比較百分比(percentage)

卡方檢定(Chi-square tests)

比較百分比

觀察一群人(N)中，產生某一現象之人數(n)

盛行率(p= $\frac{n}{N}$)或是百分比($\frac{n}{N}$ X100%)

●盛行率p與另一固定盛行率p_0比較，做一單樣本的檢定：

百分比的單樣本檢定，是指盛行率p與一固定數p_0比較，是否有顯著性差異。其檢定公式 $z_s=\frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$，其餘步驟可參考Z檢定；檢定公式中，$p$是代表母全體的盛行率，$\hat{p}$ 則是代表由樣本所計算出來的盛行率。

●資料中分成兩組後比較此兩組盛行率的差別(p_1-P2)，做一雙樣本的比較

如果是資料中分成兩組後，比較此兩組盛行率的差別(p_1-p_2)，其檢定公式為 $z_s=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1}+\frac{1}{n_2})}}$ 呈常態分配，其中 $\hat{p}=\frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$ $\hat{q}=1-\hat{p}$。其餘檢定步驟亦可參考Z檢定，需特別注意的是使用這種方法時必須 $n_1\hat{p}\hat{q}\geq 5$ 且 $n_2\hat{p}\hat{q}\geq 5$才適用 。

卡方檢定

- 探討兩類別變項 (categorical variables) 是否相關

		Design	
		Continuous	Categorical
Outcome (y)	Continuous	Pearson Correlation coeff. Linear regression *Spearman Correlation coeff.*	t-test / ANOVA Linear models *Wilcoxon Rank-Sum test* *Kruskal-Wallis test*
	Categorical	Logistic Categorical analysis	χ^2 Fisher's Exact Categorical analysis Logistic analysis

卡方檢定

假設變項A有r個選項，變項B有c個選項，那麼這兩個變項的資料就可以被整理成

變項B(有c個選項) → 每人僅可單選

變項A (有r個選項) ↓ 每人僅可單選

	1	2	3		c
1	n_{11}	n_{12}	n_{13}	---	n_{1c}
2	n_{21}	n_{22}	n_{23}	---	n_{2c}
¦	¦	¦	¦	¦	¦
r	n_{r1}	n_{r2}	n_{r3}	---	n_{rc}

H_0：變項A 及變項B 互相獨立

H_1：變項A 及變項B 沒有互相獨立

或

H_0：變項A 及變項B 沒有關係

H_1：變項A 及變項B 有關係

原始資料所反映之觀察值:

變項B

		1	2	3	……	c	
變項A	1	n_{11}	n_{12}	n_{13}	……	n_{1c}	$n_{1.}$
	2	n_{21}	n_{22}	n_{23}	……	n_{2c}	$n_{2.}$
	…	…	…	…	…	…	…
	r	n_{r1}	n_{r2}	n_{r3}	……	n_{rc}	$n_{r.}$
		$n_{.1}$	$n_{.2}$	$n_{.3}$	……	$n_{.c}$	N

計算出之期望值：

$\frac{n1.\times n.1}{N}$	$\frac{n1.\times n.2}{N}$	……	$\frac{n1.\times n.c}{N}$
$\frac{n2.\times n.1}{N}$	$\frac{n2.\times n.2}{N}$	……	$\frac{n2.\times n.c}{N}$
…	…	……	…
$\frac{nr.\times n.1}{N}$	$\frac{nr.\times n.2}{N}$	……	$\frac{nr.\times n.c}{N}$

由$O_{ij}(=n_{ij})$代表，期望值由Eij(＝$\frac{ni.\times n.j}{N}$)代表，則

卡方檢定值為 $\chi_s^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$

呈自由度(r-1)(c-1)的卡方分配$\chi^2_{(r-1)(c-1)}$，

臨界值為$\chi^2_{(r-1)(c-1)1-\alpha}$ ，

P-值為為Pr($\chi^2_{(r-1)(c-1)} > \chi^2$)

<舉例>

假設要探討職位分類是否與贊成週休二日有關，經過調查後可整理成以下的表格:

是否贊成週休2日

職位分類	贊成	反對	未決定	合計
職員	30	15	15	60
教員	40	50	10	100
主管	10	25	5	40
合計	80	90	30	200

因此虛無假設與對立假設之寫法為
H_0:對週休二日的意見與職位分類無關
H_1:對週休二日的意見與職位分類有關
根據上面表格計算出期望值:

是否贊成週休2日

職位分類	贊成	反對	未決定	合計
職員	$\frac{60\times80}{200}=24$	$\frac{60\times90}{200}=27$	$\frac{60\times30}{200}=9$	
教員	$\frac{100\times80}{200}=40$	$\frac{100\times90}{200}=45$	$\frac{100\times30}{200}=15$	
主管	$\frac{40\times80}{200}=16$	$\frac{40\times90}{200}=18$	$\frac{40\times30}{200}=6$	

因此假設檢定值為 $\chi_s^2=\frac{(30-24)^2}{24}+\frac{(15-27)^2}{27}+\cdots+\frac{(5-6)^2}{6}=18.19$

自由度為df = (3-1) x (3-1)=4，因此臨界值為 $\chi^2_{4.095}$=9.49，因為18.19 ＞9.49

所以推翻H_0；檢定結果顯示職位與是否贊成週休二日有顯著性相關。

針對 r x c table 使用 χ^2 test 之注意事項

Page 396,use this test only if the following two Conditions are satisfied:

(a) No more than 20% of the cells should have expected values< 5.

(b) No cell should be expected value < 1.

一般常用於檢定2x2 tables者有

- χ^2-test
- χ^2-test with Yates correction
- Fisher's exact test

1. 一般而言r或c的項目不可過多

2. Eij≧ 5時χ^2 -test所給的p-value才會較正確

3. 當nij或Eij過小時,可考慮combine categories

4. Fisher's exact test for RxC tables

5. Chi-square testing for trend

Chi-square testing for trend

檢定2個categorical variables之association時,χ^2 -test是基本方法。若2個變數都是nominal則僅可用χ^2 -test(df=(r-1)(c-1))

若其中1個或2個是ordinal則可用更好(powerful)的方法(df較小的方法)。所謂更好的方法是找出適當的分數目來取代ordinal variables

例如:

VAR	項目	意義	分數（score）
AGECAT	1	0-20	10
	2	20-30	25
	3	30-40	35
	4	40-50	45
	5	50→	70

↑
亦可用其他種分數取代

以整數來當分數者,僅continuous variable 轉換而來之categorical variable才適合。

例如:

項目	滿意度分數	人數	Standardized midrank score
非常滿意	1	30	0.1026
滿意	2	40	0.3344
中立	3	50	0.6325
不滿意	4	20	0.8642
非常不滿意	5	10	0.9636

1. 若僅是程度上的差異則建議用standardized midrank score
2. 相當於執行無母數分析(Y:滿意度分數)

Chi-square tests

如何做表?

X \ Y	reflex		Non-reflex		
	n	%	n	%	total
Light-eyed	542	49.0%	564	51.0%	1106
Dark-eyed	312	47.7%	342	52.3%	654
total	854	48.5%	906	51.5%	1760

$\chi^2_{1df}=0.28$, $p=0.6000$

Sample presentation:

From the sample of 1760 patients, 542 of the 1106 (49.0%) light-eyed participants and 312 of 654 (47.7%) dark-eyed participants exhibited the reflex response. The chi-square test revealed that reflex response and eye color were not statistically significantly associated($\chi^2_{1df}=0.28$, $p=0.6000$).

From: Lang & Secic, How to report statistics in medicine. 2nd (2006)

Community survey

Whether X genotype related to oral cancer/precancer?

		Cancer		PreCancer		Normal		p-value of
	total	n	%	n	%	n	%	chi-square
total	213	104	48.83	21	9.86	88	41.31	
GG	60	30	50.00	8	13.33	22	36.67	0.4595
TG	98	51	52.04	9	9.18	38	38.78	
TT	55	23	41.82	4	7.27	28	50.91	

% horizontally sum up to 100%

Subjects with “TT” genotype had higher percentage free of diseases (cancer or precancer) (50.91%) then “GG” (36.67%) and “TG” (38.78%). The difference (or association) was not statistically significant (p=0.4595).

Case-control study

Whether X genotype related to oral cancer/precancer?

	total	Cancer n	Cancer %	PreCancer n	PreCancer %	Normal n	Normal %	p-value of chi-square
total	213	104		21		88		
GG	60	30	28.85	8	38.10	22	25.00	0.4595
TG	98	51	49.04	9	42.86	38	43.18	
TT	55	23	22.14	4	19.05	28	31.81	

% vertically sum up to 100%

Normal subjects had higher percentage of "TT" (31.81%) than cancer (22.14%) or precancer (19.05%) patients. The difference (or association) was not statistically significant (p=0.4595).

X vs Y 之交叉分析

Table 1 Clinical parameters in patients with OPL and control subjects

Parameters	Patients with OPL	Betel chewer control subjects	P-value
Age (years)	56.7±11.3	56.4±9.7	0.8839
Gender (male/female)	28/33	26/35	0.7154
BMI (kg/m²)	28.0±5.1	27.4±5.0	0.5336
Systolic BP (mmHg)	144.3±23.4	141.9±25.6	0.5991
Diastolic BP (mmHg)	86.8±13.0	83.1±14.0	0.1493
Betel chewing duration (years)			
1–20	16 (26.2)	16 (26.2)	1.0000
20–30	14 (23.0)	14 (23.0)	
30+	31 (50.8)	31 (50.8)	
Cumulative amount of quid consumption			
1–50 000	11 (18.6)	11 (19.3)	0.2923
50 000–100 000	7 (11.9)	14 (24.6)	
100 000–200 000	14 (23.7)	13 (22.8)	
200 000+	27 (45.8)	19 (33.3)	
Alcohol drinking (%)	45 (73.8)	41 (68.3)	0.5096
Drinking duration (years)	26.3±10.2	26.6±11.4	0.9040
Smoking (%)	17 (27.9)	19 (31.2)	0.6914
Smoking duration (years)	27.5±13.2	29.0±11.8	0.7190

OPL = oral precancerous lesion; BMI = body mass index; BP = blood pressure. Data are expressed as mean±s.d.; comparisons performed by unpaired *t*-test or χ^2 test when appropriate.

Chung et al, British Journal of Cancer (2005) 93:602-606

範例一：各種白斑症病理組織分析結果

	沒有上皮變異 個案數	沒有上皮變異 百分比(%)	輕微或中等的上皮變異 個案數	輕微或中等的上皮變異 百分比(%)	嚴重的上皮變異與原位癌及上皮癌 個案數	嚴重的上皮變異與原位癌及上皮癌 百分比(%)	合計
均質性白斑症	73	77.7	21	22.3	0	0.0	94
疣狀白斑症	6	33.3	7	38.9	5	27.8	18
紅白斑症	4	21.1	10	52.6	5	26.3	19
結節狀白斑症	4	33.3	2	16.7	6	50.0	12
合計	87	60.8	40	28.0	16	4.2	143

X^2 =54.5, df=6, p<0.001

範例二：瑞典扁平苔癬的型態分佈

型態(依總合之百分比排序)	男 病患人數(N)	男 病患中所佔之比例(%)	女 病患人數(N)	女 病患中所佔之比例(%)	總和 病患人數(N)	總和 病患中所佔之比例(%)
合計	249		453		702	
網狀	112	45.0	205	45.3	317	45.2
斑狀	67	26.9	110	24.3	177	25.2
萎縮狀	42	16.9	92	20.3	134	19.1
丘疹狀	19	7.6	20	4.4	39	5.6
潰瘍狀	8	3.2	24	5.3	32	4.6
皰狀	1	0.4	2	0.4	3	0.4

χ^2 =5.97, df=5, p=0.3094

Chi-square tests

Odds Ratios and

Matntel-Haenszel Test

Measures of Effect in 2x2 Tables

Disease vs Exposure

Pe:Exposed者之得病率

Pue:Un-exposed者之得病率

Risk Ratio (RR)= $\frac{Pe}{Pue}$

Odds of exposure= $\frac{Pe}{(1-Pe)}$

Odds of un-exposure= $\frac{Pue}{(1-Pue)}$

Odds ratio (OR)= $\frac{Pe/(1-Pe)}{Pue/(1-Pue)}$

注意:

當Pe&Pue很小時, (1-Pe) → 1, (1-Pue) → 1

則OR= $\frac{Pe/(1-Pe)}{Pue/(1-Pue)} \fallingdotseq \frac{Pe}{Pue} = RR$

Measure of effect size in cross tab

- Odds ratios (OR)
- = (a x d)/(b x c)

	Disease	no disease
exposure	a	b
no exposure	c	d

Compute Odds Ratio (crude)

	Cancer	PreCancer	Normal	Cancer/Normal (Crude)	PreCancer/Normal (Crude)
	n	n	n	OR	OR
GG	30	8	22	1.66	2.55
TG	51	9	38	1.63	1.66
TT	23	4	28	1.00	1.00

	Cancer	PreCancer	Normal	Cancer/Normal (Crude)	PreCancer/Normal (Crude)
	n	n	n	OR	OR
GG	a	b	c	(a x m)/(c x g)	(b x m)/(c x h)
TG	d	e	f	(d x m)/(f x g)	(e x m)/(f x h)
TT	g	h	m	1.00	1.00

Compute Odds Ratio (crude)

	PreCancer/Normal (Crude)						
	OR	(	95%CI			)	p-value
GG	2.55	(	0.68	,	9.57	)	0.1666
TG	1.66	(	0.46	,	5.93	)	0.4371
TT	1.00						

Comparing to genotype “TT”, people with “GG” had 2.55 times (95% CI=2.68, 9.57) of the chance for having precancer.

95% confidence intervals

- Confidence intervals is important for people to see the efficiency
- Can blow up by small cell size

Confounders & Stratification

干擾因素與分類因素

	LC	$\overline{LC}$	
Dr	0.0194		1
$\overline{Dr}$	0.0117		1

BUT!

Smokers

	LC	$\overline{LC}$
Dr	0.03	
$\overline{Dr}$	0.03	

Non-smokers

	LC	$\overline{LC}$
Dr	0.01	
$\overline{Dr}$	0.01	

要去掉干擾或分類因素的影響,可用Mantel-Haenszel Test 做法是將原來的一個表格,依因素分成多個表再針對每一個表算出 χ^2 test statistics之後再整合起來。

相關係數 & 迴歸係數

Correlation Coefficient & Regression coefficient
Chapter 11
Sections 11.1-11.8

Design

		Continuous	Categorical
Outcome (y)	Continuous	Pearson Correlation coeff. Linear regression *Spearman Correlation coeff.*	t-test / ANOVA Linear models *Wilcoxon Rank-Sum test* *Kruskal-Wallis test*
	Categorical	Logistic Categorical analysis	χ^2 Fisher's Exact Categorical analysis Logistic analysis

相關係數(Correlation Coefficient)

要了解兩個數值變項(等距尺度、等比尺度)之相關性時，可以利用皮爾森相關係數(Pearson Correlation Coefficient)來探討，其中母全體的真值以 ρ 來代表。**皮爾森相關係數主要是測量兩變數間之線性(linear)關係，因此兩變項間是具有曲線關係時，皮爾森相關係數則無法測量**。針對變項x與變項y之皮爾森相關係數的樣本值r

計算公式為：

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{[\sum_{i=1}^{n}(x_i-\bar{x})^2][\sum_{i=1}^{n}(y_i-\bar{y})^2]}}$$

若已知 Σx_i, Σx_i^2, Σy_i, Σy_i^2, $\Sigma x_i y_i$則較簡化的計算公式為

$$r=\frac{\sum x_i y_i-\frac{1}{n}(\sum x_i)(\sum y_i)}{\sqrt{[\sum x_i^2-\frac{(\sum x_i)^2}{n}][\sum y_i^2-\frac{(\sum y_i)^2}{n}]}}$$

若已知 $\bar{x}$ ，$\bar{y}$， , S_x , S_y , $\Sigma x_i y_i$ 之計算公式

$$r=\frac{\sum x_i y_i-n\bar{x}\bar{y}}{(n-1)S_x S_y}$$

相關係數是一個-1~1的數字，正值表正相關，負值表負相關，零表沒有相關，離零越遠則相關性越強。一般來說，若相關係數大於0.75則可視為非常相關，0.5~0.75則為普遍相關。

Hypothesis Testing for ρ

①一般統計軟體提供的p-value是當x、 y呈normal distribution要檢定ρ是否不同於0時

②當x、 y不一定呈normal distribution要檢定ρ是否 不同於ρ_0時, 先用 Fisher'Z Transformation of the r, 再作檢定

Pearson Correlation Coefficient常被用來當作regression之前置步驟

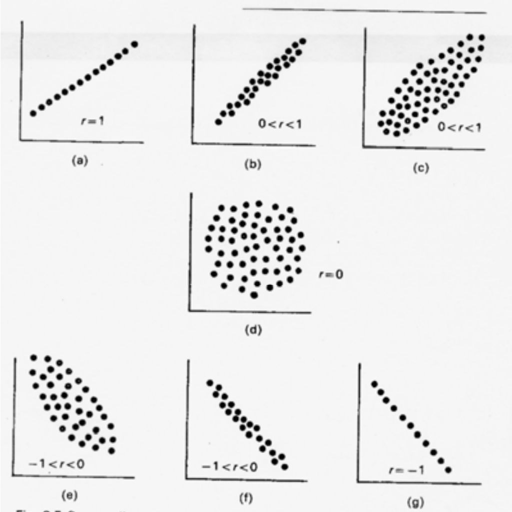


Fig. 8.5 Scatter diagrams and correlation. Typical patterns and the range of the correlation coefficient.

Page 107, Pearson and Turton: Statistical methods in environmental health

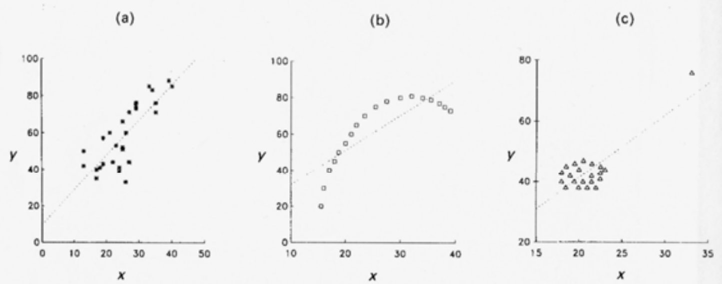


Fig. 8.6 'Correlation coefficient = 0.83'. Three sets of data with the same apparent value of the correlation coefficient, but only set (a) is valid. The other sets violate the conditions for the use of the product moment correlation coefficient.

Page 108, Pearson and Turton: Statistical methods in environmental health

相關係數的表示

Sample Presentation:

Dentene lead levels correlated well and inversely with family income, indicating that poorer children have higher levels of lead in their systems (n=39; Pearson's r=- 0.62; P=0.001).

From: Lang & Secic, How to report statistics in medicine. 2nd (2006)

TABLE 6.2. Sample Correlation Matrix.*

	Variables				
	1	2	3	4	5
	r *P* n	*r* *P* n	*r* *P* n	*r* *P* n	*r* *P* n
Variable 1	—	−0.243[†] 0.20 29	−0.177 0.37 27	0.013 0.94 30	0.009 0.96 30
Variable 2	—	—	−0.226 0.24 28	−0.383 0.03 31	0.038 0.83 31
Variable 3	—	—	—	0.327 0.08 29	−0.119 0.53 29
Variable 4	—	—	—	—	0.289 0.10 32
Variable 5	—	—	—	—	—

*Duplicate cells are usually left blank (indicated by the dashes) to simplify the presentation.
†Here, the correlation for variable 1 and variable 2 is $r = 0.243$ ($P = 0.20$) for the 29 subjects who expressed both variables.
r = correlation coefficient
P = probability value
n = sample size

From: Lang & Secic, How to report statistics in medicine. 2nd (2006)

Original table

	Age	Weight	Oxy	Runtime	RunPulse	RstPulse	MaxPulse
Age	1.0000	-0.2405	-0.3118	0.1952	-0.3161	-0.1509	-0.4149
Weight	-0.2405	1.0000	-0.1628	0.1435	0.1815	0.0440	0.2494
Oxy	-0.3118	-0.1628	1.0000	-0.8622	-0.3980	-0.3994	-0.2367
Runtime	0.1952	0.1435	-0.8622	1.0000	0.3136	0.4504	0.2261
RunPulse	-0.3161	0.1815	-0.3980	0.3136	1.0000	0.3525	0.9298
RstPulse	-0.1509	0.0440	-0.3994	0.4504	0.3525	1.0000	0.3051
MaxPulse	-0.4149	0.2494	-0.2367	0.2261	0.9298	0.3051	1.0000

2 digits

	Age	Weight	Oxy	Runtime	RunPulse	RstPulse	MaxPulse
Age	1.00	-0.24	-0.31	0.20	-0.32	-0.15	-0.41
Weight	-0.24	1.00	-0.16	0.14	0.18	0.04	0.25
Oxy	-0.31	-0.16	1.00	-0.86	-0.40	-0.40	-0.24
Runtime	0.20	0.14	-0.86	1.00	0.31	0.45	0.23
RunPulse	-0.32	0.18	-0.40	0.31	1.00	0.35	0.93
RstPulse	-0.15	0.04	-0.40	0.45	0.35	1.00	0.31
MaxPulse	-0.41	0.25	-0.24	0.23	0.93	0.31	1.00

sorting

	RunPulse	MaxPulse	Oxy	Runtime	RstPulse	Weight	Age
RunPulse	1.00	0.93	-0.40	0.31	0.35	0.18	-0.32
MaxPulse	0.93	1.00	-0.24	0.23	0.31	0.25	-0.41
Oxy	-0.40	-0.24	1.00	-0.86	-0.40	-0.16	-0.31
Runtime	0.31	0.23	-0.86	1.00	0.45	0.14	0.20
RstPulse	0.35	0.31	-0.40	0.45	1.00	0.04	-0.15
Weight	0.18	0.25	-0.16	0.14	0.04	1.00	-0.24
Age	-0.32	-0.41	-0.31	0.20	-0.15	-0.24	1.00

	RunPulse	MaxPulse	Oxy	Runtime	RstPulse	Weight	Age
RunPulse	1.00	0.93	-0.40	0.31	0.35	0.18	-0.32
MaxPulse	0.93	1.00	-0.24	0.23	0.31	0.25	-0.41
Oxy	-0.40	-0.24	1.00	-0.86	-0.40	-0.16	-0.31
Runtime	0.31	0.23	-0.86	1.00	0.45	0.14	0.20
RstPulse	0.35	0.31	-0.40	0.45	1.00	0.04	-0.15
Weight	0.18	0.25	-0.16	0.14	0.04	1.00	-0.24
Age	-0.32	-0.41	-0.31	0.20	-0.15	-0.24	1.00

Pair-wise

Variable	by Variable	Count	Correlation	p-value
Weight	Age	31	-0.2405	0.1925
Runtime	Age	31	0.1952	0.2926
Runtime	Weight	31	0.1435	0.4412
RunPulse	Age	31	-0.3161	0.0832
RunPulse	Weight	31	0.1815	0.3284
RunPulse	Runtime	31	0.3136	0.0858
RstPulse	Age	31	-0.1509	0.4178
RstPulse	Weight	31	0.0440	0.8143
RstPulse	Runtime	31	0.4504	0.0110
RstPulse	RunPulse	31	0.3525	0.0518
MaxPulse	Age	31	-0.4149	0.0203
MaxPulse	Weight	31	0.2494	0.1761
MaxPulse	Runtime	31	0.2261	0.2213
MaxPulse	RunPulse	31	0.9298	<.0001
MaxPulse	RstPulse	31	0.3051	0.0951
Oxy	Age	31	-0.3118	0.0878
Oxy	Weight	31	-0.1628	0.3817
Oxy	Runtime	31	-0.8622	<.0001
Oxy	RunPulse	31	-0.3980	0.0266
Oxy	RstPulse	31	-0.3994	0.0260
Oxy	MaxPulse	31	-0.2367	0.1997

Pair-wise and sorted by r

Variable	by Variable	Count	Correlation	p-value
MaxPulse	RunPulse	31	0.9298	<.0001
RstPulse	Runtime	31	0.4504	0.0110
RstPulse	RunPulse	31	0.3525	0.0518
RunPulse	Runtime	31	0.3136	0.0858
MaxPulse	RstPulse	31	0.3051	0.0951
MaxPulse	Weight	31	0.2494	0.1761
MaxPulse	Runtime	31	0.2261	0.2213
Runtime	Age	31	0.1952	0.2926
RunPulse	Weight	31	0.1815	0.3284
Runtime	Weight	31	0.1435	0.4412
RstPulse	Weight	31	0.0440	0.8143
RstPulse	Age	31	-0.1509	0.4178
Oxy	Weight	31	-0.1628	0.3817
Oxy	MaxPulse	31	-0.2367	0.1997
Weight	Age	31	-0.2405	0.1925
Oxy	Age	31	-0.3118	0.0878
RunPulse	Age	31	-0.3161	0.0832
Oxy	RunPulse	31	-0.3980	0.0266
Oxy	RstPulse	31	-0.3994	0.0260
MaxPulse	Age	31	-0.4149	0.0203
Oxy	Runtime	31	-0.8622	<.0001

簡單線性迴歸
Simple Linear Regression

$y = \alpha + \beta x + e$

y dependent variable,the value of the response variable to be predicted

x independent variable,the explanatory variable used to predict the value of y

α the point at which the regression line crosses the y axis(the y intercept point)

β the slope of the regression line

e is normally distributed with mean 0 and variance σ^2(注意!迴歸分析的normal assumption是在e不是在y)

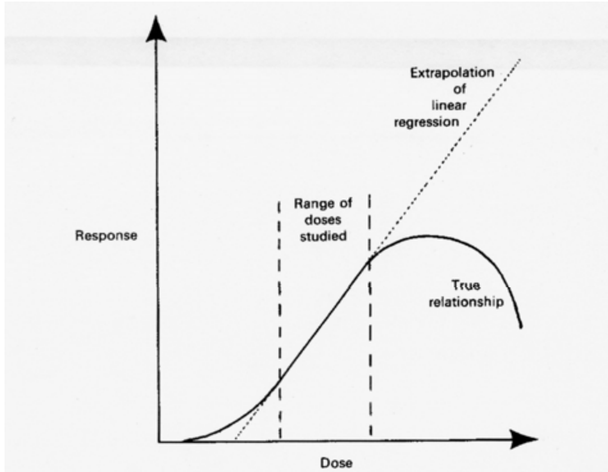


Fig. 8.4 Comparison of a possible true relationship between dose and response with the regression line obtained from a study of a limited range of doses. The linear relationship is adequate within the range of doses studied, but cannot be extrapolated.

Page 106, Pearson and Turton: Statistical methods in environmental health

Estimation of the Least-Squares Line

$y=\alpha+\beta x+e$

其中 α & β 可由統計軟體中計算出來

R^2 the proportion of the variance of y that can be explained by the variable x

在simple linear regression 中square root of R^2就是 Pearson correlation coefficient

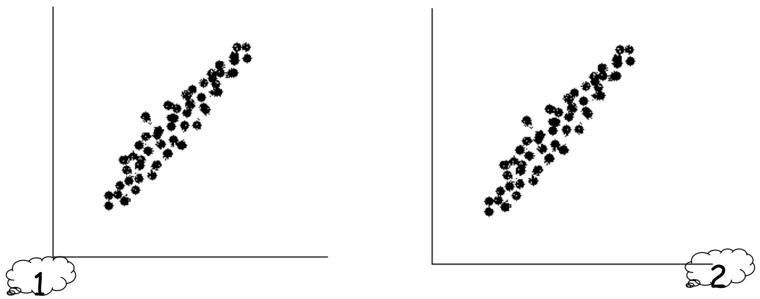
Confidence intervals of parameters
Confidence intervals for prediction

Sample Presentation: From our 453 participants,we attempted to predict serum levels from weight using simple linear regression analysis. The slope of the regression line was significantly greater than zero,indicating that serum level tends to increase as weight increases (slope=0.25; 95%CI=0.19 to 0.31;t_{451}=8.3; $p<0.001$; $y=12.6+0.25X$;R^2=0.67).

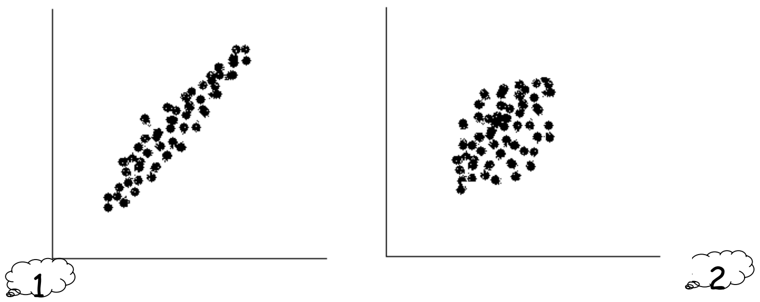
From: Lang & Secic, How to report statistics in medicine. 2nd (2006)

Difference between

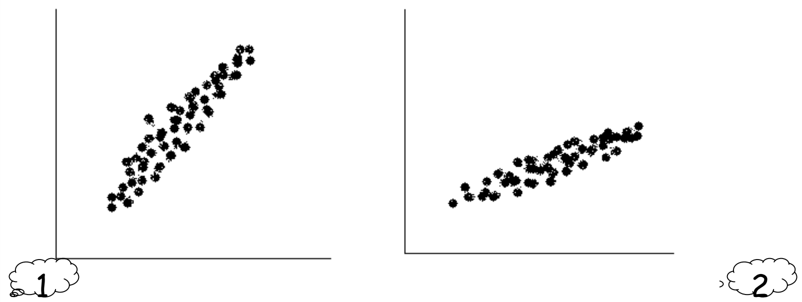
Correlation coefficient (r) &
Regression Coefficient (β)



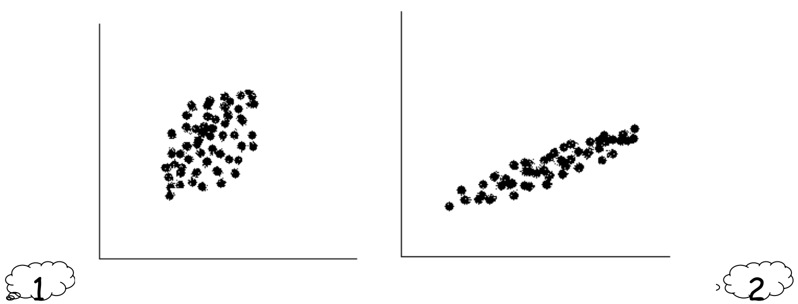
- $r_1 = r_2$
- $\beta_1 = \beta_2$



- $r_1 > r_2$
- $\beta_1 = \beta_2$



- $r_1 = r_2$
- $\beta_1 > \beta_2$



- $r_1 < r_2$
- $\beta_1 > \beta_2$

			Correlation		Regression	
Variable	by Variable	Count	r	p-value	β	p-value
Oxy	Runtime	31	-0.86	<.0001	-3.31	<.0001
Oxy	RstPulse	31	-0.40	0.0260	-0.28	0.0260
Oxy	RunPulse	31	-0.40	0.0266	-0.21	0.0266
Oxy	Age	31	-0.31	0.0878	-0.31	0.0878
Oxy	MaxPulse	31	-0.24	0.1997	-0.14	0.1997
Oxy	Weight	31	-0.16	0.3817	-0.10	0.3817

Any questions?

引用圖文出處:

Rosner: Fundamentals of Biostatistics, 6th. Wadsworth Publishing Company.

公共衛生學: 4th ed., 邱清華總校閱,華杏出版社.

Lang & Secic: How to report statistics in medicine. 2nd ed (2006)

Perason & Turton: Statistical methods in environmental health. Chapman and Hall