

統計學作業 #3: 回歸分析深度解析

# 變數選擇方法與習題解答

# 為什麼需要變數選擇？



## 效率 (Efficiency)

減少不必要的計算與資料收集成本，專注於最具影響力的因子。



## 預測力 (Accuracy)

避免過度擬合 (Overfitting)，提升模型在未知資料上的表現。



## 精簡原則 (Parsimony)

以最簡單的模型解釋最複雜的現象，提高模型的可解釋性。

# 方法比較: 選取適合的分析工具

特性	逐步回歸 (Stepwise)	最佳子集 (Best Subset)
基本邏輯	貪婪算法 (每次加/減最顯著變數)	窮舉算法 (評估所有組合)
最優性	局部最優 (Local Optimum)	全域最優 (Global Optimum)
計算速度	極快	視變數數量而定 (較慢) $2^k$
適用情境	大數據集、初步探索	精確研究、變數數量中等

---

# 第三題：模型比較與決策建議

探討是否應將變數  $X_3$  加入現有模型

# 問題背景: 研究者的困境

## 模型 A (原始模型)

自變數:  $X_1, X_2$  ( $k=2$ )

SSE = 1425

SSR = 1320

$n = 28$

## 模型 B (全模型)

自變數:  $X_1, X_2, X_3$  ( $k=3$ )

SSE = 1350

SSR = 1395

增加  $X_3$  後的改善幅度:  $\Delta\text{SSE} = 75$

# | 局部 F 檢定: 計算結果

**1.33**

計算出的 F 統計量

**2.93**

臨界值  $F(0.10, 1, 24)$

F-table of Critical Values of  $\alpha = 0.10$  for F(df1, df2)

DF1=	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$
<b>DF2=1</b>	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
<b>2</b>	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
<b>3</b>	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
<b>4</b>	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
<b>5</b>	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.11
<b>6</b>	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
<b>7</b>	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
<b>8</b>	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
<b>9</b>	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
<b>10</b>	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
<b>11</b>	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
<b>12</b>	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
<b>13</b>	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
<b>14</b>	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
<b>15</b>	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
<b>16</b>	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
<b>17</b>	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
<b>18</b>	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
<b>19</b>	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
<b>20</b>	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
<b>21</b>	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
<b>22</b>	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
<b>23</b>	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
<b>24</b>	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
<b>25</b>	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
<b>26</b>	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
<b>27</b>	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
<b>28</b>	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
<b>29</b>	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
<b>30</b>	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
<b>40</b>	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
<b>60</b>	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
<b>120</b>	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.42	1.38	1.34	1.30	1.25
<b><math>\infty</math></b>	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

# Partial F-Test 逐步推導

$$F = \frac{(SSE_R - SSE_F) / 1}{SSE_F / (n - k - 1)} = \frac{75 / 1}{1350 / 24} = \frac{75}{56.25} = 1.333$$

✓ **自由度計算**：分子  $df = 3 - 2 = 1$ ；分母  $df = 28 - 3 - 1 = 24$ 。

📊 **查表**：在  $\alpha = 0.10$  下，F 臨界值為 2.93。

⚖️ **結論**：由於  $1.33 < 2.93$ ，不拒絕虛無假設。

## 研究建議與總結

在顯著水準 0.10 下，統計證據不足以證明增加變數 X3 能顯著提升模型的解釋能力。

— 最終建議：保留結構較簡單的原始模型，不建議加入 X3。