

生物統計報告

多子多孫多福氣

第六組

B935020011 桂德豪

B935020013 郭亮鈞

B935020021 翁偉然

B935020036 方士碩

前言

「恭喜您，是個胖娃兒呢！」現在的婦產科醫生已經少了很多機會講這句話了，因為台灣的生育率近十年有下滑趨勢。政府爲了提高生育率，策定了許多政策，到底我們的阿公阿媽和我們的爹娘，在對於生育數有什麼不同，於是我們想進行兩個世代之間的生育率的比較。

目的

藉由比較(外)祖父母的子代數以及父母親的子代數，檢驗兩世代(50年代與70年代)之間出生率是否有下滑趨勢。

方法

- 1.從海資系大一到大四，隨機抽樣選出 27 位同學。
- 2.詢問受訪者及父母的兄弟姊妹數，並保證不外流做不正當用途。

數據

X(年次)	Y(家庭子女數)	X(年次)	Y(家庭子女數)	X(年次)	Y(家庭子女數)
37	3	48	6	74	2
39	5	49	11	74	2
40	5	50	4	75	3
41	4	50	6	75	1
43	7	52	7	75	2
44	5	52	3	75	3
45	8	52	4	75	3
45	9	53	3	75	4
46	9	56	5	75	3
46	5	72	3	75	2
46	6	74	6	75	2
46	4	74	2	76	3
46	5	74	3	76	2
46	5	74	3	76	2
47	4	74	3	76	2
47	3	74	3	76	1
47	5	74	2	76	2
47	3	74	4	77	3

數據分析

40年代 **F-test**

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

Assumption :

1. $E_{ij} \sim NID(0, \sigma^2)$ for all i, j .

2. 採樣為 **Random Sample**

3. 數據呈 **linear**

4. X measure are without error

$$\alpha = 0.05$$

$$\bar{X} = 46.667$$

$$\bar{Y} = 5.333$$

$$b = \frac{\sum xy}{\sum x^2} = -0.00126$$

$$a = \bar{Y} - b\bar{X} = 5.392$$

$$\hat{Y} = a + bX = 5.392 - 0.00126X$$

$$SSR = b^2 \sum x^2 = b \sum xy = 0.00079$$

$$SSE = \sum y^2 - b \sum xy = 110$$

$$SST = SSR + SSE = \sum y^2 = 110$$

$$R^2 = SSR/SST = 7.2 \times 10^{-6}$$

SOV	SS	df	MS	F
Regression	0.00079	1	0.00079	<<1
Error	110	25	4.4	
Total	110	26		

$$MS = \frac{SS}{df}$$

$$F = \frac{MSR}{MSE}$$

$F_{0.95}(1, 25) = 4.24 >$ 所算之 F 值

故接受 H_0 ，拒絕 H_a ，即 40 年代生育數沒有顯著線性關係，有犯 type I error 的 $P (P \leq 5\%)$ 。

70年代 F-test

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

Assumption :

1. $E_{ij} \sim NID(0, \sigma^2)$ for all i, j .

2. 採樣為 Random Sample

3. 數據呈 linear

4. X measure are without error

$$\bar{X} = 74.815$$

$$\bar{Y} = 2.63$$

$$b = \frac{\sum xy}{\sum x^2} = -0.315$$

$$a = \bar{Y} - b\bar{X} = 26.197$$

$$\hat{Y} = a + bX = 26.197 - 0.315X$$

$$\alpha = 0.05$$

$$SSR = b^2 \sum x^2 = b \sum xy = 2.786$$

$$SSE = \sum y^2 - b \sum xy = 23.51$$

$$SST = SSR + SSE = \sum y^2 = 26.296$$

$$R^2 = \frac{SSR}{SST} = 0.106$$

SOV	SS	df	MS	F
Regression	2.786	1	2.786	2.963
Error	23.51	25	0.9404	
Total	26.296	26		

$$MS = \frac{SS}{df}$$

$$F = \frac{MSR}{MSE}$$

$$F_{0.95}(1, 25) = 4.24 > F = 2.963$$

故接受 H_0 ，拒絕 H_a ，即 70 年代生育數沒有顯著線性關係，有犯 type I error 的 P ($P \leq 5\%$)。

綜合 F-test

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

Assumption :

1. $E_{ij} \sim NID(0, \sigma^2)$ for all i, j .

2. 採樣為 *Random Sample*

3. 數據呈 *linear*

4. X measure are without error

$$\alpha = 0.05$$

$$\bar{X} = 60.741$$

$$\bar{Y} = 3.981$$

$$b = \frac{\sum xy}{\sum x^2} = -0.092$$

$$a = \bar{Y} - b\bar{X} = 9.569$$

$$\hat{Y} = a + bX = 9.569 - 0.092X$$

$$SSR = b^2 \sum x^2 = b \sum xy = 95.003$$

$$SSE = \sum y^2 - b \sum xy = 139.978$$

$$SST = SSR + SSE = \sum y^2 = 234.981$$

$$R^2 = SSR/SST = 0.404$$

SOV	SS	df	MS	F
Regression	95.003	1	95.003	35.291
Error	139.978	52	2.692	
Total	234.981	53		

$$MS = \frac{SS}{df}$$

$$F = \frac{MSR}{MSE}$$

$$F_{0.95}(1, 52) = 4.032 < F = 35.291$$

故接受 H_a , 拒絕 H_0 , 即 40 年代到 70 年代生育數有顯著線性關係, 有犯 type II error 的 P

t-statistic

$$H_0: \beta = 0$$

$$H_a: \beta \neq 0$$

assumption

- ① $\varepsilon_{ij} \sim NID(0, \sigma^2)$ for all i, j
- ② Random Sampling
- ③ X are measured without error
- ④ $My \cdot x_i$ is linear

$$t = \sqrt{\frac{MSR}{MSE}}$$

$$t_{40} = \sqrt{\frac{0.000944}{4.4}} = 0.013$$

$$t_{70} = \sqrt{\frac{2.786}{0.9409}} = 1.721$$

$$t_{mix} = \sqrt{\frac{95.003}{2.692}} = 5.941$$

$$t_{\frac{0.05}{2}}(27-2) = 2.060$$

$$t_{\frac{0.05}{2}}(54-2) = 1.960$$

$$\begin{cases} 0.013 < 2.060 \\ 1.721 < 2.060 \end{cases}$$

故接受 $H_0: \beta = 0$

在 40 年代和 70 年代, linear regression is insignificant
Type II error is possible.

$$5.941 > 1.960, \text{ 故接受 } H_a: \beta \neq 0$$

將兩組數據混合, linear regression is significant
Type I error is possible ($p \leq 5\%$)

討論

1. 將 40 年代與 70 年代綜合來看, 呈現明顯線性關係, 並有負相關的趨勢, 但因中間有年代並未取得資料, 故不適合下「生育率年年下降」的結論。
2. 生育率下降的原因可能是因為結婚的人變少, 而低結婚率與高離婚率對提昇生育率更是大大不利。
3. 另一個原因則是現代人都對未來想太多, 但又不夠樂觀。一旦評估未來環境可能不如期待中美好, 就寧願不生孩子。

問題回答與檢討

1. 在本組採樣上因為 50 年代之人數有所缺失, 採樣樣本中斷, 無連續性, 對我們所做出的線性關係圖有所影響, 故採樣必需完整齊全。

2. 由於本組所採用的採樣方法為計算年次，將導致數據過於繁多且不易看出是否有明顯線性關係，故同學建議以年區間，即每 5 年或 10 年作為 X 軸，這樣所得的數據將比較有合理性以及正確性，且較易觀察期 X 與 Y 之關係。
3. 若以出生率為題，採樣時是否該曲祖父母及父母的出生年代，而非父母及我們至一代，因為數據並非代表此年出生的人之生育數（ex.47.4 並非代表此人 47 年出生而生了 4 個子代），故此採樣方法所得數值之解釋為 47 年出生者有 4 個兄弟姐妹，而也可以代表其父母所擁有的子代數，故我們所採集這樣的數據為 data。