

ECON2001 总复习

#fleeting

我的想法

- Analysis of Variance
- Chi-squared test
- Simple Linear Regression
- Multiple Regression
- Model Building
- Non-parameter Statistic
- Time-Series Analysis
- Statistical Process Control
- Decision Analysis

所有例题大盘点

14.1 投资于股票市场的比例

14.2 比较维修缓冲器的作用

14.3 比较降低胆固醇的药物

15.1 检验市场份额

15.2 大学学位与 MBA 专业之间的关系

16.1 工作年限和年终奖

16.2 里程数和二手丰田凯美瑞的价格 I

一百个数据，自变量为里程数，因变量为价格，要会算自变量的方差和协方差：

协方差：

$$s_{xy} = \frac{1}{n-1} \left[\sum_{i=1}^n x_i y_i - \frac{\sum x_i \sum y_i}{n} \right]$$

方差：

$$s_x = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n} \right]$$

斜率计算：

$$b_1 = \frac{s_{xy}}{s_x^2}$$

可以得出回归方程

16.3 里程数和二手丰田凯美瑞的价格 II

通过计算标准误，评价拟合效果

标准误：

$$SSE = (n-1) \left(s_x^2 - \frac{s_{XY}^2}{s_x^2} \right)$$

估计标准误：

$$s_\epsilon = \sqrt{\frac{SSE}{n-2}}$$

用来直接输出标准误。 s_ϵ 最小值可以是 0，那么 $SSE = 0$ 这个时候所有的点都在回归直线上，说明 s_ϵ 越小，拟合效果越好。

16.4 里程数和二手丰田凯美瑞的价格相关吗？

如何验证假设是否为真。需要计算出检验统计量 t 然后和用 t 来计算出对应的值。

$$s_{b_1} = \frac{s_\epsilon}{\sqrt{(n-1)s_x^2}}$$

然后用这个值算出检验统计量：

$$t = \frac{b_1 - \beta_1}{s_{b_1}}$$

双尾检验的拒绝域为

$$t < -t_{\alpha/2, v} \text{ or } t > t_{\alpha/2, v}$$

16.5 衡量里程数和二手车价格线性关系的强度

记住公式：

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2}$$

R^2 是所谓的判定系数。可以用来衡量线性关系的强弱。

16.6 衡量里程数和二手车价格线性关系的强度

检验是否呈线性相关：

相关系数为：

$$r = \frac{s_{xy}}{s_x s_y}$$

检验统计量为：

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

(用来检验二元正态分布)

16.7 估计二手车的平均价格并预测

感觉不是很重要，如果真的遇见就直接套公式，会有一个专门的预测公式

需要注意的是，预测区间和置信区间并不是一个东西，详情见：

<https://juejin.cn/post/7116376552513732616>

在十六章的最后还涉及到了回归诊断的说明，但是并没有相关的例题。

17.1 圣诞周缆车票销售情况

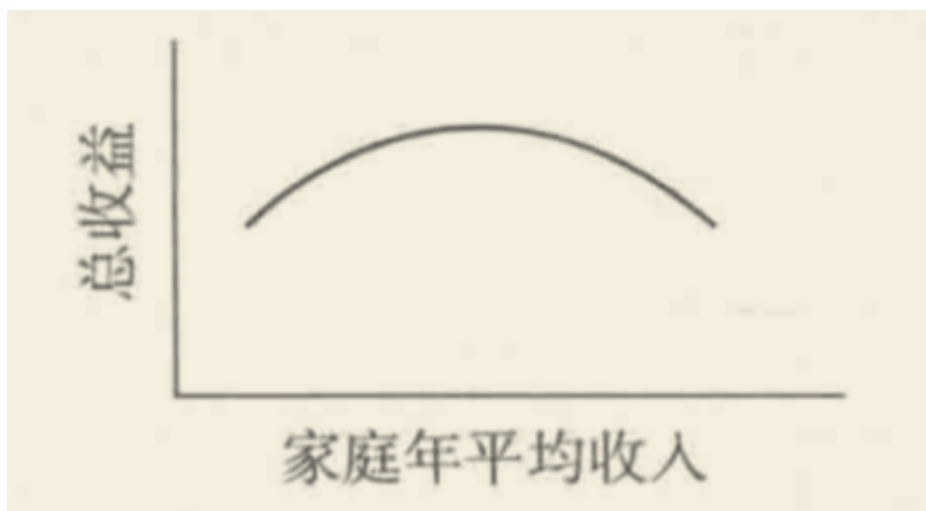
建立的多元回归模型：参考：[Multiple Regression](#)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

通过找附录，找 d_L 和 d_U 计算出 DW 统计量，因此可以用来判断能否拒绝原假设。

18.1 为快餐店选址

可能呈现的一个关系：



因为影响的因素很多，很难断言不同变量和总收益/变量之间是否会有何种关系。

因此直接建模为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$$

18.2 为快餐店选址 II

用计算器直接输出拟合结果 R^2 了

18.3 工资平等：同工同酬

模型的拟合效果相当不错。判定系数是 0.693 2，这说明薪水中 69.32% 的变异可以由模型中自变量的变异解释。 F 统计量的值是 72.29， p 值为 0，因此有足够的证据可以推断，模型是有效的。

决定是否有证据表明薪水与受教育年限、工作年限和性别之间存在线性关系的 t 检验的 p 值分别为 0.040 1、0 和 0.618 3，这说明受教育年限和工作年限都与薪水相关。但是，性别的系数的 t 检验则表明，没有证据可以推断男性与女性管理者的薪水之间存在差别。因此，这与均值之差的 t 检验的结果是相互矛盾的。进一步分析可以发现，薪水不同应归咎于工作年限。一般而言，女性管理者的工作年限较短，因而薪水相对较低。在进行回归分析之前，我们知道薪水的样本均值之差是 $97\,832 - 76\,189 = 21\,643$ 美元。剔除受教育年限和工作年限的影响之后，这个差值降为 1 851 美元，在统计上是不显著的。

19.1 威尔克科森秩和检验

Wilcoxon Rank Sum Test

两个样本，不同观测值，在 5% 的显著性水平下判断总体

秩，即为排序，最小的为 1，最大的为 6。

秩和，即将所有的排序加起来得到 T_1 和 T_2

当样本量大于 10 的时候，检验统计量就服从正态分布：

$$E(T) = \frac{n_1(n_1 + n_2 + 1)}{2}$$
$$\sigma_T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

检验统计量为：

$$z = \frac{T - E(T)}{\sigma_T}$$

19.2 止痛药的疗效

两种药物，不同的疗效，（一种是新药一种是旧的）

注意，当排名相同的时候，比如有 3 个“1”，时，也就是三个数据占据第一名到第三名，此时此刻，求平均值，也就是第二名。

当备择假设 (H_1) 为总体 1 的位置在总体 2 位置的右边的时候, 拒绝域:

$z > z_\alpha$, 就是希望看 z 是否足够大到我们可以拒绝原假设而反而去接受备择假设。

19.3 留住员工

注意, 题设的备择假设 (H_1) 为“总体 1 和总体 2 的位置不同”, 所以拒绝域是 $z < -z_{\alpha/2}$ 或者 $z > z_{\alpha/2}$

19.4 两款中型车舒适度比较

由于是同样的 n 个人评价两个总体, 所以应该设计为配对实验, 使用符号检验。

$$z = \frac{x - np}{\sqrt{np(1-p)}}$$

在符号检验中, 一般会把插值为 0 的配对观测值去掉, 因此, n 是样本中非零差值的个数。

19.5 “弹性工作制”和“固定时间工作制”的比较

考察的是 [Wilcoxon Signed Rank Sum Test](#), 当 $n > 30$ 的时候可以被定义为大样本, 此时 T 服从正态分布, 均值为:

$$E(T) = \frac{n(n+1)}{4}$$

标准差为:

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$
$$z = \frac{T - E(T)}{\sigma_T}$$

通勤时间						
员工	早上 8:00	弹性工作制	差值	差值	秩	秩
1	34	31	3	3	21.0	
2	35	31	4	4	27.0	
3	43	44	-1	1		4.5
4	46	44	2	2	13.0	
5	16	15	1	1	4.5	
6	26	28	-2	2		13.0
7	68	63	5	5	31.0	
8	38	39	-1	1		4.5
9	61	63	-2	2		13.0
10	52	54	-2	2		13.0
11	68	65	3	3	21.0	
12	13	12	1	1	4.5	
13	69	71	-2	2		13.0
14	18	13	5	5	31.0	
15	53	55	-2	2		13.0
16	18	19	-1	1		4.5
17	41	38	3	3	21.0	
18	25	23	2	2	13.0	
19	17	14	3	3	21.0	
20	26	21	5	5	31.0	
21	44	40	4	4	27.0	
22	30	33	-3	3		21.0
23	19	18	1	1	4.5	
24	48	51	-3	3		21.0
25	29	33	-4	4		27.0
26	24	21	3	3	21.0	
27	51	50	1	1	4.5	
28	40	38	2	2	13.0	
29	26	22	4	4	27.0	
30	20	19	1	1	4.5	
31	19	21	-2	2		13.0
32	42	38	4	4	27.0	
				$T^+ = 367.5 \quad T^- = 160.5$		

19.6 比较经理对应聘者的评估结果

有八名应聘者，而有四名经理对他们进行评估，用 [Friedman Test](#) 来验证是否是评估系统出现了问题（也就是经理之间的评分是否有显著差异）

备择假设为： H_1 ：至少有两个总体的位置不同。

$F_r > \chi_{\alpha, k-1}^2$ 通过跨行移动来确定秩。（横着看）

$$F_r = \left[\frac{12}{bk(k+1)} \sum_{j=1}^k T_j^2 \right] - 3b(k+1)$$

符合Chi-squared test

19.7 检验工人能力测验得分和表现之间的关系

问题的目标是分析两个变量之间的关系。需要用到Spearman rank correlation coefficient

需要计算秩的协方差

$$s_{ab} = \frac{1}{n-1} \left[\sum a_i b_i - \frac{\sum a_i \sum b_i}{n} \right]$$

单个样本的协方差是：

$$s_{a^2} = \frac{1}{n-1} \left[\sum a_i^2 - \frac{(\sum a_i)^2}{n} \right]$$

$$r_s = \frac{s_{ab}}{s_a s_b}$$

$$z = r_s \sqrt{n-1}$$

20.1 汽油销量 I

来到这道题，也就正式进入了时间序列！需要运用到的方法有 Moving Averages,

移动平均值，即不断地替换

注意四期移动平均值和四期中心移动平均值的结果是不一样的

20.2 汽油销量 II

指数平滑法（一般是给了 w 的）

$$S_t = w y_t + (1-w) S_{t-1}$$

20.3 旅馆季度入住率

计算指数。通过剔除入住率的季节效应，得到真正的增长或者下降

20.4 预测模型的比较

$$MAD = \frac{\sum_{i=1}^n |y_t - F_t|}{n}$$

$$SSE = \sum_{i=1}^n (y_t - F_t)^2$$

20.5 预测旅馆入住率

得出了趋势线方程之后，直接算出预测值 \hat{y}

$$F_t = (b_0 + b_1 t) * SI_t$$

算出四个季度的预测值（根据季节指数）

个人理解：根据季节指数调整趋势线，得到预测值

20.6 预测消费价格指数的变化

好像无法手算，遂罢。

21.1 李尔公司的统计过程控制 I

每小时随机抽取 4 根弹簧

中线 = $\bar{\bar{x}}$

控制上下限 = $\bar{\bar{x}} \pm 3 \frac{s}{\sqrt{n}}$

S 是联合标准差

21.2 李尔公司的统计过程控制 II

无

21.3 李尔公司的统计过程控制 III

没有讲这个

Reference