

第十二組 Data set-3

CKD 慢性腎臟疾病

組頭(彙整者)： B101099101 曾士剛

B101099045 周孜容 B101099052 吳紋綾 B101099080 陳永安

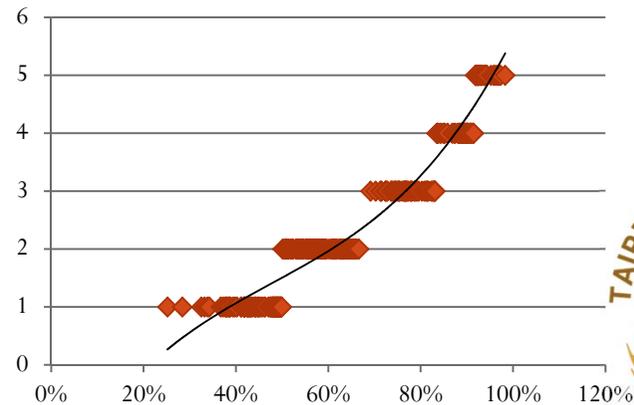
B101099098 鄒雙印 B101099105 陳怡蓁 B101099139 李佳駿

B101099142 莊登棋 B101099152 倪漢斌 B101099154 吳克宣



CKD 慢性腎臟疾病

- 全名：Chronic Kidney Disease
- 中文常稱為「腎衰竭」
- 台灣人十大死因之一
- 主要以腎絲球過濾率(GFR)分為五期
- 常由高血壓、糖尿病等慢性病併發而生
- 控制的方式主要為飲食控制



數據分析

類別變項	
CKD	有、無
性別	男、女
教育程度	不識字、小學、國中、高中、大學、研究所
飲酒	有、無、偶爾
消炎藥	有、無、有需要才用
糖尿病	有、無、不清楚
高血壓	有、無、不清楚
腎絲球過濾率分期	1、2、3、4、5

連續變項
年齡
腎絲球過濾率
α -tocopherol ($\mu\text{g}/\text{mL}$)



初步統計和分布

有CKD：167筆

男：146筆

無糖尿病：343筆

無高血壓：289筆

無飲酒：308筆

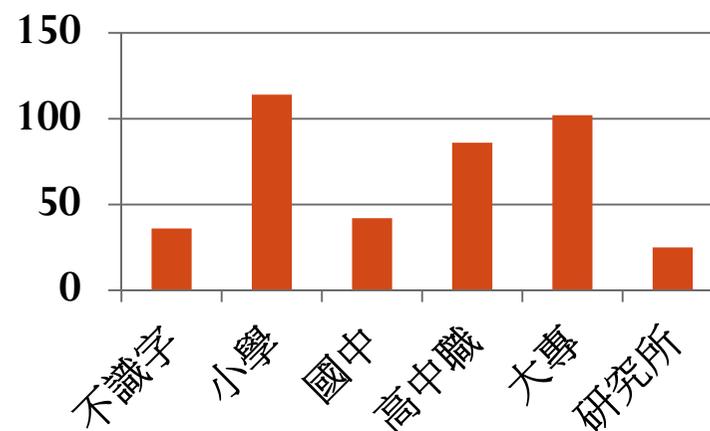
無CKD：238筆

女：259筆

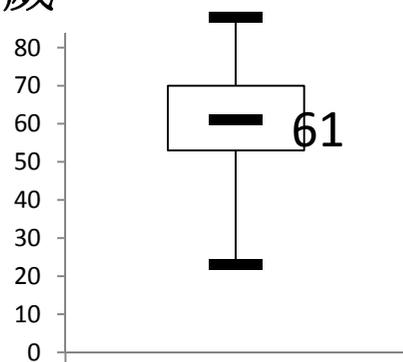
有糖尿病：62筆

有高血壓：116筆

有：24筆 偶爾：73筆

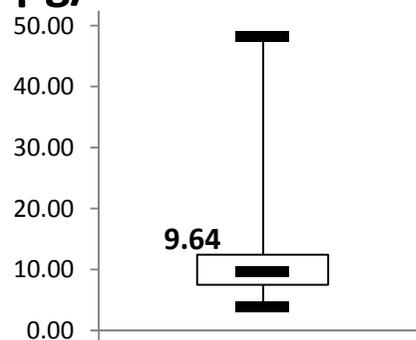


歲



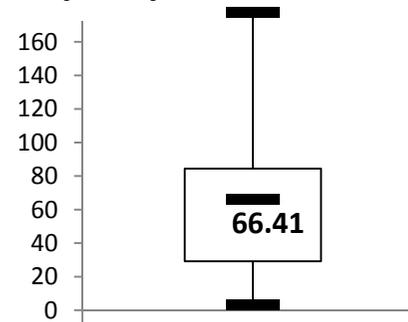
年齡箱鬚圖

$\mu\text{g/mL}$



Toco箱鬚圖

$\text{ml/min}/1.73\text{m}^2$



GFR箱鬚圖

探討主題與使用的檢定

科學假說

有無患有CKD者之血中Toco值是否有顯著差異？

性別對於罹患慢性腎臟病的機率是否有顯著差異？

有飲酒、無飲酒與飲酒頻率對於罹患慢性腎臟病的觀察值與期望值是否有差異存在？

糖尿病、高血壓與CKD是不相關？

不同教育程度的人，其腎絲球過濾率(GFR)是否有顯著差異？

年齡與GFR是否有關聯？

檢定方法

- t 檢定

- Z檢定

- 卡方檢定

- 卡方檢定

- ANOVA

- 卡方檢定、Regression

t檢定主題：
有無患有CKD者之血中Toco值
是否有顯著差異

上台報告：B101099101 曾士剛
PPT 製作：B101099080 陳永安



背景資料

- **Toco:**
- 血清中 α -tocopherol(即維他命E)的濃度
- 單位： $\mu\text{g}/\text{mL}$

- **GFR：腎絲球過濾率**
- **Glomerular Filtration Rate**
- 單位： $\text{mL}/\text{min}/1.73\text{m}^2$



科學假說

- 科學假說：
有無患有CKD者之血中Toco值是否有顯著差異？
- 資料整理：

	人數	血中Toco平均值 ($\mu\text{g}/\text{mL}$)	變異數 (cm^2)	標準差 (cm)
患CKD	167	10.015	25.734	5.073
無CKD	238	12.126	21.638	4.652

第一步 F test

- 檢定有患有CKD與無CKD的樣本中，
血液中Toco的變異數是否有顯著差異。
- F test: 統計假說：
- $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$ set $\alpha=0.05$
- $F = \frac{S_1^2}{S_2^2} = \frac{25.734}{21.638} = 1.189 < F_{Vn=166, Vd=237, \alpha=0,05} = 1.3277$
- Do not reject H_0 使用 equal-variance t test

第二步 t test

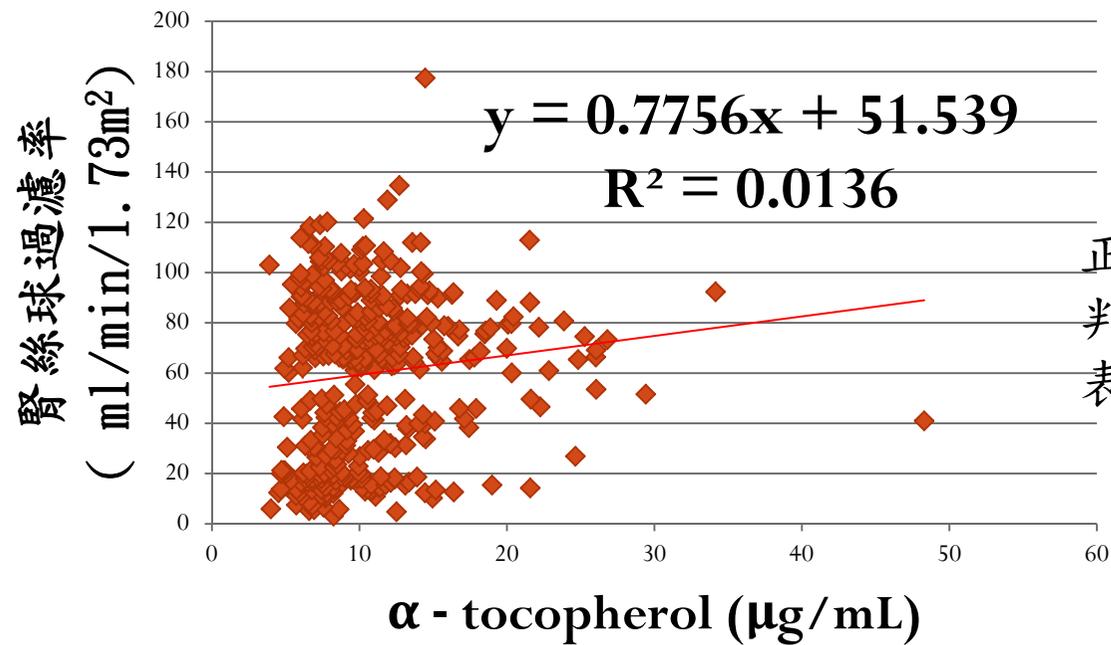
- 檢定血液中Toco值的平均值是否有顯著差異
- Pooled-variance t test:
- 統計假說： $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
set $\alpha=0.05$
- $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = 23.325$
- $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}} = 0.02 < t_{403, \alpha=0.05} = 1.96$
- Do not reject H_0
是否患有CKD，與血中Toco值的高低
沒有顯著差異



Toco和GFR的關係

- Correlation

Toco 對 GFR作圖



正相關
判定係數0.0136
表示是低度的相關

Toco和GFR的關係

論文支持：「Plasma levels of alpha-tocopherol (vitamin E) in chronic renal insufficiency (CRI) patients may be decreased, normal or elevated.」

引自 [Vitamin E disturbances in chronic renal failure]. Yukawa S, Mune M, Otani H. Nihon Rinsho. 1999 Oct. Review. Japanese.



Toco和GFR的關係

引申：「 an abnormal distribution of vitamin E in each lipoprotein has been reported: patient LDL contained less vitamin E 」

所以本實驗如進一步測量患者血液中LDL所含的vitE
濃度或許可有進一步發現

引自 [Vitamin E disturbances in chronic renal failure]. Yukawa S, Mune M, Otani H. Nihon Rinsho. 1999 Oct. Review. Japanese.



結論

- 是否患有CKD，與血中Toco值的高低沒有顯著差異



Z TEST AND CHI-SQUARE TEST

性別與飲酒

報告：B101099098 鄒雙印

DDT：B101099052 吳紋綾



命題一：
不同性別對於罹患慢性腎臟病的機率是
否有顯著差異？

採用統計方法：Z檢定



假說

- 研究問題：不同性別對於罹患慢性腎臟病的機率是否有顯著差異？
- 科學假說：性別與罹患慢性腎臟病的機率有關



數據整理

	有腎臟病	無腎臟病	總人數
男性	79	67	146
女性	88	171	259
總人數	167	238	405

- 統計假說：

$$H_0 : \hat{p}_1 = \hat{p}_2 ;$$

$$H_1 : \hat{p}_1 \neq \hat{p}_2$$

• set $\alpha = 0.05$



機率計算

- 男性有腎臟病的機率： $\hat{p}_1 = 0.54$
- 女性有腎臟病的機率： $\hat{p}_2 = 0.34$
- 總人數有腎臟病之機率： $\hat{p} = \frac{79+88}{146+259} = 0.412$
- 兩組樣本的 $n\hat{p}$ 為
 - $0.412(146) = 60.152 > 5$
 - $0.412(259) = 106.708 > 5$



(1)不用 Yates修正

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\frac{0.54 - 0.34}{\sqrt{0.412(0.588)\left(\frac{1}{146} + \frac{1}{259}\right)}}$$

$$= 0.2 / 0.05 = 3.926 > Z_{\alpha=0.05} = 1.96$$

(2) 使用 Yates 修正

$$Z = \frac{|\hat{p}_1 - \hat{p}_2| - \frac{1}{2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\hat{p}(1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$= 0.19/0.05 = 3.73 > Z_{\alpha=0.05} = 1.96$$

Reject H_0

性別對於罹患慢性腎臟病的
機率具有有顯著差異



命題二：

有無飲酒與飲酒頻率對於罹患慢性腎臟病的觀察值與期望值是否有差異存在？

採用統計方法：Chi-square檢定



假說

- 科學假說：有無飲酒與飲酒頻率對於罹患慢性腎臟病的觀察值與期望值具有顯著差異。
- (有無飲酒與飲酒頻率與罹患慢性腎臟病相關)

統計假說

H_0 : $P(A \cap B) = P(A) \times P(B) \rightarrow$ independent

→ 有無飲酒與飲酒頻率對於罹患慢性腎臟病無關

H_1 : $P(A \cap B) \neq P(A) \times P(B) \rightarrow$ dependent

→ 有無飲酒與飲酒頻率對於罹患慢性腎臟病有關



表一：觀察值

	有腎臟病	無腎臟病	總人數
0(沒有飲酒習慣)	140	168	308
1(有飲酒習慣， 包含已經戒酒者)	15	10	25
2(偶爾飲酒者)	12	60	72
總人數	167	238	405

表二：期望值

	有腎臟病	無腎臟病	總人數
0(沒有飲酒習慣)	127.00	181.00	308
1(有飲酒習慣， 包含已經戒酒者)	10.31	14.69	25
2(偶爾飲酒者)	29.69	42.31	72
總人數	167	238	405

數據分析

- 小於5的格數=0%
- 自由度 $= (r-1)(c-1) = (3-1)(2-1) = 2$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \sum \frac{(\text{觀察值} - \text{期望值})^2}{\text{期望值}}$$

$$\begin{aligned} \chi^2 &= \frac{(140-127)^2}{127} + \frac{(15-10.31)^2}{10.31} + \frac{(12-29.69)^2}{29.69} + \frac{(168-181)^2}{181} + \frac{(10-14.69)^2}{14.69} + \frac{(60-42.31)^2}{42.31} \\ &= 23.832 > \chi^2_{v=2, \alpha=0.01} = 9.21 \end{aligned}$$

- χ^2 極大且大於以1%為顯著水準(即 $\alpha=0.01$)查表所得的臨界值9.21 → **REJECT H_0**



Reject H_0

有無飲酒與飲酒頻率對於罹患
慢性腎臟病的觀察值與期望值
具有顯著差異



卡方檢定

糖尿病(DM)與慢性腎臟病(CKD)是否相關?

高血壓(HTN)與慢性腎臟病(CKD)是否相關?

ppt製作:周孜容B101099045

上台報告:陳怡蓁B101099105



研究問題

- 糖尿病(DM)與慢性腎臟病(CKD)是否相關?



- 科學假說:

糖尿病(DM)與慢性腎臟病(CKD)有關

- 統計假說:

$H_0 : P(A \cap B) = P(A) \times P(B) \rightarrow \text{independent}$

\rightarrow 糖尿病(DM)與慢性腎臟病(CKD)無關

$H_1 : P(A \cap B) \neq P(A) \times P(B) \rightarrow \text{dependent}$

\rightarrow 糖尿病(DM)與慢性腎臟病(CKD)有關



慢性腎臟病與糖尿病列聯表

組別	有CKD	有CKD 期望值	無CKD	無CKD 期望值	總數
有DM	42	25.98	21	37.02	63
無DM	125	141.02	217	200.98	342
總數	167		238		405

31

自由度: $v = (2-1)(2-1) = 1$

$\alpha = 0.05$



卡方檢定過程

$$\chi^2 = \sum \frac{[|O - E| - \frac{1}{2}]^2}{E} \quad (2 \times 2 \text{ 列聯表須進行 Yates 校正})$$

$$\chi^2 = \frac{[|42 - 25.98| - \frac{1}{2}]^2}{25.98} + \frac{[|125 - 141.02| - \frac{1}{2}]^2}{141.02} + \frac{[|21 - 37.02| - \frac{1}{2}]^2}{37.02}$$

$$+ \frac{[|217 - 200.98| - \frac{1}{2}]^2}{200.98}$$

$$= 9.271 + 1.708 + 6.506 + 1.198$$

$$= 18.683 > \chi^2_{\alpha=0.05, v=1} = 3.841$$

→ Reject H_0



結論

- 糖尿病(DM)與慢性腎臟病(CKD)有關



研究問題

- 高血壓(HTN)與慢性腎臟病(CKD)是否相關?



- 科學假說:

高血壓(HTN)與慢性腎臟病(CKD)有關

- 統計假說:

$H_0 : P(A \cap B) = P(A) \times P(B) \rightarrow \text{independent}$

→ 高血壓(HTN)與慢性腎臟病(CKD)無關

$H_1 : P(A \cap B) \neq P(A) \times P(B) \rightarrow \text{dependent}$

→ 高血壓(HTN)與慢性腎臟病(CKD)有關



慢性腎臟病與高血壓列聯表

組別	有CKD	有CKD 期望值	無CKD	無CKD 期望值	總數
有HTN	69	48.24	48	68.76	117
無HTN	98	118.76	190	169.24	288
總數	167		238		405

36

自由度: $v = (2-1)(2-1) = 1$

$\alpha = 0.05$



卡方檢定過程

$$\chi^2 = \sum \frac{[|O - E| - \frac{1}{2}]^2}{E} \quad (2 \times 2 \text{ 列聯表須進行 Yates 校正})$$

$$\begin{aligned} \chi^2 = & \frac{[|69 - 48.24| - \frac{1}{2}]^2}{48.24} + \frac{[|98 - 118.76| - \frac{1}{2}]^2}{118.76} + \frac{[|48 - 68.76| - \frac{1}{2}]^2}{68.76} \\ & + \frac{[|190 - 169.24| - \frac{1}{2}]^2}{169.24} \end{aligned}$$

$$= 8.509 + 3.456 + 5.970 + 2.425$$

$$= 20.360 > \chi^2_{\alpha=0.05, v=1} = 3.841$$

→ Reject H_0



結論

高血壓(HTN)與慢性腎臟病(CKD)有關



ONE-WAY ANOVA

比較不同教育程度以及腎絲球過濾率有無
差異？如有，其差異程度有多少？

ppt製作：倪漢斌 B101099152

上台報告：李佳駿 B101099139



假說

- 科學假說

不同教育程度的人，其腎絲球過濾率（GFR）有顯著差異。

- 統計假說

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$

$$H_1 : \mu_i \neq \mu_j$$

$$\text{set } \alpha = 0.05$$



資料整理

	人數	GFP的平均值	SD
不識字、無	36	44.23	29.45
小學	114	48.41	30.70
國中	42	60.41	30.79
高中職	86	61.61	32.42
大專、大學	102	69.81	30.77
研究所以上	25	86.29	19.56



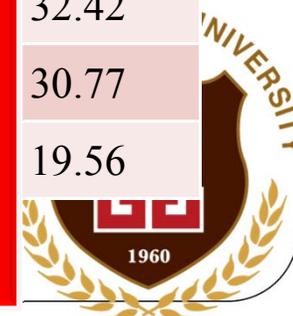
計算過程

總平均值

$$= \frac{36 \times 44.23 + 114 \times 48.41 + 42 \times 60.61 + 86 \times 61.61 + 102 \times 69.81 + 25 \times 86.29}{36 + 114 + 42 + 86 + 102 + 25}$$

$$= 59.81$$

	人數	GFP的平均值	$(x_i - \text{總平均值})^2$	
不識字、無	36	44.23	242.74	29.45
小學	114	48.41	129.96	30.70
國中	42	60.41	0.36	30.79
高中職	86	61.61	3.24	32.42
大專、大學	102	69.81	100	30.77
研究所以上	25	86.29	701.19	19.56



計算過程

- 組內變異 S_{wit}^2

$$= \frac{(36 - 1) \times 29.45^2 + \dots + (25 - 1) \times 19.56^2}{(36 - 1) + \dots + (25 - 1)}$$
$$= \underline{927.00}$$

- 組間變異 S_{bet}^2

$$= \frac{36 \times 242.74 + \dots + 25 \times 701.19}{6 - 1}$$
$$= \underline{10315.5}$$



計算過程

$$Vn=6-1=5$$

$$\begin{aligned} Vd &= (36 - 1) + (114 - 1) + (42 - 1) + (86 - 1) \\ &\quad + (102 - 1) + (25 - 1) \\ &= \underline{\underline{399}} \end{aligned}$$

$$F = \frac{S_{bet}^2}{S_{wit}^2} = \frac{10315.5}{927.00} = \underline{\underline{11.13}} > F_{5,399, \alpha=0.05} = \underline{\underline{2.21}}$$

\therefore reject H_0



結論～1

- 不同教育程度的人其腎絲球過濾率
有顯著差異

- 哪些組有差異？其差異程度呢？
95%信賴區間



資料整理

	人數	GFP的平均值	SD
不識字、無	36	44.23	29.45
小學	114	48.41	30.70
國中	42	60.41	30.79
高中職	86	61.61	32.42
大專、大學	102	69.81	30.77
研究所以上	25	86.29	19.56

- 平均值按大小排列，由最小的兩組開始兩兩比較。

計算~一、二組比較

	人數	GFP的平均值	SD
不識字、無	36	44.23	29.45
小學	114	48.41	30.70

$$S^2 = (35 \times 29.87^2 + 113 \times 30.70^2) / (35 + 113) = \underline{930.6}$$

$$S_{X_{\text{國中}} - X_{\text{不識字、無}}}^2 = \sqrt{(930.6/36 + 930.6/114)} = \underline{5.8}$$

$$V = 36 + 114 - 2 = \underline{148}$$

$$t_{0.05, v=148} = \underline{1.96}$$

$$X_{\text{國中}} - X_{\text{不識字、無}} = \underline{4.18}$$



計算～一、二組比較

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha} S_{\bar{X}_1 - \bar{X}_2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\alpha} S_{\bar{X}_1 - \bar{X}_2}$$

$$4.18 - 1.96 \times 5.8 < \mu_{\text{小學}} - \mu_{\text{不識字}} < 4.18 + 1.96 \times 5.8$$

$$\underline{-7.19 < \mu_{\text{小學}} - \mu_{\text{不識字}} < 15.548}$$

小結：其信賴區間包含0，所以這兩組不同教育程度的人，其GFP 沒有顯著差異。



計算~二、三組比較

	人數	GFP的平均值	SD
不識字、無	36	44.23	29.45
國中	42	60.41	30.79

$$S^2 = (35 \times 29.87^2 + 41 \times 30.79^2) / (35 + 41) = \underline{922.32}$$

$$S_{X_{\text{國中}} - X_{\text{不識字、無}}}^2 = \sqrt{(922.32/36 + 922.32/42)} = \underline{6.90}$$

$$V = 36 + 42 - 2 = \underline{76}$$

$$t_{0.05, v=76} = \underline{1.99}$$

$$X_{\text{國中}} - X_{\text{不識字、無}} = \underline{16.18}$$



計算~二、三組比較

$$(\bar{X}_1 - \bar{X}_2) - t_\alpha S_{\bar{X}_1 - \bar{X}_2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_\alpha S_{\bar{X}_1 - \bar{X}_2}$$

$$16.18 - 1.99 \times 6.90 < \mu_{\text{國中}} - \mu_{\text{不識字}} < 16.18 + 1.99 \times 6.90$$

$$\underline{2.45 < \mu_{\text{國中}} - \mu_{\text{不識字}} < 29.91}$$

小結：其信賴區間不包含0，所以這這兩組不同教育程度的人，其GFP 有顯著差異。



Age 與 Glomerular filtration rate (GFR) 的關係

(PPT) B101099142 莊登棋
(報告) B101099154 吳克宣



可用分析

- 可把年齡分組: 用ANOVA分析
- 以年齡分層後, 可把GFR跟標準值比較便分類: 用 χ^2 或 **proportion**分析

共同缺點: 分層過程造成Data Lost
(continuous/cardinal \rightarrow discrete/ordinal)

Test for trend
(分析趨勢)



Hypothesis

$$H_0: \rho = 0$$

$$H_A: \rho \neq 0$$

雙尾鑑定

ρ 是母群體的regression的意思。等於0的話就代表兩個變像是沒有關係的；不等於0的話就是有關係。差別在關係大或小。



資料整理

	X (age)	X ²	Y (GFR)	XY
	59	3481	60.054643	3543.165
	38	1444	60.281649	2290.703

	46	2116	53.550532	2463.324
	57	3249	55.518611	3164.561
<u>n</u>	405	405	405	405
<u>SUM</u>	24285	1525299	24224.43	1419197



計算

$$\begin{aligned} a &= \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum X^2) - (\sum X)^2} \\ &= \frac{(24224.43)(1525299) - (24285)(1419197)}{(405)(1525299) - (24285)^2} \\ &= 88.77296 \dots \approx \mathbf{88.8} \end{aligned}$$

$$\begin{aligned} b &= \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \\ &= \frac{(405)(1419197) - (24285)(24224.43)}{(405)(1525299) - (24285)^2} \\ &= -0.48296 \dots \approx \mathbf{-0.483} \end{aligned}$$



$$\begin{aligned}
 s_x &= \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} \\
 &= \sqrt{\frac{(59 - 59.963)^2 + (38 - 59.963)^2 + \dots + (46 - 59.963)^2 + (57 - 59.963)^2}{405 - 1}} \\
 &= 13.07806 \dots \approx 13.1
 \end{aligned}$$

$$\begin{aligned}
 s_y &= \sqrt{\frac{\sum(Y - \bar{Y})^2}{n - 1}} \\
 &= \sqrt{\frac{(60.055 - 59.813)^2 + (60.282 - 59.813)^2 + \dots + (53.551 - 59.813)^2 + (55.519 - 59.813)^2}{405 - 1}} \\
 &= 32.3307 \approx 32.3
 \end{aligned}$$

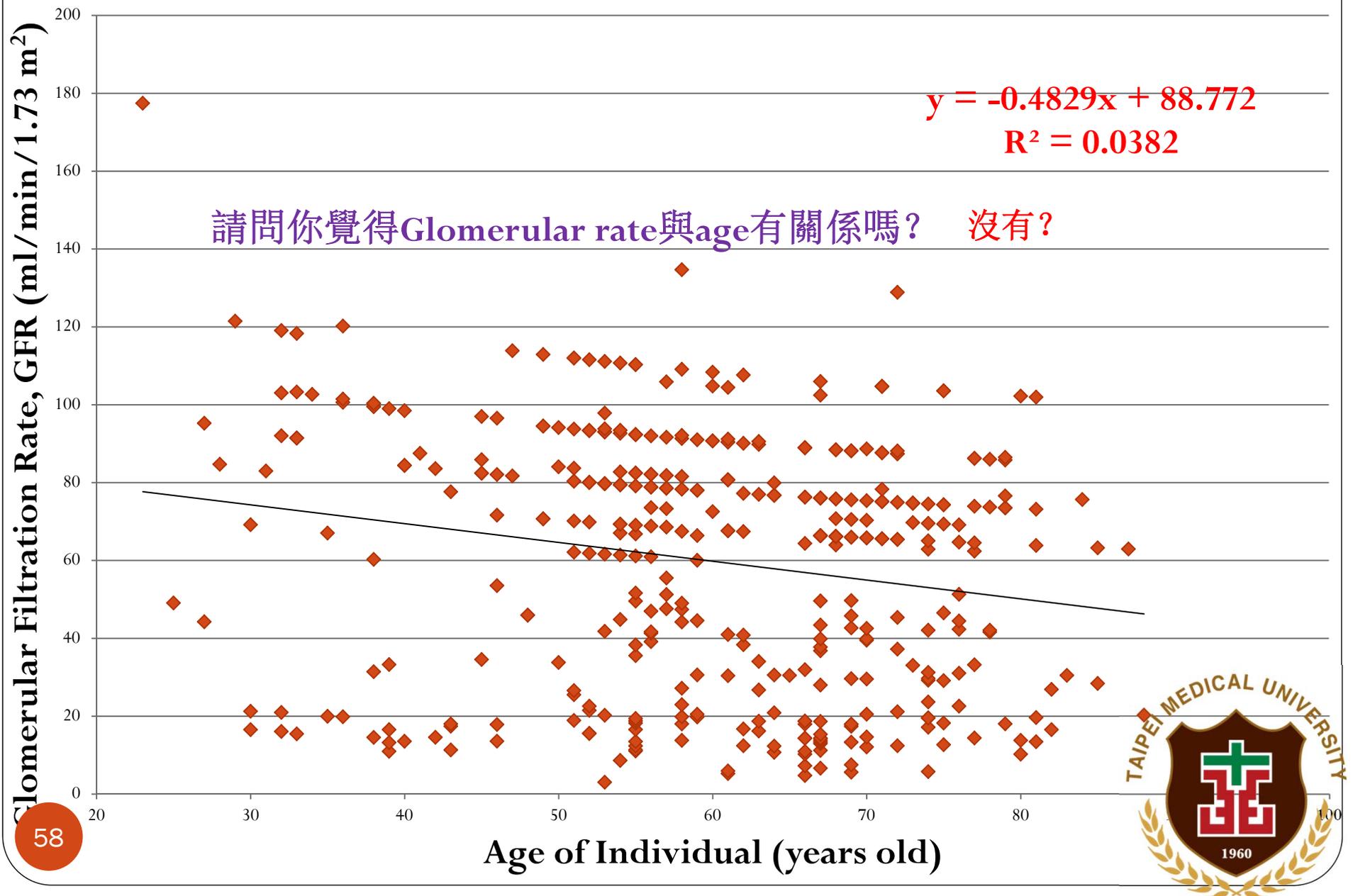
$$r = b \frac{s_X}{s_Y}$$

$$= (-0.483) \left(\frac{13.1}{32.3} \right)$$

$$= -0.19536 \dots \approx -0.195$$



The Age of an Individual vs. the Glomerular Filtration Rate (GFR) of the Individual



t-test

$$t_{n-2} = \frac{r}{\sqrt{(1-r^2)/(n-2)}} = 3.999$$

查表:

$$t_{\alpha=0.05, \nu = n-2}$$

$$= t_{\alpha=0.05, \nu = 403} \approx Z_{\alpha=0.05}$$

$$= 1.960$$

$$t_{n-2} = 3.999 > Z_{\alpha=0.05} = 1.960$$

Null hypothesis (H_0) rejected.

There is a significant correlation between an individual's age and his glomerular filtration rate. (但是關係不大)



Research

“It had been shown through clinical observations that glomerular filtration rate decreases as age increases.”

Davies, D.F., Shock, N.W. “Age Changes in Glomerular Filtration Rate, Effective Renal Plasma Flow, and Tubular Excretory Capacity in Adult Males.” *Journal of Clinical Investigation*. 1950 May; 29(5): 496-507.



Further analysis

CKD是根據GFR來劃定。照理來說，CKD level與年齡會有關係。5個CKD level，有5個不同的R，個別做t-test。以下為資料整理：(t value計算步驟省略)

CKD Level	Sample Size	R	t-value
1	78	-0.17396	1.540
2	160	-0.14254	1.810
3	61	-0.17262	1.346
4	62	0.31734	2.592
5	44	-0.07593	0.493

除了CKD level 4，其他都小於 $z_{\alpha=0.05} = 1.960$ 。所以CKD level 1、2、3、5都不能拒絕null hypothesis (age和GFR沒有關係的意思)，但4卻可以拒絕。



解釋

- 樣本GFR不只和age有關係，還可能和健康狀況、服藥狀況、和其他生活習慣有關。(其他組員有測其它factor與GFR的關係)
- 因為CKD分層后可能會有造成不同age的GFR overlap，造成其年齡與GFR的關係變小，進而影響CKD與age的關係。



總結 test for trend

R 不等于 0 時，不代表 ρ 不等于 0，反之亦然。所以要檢定假說時，需要用 t-test。

樣本數的重要性。

$$t_{n-2} = \frac{r}{\sqrt{(1-r^2)/(n-2)}}$$



總結論

- 是否患有CKD，與血中Toco值的高低沒有顯著差異
- 性別對於罹患慢性腎臟病的機率具有有顯著差異
- 有無飲酒與飲酒頻率對於罹患慢性腎臟病的觀察值與期望值具有顯著差異
- 糖尿病(DM)與慢性腎臟病(CKD)有關
- 高血壓(HTN)與慢性腎臟病(CKD)有關
- 不同教育程度的人其腎絲球過濾率有顯著差異
- Age 與 Glomerular filtration rate (GFR) 有關係



謝謝您的聆聽

組頭(彙整者)： B101099101 曾士剛

B101099045 周孜容 B101099052 吳紋綾 B101099080 陳永安

B101099098 鄒雙印 B101099105 陳怡蓁 B101099139 李佳駿

B101099142 莊登棋 B101099152 倪漢斌 B101099154 吳克宣

