

小 结

本章详细介绍了一元线性回归模型,包含它的未知参数的估计和模型的假设检验。给出了未知参数的最小二乘估计以及极大似然估计,发现它们是一致的;讨论了最小二乘估计的优良性质。模型的假设检验包含 F -检验、 t -检验和相关系数检验,由于它们在一元线性回归模型的假设检验问题上是一致的,使用哪一个都是可以的。本章还讨论了一元线性回归模型的预测问题以及数据变换问题。

习 题 二

- 1. 一元线性回归模型有哪些基本假设?
- 2. 证明: (1) $\text{Var}(\hat{\beta}_0) = \left[\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}} \right] \sigma^2$;
(2) $\text{Var}(e_i) = \left[1 - \frac{1}{n} - \frac{(x - \bar{x})^2}{l_{xx}} \right] \sigma^2$ 。
- 3. 验证三种检验的关系,即验证:
(1) $t = \frac{\hat{\beta}_1 \cdot \sqrt{l_{xx}}}{\hat{\sigma}} = \frac{\sqrt{n-2} \cdot r}{\sqrt{1-r^2}}$;
(2) $F = \frac{\hat{\beta}_1^2 \cdot l_{xx}}{\hat{\sigma}^2} = t^2$ 。
- 4. 随机抽取某地 12 个居民家庭为样本,调查得到有关人均收入与食品支出的资料如下:

单位: 元

编 号	家庭人均 收入 x	人均食品 支出 y	编 号	家庭人均 收入 x	人均食品 支出 y
1	820	750	7	1 600	1 300
2	930	850	8	1 800	1 450
3	1 050	920	9	2 000	1 560
4	1 300	1 050	10	2 700	2 000
5	1 440	1 200	11	3 000	2 000
6	1 500	1 200	12	4 000	2 200

要求:

- (1) 画出散点图。
- (2) x 与 y 之间是否大致呈线性关系?
- (3) 用最小二乘估计求出回归方程。
- (4) 对回归方程作方差分析。

5. 某企业生产某种产品的产量和单位成本资料如下:

月 份	1	2	3	4	5	6
产量 x (千件)	4	6	8	7	8	9
单位成本 y (元/件)	73	72	71	72	70	69

要求:

- (1) 确定单位成本对产量的线性回归模型。
- (2) 对该模型的回归系数进行检验。
- (3) 以 95% 的置信度估计产量为 10 千件时单位成本特定值的置信区间。

6. 在钢线碳含量对于电阻的效应的研究中,得到以下数据:

碳含量(x)	0.10	0.30	0.40	0.55	0.70	0.80	0.95
电阻(y)(微欧)	18	18	19	21	22.6	23.8	26

要求:

- (1) 画出散点图。
- (2) x 与 y 之间是否大致呈线性关系?
- (3) 用最小二乘估计求出回归方程。
- (4) 对回归方程作方差分析。
- (5) 给出 β_1 的置信度为 0.95% 的置信区间。
- (6) 求 $x = 0.50$ 处的置信水平为 0.95 的预测区间。

7. 假定保险公司希望确定居民住宅区火灾造成的损失数额与该住户到最近的消防站之间的距离的关系,以便准确地定出保险金额。保险公司收集了如下数据:

火灾损失与距离消防站距离数据

到消防站的距离(km)	3.0	2.6	4.3	2.1	1.1	6.1	4.8	3.8
火灾损失(千元)	22.3	19.6	31.3	24.0	17.3	43.2	36.4	26.1

(续表)

到消防站的距离(km)	3.4	1.8	4.6	2.3	3.1	5.5	0.7	
火灾损失(千元)	26.2	17.8	31.3	23.1	27.5	36.0	14.1	

要求:

- (1) 画出散点图。
- (2) 建立回归方程。
- (3) 对回归方程作方差分析。
- (4) 给出回归系数的置信度为 95% 的置信区间。
- (5) 求到消防站距离 2.5 km 处的置信水平为 95% 的预测区间。
- (6) 给保险公司提出一个合理的建议。

8. 一家保险公司十分关心其总公司营业部加班的程度,决定认真调查一下现状。经过 10 周时间,收集了每周加班工作时间的数据和签发的新保单数目, x 为每周签发的新保单数目, y 为每周加班工作时间(小时),见下表。

周序号	1	2	3	4	5	6	7	8	9	10
x	825	215	1 070	550	480	920	1 350	325	670	1 215
y	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

要求:

- (1) 画出散点图。
- (2) x 与 y 之间是否大致成线性关系?
- (3) 用最小二乘估计求出回归方程。
- (4) 求回归标准误差 $\bar{\sigma}$ 。
- (5) 给出回归系数的置信度为 95% 的区间估计。
- (6) 计算 x 与 y 的决定系数。
- (7) 对回归方程作方差分析。
- (8) 作回归系数 β_1 显著性的检验。
- (9) 作相关系数的显著性检验。
- (10) 对回归方程作残差图并作相应的分析。
- (11) 该公司预计下一周签发新保单 $x_0 = 1\,000$ 张,需要的加班时间是多少?
- (12) 给出置信水平为 95% 的近似预测区间。

9. 现代投资分析的特征线涉及如下回归方程:

$$r_i = \beta_0 + \beta_1 r_{mt} + \epsilon_i$$

其中, r 表示股票或债券的收益率, r_m 表示有价证券的收益率(用市场指数表示,如标准普尔 500

$$\hat{r}_t = 0.7264 + 1.0598r_{mt} \quad r^2 = 0.4710$$

(0.3001) (0.0728)

(2) 如何解释 r^2 ?

人均国民收入与人均消费金额

年 份	人均国民 收入(元)	人均消费 金额(元)	年 份	人均国民 收入(元)	人均消费 金额(元)
1980	460	234.75	1990	1 634	797.08
1981	489	259.26	1991	1 879	890.66
1982	525	280.58	1992	2 287	1 063.39
1983	580	305.97	1993	2 939	1 323.22
1984	692	347.15	1994	3 923	1 736.32
1985	853	433.53	1995	4 854	2 224.59
1986	956	481.36	1996	5 576	2 627.06
1987	1 104	545.40	1997	6 053	2 819.36
1988	1 355	687.51	1998	6 392	2 958.18
1989	1 512	756.27			

年 份	CPI	S&P500 指数	年 份	CPI	S&P500 指数
1990	130.7	334.59	1994	148.2	460.33
1991	136.2	376.18	1995	152.4	541.64
1992	140.3	415.74	1996	159.6	670.83
1993	144.5	451.41			

44

要求:

- (1) 以 CPI 指数为横轴、S&P 指数为纵轴作图。
- (2) 你认为 CPI 指数与 S&P 指数之间的关系如何?
- (3) 考虑下面的回归模型:

$$(S\&P)_t = \beta_0 + \beta_1 CPI_t + \epsilon_t$$

根据表中的数据,运用最小二乘估计上述方程,并解释你的结果;你的结果有经济意义吗?

12. 下表给出 1988 年 9 个工业国的名义利率(y)与通货膨胀率(x)的数据:

国 家	y(%)	x(%)	国 家	y(%)	x(%)
澳大利亚	11.9	7.7	墨西哥	66.3	51.0
加拿大	9.4	4.0	瑞 典	2.2	2.0
法 国	7.5	3.1	英 国	10.3	6.8
德 国	4.0	1.6	美 国	7.6	4.4
意大利	11.3	4.8			

* 资料来源:原始数据来自国际货币基金组织出版的《国际金融统计》。

要求:

- (1) 以利率为纵轴、通货膨胀率为横轴作图。
- (2) 用最小二乘法进行回归分析,写出求解步骤。
- (3) 如果实际利率不变,名义利率与通货膨胀率的关系如何?

13. 某地区有 10 个百货商店,它们的销售额和流通费率资料如下:

商店编号	销售额 x (百万元)	流通费率 y (%)	商店编号	销售额 x (百万元)	流通费率 y (%)
1	0.7	6.4	6	4.3	1.5
2	1.5	4.5	7	5.5	1.4
3	2.1	2.7	8	6.4	1.3
4	2.9	2.1	9	6.9	1.3
5	3.4	1.8	10	7.8	1.2

要求:

- (1) 试用散点图观察销售额 x 与流通费率 y 的相关关系。
- (2) 拟合双曲线回归模型。
- (3) 检验该模型的显著性,并预计 $x_0 = 900$ 万元时的流通费率。

14. 下表中的数据是 1880—1988 年间的世界原油年产量。

1880—1988 年间石油产量数据 单位：百万桶

年 份	石油产量	年 份	石油产量	年 份	石油产量
1880	30	1940	2 150	1972	18 584
1890	77	1945	2 595	1974	20 389
1900	149	1950	3 803	1976	20 188
1905	215	1955	5 626	1978	21 922
1910	328	1960	7 674	1980	21 722
1915	432	1962	8 882	1982	19 411
1920	689	1964	10 310	1984	19 837
1925	1 069	1966	12 016	1986	20 246
1930	1 412	1968	14 104	1988	21 338
1935	1 655	1970	16 690		

- 利用上面数据讨论下列问题：
- (1) 构造一个石油产量对年份的散点图，并观察该图像中的点散布呈非线性状。为了对这些数据拟合线性回归模型，必须对石油产量数据作变换。
- (2) 构造 $\log(\text{石油产量})$ 对年份的散点图，作线性回归，评价模型的拟合效果。
15. 为研究某都市报开设周日报的可行性，得到 35 种报纸平日和周日的发行量数据(以千为单位)，数据如下表：

报纸序号	平日发行量	周日发行量	报纸序号	平日发行量	周日发行量
1	391.952	488.506	15	1 164.388	1 531.527
2	516.981	798.198	16	444.581	553.479
3	355.628	235.084	17	412.871	685.975
4	238.555	299.451	18	272.280	324.241
5	391.952	488.506	19	781.796	983.240
6	537.780	559.093	20	1 209.225	1 762.015
7	733.775	1 133.249	21	825.512	960.038
8	198.823	348.741	22	223.748	284.611
9	252.624	417.779	23	354.843	407.760
10	206.204	344.552	24	515.523	982.663
11	231.177	323.084	25	220.465	557.000
12	449.775	620.752	26	337.672	440.923
13	288.571	423.305	27	197.120	268.060
14	185.736	202.614	28	133.239	262.048

(续表)

报纸序号	平日发行量	周日发行量	报纸序号	平日发行量	周日发行量
29	374.009	482.052	33	201.860	267.781
30	273.884	338.355	34	321.626	480.343
31	570.364	704.322	35	838.902	1 165.567
32	391.286	585.681			

* 资料来源：Gale Directory of Publications, 1994。

要求：

- (1) 画一个周日发行量关于平日发行量的散点图，观察该散点图中两个变量之间是否具有线性关系。
- (2) 拟合一条回归直线，由平日发行量去预测周日发行量。
- (3) 给出 β_0 和 β_1 的 95% 的置信区间。
- (4) 周日发行量和平日发行量有显著关系吗？使用统计检验来回答这一问题，并写明你所作的假设检验以及结论。
- (5) 某一报纸正考虑推出周日版，平日发行量为 500 000 份，给出该报纸周日发行量的 95% 的置信区间。