 对因变量 y 的影响是否显著，即对回归系数 β_1 的显著性进行检验，则需要用用 t 统计量。

根据回归系数的计算公式：

$$\hat{\beta}_1 = \frac{L_{xy}}{L_{xx}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum(x-\bar{x})y-\bar{y}\sum(x-\bar{x})}{\sum(x-\bar{x})^2}$$

由于 $\sum(x-\bar{x})=0$ 则：

$$\hat{\beta}_1=\frac{\sum(x-\bar{x})y}{\sum(x-\bar{x})^2}=\sum\frac{x-\bar{x}}{\sum(x-\bar{x})^2}y=\sum\omega y \tag{1-11}$$

因此，$\hat{\beta}_1$ 是 y 的线性组合。在一元回归模型 $y=\beta_0+\beta_1x+\varepsilon$ 满足 $\varepsilon\sim N(0,\sigma^2)$ 、x 为确定性变量的情况下，y 服从正态分布，所以 $\hat{\beta}_1$ 也服从正态分布：

$$\hat{\beta}_1\sim N(\hat{\beta}_1,S_{\hat{\beta}_1}^2)$$

其中 $S_{\hat{\beta}_1}^2=\frac{\sigma^2}{L_{xx}}$（证明略）。由于 σ^2 是未知的总体参数，应该时需要用它的优良估计量：

$$\hat{\sigma}^2=\frac{\sum(y-\hat{y})^2}{n-2}=S_y^2$$

来代替，因此可以构造 t 统计量：

$$t=\frac{\hat{\beta}_1-\beta_1}{S_{\hat{\beta}_1}}\sim t(n-2) \tag{1-12}$$

来检验原假设 $H_0:\beta_1=0$ 是否成立。

给定显著性水平 a，以 a 和自由度 $n-2$ 查附表 2，得到临界值 $t_{a/2}(n-2)$。计算样本 t 值，如果 $|t|\geqslant t_{a/2}(n-2)$，拒绝原假设 $H_0:\beta_1=0$，此时认为 β_1 显著不为零，即变量 x 对 y 的影响是显著的，如果 $|t|\leqslant t_{a/2}(n-2)$，则没有足够的理由拒绝原假设，此时认为 β_1 显著为零，变量 x 对 y 的影响不显著。

（四）利用回归方程进行预测与控制

如果回归方程显著性高，则可利用它对变量 y 作预测或对自变量 x 进行控制。预测就是根据自变量 x 的某一已知值 x_0，估计因变量 y 的相应值 y_0 的可能范围。

对于任一给定的 x_0，根据样本建立的回归方程 $\hat{y}=\hat{\beta}_0+\hat{\beta}_1x$ 所得到的 $\hat{y}_0$，可以作为 y_0 的一个点估计值，由于不同的样本会得到不同的 $\hat{\beta}_0$、$\hat{\beta}_1$，$\hat{y}_0$ 与 y_0 之间总存在一定的抽样误差，因此，在对 y_0 的实际值进行预测时，通常是在一定的置信水平 $1-a$ 下，给出 y_0 的预测区间（或置信区间）：

$$\hat{y}_0-t_{a/2}(n-2)S_y\sqrt{1+\frac{1}{n}+\frac{x_0-\bar{x}}{\sum(x-\bar{x})^2}}\leqslant y_0\leqslant$$

$$\hat{y}_0+t_{a/2}(n-2)S_y\sqrt{1+\frac{1}{n}+\frac{x_0-\bar{x}}{\sum(x-\bar{x})^2}}$$

$t_{1-a/2}$ 查自由度为 $n-2$ 的 t 分布临界值，区间范围如图 1-8 所示。