

2.2 F 检验与两个 t 检验的等价

F 统计量 (2) 式是 t 检验量的平方,事实上

$$F = \frac{SSR/1}{SSE/(n-2)} = \frac{(n-2) \sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2} = \frac{(n-2) \beta_1^2 L_{xx}}{\sum (y_i - \hat{y}_i)^2} = t^2$$

由此可见,对一元线性回归三种检验实际只需做一种即可,三种检验只是从不同角度做了分析,结果是一致的。

3 决定系数与残差分析的作用

一个线性回归方程通过了 t 检验、F 检验或相关系数的显著性检验中任何一种,就表明变量 x 与 y 之间线性关系是显著的,或说线性回归方程有效,但这并不能保证拟合的效果好,要用这个方程进行分析或预测之前还必须看他拟合效果如何,这就要考察决定系数与残差图。

3.1 决定系数 (coefficient of determination)

定义 回归平方和与总离差平方和之比为决定系数,也称为判定系数。记为 r^2 ,即

$$r^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

决定系数 r^2 是一个回归直线与样本观测值拟和优度的相对指标,反映了因变量的变异中能用自变量解释的比例, $0 < r^2 < 1$ 。一般地,决定系数与 1 越接近,因变量不确定性的绝大部分能由回归方程解释,回归方程拟和优度就好,反之效果就不好,应该修改。

3.2 残差 $e_i = y_i - \hat{y}_i = y_i - \beta_0 - \beta_1 x_i$

标准化残差 $ZRE_i = \frac{e_i}{\sigma}$,

学生化残差 $SRE_i = \frac{e_i}{\sigma \sqrt{1-h_{ii}}}$, 其中 $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{L_{xx}}$

以自变量 x (或以 $\hat{y}$) 做横轴,以残差 (标准化残差或学生化残差) 做纵轴,做出相应的残差点得到残差图,一般来说残差如果围绕 $e=0$ 随机波动,拟合效果较好,回归方程满意。否则的话要具体分析。

4 实例验证

表 1 1992-2006 河北省城镇居民人均可支配收入与消费情况

年份	人均消费支出 (元) y	人均可支配收入 (元) x
1992	1547	1872
1993	1898	2334
1994	2513	3177
1995	3612	3921
1996	3424	4443
1997	4004	4959
1998	3834	5085
1999	4026	5365
2000	4348	5661
2001	4479	5985
2002	5069	6679
2003	5439	7239
2004	5819	7951
2005	6699	9107
2006	7343	10304

数据来源: 国家统计数据库。

随着居民人均可支配收入的增加,人均消费额也在增加,选用直线回归模型,采用 SPSS 软件,选中统计量,得输出结果。

表 2 相关性

		支出	可支配收入
Pearson 相关性	支出	1.000	.995
	可支配收入	.995	1.000
Sig. (单侧)	支出	.	.000
	可支配收入	.000	.
N	支出	15	15
	可支配收入	15	15

表 3 模型汇总^b

模型	R	R 方	调整 R 方	标准 估计的误差	更改统计量				
					R 方更改	F 更改	df1	df2	Sig.F 更改
1	.995 ^a	.990	.989	171.30362	.990	1261.316	1	13	.000

a. 预测变量: (常量), x。 b. 因变量: y。

表 4 Anova^b

模型	平方和	df	均方	F	Sig.
1					
回归	3.701E7	1	3.701E7	1261.316	.000 ^a
残差	381484.077	13	29344.929		
总计	3.739E7	14			

a. 预测变量: (常量), x。 b. 因变量: y。

表 5 系数^a

模型		非标准化系数		标准系数	t	Sig.	相关性		
		B	标准误差				零阶	偏	部分
1	(常量)	440.196	116.562		3.777	.002			
	x	.683	.019	.995	35.515	.000	.995	.995	.995

a. 因变量: y。

表 6

y	x	RES	ZRE	SRE
1547.00	1872.00	-172.28499	-1.00573	-1.15554
1898.00	2334.00	-136.95770	-0.79950	-0.89480
2513.00	3177.00	-97.95790	-0.57184	-0.61701
3612.00	3921.00	492.68605	2.87610	3.03582
3424.00	4443.00	-51.98311	-0.30346	-0.31701
4004.00	4959.00	175.44737	1.02419	1.06314
3834.00	5085.00	-80.64519	-0.47077	-0.48819
4026.00	5365.00	-79.96198	-0.46679	-0.48336
4348.00	5661.00	39.78884	0.23227	0.24043
4479.00	5985.00	-50.59202	-0.29534	-0.30600
5069.00	6679.00	65.21564	0.38070	0.39717
5439.00	7239.00	52.58205	0.30695	0.32361
5819.00	7951.00	-53.90922	-0.31470	-0.33857
6699.00	9107.00	36.22573	0.21147	0.23965
7343.00	10304.00	-137.65356	-0.80356	-0.99298

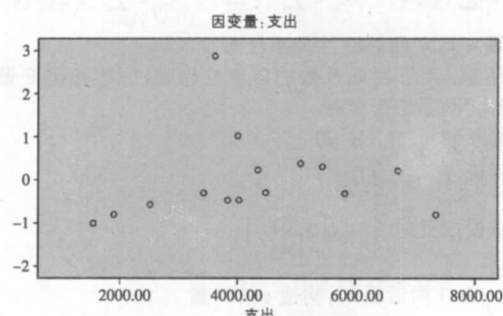


图 1 散点图

结果分析:

①从相关表中看到,相关系数 $r=0.995$,单侧检验显著性 $Sig \approx 0.000$,说明 y 与 x 有显著线性关系。②从模型汇总表中看到,决定系数 $r^2=0.995$ 。从相对水平上看,回归方程能够减少因变量 y 的 99.5% 的方差波动。③从方差分析表中看到, $F=1261.316$,显著性 $Sig \approx 0.000$,说明 y 对 x 的线性回归高度显著,与相关系数的检验结果一致。④从系数表中得到回归方程为 $\hat{y}=440.196+0.683x$,回归系数 β_1 检验的 t 值 $=35.515$,显著性 $Sig \approx 0.000$,与 F 检验和相关系数的检验结果一致。⑤从计算出的残差表,及所画的残差图知道,本例数据基本正常,模型基本合适。

在实际应用中,要得到可靠的结果,需要 F 检验、散点图、残差分析等一起使用,得到一致的结果时才可以下结论。

参考文献:

- [1]刘次华,万建平.概率论与数理统计[M].2版.北京:高等教育出版社,2008.
- [2]何晓群,刘文卿.应用回归分析[M].2版.北京:中国人民大学出版社,2007.
- [3]王天营.一元线性回归分析中三种检验的等价性研究[J].统计与决策,2011(3):8-11.