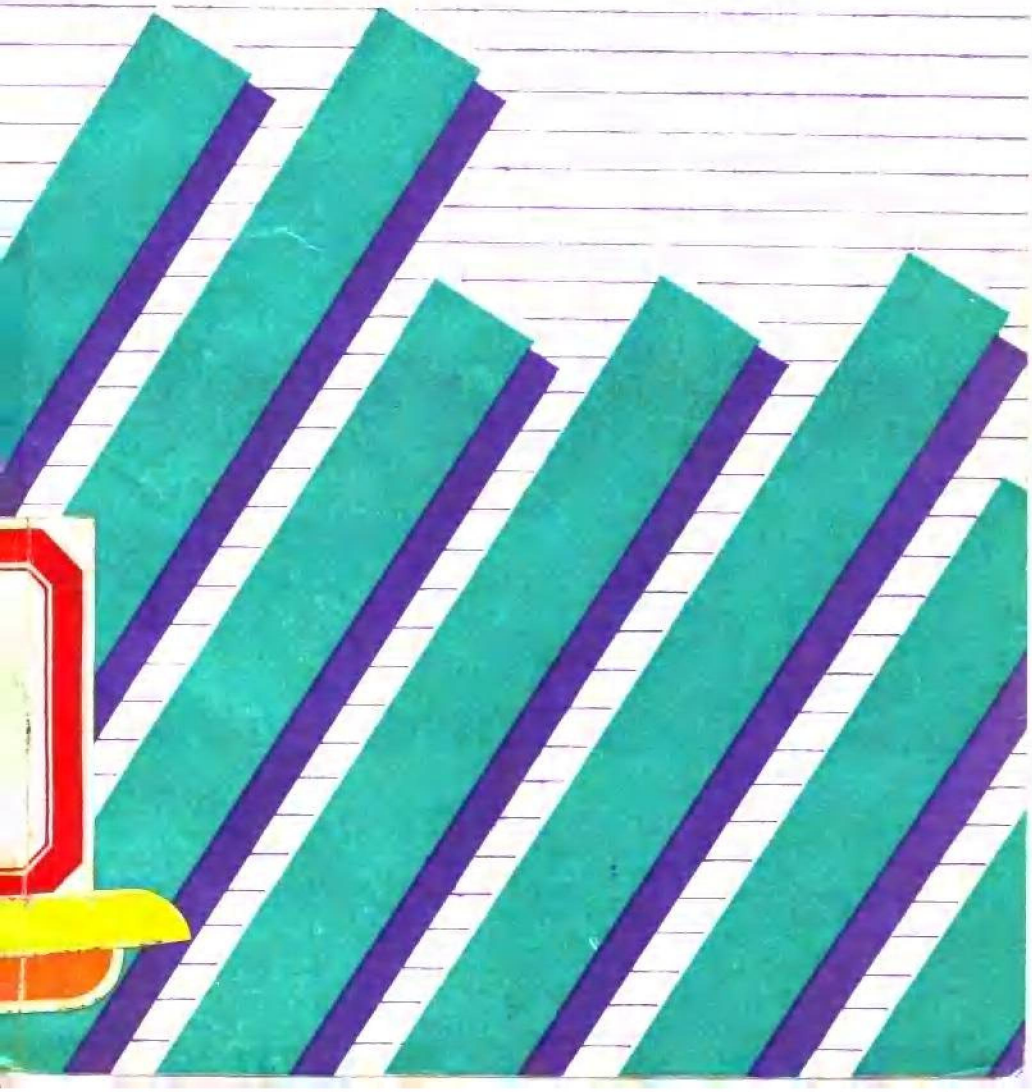
易丹辉　编著

统计预测
——方法与应用



统计预测
——方法与应用

易丹辉　编著

中国人民大学出版社

统计预测——方法与应用

易丹辉　编著

*

中国人民大学出版社出版发行

（北京西郊海淀路39号）

中国人民大学出版社印刷厂印刷

（北京鼓楼西大石桥胡同61号）

新 华 书 店 经 销

*

开本：850×1168毫米32开　印张：10.25

1990年11月第1版　1990年11月第1次印刷

字数：257 000　册数：1-3 500

*

ISBN 7-300-00870-4

F·257　定价：4.60元

前　言

在社会经济活动中，无论从宏观的角度还是从微观的角度，都存在着许多未知的因素，影响着各级的管理决策。为了克服未知因素可能带来的消极后果，必须进行有科学根据的预测。所谓预测，是人们在观察和分析客观事物发展过程的历史及现状的基础上，通过对客观事物发展规律的认识，进而推断其未来状况的过程。为了收到预期的预测效果，对于预测对象最好提出几种不同的预测方案，在各种方案中，充分衡量预测对象变化的条件以及可能变化的幅度，相应地采取有关措施，以便保持最佳的管理过程。换句话说，预测是在制定切实可行的计划时，为了避免可能产生的缺点和失误，而对事物的未来发展预先进行的多种方案的设计和研究。

预测可以按不同的标准进行分类。预测方法基本上分为两大类，即定性分析法和定量分析法。本书比较详尽地介绍了用于预测的定量分析方法：因果回归分析法和时间序列分析法。为了将每种具体方法与我国的社会经济实际相结合，在每一方法介绍之后，都配有实例说明其应用，书中所有计算均应用电子计算机完成。为帮助读者掌握和运用各种方法，特别是无法进行手工计算的方法，书后附有TSP软件的使用说明，它适用于IBM—PC机以及与它兼容的微型机，如长城0520。介绍方法时，涉及到的比较复杂的数学公式推导和证明，均列入各章附录中，供读者参考。

本书编写的过程中，得到中国人民大学计划统计学院计算机

室刘延军、陈虹同志，计划经济学系成晓梅同志以及校信息中心的同志们的帮助与支持。书中采用的某些实例，是我系卫袁同志在硕士研究生学习期间收集的资料，他为编写此书提出了不少建议。在此一并表示衷心的感谢。

本书试图将各种预测方法与我国的实际结合运用，由于水平有限，编写时间又较仓促，一定存在不少缺点，殷切期望读者们随时给予批评指教。

1988年2月

目　录

第一章　简单回归分析法

客观事物之间常存在着某种因果关系，如工业产品成本的降低常导致利润的上升；某种消费品价格的提高往往造成销售量的下降，等等。这种因果关系往往无法用精确的数学表达式来描述，只有通过对大量观察数据的统计处理，才能找到它们之间的关系和规律。回归分析就是通过对观察数据的统计分析和处理，研究与确定事物间相关关系和联系形式的方法。运用回归分析法寻找预测对象与影响因素之间的因果关系，建立回归模型进行预测的方法，称为因果回归分析法。其特点是，将影响预测对象的因素分解，在考察各个因素的变动中，估计预测对象未来的数量状况。它建立的是预测对象与影响因素之间的单一方程，因此也被称为单方程模型分析。按方程中影响预测对象因素的多少。分为简单回归分析法和多重回归分析法。

回归分析法在预测中主要用以解决下面的问题：

（1）分析所获得的统计数据，确定几个特定变量之间的数学关系形式，即建立回归模型。

（2）对回归模型的参数进行估计和统计检验，分析影响因素对预测对象的影响程度，确定预测模型。

（3）利用确定的回归模型和自变量的未来可能值，估计预测对象的未来可能值，并分析研究预测结果的误差范围及精度。

第一节　模型和参数估计

如果影响预测对象的主要因素只有一个，并且它们之间呈线性关系，那么可以采用简单回归分析法预测。由于这种方法只涉及一个自变量，也称为一元线性回归分析法。

1.1.1.理论回归模型

将预测对象作为因变量Y，主要影响因素为自变量X，它们之间的线性关系，从理论上说，能够表述为下面的形式

$$Y=\alpha+\beta X+\varepsilon \tag{1.1}$$

式中：α和β是固定的但未知的参数，它们反映了变量X与Y之间应该有的一种线性关系；α是常数项，β是理论回归系数；ε是那些除X以外，被忽略和（或）无法考虑到的因素，被称为随机项。对于每一组可以观察到的因变量、自变量数值Y_i、X_i，（1.1）式可以写成

$$Y_i=\alpha+\beta X_i+\varepsilon_i \tag{1.2}$$

式中：ε_i满足

$$E(\varepsilon_i)=0$$

$$\mathrm{Cov}(\varepsilon_i,\ \varepsilon_j)=\begin{cases}\sigma^2 & i=j\\ 0 & i\neq j\end{cases} \tag{1.3}$$

1.1.2.实际回归模型

要得到（1.1）式中参数α和β的精确值几乎不可能，因为通常只有有限的样本数据和情报。利用有限的资料，只能得到参数α和β的估计值a和b。实际上，因变量Y和自变量X之间的简单线性关系能够表述为

$$Y=a+bX+e \tag{1.4}$$

这里，a、b不是象α、β那样固定的数值，而是能够取多个数值的统计估计值；e是残差项，也被称为回归余项，它是由于

用$a+bX$估计因变量Y的数值所造成的，是估计值与实际数值之间的离差。

相对于（1.2）式，实际回归模型也可以写成

$$Y_i = a + bX_i + e_i \qquad (1.5)$$

这里，e_i是$a+bX_i$的估计值$\hat{Y}_i$与实际观察值Y_i的离差，即$e_i = Y_i - \hat{Y}_i$。

1.1.3.预测模型

实际预测时，残差项e_i是无法预测的，我们的目的是借助$a+bX$得到预测对象Y的估计值，所以预测模型为

$$\hat{Y} = a + bX \qquad (1.6)$$

式中：a为回归常数，是回归直线的高度。其实际含义为，若在某一刻不考虑自变量时，因变量所能达到的数值。b为回归系数，是回归直线的斜率。其实际含义为，当自变量X每变动一个单位时，因变量Y的平均变动量。

可以看出，（1.6）式$\hat{Y} = a + bX$实际上是（1.1）式$Y = \alpha + \beta X + e$的系统部分$Y = \alpha + \beta X$的一个估计，也是（1.4）式$Y = a + bX + e$的主体部分。相对于（1.5）式，（1.6）式也可以写成

$$\hat{Y}_i = a + bX_i \qquad (1.7)$$

1.1.4.参数估计

对（1.1）式中α、β进行估计，依照不同的准则，采用不同的统计方法，可以得到不同的数值，因而（1.4）式中的a、b不是唯一确定的.预测中,通常采用最小平方法〔Least Sguares〕，亦称最小二乘法。其准则是，选择的参数a、b要使因变量Y的观察值Y_i与估计值$\hat{Y}_i$之间的离差平方和最小，即$\min \Sigma (Y_i - \hat{Y}_i)^2 = \min \Sigma e_i^2$。①

采用最小二乘法得到a、b的计算公式为

① 有关参数a、b的确定参见本章附录1—A。

$$\begin{cases} b = \dfrac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2} \\ a = \bar{Y} - b\bar{X} \end{cases} \tag{1.8}$$

式中：X_i为自变量X的第i个观察值；Y_i为因变量Y的第i个观察值；n为观察值的个数亦称样本数据个数；$\bar{X}$为n个自变量观察值的平均数；$\bar{Y}$为n个因变量观察值的平均数。

（1.8）式还可以写成

$$\begin{cases} b = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})\ (Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \\ a = \bar{Y} - b\bar{X} \end{cases} \tag{1.9}$$

【例1.1】根据表1.1的数据，分析可能建立的预测国民收入中消费额的模型。

分析：设国民收入总额与国民收入中消费额分别作为自变量X和因变量Y。将表1.1中有关数据绘制成图（见图1.1）。从图中可以看出，国民收入与消费之间大致是线性关系，计算它们之间的简单相关系$r_{x,y} = 0.9954$这说明，X与Y之间，可以建立线性回归模型

$$\hat{Y} = a + bX$$

根据表1.1中的数据，采用（1.9）式，得到

$$b = 0.6835 \qquad a = 31.5327$$

我国国民收入中消费额的预测模型可以是

$$\hat{Y} = 31.5327 + 0.6835X$$

表1.1 国民收入总额与消费额

单位：亿元

年份	国民收入总额 (X)	国民收入消费额 (Y)	年份	国民收入总额 (X)	国民收入消费额 (Y)
1952	589	477	1969	1 617	180
1953	709	559	1970	1 926	1 258
1954	748	570	1971	2 077	1 324
1955	788	622	1972	2 136	1 404
1956	882	671	1973	2 318	1 511
1957	908	702	1974	2 348	1 550
1958	1 118	738	1975	2 503	1 621
1959	1 222	716	1976	2 427	1 676
1960	1 220	763	1977	2 644	1 741
1961	996	818	1978	3 010	1 888
1962	924	849	1979	3 350	2 195
1963	1 000	864	1980	3 688	2 531
1964	1 166	921	1981	3 940	2 799
1965	1 387	982	1982	4 261	3 054
1966	1 586	1 065	1983	4 73	3 358
1967	1 487	1 124	1984	5 630	3 895
1968	1 415	1 111	1985	6 822	4 820

资料来源：《中国统计年鉴（1986）》，中国统计出版社1986年版。

这个模型表明，在1952—1981年30年间，我国每增加1元的国民收入，平均就有0.68元用于消费。这就是回归系数b在这里所提供的经济意义。

模型中的参数，对不同的预测对象有不同的含义。参数估计值的符号和大小，要符合它的实际意义。例1.1中当国民收入增加时，消费额一般也有增加，因此b必须大于0，若得到的估计值b小于0，则模型应否定。b的变动范围是否适当，主要根据预测人员的经验确定。参数估计值的符号和大小不符合其实际含

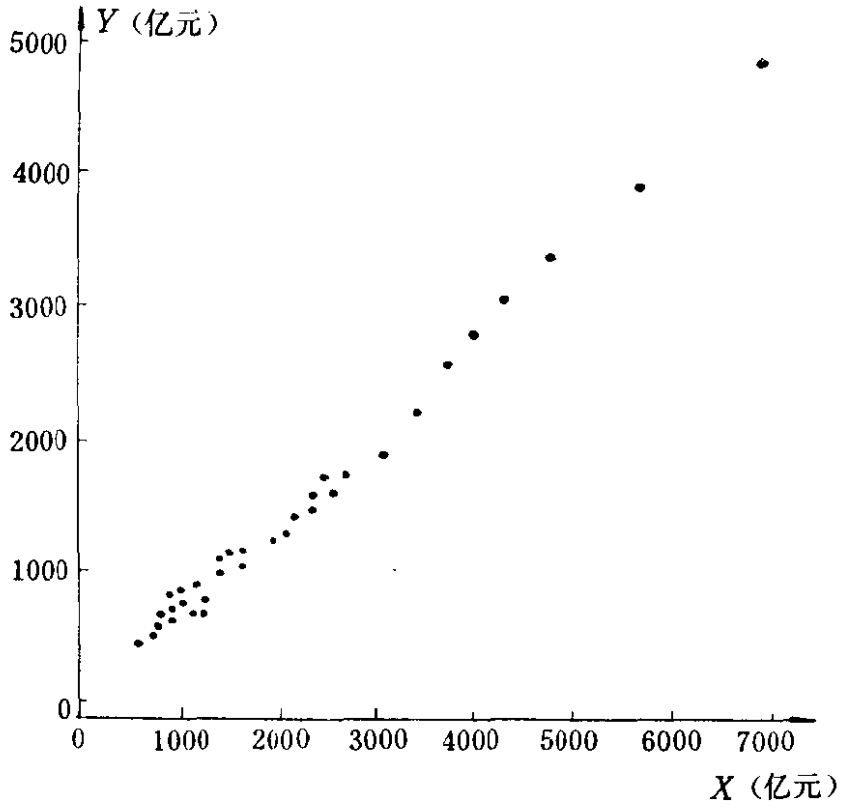


图1.1 国民收入总额与消费额

义，其主要原因可能是：所选用的模型不能代表变量之间的关系；统计数据不足或口径不一致；违反了最小二乘法的某些假定。

预测模型中的回归系数b，反映了因变量Y和自变量X之间的一种变动结构关系。这种变动结构对未来是否合适，决定着模型能否用于预测。这一点是预测时应该予以考虑的。

第二节 模型的检验

数理统计理论证明，采用最小二乘法得到的估计值a、b是α、β的最小方差无偏估计①，它们是较为理想和实用的估计

① 有关证明参看参考书目[10]。

值。在这一过程中，实际上我们是承认了几点假设：

（1）变量X与Y之间为线性关系；

（2）回归余项线性独立，即$E(e_i e_j)=0 \quad (i \neq j)$；

（3）回归余项服从正态分布，即$e_i \sim N(0, \sigma^2)$。

当利用变量的样本数据（实际观察值）建立起预测模型后，需要判断我们所做的各种假设的合理性以及模型的优劣。模型检验就是利用各种统计检验来判别模型的适用性。

1.2.1.回归系数的显著性检验

对于预测模型$\hat{Y}=a+bX$，变量X、Y之间的线性假设是否合理，可以通过回归系数的显著性检验得到判别。回归系数的显著性检验由于要用参数的t值，因而也称为参数的t检验。

检验假设

$$H_0: b=0 \text{①}$$

计算参数b的t值

$$t_b=\frac{b}{S_b} \tag{1.10}$$

式中：S_b是参数b的标准差，$S_b=S_y/\sqrt{\sum(x-\bar{x})^2}$②，这样，（1.10）式可以写成

$$t_b=(b\cdot\sqrt{\sum(x-\bar{x})^2})/S_y \tag{1.11}$$

式中：S_y为回归标准差，$S_y^2=\sum(Y-\hat{Y})^2/n-2$，$n$是样本数据个数，2是参数个数。

$t_b=b/S_b$服从t分布，即$t_b \sim t(n-2)$，因此，可以通过t分布

① 检验回归常数a是否为0的意义不大，故通常只检验参数b。

② 为方便，本书中凡有求和符号$\sum$若无注明均表示$\sum_{i=1}^{n}$。

表查得显著性水平为a①，自由度为$n-2$的数值t_c。将t_b与t_c比较，可决定是接受还是否定H_0假设。

若　$|t_b|>t_c$

可以否定H_0。它表明回归系数显著不为0，参数的t检验通过。回归系数显著，说明变量X与Y之间的线性假设合理，这意味着，所选择的自变量能比较有效地解释预测对象的变化。

若　$|t_b|\leqslant t_c$

则接受H_0，它表明回归系数为0的可能性较大，参数的检t验未通过。回归系数不显著，说明对于变量X与Y之间的线性假设不合理，意味着模型中的自变量无法较好地解释预测对象的变化，应重新考虑。

【例1.2】 国民收入中消费额的预测模型的t检验（续〔例1.1〕）。

分析：例1.1中建立的回归模型为

$$\hat{Y}=31.5327+0.6835X$$

对这个模型的回归系数进行显著性检验。

$$t_b=59.0858$$

查t分布表，$a=0.05$，自由度$df=34-2=32$，得

$$t_c=2.0369$$

显然$t_b=59.0858>t_c=2.0369$

因此，参数的t检验通过。这说明国民收入总额对消费额有显著影响。

① 预测时，只需检验b是否为0，故为双侧假设检验。查t分布表时，则应以显著水平为$\frac{a}{2}$查找。有时t_c也写成$t_a(n-2)$，或$t_{\frac{a}{2}}(n-2)$。

1.2.2.回归方程的显著性检验

参数的t检验，考察的是自变量X与因变量Y之间线性假设的合理性。但预测模型$\hat{Y}=a+bX$作为一个整体，在一定程度上也反映了变量X与Y之间的统计线性关系，其是否适用于预测，仍需检验。回归方程的显著性检验，是利用方差分析所提供的F统计量，检验预测模型的总体线性关系的显著性，也被称为方程的F检验。

检验假设

$$H_0:\ b=0$$

计算回归方程的F值①

$$F(1,n-2)=\frac{\sum(\hat{Y}-\bar{Y})^2/1}{\sum(Y-\hat{Y})^2/n-2} \qquad (1.12)$$

F $(1,\ n-2)$ 服从F分布，即 $F(1,n-2)\sim F_\alpha(1,n-2)$。在$F$分布表中，查找显著性水平为$\alpha$，自由度$n_1=1,n_2=n-2$的$F$值$F_\alpha(1,\ n-2)$。将$F$ $(1,\ n-2)$与$F_\alpha(1,n-2)$比较，能够判定接受H_0还是否定H_0。

若 $F(1,n-2)>F_\alpha(1,n-2)$

否定H_0，回归方程较好地反映了变量X与Y之间的线性关系，回归效果显著，方程的F检验通过。这意味着，预测模型从整体上看适用。若

$$F(1,n-2)\leqslant F_\alpha(1,n-2)$$

接受H_0，回归方程不能很好地反映变量X与Y之间的关系，回归效果不显著，方程的F检验未通过。这意味着，预测模型不能被采用。

【例1.3】 国民收入中消费额的预测模型的F检验（续〔例1.1〕）。

---

① 有关F检验的详细内容参见本章附录1—B。