项目二：相关与回归分析

一、实验目的

1、掌握简单相关分析方法，并根据相关系数判断两变量的相关关系。

2、掌握回归分析方法，并对回归结果进行分析。

二、实验内容和要求

现有杭州市区1978－2002年的GDP、居民年人均可支配收入和年人均消费支出的数据资料，如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 年份 | GDP(万元） | 年人均可支配收入(元) | 年人均消费支出(元) |
| 1978 | 141995 | 338 | 301 |
| 1979 | 167206 | 396 | 365 |
| 1980 | 208220 | 521 | 491 |
| 1981 | 229243 | 540 | 513 |
| 1982 | 248297 | 532 | 532 |
| 1983 | 282171 | 578 | 535 |
| 1984 | 353781 | 729 | 679 |
| 1985 | 448574 | 1026 | 908 |
| 1986 | 513639 | 1169 | 1072 |
| 1987 | 605234 | 1260 | 1118 |
| 1988 | 708474 | 1565 | 1515 |
| 1989 | 772208 | 1764 | 1615 |
| 1990 | 896496 | 1985 | 1685 |
| 1991 | 1092037 | 2128 | 1894 |
| 1992 | 1400617 | 2580 | 2296 |
| 1993 | 2057297 | 3525 | 3183 |
| 1994 | 2731046 | 5249 | 4559 |
| 1995 | 3600744 | 6301 | 5559 |
| 1996 | 4565711 | 7206 | 6095 |
| 1997 | 5238807 | 7896 | 6766 |
| 1998 | 5687426 | 8465 | 7235 |
| 1999 | 6049378 | 9085 | 7424 |
| 2000 | 6778498 | 9668 | 7790 |
| 2001 | 11951694 | 10896 | 8968 |
| 2002 | 13668231 | 11778 | 9215 |

要求：

1、分别求人均可支配收入与GDP、人均消费性支出与GDP、人均可支配收入与人均消费支出的相关系数。

2、画出人均可支配收入与人均消费支出的散点图，求人均消费支出倚人均可支配收入的直线回归方程，解释方程结果，并找出方程的估计标准误差。

3、画出GDP与人均可支配收入的散点图，求人均可支配收入倚GDP的直线回归方程。解释方程结果，并找出方程的估计标准误差。

4、画出GDP与人均消费支出的散点图，求人均消费支出倚GDP的直线回归方程。解释方程结果，并找出方程的估计标准误差。

5、若将GDP的单位改为亿元，再做第3和第4题，观察单位变化对回归方程的影响。

6、求人均可支配收入倚GDP的二次回归方程，并与直线回归方程比较，选出最适合的方程。

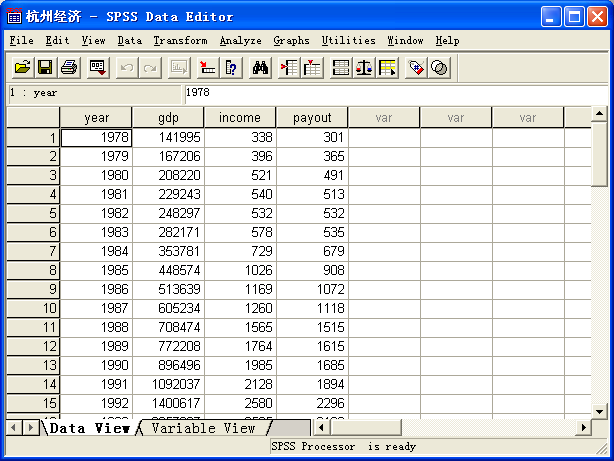
7、求人均消费支出倚GDP的二次回归方程，并与直线回归方程比较，选出最适合的方程。

\*8、求人均可支配收入对GDP的弹性系数和人均消费支出对GDP的弹性系数（提示：一个变量Y对另一个变量X的弹性系数E定义为：E=Y的增长率÷X的增长率）。

三、实验步骤

建立数据文件，本数据中包括变量：year年份、gdp国内生产总值（万元）、income人均可支配收入（元）、payout人均消费支出（元）。如图2－1所示。

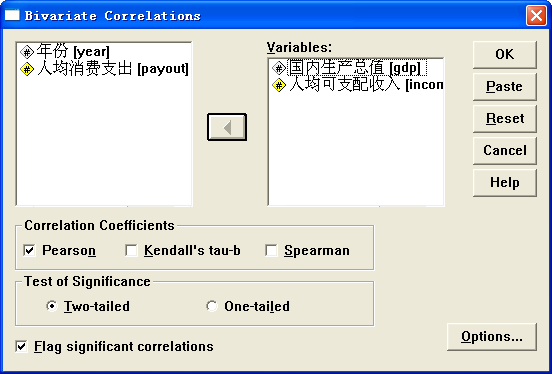
图2－1：杭州市区GDP、居民年人均可支配收入和年人均消费支出数据



1、计算相关系数。

按Analyze→Correlate→Bivariate顺序单击菜单项，弹出简单相关分析主对话框，如图2－2所示。从主对话框左边的变量表将变量income和gdp选入右边的Variables框，对人均可支配收入和GDP进行相关分析。

图2－2：简单相关分析主对话框

点击OK，输出结果见表2－1。

****

由表2－1可知，在99％的置信度下，GDP与人均可支配收入的相关系数为0.944，属于高度相关。

按上述方法也可求出人均消费性支出与GDP、人均可支配收入与人均消费支出的相关系数。结果如表2－2和表2－3所示。

****

****

由表2－2和表2－3可知，在99％的置信度下，人均消费性支出与GDP的相关系数为0.931，人均可支配收入与人均消费支出的相关系数为0.999，均属于高度相关。

2、求直线回归方程。

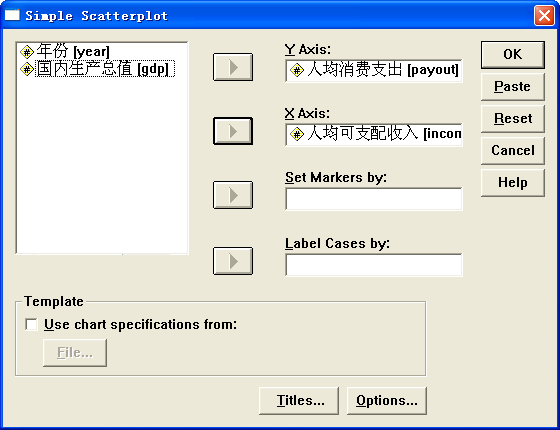
（1）首先做散点图，按Graphs→Scatter→Simple顺序展开如图2－3所示的对话框。将变量payout选入Y Axis作为y轴，变量income选入X Axis作为x轴。

图2－3：散点图对话框

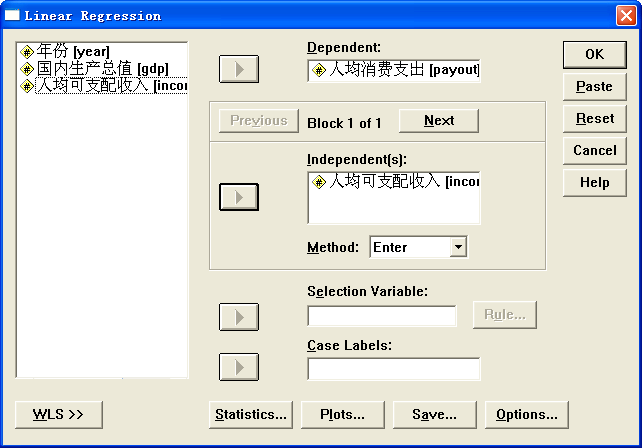
点击OK，生成的图形见图2－4，其中Y轴为人均消费支出（payout），X轴为人均可支配收入（income）。



图2－4：人均消费支出与人均可支配收入的散点图

由图2－4可以看出人均消费支出与人均可支配收入接近于直线关系，可以做直线回归方程。

（2）做直线回归，按Analyze→Regression→Linear顺序单击菜单项，弹出直线回归分析主对话框，如图2－5所示。选择人均消费支出payout作为因变量选入右边的Dependent框中，选择人均可支配收入income作为自变量进入Independent(s)框中。

图2－5：线性回归分析主对话框

点击OK，输出结果见表2－4和表2－5。

****

****

从表2－4可看出，判定系数R2为0.997，修正R2为0.997，估计标准误差值为162.89。从表2－5结果可得出线性回归方程为：人均消费支出＝144.717＋0.81×人均可支配收入。说明人均可支配收入每增加一元，人均消费支出平均增加0.81元。人均可支配收入的t值为92.735，且Sig值小于0.01，即在99％的置信度下具有显著的统计意义，说明线性回归方程拟合效果很好。

同理也可求出人均可支配收入与GDP、人均消费性支出与GDP的直线回归方程，结果如表2－6至表2－9所示。

****

****

从表2－6可看出，判定系数R2为0.891，修正R2为0.886，估计标准误差值为1286.22。从表2－7结果可得出线性回归方程为：人均可支配收入＝1139.791＋0.0009757×GDP。说明GDP每增加一万元，人均可支配收入平均增加0.0009757元。

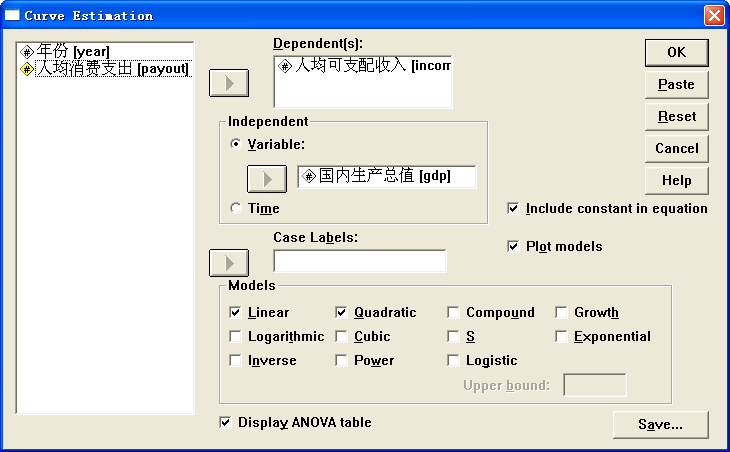
****

****

从表2－8可看出，判定系数R2为0.867，修正R2为0.861，估计标准误差值为1151.11。从表2－9结果可得出线性回归方程为：人均消费支出＝1094.767＋0.0007805×GDP。说明GDP每增加一万元，人均可支配收入平均增加0.0007805元。

3、求曲线回归方程

按Analyze→Regression→Curve Estimation顺序打开曲线回归分析主对话框，如图2－6所示。选择人均可支配收入作为因变量选入右边的Dependent框中，选择GDP作为自变量进入Independent框中。在Models框中选择Quadratic二次方程，并选择Linear直线方程以作比较。

图2－6：曲线回归分析主对话框

点击OK，输出结果见表2－10和图2－7。

表2－10 二次方程拟合结果

① Dependent variable.. INCOME Method.. QUADRATI

Listwise Deletion of Missing Data

② Multiple R .99823

R Square .99646

Adjusted R Square .99614

Standard Error 236.58183

③ Analysis of Variance:

DF Sum of Squares Mean Square

Regression 2 346747212.9 173373606.4

Residuals 22 1231361.1 55971.0

F = 3097.56356 Signif F = .0000

④ -------------------- Variables in the Equation --------------------

Variable B SE B Beta T Sig T

GDP .001949 4.0143E-05 1.884999 48.547 .0000

GDP\*\*2 -8.28611756E-11 3.2307E-12 -.995875 . .

(Constant) 135.732797 71.655655 1.894 .0714

图2－7：人均可支配收入与GDP的直线方程和二次方程拟合图形

表2－10的结果分析：

①回归方程的因变量是人均可支配收入INCOME，使用的方法是QUADRATIC二次方程回归。

②判定系数R2为0.99646，修正R2为0.99614，标准误为236.58183。而前面算出的人均可支配收入与GDP直线方程的判定系数R2为0.891，修正R2为0.886，估计标准误为1286.22。可见二次方程的拟合效果更好。

③ 方差分析的结果：F值为377.20937，显著性水平小于0.0001，方程具有显著统计意义。

④ 得出的二次方程模型为：

INCOME=135.733＋0.001949GDP－(8.286E-11)GDP2

从图2－7中也可看出，二次方程的拟合效果比直线方程要好，综上所述，人均可支配收入与GDP的回归方程应选择二次方程。

按上述方法，亦可求出人均消费支出与GDP的二次方程，结果如表2－11和图2－8所示。

表2－11 二次方程拟合结果

① Dependent variable.. PAYOUT Method.. QUADRATI

Listwise Deletion of Missing Data

② Multiple R .99837

R Square .99675

Adjusted R Square .99645

Standard Error 183.89654

③ Analysis of Variance:

DF Sum of Squares Mean Square

Regression 2 228054853.6 114027426.8

Residuals 22 743994.6 33817.9

F = 3371.80322 Signif F = .0000

④ -------------------- Variables in the Equation --------------------

Variable B SE B Beta T Sig T

GDP .001655 3.1203E-05 1.974142 53.038 .0000

GDP\*\*2 -7.44611260E-11 2.5112E-12 -1.103654 . .

(Constant) 192.495683 55.698390 3.456 .0022

表2－11结果分析：

① 标示出回归方程的因变量是人均消费支出PAYOUT，使用的方法是QUADRATIC二次方程回归。

②判定系数R2为0.99675，修正R2为0.99645，标准误为183.89654。而前面算出的人均消费支出与GDP直线方程的判定系数R2为0.867，修正R2为0.861，估计标准误差值为1151.11。可见二次方程拟合效果更好。

③ 方差分析的结果：F值为377.20937，显著性水平小于0.0001，二次方程具有显著统计意义。

④ 得出二次方程模型为：

PAYOUT=192.496＋0.001655GDP－(7.446E-11)GDP2



图2－8：人均消费支出与GDP的直线方程和二次方程拟合图形

从图2－8可见二次方程的拟合效果比直线方程要好，因此，人均消费支出与GDP的回归方程应选择二次方程。

\*4、求弹性系数。

考虑对数线性模型：

将上式两边对x求导，可得

化简可得：

因为一个变量Y对另一个变量X的弹性系数E定义为：E=Y的增长率/X的增长率，即：



所以对数线性模型的参数可表示Y对X的弹性系数。

按Analyze→Regression→Curve Estimation顺序打开曲线回归分析主对话框，如图2－9所示。选择人均可支配收入和人均消费支出作为因变量选入右边的Dependent框中，选择GDP作为自变量进入Independent框中。在Models框中选择Power对数线性方程，同时做人均可支配收入对GDP和人均消费支出对GDP的对数线性回归。

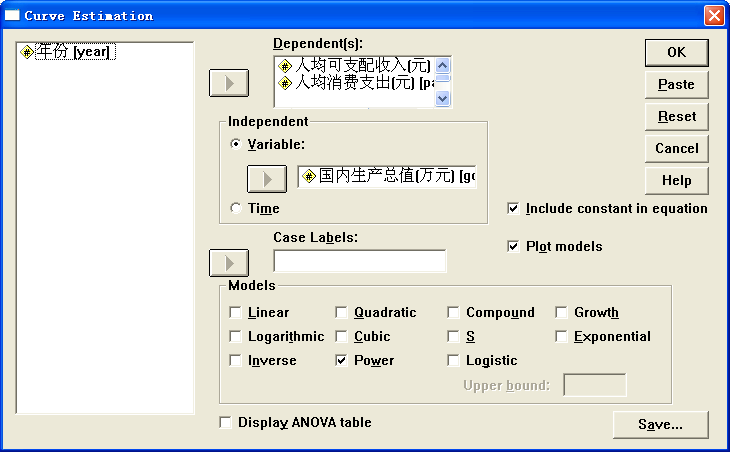


图2－9：曲线回归分析主对话框

点击OK，得出的结果如表2－12所示。

表2－12：对数线性回归结果

Independent: GDP

Dependent Mth Rsq d.f. F Sigf b0 b1

INCOME POW .988 23 1888.64 .000 .0207 .8271

PAYOUT POW .985 23 1489.52 .000 .0295 .7930

由表2－12可知两个方程拟合效果都较好，具有统计显著性。表中的b1即为对数线性方程中的参数。可见人均可支配收入对GDP的弹性系数为0.8271，人均消费支出对GDP的弹性系数为0.7930。也就是说GDP每增长1％，人均可支配收入和人均消费支出分别增长0.8271％和0.7930％。