MATLAB統計分析工具

盧家鋒 助理教授

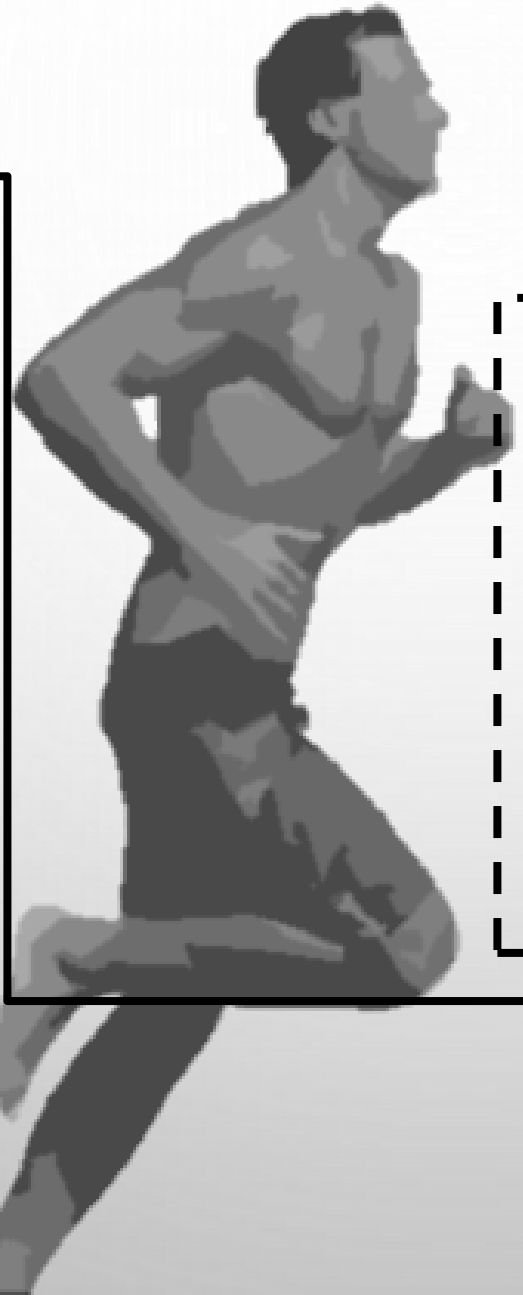
國立陽明大學 物理治療暨輔助科技學系

alvin4016@ym.edu.tw

請先下載本週上課資料

- http://www.ym.edu.tw/~cflu
- 點選左欄〔課程資料〕
- 下載第14週上課資料〔demodata_L12.zip〕，檔案大小約15KB

訊號分析方法



資料輸入(data input)

- 濾波(filtering)
- 重新取樣 (resampling)
- 多變數分析 去雜訊(denoising)
- 訊號切割 (epochs, blocks...)
- 閥值偵測 (thresholding)
- 訊號標準化 (normalizing)
- 事件偵測 (event detecting)
- 其他前處理 (pre-processing)

- 時頻分析 (time-frequency analysis)
- 相關性分析 (coupling/ causality analysis)
- 統計分析 (statistics analysis)

資料儲存/輸出 (data output)

本週課程內容

- 描述性統計
- t檢定
- 變異數分析
- 卡方檢定
- 無母數分析
- 相關分析
- 信度分析

- H_0（null hypothesis）：為介入治療無效
- H_1（alternative hypothesis）：為介入治療有助於功能恢復

	H_0正確	H_0錯誤
拒絕H_0	第一類型錯誤 (α)	正確推論 (1-β)
接受H_0	正確推論 (1-α)	第二類型錯誤 (β)

範例參考書籍：1. SPSS for 生物統計 呂桂雲審閱、張梁治、蔡志堅編著 啟英文化事業有限公司

2. Principles of biostatistics, 2nd edition. M. Pagano and K. Gauvreau. Thomson Learning, Inc.

描述性統計

集中量數與變異量數

請參照demodata_L12\example_descriptive.m

< Central location >

- 算術平均數(mean)
- 中位數(median)
- 眾數(mode)

< variation>

- 全距(range)
- 標準差(standard deviation)

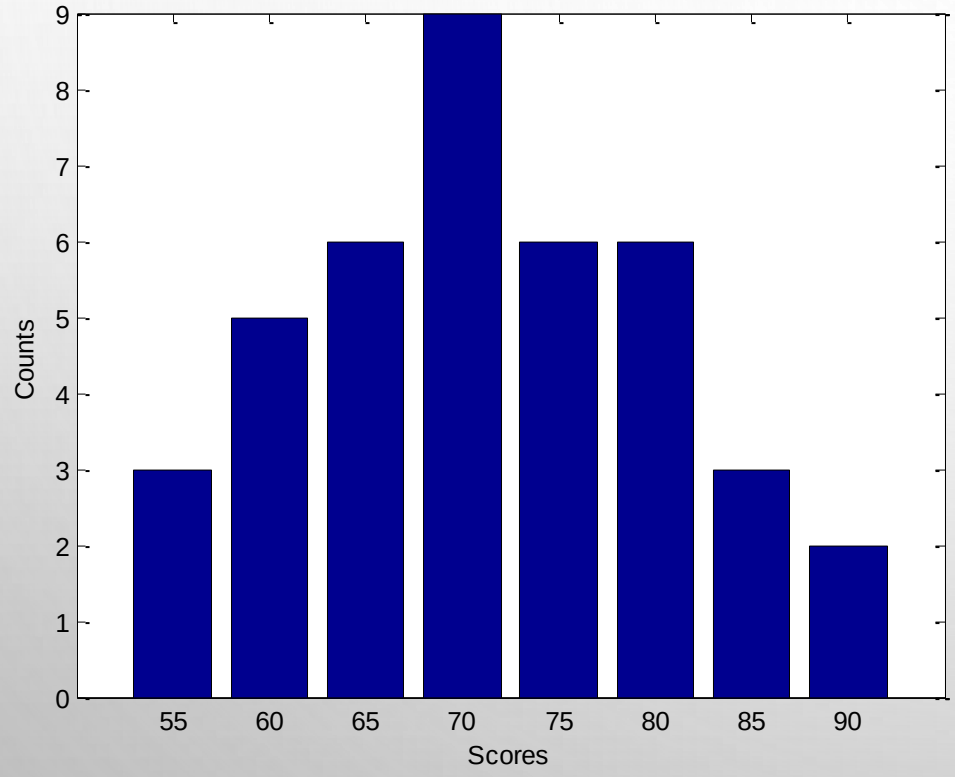
- mean(X,dim)
- median(X,dim)
- [M,F]=mode(X,dim)

- range(X,dim)
- std(X,dof,dim)
  - dof=1➔ 除以N
  - dof=0➔ 除以N-1 (樣本標準差)

次數統計

請參照demodata_L12\example_descriptive.m

- 次數分配長條圖



- [Count,Center]=hist(X,binNumber)
- [Count,Center]=hist(X,binCenter)

t檢定

平均數差異檢定(連續變數)

- 獨立樣本t檢定(two-sample t test)
- 單一樣本t檢定(one-sample t test)
- 成對樣本t檢定(paired-sample t test)

- ttest2(X,Y,alpha)
- ttest(X,mean,alpha)
- ttest(X1,X2,alpha)

✓ 母群體中每一樣本的抽取機率均等且獨立
✓ 樣本為常態分佈
✓ 用於比較的群體分散情形相似

獨立樣本T檢定(TWO-SAMPLE T TEST)

請參照demodata_L12\example_2samplettest.m

- F檢定兩群體變異數同質性
  - [H,P]=vartest2(X,Y,alpha)
- 如兩群體變異數相同➔
  - [H,P,CI,STATS]=ttest2(X,Y,alpha,'vartype','equal')
- 如兩群體變異數不同➔
  - [H,P,CI,STATS]=ttest2(X,Y, alpha,'vartype','unequal')

- 單尾檢定
  - ttest2(X,Y, 'tail','right') (test if X>Y)
- 信賴區間均正或均負為顯著

成對樣本T檢定(PAIRED-SAMPLE T TEST)

請參照demodata_L12\example_pairttest.m

- 成對樣本X1與X2
- [H,P,CI,STATS]=ttest(X1,X2, alpha)
- 單尾檢定
- ttest(X1,X2, 'right') (test if X1>X2)
- 信賴區間均正或均負為顯著

單一樣本T檢定(ONE-SAMPLE T TEST)

請參照demodata_L12\example_1samplettest.m

- 檢定X之平均值是否與M相同
  - [H,P,CI,STATS]=ttest(X, M, alpha)
- 單尾檢定
  - ttest(X,M, 'tail','right') (test if X>M)
- 信賴區間均正或均負為顯著

變異數分析

變異數分析(檢定多個群體/因子的平均數差異)

- 單因子變異數分析(one-way ANOVA)
- 雙因子變異數分析(two-way ANOVA)
- 多因子變異數分析(N-way ANOVA)
- 事後檢定

  (posteriori comparisons, post-hoc)

- anova1(X,group)
- anova2(X,repetition)
- anovan(X,{factor1 factor2 ...})
- multcompare(Stats,'ctype','scheffe')

✓ 母群體中每一樣本的抽取機率均等且獨立
✓ 樣本為常態分佈
✓ 用於比較的群體分散情形相似
✓ 各變異來源的變異量須相互獨立

單因子變異數分析(ONE-WAY ANOVA)

請參照demodata_L12\example_1wayANOVA.m

- 檢定多個群體的平均數差異
- [p,ANOVAstats,Stats]=anova1(X,group)
- **學會輸入引數的格式!!**

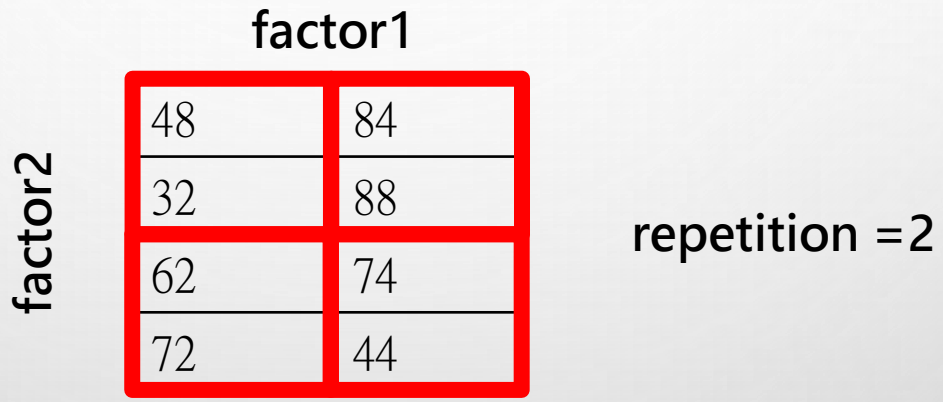
X	18	12	…	15	19	16	…	21	10	16
group	1	1	…	2	2	2	…	3	3	3

- Post-hoc事後檢定
- [comp,m] = multcompare(Stats,'ctype','scheffe')

雙因子變異數分析(TWO-WAY ANOVA)

請參照demodata_L12\example_2wayANOVA.m

- 檢定多個變因群體平均數差異
- [p,table,Stats]=anova2(X,rep)
- **學會輸入引數的格式!!**



- 注意雙因子間是否有交互關係
- 視需要進行Post-hoc事後檢定

多因子變異數分析(N-WAY ANOVA)

請參照demodata_L12\example_NwayANOVA.m 以雙因子為例, 各情況量測數量不同時

- 檢定多個群體的平均數差異
- [p,table,Stats]=anovan(X,{factor1 factor2})
- **學會輸入引數的格式!!**

X	47	45	…	48	49	47	…	49	46	45
factor 1	0	0	…	0	0	1	…	1	1	1
factor 2	2	2	…	1	1	1	…	0	0	0

- Post-hoc事後檢定
- [comp,m] = multcompare(Stats,'ctype','scheffe','dimension',dim)

相依樣本多因子變異數分析 (REPEAT MEASURE N-WAY ANOVA)

請參照demodata_L12\example_rm1wayANOVA.m　demodata_L12\example_rm2wayANOVA.m

- 檢定多個群體的平均數差異　[p,table,Stats]=anovan(X,{factor1 factor2} ,'random',2)
- **學會輸入引數的格式!!**　**（以相依單因子為例）**

X	29	32	…	34	41	36	…	53	50	48
factor 1	1	1	…	2	2	2	…	3	3	3
factor 2 (subject)	1	2	…	1	2	3	…	1	2	3

- Post-hoc事後檢定　[comp,m] = multcompare(Stats,'ctype','scheffe','dimension',1)

卡方檢定/無母數檢定

卡方檢定(Chi-square Test)-類別變數

- 適合度檢定(goodness of fit)
- 百分比同質性檢定
  (homogeneity of proportions)
- 獨立性檢定(independence)
- 改變的顯著性檢定McNemar's
  (significance of change)

- ChiSquareTest(X,alpha)
- mcnemar(X,alpha)

卡方檢定(Chi-square Test)

請參照demodata_L12\example_chi2test.m

- 2x2列聯表

- [H,P,chisquare]=ChiSquareTest(X,alpha)
- [H,P,chisquare]=mcnemar(X, alpha)

	實際	預期
正	52	50
反	48	50

	贊成	否決
男	3	10
女	8	7

	無感染	感染
刷手	47	6
沒刷	5	8

改革前後	公平	不公
公平	15	11
不公	32	18

無母數分析(NON-PARAMETRIC)

請參照demodata_L12\example_signrank.m

- Wilcoxon符號化等級檢定(sign rank)
  - 比較相依樣本差距，中位數是否為0
- [P,H,Stats]=signrank(X1,X2,'alpha',alpha)

請參照demodata_L12\example_ranksum.m

- Wilcoxon等級和檢定(rank sum)
  - 比較獨立兩樣本，中位數是否相等
- [P,H,Stats]=ranksum(X,Y,'alpha',alpha)

✓ 母群體不需要假設為常態分布或近似常態分布

相關分析

相關分析

請參照demodata_L12\example_corrtest.m

- Pearson相關係數
- Spearman等級相關係數
  - 無母數統計方法

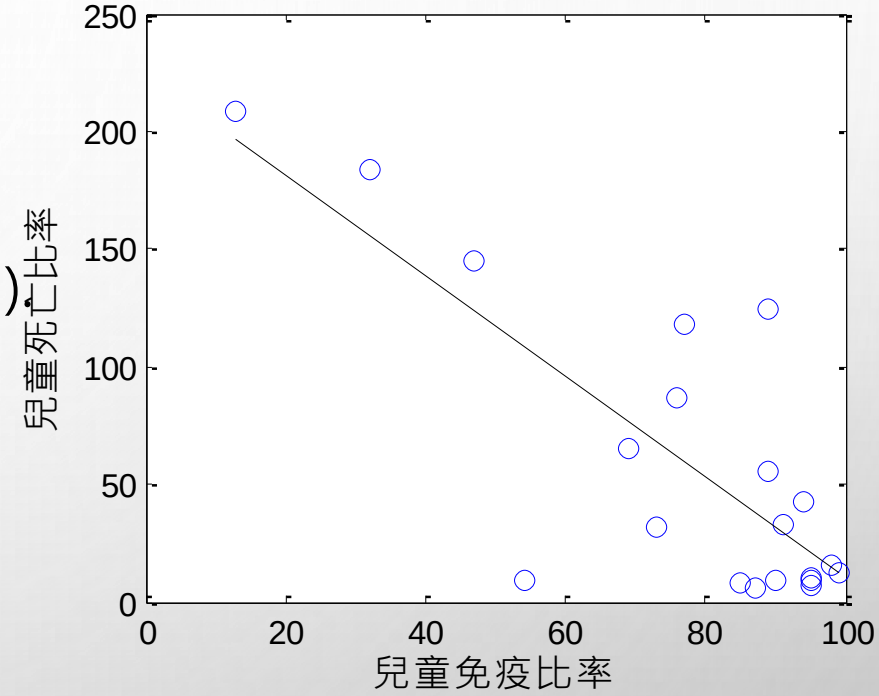
- [R,P]=CORR(X)
- [R,P]=CORR(X,Y)
- [R,P]=CORR(... ,'type','spearman')
- [R,P]=CORR(... ,'type','Kendall')

✓ 兩變相間互為獨立
✓ Pearson相關係數對於異常值較為敏感

線性迴歸

請參照demodata_L12\example_corrtest.m

- P = POLYFIT(X,Y,N)
  - Y= P(1)*X^N + P(2)*X^(N-1) +…+ P(N)*X + P(N+1)
  - 當N=1 ➔ Y=P(1)*X+P(2)



信度分析

組內相關係數(INTRACLASS CORRELATION COEFFICIENT)

- 六種不同模型的ICC

Shrout and Fleiss convention	Name in SPSS
ICC(1,1)	One-way random single measures
ICC(1,k)	One-way random average measures
ICC(2,1)	Two-way random single measures (Consistency/Absolute agreement)
ICC(2,k)	Two-way random average measures (Consistency/Absolute agreement)
ICC(3,1)	Two-way mixed single measures (Consistency/Absolute agreement)
ICC(3,k)	Two-way mixed average measures (Consistency/Absolute agreement)

✓計算重複量測的依變項相關性，兩個變項間可以是相依的（dependent）。
✓可計算小族群的相關性，受試者<15，不會受到小族群的影響。

組內相關係數(INTRACLASS CORRELATION COEFFICIENT)

請參照demodata_L12\example_ICC.m

- 針對同一測量做不同時間點重複量測的信度
  - 假設兩時間點中間無影響測量值的關鍵變因，僅就量測方式之信度做探討

受試者編號	前測	後測
1	10.5	10.1
2	8.4	9.0
⋮	⋮	⋮
10	13.6	13.2

變異來源	ICC(1,1) One-way random single measure	ICC(2,1) Two-way random single measure	ICC(3,1) Two-way mixed single measure
受試者差異	random	random	random
時間點差異	--	random	fixed

THE END

http://www.ym.edu.tw/~cflu

Table 4

Single Score Intraclass Correlation Coefficients (ICCs) for One-Way and Two-Way Models

Definitions of ICCs ρ	Formulas for calculating $\hat{\rho}$	Designation	Interpretation of ICC
	Row effects random		
One-way model *Case 1 model* $\frac{\sigma_r^2}{\sigma_r^2 + \sigma_w^2}$	$\frac{MS_R - MS_W}{MS_R + (k-1)MS_W}$	ICC(1)	The degree of absolute agreement among measurements made on randomly selected objects. It estimates the correlation of any two measurements.
	Column and row effects random (two-way random effects model)		
Two-way models[a] *Case 2 model* $\frac{\sigma_r^2}{\sigma_r^2 + (\sigma_{rc}^2 + \sigma_e^2)}$ *or Case 2A model* $\frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$	ICC(C,1)	The degree of consistency among measurements. Also known as norm-referenced reliability and as Winer's adjustment for anchor points (Winer, 1971). In generalizability theory, this ICC estimates the squared correlation of individual measurements and universe scores.
Case 2 model $\frac{\sigma_r^2}{\sigma_r^2 + \sigma_c^2 + (\sigma_{rc}^2 + \sigma_e^2)}$ *or Case 2A model* $\frac{\sigma_r^2}{\sigma_r^2 + \sigma_c^2 + \sigma_e^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$	ICC(A,1)	The degree of absolute agreement among measurements. Also known as criterion-referenced reliability. Estimates the Type 1 ICC for one-way, unmatched data (Rajartnam, 1960).

Column effects fixed, row effects random (two-way mixed effect model)			
Case 3 model $\frac{\sigma_r^2 - \sigma_{rc}^2/(k-1)}{\sigma_r^2 + (\sigma_{rc}^2 + \sigma_e^2)}$ *or Case 3A model* $\frac{\sigma_r^2}{\sigma_r^2 + \sigma_e^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$	ICC(C,1)	The degree of consistency among measurements made under the fixed levels of the column factor. This ICC estimates the correlation of any two measurements, but when interaction is present, it underestimates reliability.
Case 3 model $\frac{\sigma_r^2 - \sigma_{rc}^2/(k-1)}{\sigma_r^2 + \theta_c^2 + (\sigma_{rc}^2 + \sigma_e^2)}$ *or Case 3A model* $\frac{\sigma_r^2}{\sigma_r^2 + \theta_c^2 + \sigma_e^2}$	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$	ICC(A,1)	The absolute agreement of measurements made under the fixed levels of the column factor.

Note. MS_R = mean square for rows; MS_W = mean square for residual sources of variance; MS_E = mean square error; MS_C = mean square for columns.

[a] In the event of data with a two-way classification for which the column variance is zero (i.e., $\sigma_c^2 = 0$ or $\theta_c^2 = 0$, depending on the model), a one-way model should be used. Thus even though test scores on k parallel tests can be classified by test and test taker, he column variance by definition is zero, which means that a one-way model applies.

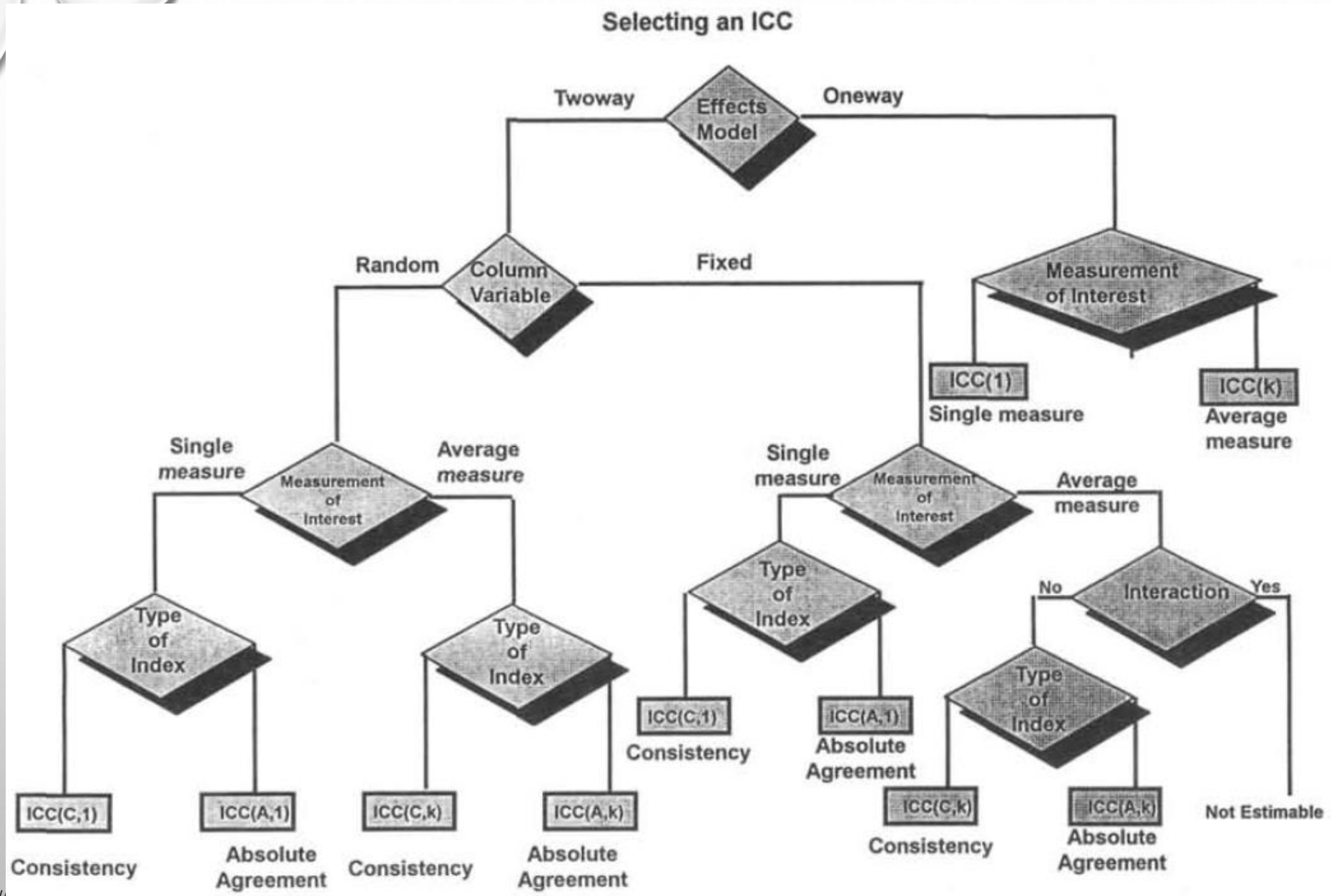


Figure 1. Flow chart for selecting an appropriate intraclass correlation coefficient (ICC).