一元线性回归

一元线性回归

一元线性回归参数估计

一元线性回归可用来分析自变量x取值与因变量Y取值的内在联系，不过这里的自变量x是确定性的变量，因变量Y是随机性的变量。

进行n次独立试验，测得数据如下：

X	x_1	x_2	$\cdots$	x_n
Y	y_1	y_2	$\cdots$	y_n

一元线性回归

力图建立回归方程的估计式或经验回归方程

$$\hat{y} = \hat{\alpha} + \hat{\beta}x,$$

$$\hat{\alpha} = a, \hat{\beta} = b \text{及} \hat{y}_i = a + bx_i \text{使}$$

使用最小二乘法进行参数估计

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2$$

的值最小，所求出的a称为经验截距，简称为截距，b称为经验回归系数，简称为回归系数。

一元线性回归

根据最小二乘法的要求由

$$\frac{\partial Q}{\partial a}=0,\frac{\partial Q}{\partial b}=0,\text{得}$$

$$a=\bar{Y}-b\bar{x},b=\frac{l_{xy}}{l_{xx}},\quad l_{xy}=\sum(x-\bar{x})(y-\bar{y})$$

$$l_{xy}=\sum(x-\bar{x})^{2}$$

一元线性回归

一元回归方程检验

（1）F检验法： $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$,

当H_0为真时，$\frac{SSE}{\sigma^2} \sim \chi^2(1)$;

且SSR与SSE相互独立；因此，当H_0为真时，

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

当$F \geq F_{1-\alpha}(1,n-2)$时应该放弃原假设$H_0$。

一元线性回归

(2) t检验法：

$$\because b \sim N(\beta, \frac{\sigma^2}{l_{xx}}), \frac{SSE}{\sigma^2} \sim \chi^2(n-2),$$

当H_0为真时，

$$t = b\sqrt{\frac{l_{xx}}{SSE/(n-2)}} \sim T(n-2),$$

当$|t| \geq t_{1-0.5\alpha}(n-2)$时应该放弃原假设$H_0$。

一元线性回归

(3) r检验法：

根据x与Y的观测值的相关系数

$$r=\frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}},r^2=\frac{l_{xy}^2}{l_{xx}l_{yy}},$$

可以推出 $$r^2=\frac{SSR}{SST}.$$

当H_0为真时，

$$F=\frac{r^2}{(1-r^2)/(n-2)}\sim F(1,n-2),$$

一元线性回归

当$F \geq F_{1-\alpha}(1,n-2)$或$|r| \geq r_{\alpha}(n-2)$时应该放弃原假设$H_0$，式中的

$$r_{\alpha}(n-2)=\sqrt{\frac{F_{1-\alpha}(1,n-2)}{F_{1-\alpha}(1,n-2)+(n-2)}}$$

可由r检验用表中查出。

$$\because r^2=\frac{SSR}{SST},$$

因此，r常常用来表示x与Y的线性关系在x与Y的全部关系中所占的百分比，又称为x与Y的观测值的决定系数。

一元线性回归

利用回归方程进行点预测和区间预测

若线性回归作显著性检验的结果是放弃H_0，也就是放弃回归系数$\beta=0$的假设，便可以利用回归方程进行点预测和区间预测，这是人们关注线性回归的主要原因之一。

(1) 当$x=x_0$时，用$\hat{y}_0=a+bx_0$预测Y_0的观测值y_0称为点预测。

由于$E(\hat{y}_0)=\alpha+\beta x_0=E(Y_0)$，

Y_0的观测值y_0的点预测是无偏的。

一元线性回归

(2) 当$x=x_0$时，用适合不等式$P\{Y_0\in(G,H)\}\geq 1-\alpha$的统计量G和H所确定的随机区间(G,H)预测$Y_0$的取值范围称为区间预测，而(G,H)称为$Y_0$的$1-\alpha$预测区间。

若Y与样本中的各Y相互独立，则根据$Z=Y_0-(a+bx_0)$服从正态分布，$E(Z)=0$，

$$D(Z)=\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}\right),$$

及$\frac{SSE}{\sigma^2}\sim\chi^2(n-2)$，　Z与SSE相互独立，

一元线性回归

可以导出

$$t=\frac{Z}{\sqrt{\frac{SSE}{n-2}\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}\right)}}\sim t(n-2).$$

因此，Y_0的1-α预测区间为 $a+bx_0\pm\Delta(x_0)$，

$$\Delta(x_0)=t_{1-0.5\alpha}(n-2)\sqrt{\frac{SSE}{n-2}\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{l_{xx}}\right)}.$$

一元线性回归

例 《吸附方程》某种物质在不同温度下可以吸附另一种物质，如果温度x(单位：℃)与吸附重量Y(单位：mg)的观测值如下表所示：

温度x	1.5	1.8	2.4	3.0	3.5	3.9	4.4	4.8	5.0
重量y	4.8	5.7	7.0	8.3	10.9	12.4	13.1	13.6	15.3

试求线性回归方程并用三种方法作显著性检验，若$x_0=2$，求Y_0的0.95预测区间。

解：根据上述观测值得到n＝9，

一元线性回归

```
/*代码以及结果的解释见教材*/
data ex;
input x y@@;
cards;
1.5 4.8 1.8 5.7 2.4 7 3 8.3 3.5 10.9 3.9 12.4 4.4 13.1 4.8
13.6 5 15.3 2
.;
proc gplot;plot y*x;symbol i=rl v=dot;proc
reg;model y=x/cli;
run;
```

一元线性回归

The SAS System 23:52 Tuesday, March 13, 2007

The REG Procedure
Model: MODEL1
Dependent Variable: y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	112.48368	112.48368	387.52	<.0001
Error	7	2.03188	0.29027		
Corrected Total	8	114.51556			

Root MSE	0.53877	R-Square	0.9823
Dependent Mean	10.12222	Adj R-Sq	0.9797
Coeff Var	5.32260		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.25695	0.53235	0.48	0.6441
x	1	2.93028	0.14886	19.69	<.0001

一元线性回归

The REG Procedure
Model: MODEL1
Dependent Variable: y

Output Statistics

Obs	Dep Var y	Predicted Value	Std Error Mean Predict	95% CL Predict		Residual
1	4.8000	4.6524	0.3308	3.1574	6.1474	0.1476
2	5.7000	5.5315	0.2943	4.0797	6.9832	0.1685
3	7.0000	7.2896	0.2301	5.9043	8.6749	-0.2896
4	8.3000	9.0478	0.1877	7.6987	10.3969	-0.7478
5	10.9000	10.5129	0.1807	9.1692	11.8566	0.3871
6	12.4000	11.6850	0.1964	10.3291	13.0410	0.7150
7	13.1000	13.1502	0.2365	11.7589	14.5415	-0.0502
8	13.6000	14.3223	0.2789	12.8878	15.7568	-0.7223
9	15.3000	14.9083	0.3023	13.4476	16.3691	0.3917
10	.	6.1175	0.2714	4.6911	7.5440	.

Sum of Residuals	0
Sum of Squared Residuals	2.03188
Predicted Residual SS (PRESS)	3.13772

一元线性回归