

平時表現對期中和期末成績的影響

沈侑穎¹ 胡佑熹² 仲偉浩³

¹ 中山大學海洋環境及工程學系

² 大仁科技大學觀光事業系

³ 陸軍軍官學校土木系

摘要

本文將探討學生之平時表現對期中和期末成績的影響，平時表現包括了作業，出席率和平時測驗，期中和期末考採閉卷筆試。為了要瞭解自由度或樣本大小是否影響信賴區間的大小和預測實際族群間平均值的準確性，分別針對中山大學水文學以及陸軍官校工程數學的成績進行分析。分析方法採用 t -檢定,單因子變異數分析(ANOVA)和 Duncan 多重變域檢定，所用分析軟體為 SPSS。

關鍵詞: 自由度， t -檢定，變異數分析，Duncan 多重變域檢定

一、簡介

分析資料取自中山大學和陸軍官校的學生成績，成績包含平時成績，期中考和期末考，除平時成績外均採閉卷考試，以期降低學生在臨考時對外界資訊的倚賴，並凸顯課堂中之學習成效和記憶能力。中山大學雖有點名制以做為平時成績的一部份，但校風自由，陸軍官校則除特殊原因學生均得到課，校風嚴謹，在英語教學和測驗上常採取密集式的訓練。成為鮮明的對比，兩校之取樣成績則均為獨立作答之結果。

若單純求取取樣個數或檢定若干數據的平均值為何，且母體(population)之平均值 μ 和標準偏差(standard deviation) σ 為已知，則引用單位常態分佈一般即可滿足，當 σ 為未知時，通過 t 檢驗常可得到所要的效果，否則兩組數據的差異性用 t -pair 檢定亦足矣。

機率紙和卡方檢定則常用在判定數據是否成常態分佈[1][2]。本研究以 t 檢定和 ANOVA(analysis of variance)[3]為主，迴歸

分析和 Duncan 多重變域檢定[4][5]為輔。ANOVA 的使用前提應滿足常態性，獨立性和同質性，同質性的成立與否可用 Levene 檢定，若非同質則可透過對數或自由度的轉變，再進行同樣的檢定，如 Welch's t -test [6]，Brown-Forsythe [7] 或 Games-Howell 檢定法[5]。

ANOVA 和 t 檢定於使用時應先滿足五個主要假設：尺度連續，取樣隨機，觀測值獨立(independence)，常態性(normality)，同質性(homogeneity of variance) [8]。其中樣本大小(sample size)在 30 以上時常態分佈之假設可忽略而無大礙，而兩樣本大小之比介於 $2/3 \sim 3/2$ 時[8]，ANOVA 和 t -test 也仍有其相當的可信度。

二、研究方法

統計方法一般探討幾個變數(variable)間的關係，或幾個組群(group)是否為顯著相異，常見的前者有如迴歸分析(regression analysis)，後者則如 t 檢驗。 t 檢驗係用來比較連續狀態下兩族群間的平均值，當此

兩族群為獨立時， t 檢驗成為獨立樣本 (independent-samples) t 檢驗，若為探討一組群在兩不同條件下之差異則為成對 (paired) t -test。當組群數為兩個或兩個以上，且欲檢驗之目標為某一自變數 (independent variable) 對這些組群的因變數 (dependent variable) 之影響時，單向 (one way) ANOVA 常被選用[3]，一般視為 t -test 的擴充。

按照前述原則，本研究將引用 ANOVA 來探討組群間平均值的顯著性，從而得知學生的平時表現對期中和期末成績的影響。基本上 ANOVA 是利用組間差異除以組內差異所得之 F 值來研判虛假假設的真與偽，當 ANOVA 得出顯著性時，即表示樣本平均值間有差異存在，虛假假設應予拒絕，並進行事後檢定 (post hoc) 以得知哪些樣本間的差異為顯著。不論新舊，各檢定方法的原理均大同小異並以 t 檢定為基石。

在事後檢定的方法中，Fisher's LSD [9]，Scheffe's S 檢定[10]，Duncan's DMRT [4][5]三種檢定可為代表，均應用於兩兩樣本間的平均值 (μ_i 和 μ_j) 比較，以找出何者不同於其它。其中 LSD 過於寬鬆，在樣本選取極端和組數愈多的情況下特別容易達到顯著 (型一誤差 type-I error)，即此時預設的信賴水準 α 無形中已提升為 $1-(1-\alpha)^k$ ， k =組數 (treatment level)。Scheffe 為了降低 LSD 之 type I error，提出了 LSD 之修正式。Duncan 多重變域檢定 (DMRT) 則為十分普遍且為生物學上廣為應用的統計方法，其嚴格度介於 LSD 和 Scheffe 之間，檢定步驟概述如下[5]：

1. 由小至大排列各處理 (treatment) 之平均值 $\bar{Y}_i$

2. 得出各處理的變距 p 值。

3. 由 MSE 的自由度 ($df=(n-1)k$) 和信賴水準 (α) 求出各 p 值相對應的 r_p 值。

4. 最小顯著範圍 (shortest significant range) $SSR_p = r_p \sqrt{MSE / n}$

5. $|\bar{Y}_i - \bar{Y}_j| < SSR_p$ 則不拒絕虛無假設

由 Duncan 檢定的特性自然可知：如有一個包含幾個均值的不顯著變距存在，這變距內任意兩個均值差都不得為顯著。

SPSS 統計軟體包含了以上各檢定法，本研究引用之。以上各準則中預設的顯著水準 α 均已透過自由度或組數 (treatment level) 被轉換為某個臨界值，不再直接比較 α 和 p -value 的差異。

三、於教學方法之應用

本研究為了要瞭解自由度 (df) 越大時，是否影響信賴區間的大小以及預測實際族群間平均值的準確性，分別針對中山大學水文學以及陸軍官校工程數學的成績進行分析。

實驗一使用中山大學水文學的最後學期分數作為依據，本研究針對全班人數 37 人進行抽樣，分別抽取針對 10、20、30 人，每一項抽樣皆重複 3 次，並根據亂數產生器 (<http://lab.25sprout.com/nrprnd/>) 進行試驗，將抽取後的數據填入表格中，並算出該組平均數與標準差，結果如表 1。

將表 1 中各組所抽取的成績計算出平均值與標準差放置在表單末，之後將平均值、 t 值 (樣本檢驗統計值)、樣本標準差以及樣本數，代入信賴區間公式 (表 2) 進行計算。計算結果分為 μ 的上限與下限，其區間稱為信賴區間，研究發現實際 μ 值 (72.49) 皆有包括在各項抽樣群組的信賴區間內，不過從表 2 中的信賴區間差值，可以觀察到當抽樣樣本數越大時， t 值也越小，所求

得的信賴區間差值也越小，表示當樣本數較多時預測實際分數效果較佳。

此外，將中山大學水文學的成績與出席率進行關聯性分析，首先挑選出期中考、期末考和總成績進行分析，並將出席率分為 70%以下、70~80%和 90%以上三組，其中總成績有 20%是根據出席率以及依照全班課堂表現再進行微調。成績部分是選取期中考、期末考、總成績與調整後總成績共 4 項，先進行描述性統計，將個數、最小值、最大值、平均數和標準差統計如表 3，最後將出席率與期中考、期末考、總成績和調整後總成績進行單因子變異數分析 (ANOVA)。

從單因子變異數分析 (ANOVA) 發現出席率與期中考 (顯著性 0.982)、期末考 (顯著性 0.970) 並沒有相關性，反而只有總成績 (顯著性 0.004) 與調整後總成績 (顯著性 0.004) 具有相關性 (表 4)。進而觀察出席率 70%以下、70~80%和 90%以上與總成績的關係，透過 Duncan 檢定發現出席率 70%以下在總成績上的表現的確比其他兩群的表現還要差 (表 5)，同樣在調整後總成績的分析也發現 70%以下的表現較差 (表 6)。研究發現出席率與期中、期末考無關的可能原因為課堂學生均有學長姐提供歷屆講義、考古題，於是學生自行學習的狀態下，與親臨課堂學習者在學習表現上並無太大差異，不過因為出席率會納入部分總成績，導致出席較低者總成績的表現也較差。

實驗二使用陸軍官校學的成績作為依據，本研究針對全班人數 27 人進行抽樣，分別抽取針對 10、20 人，每一項抽樣皆重複 3 次，並根據亂數產生器 (<http://lab.25sprout.com/nrprnd/>) 進行試驗，將抽取後的數據填入表格中，並算出該組平均數與標準差，結果如表 7。

將表 7 中各組所抽取的成績計算出平

均值與標準差放置在表單末，之後將平均值、 t 值 (樣本檢驗統計值)、樣本標準差以及樣本數，代入信賴區間公式進行計算 (表 8)。計算結果分為 μ 的上限與下限，其區間稱為信賴區間。研究發現實際 μ 值 (80.35) 皆有包括在信賴區間內，不過從表 8 中抽樣 10 人的信賴區間差值 (12.00、12.57、18.57)、抽樣 20 人的信賴區間差值 (11.39、10.62、9.44)，研究從這兩組抽樣發現標準差可能隨著抽樣的不同而造成信賴區間差值可能有較大的變化，建議可以多抽樣幾組將標準差異較大的組別與以剔除，不過從這兩組數據中也能發現信賴區間差值，隨著抽樣數增加時，區間差值也會減少，能夠有效預測實際分數效果較佳。

研究額外分析，由陸軍官校工程數學的平時成績與總成績進行關聯性分析，因為官校生的出席率皆為 100% 所以改以平時成績作為平時課堂表現，平時成績分為 70~80 分、80~90 分、90 分以上三組進行分析，分析結果如表 9 發現顯著性為 0.20 (>0.05)，表示這三群平時在總成績的表現並沒有顯著差異，而我們也從平時成績的 Duncan 檢定分群發現，平時成績在 80~90 分在總成績的表現上還比 70~80 分還低 (表 10)，雖然在檢定上並沒有差異，不過可以將此現象列為後續觀察研究。

研究初步將平時成績分為三群，不過並沒有達到顯著差異，所以接續使用線性回歸分析總成績與平時成績顯著 0.03 (<0.05) (表 11)，兩者有達到線性相關，相關係數達 0.52 (即表 12 中 B 估計值下方之數字) 表示總成績與平時成績達到中度相關，關係式為：總成績 = 平時成績 * 0.52 + 38.37。另外同樣一組數據從 excel 所繪製而成的圖 (圖一)，所求得的方程式 $y = 0.525x + 38.37$ ，也與 SPSS 線性迴歸式相同。

四、結果與討論

在教學上評定學生之成績不外乎平時考核(包括了出席率/平時考/作業/報告之撰寫/實驗等等)、期中考和期末考。考量陸軍官校之常態出席率為 100%，故其平時成績將改為作業成績計算，而中山大學之平時成績仍以個人之出席率計之。為加深學生之學習印象，於課堂之重點處(常與期中或期末考試相關聯)常會予以特別叮嚀，又不論中山大學或陸軍官校均有考古題或題庫以供參考。此前提下，出席率或作業成績是否能忠實的反應在期中考或期末考上則有待商榷，也就是平時成績不高者(出席率不高或作業成績不高)可以在考試前極端努力或大量習做考古題而將考試成績

拉高，致使平時成績和期中或期末考之成績不存在正相關性----如本文所述，但此為一般教學之常態。平時成績若改以抽考計之，而期末和期中考試並無考古題/題庫之存在，則直覺上應能提升二者之正相關性，然此法較難激勵學生而學習效果或又不彰。

從實驗一與實驗二這兩個案例證明自由度(df)越大時，的確會讓信賴區間變小，藉由本研究的統計方法發現當樣本數較多時，學生分數的信賴區間將會較小，進而提升預測實際分數的準確性，再根據老師判斷其他表現將遺漏值填入，進而求出平均值。此方法有助於解決過去紙本資料因為年代久遠或是儲存方式疏忽造成分數遺漏，導致缺乏在評鑑上的完整性。

表 1:中山大學水文學的成績與抽樣成績

編號	總成績	抽樣 10 人 (1)	抽樣 10 人 (2)	抽樣 10 人 (3)	抽樣 20 人 (1)	抽樣 20 人 (2)	抽樣 20 人 (3)	抽樣 30 人 (1)	抽樣 30 人 (2)	抽樣 30 人 (3)
1	59			59	59			59	59	59
2	75					75	75	75	75	75
3	70		70			70	70		70	
4	83						83	83	83	83
5	52				52	52			52	52
6	57				57		57	57	57	
7	60	60	60	60		60	60		60	60
8	60						60	60	60	60
9	84	84			84	84	84	84	84	
10	81				81		81	81	81	
11	78	78		78			78		78	78
12	78						78	78	78	78
13	78		78	78			78	78	78	78
14	83	83				83	83	83	83	83
15	67						67	67	67	67
16	73		73			73	73	73	73	73
17	76	76	76		76			76	76	76
18	71			71	71	71		71	71	
19	71	71					71	71	71	71

20	77				77		77	77	77	77
21	74	74	74		74	74		74		74
22	79	79	79		79	79		79	79	79
23	85	85			85	85		85	85	85
24	79				79	79		79		
25	75	75		75		75		75	75	75
26	72				72	72		72		72
27	67						67			67
28	76				76	76				76
29	73			73	73				73	73
30	68				68	68	68	68	68	68
31	60			60	60	60		60	60	60
32	74		74	74				74	74	74
33	74					74		74	74	74
34	83			83			83	83		83
35	58		58		58	58	58	58	58	58
36	73		73		73			73	73	
37	79				79	79		79		79
平均	72.49	76.50	71.50	71.10	71.65	72.35	72.55	73.53	71.73	72.23
標準差		7.38	7.09	8.54	9.60	8.96	8.90	8.09	8.96	8.49

表 2:中山大學抽樣群組的信賴區間公式

信賴區間公式	$\bar{X} - t \frac{S}{\sqrt{n}} < \mu < \bar{X} + t \frac{S}{\sqrt{n}}$ $\bar{X}$：樣本平均值；t：樣本檢驗統計值； S：樣本標準差；n：樣本數；μ：母體平均值的信賴區間						
	$\bar{X}$	$t(\alpha=0.05)$	S	n	μ		區間差值
					下限	上限	
抽樣 10 人(1)	76.50	2.262	7.38	10	71.22	81.78	10.56
抽樣 10 人(2)	71.50	2.262	7.09	10	66.43	76.57	10.14
抽樣 10 人(3)	71.10	2.262	8.54	10	64.99	77.21	12.22
抽樣 20 人(1)	71.65	2.093	9.6	20	67.16	76.14	8.99
抽樣 20 人(2)	72.35	2.093	8.96	20	68.16	76.54	8.39
抽樣 20 人(3)	72.55	2.093	8.9	20	68.38	76.72	8.33
抽樣 30 人(1)	73.53	2.045	8.09	30	70.51	76.55	6.04
抽樣 30 人(2)	71.73	2.045	8.96	30	68.38	75.08	6.69
抽樣 30 人(3)	72.23	2.045	8.49	30	69.06	75.40	6.34

表 3:描述性統計

	個數	最小值	最大值	平均數	標準差
期中考	37	50.0	90.0	68.3	10.0
期末考	37	28.0	87.0	61.7	13.0
總成績	37	49.0	81.8	69.3	8.4
調整後總成績	37	52.0	84.8	72.3	8.4
出席率	37	21%	92%	76%	17%

表 4:單因子變異數分析

		平方和	自由度	平均平方和	F	顯著性
期中考	組間	4.0	2	2.0	.019	.982
	組內	3662.7	34	107.7		
	總和	3666.7	36			
期末考	組間	11.0	2	5.5	.031	.970
	組內	6084.6	34	178.9		
	總和	6095.7	36			
總成績	組間	707.7	2	353.8	6.453	.004
	組內	1864.5	34	54.8		
	總和	2572.2	36			
調整後總成績	組間	707.7	2	353.8	6.453	.004
	組內	1864.5	34	54.8		
	總和	2572.2	36			

表 5:總成績的 Duncan 檢定分群

	出席率	個數	alpha = 0.05 的子集	
			1	2
Duncan 檢定	70%以下	9	62.1	
	70~80 %	17		70.3
	90%以上	11		73.8
	顯著性		1.000	.268

表 6:調整後總成績的 Duncan 檢定分群

	出席率	個數	alpha = 0.05 的子集	
			1	2
Duncan 檢定 a,b	70%以下	9	65.1	
	70~80 %	17		73.3
	90%以上	11		76.8
	顯著性		1.000	.268

表 7:陸軍官校工程數學的成績與抽樣成績

編號	平時平均	學期成績	抽樣 10 人(1)	抽樣 10 人(2)	抽樣 10 人(3)	抽樣 20 人(1)	抽樣 20 人(2)	抽樣 20 人(3)
1	70.00	64.20		64.20		64.20	64.20	64.20
2	83.60	75.48	75.48			75.48	75.48	75.48
3	79.00	84.00						84.00
4	83.00	94.10		94.10	94.10	94.10		94.10
5	89.60	85.78				85.78	85.78	85.78
6	99.00	98.70	98.70			98.70	98.70	
7	94.00	86.10		86.10				86.10
8	86.00	90.80	90.80		90.80		90.80	90.80
9	85.00	86.50		86.50	86.50	86.50	86.50	86.50
1	79.00	85.30	85.30	85.30		85.30	85.30	
11	98.00	59.50				59.50		59.50
12	81.67	54.20			54.20	54.20	54.20	
13	82.60	74.58			74.58	74.58	74.58	74.58
14	98.60	97.58					97.58	97.58
15	83.00	84.70	84.70	84.70		84.70		84.70
16	95.00	89.70				89.70	89.70	89.70
17	72.00	89.20			89.20	89.20		89.20
18	88.00	81.80	81.80	81.80		81.80	81.80	81.80
19	96.60	87.78	87.78			87.78	87.78	87.78
2	87.00	93.00			93.00		93.00	93.00
21	87.60	86.68				86.68	86.68	
22	70.40	72.92		72.92		72.92	72.92	
23	95.00	96.20	96.20		96.20	96.20	96.20	96.20
24	76.00	71.70	71.70	71.70			71.70	
25	88.00	81.00		81.00		81.00	81.00	
26	94.00	95.90			95.90	95.90		95.90
27	71.60	82.98	82.98		82.98		82.98	82.98
平均		80.35	85.54	80.83	85.75	82.21	82.84	84.99
標準差		18.02	8.39	8.78	12.91	12.17	11.34	10.08

表 8:陸軍官校抽樣群組的信賴區間公式

信賴區間公式	$\bar{X} - t \frac{S}{\sqrt{n}} < \mu < \bar{X} + t \frac{S}{\sqrt{n}}$ $\bar{X}$：樣本平均值；t：樣本檢驗統計值； S：樣本標準差；n：樣本數；μ：母體平均值的信賴區間						
	$\bar{X}$	$t(\alpha=0.05)$	S	n	μ		區間差值
					下限	上限	
抽樣 10 人 (1)	85.54	2.262	8.39	10	79.54	91.55	12.00
抽樣 10 人 (2)	80.83	2.262	8.78	10	74.55	87.12	12.57
抽樣 10 人 (3)	85.75	2.262	12.91	10	76.51	94.98	18.47
抽樣 20 人 (1)	82.21	2.093	12.17	20	76.51	87.91	11.39
抽樣 20 人 (2)	82.84	2.093	11.34	20	77.53	88.15	10.62
抽樣 20 人 (3)	84.99	2.093	10.08	20	80.27	89.71	9.44

表 9:總成績與平時成績的單因子變異數分析

學期成績					
	平方和	自由度	平均平方和	F	顯著性
組間	417.47	2	208.73	1.71	.20
組內	2918.83	24	121.61		
總和	3336.31	26			

表 10:總成績與平時成績的 Duncan 檢定

組別	個數	alpha = 0.05 的子集
		1
70~80分	7	78.61
80~90分	12	82.38
90分以上	8	88.93
顯著性		.07

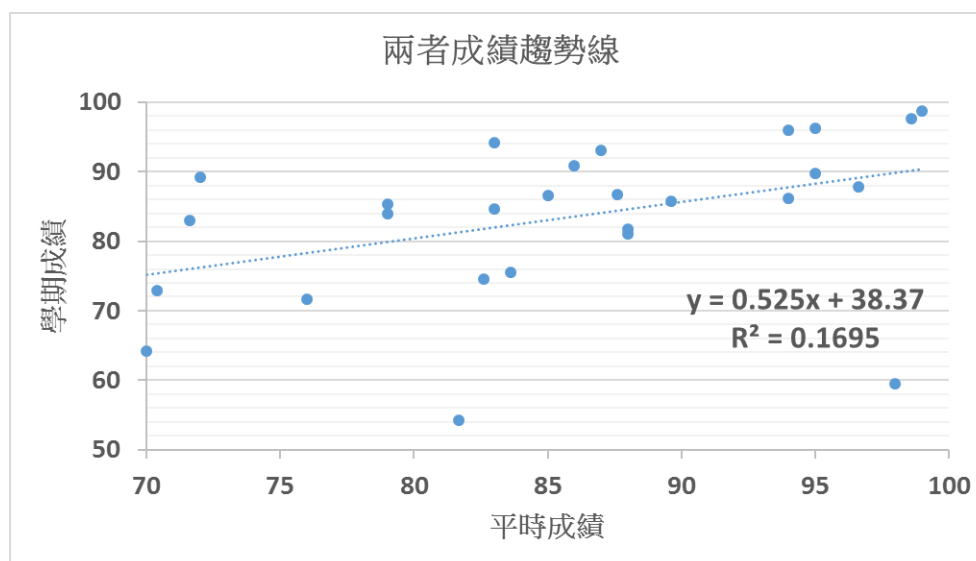
表 11:總成績與平時成績的 Anova 檢定

模式		平方和	df	平均平方和	F	顯著性
	迴歸	565.39	1	565.39	5.10	.03
	殘差	2770.91	25	110.83		
	總數	3336.31	26			

表 12:總成績與平時成績的回歸係數

模式		未標準化係數		標準化係數	t	顯著性
		B 之估計值	標準誤差	Beta 分配		
	(常數)	38.37	20.01		1.91	.06
	平時平均	.52	.23	.41	2.25	.03

圖一:平時成績與學期成績趨勢線



五、結論

從實驗一與實驗二這兩個案例證明自由度(df)越大時，的確會讓信賴區間變小，藉由本研究的統計方法發現當樣本數較多時，學生分數的信賴區間將會較小，進而提升預測實際分數的準確性，老師可再判斷其他表現將遺漏值填入，進而求出平均值。此方法有助於解決過去紙本資料

因為年代久遠或是儲存方式疏忽造成分數遺漏，導致缺乏在評鑑上的完整性。

從單因子變異數分析(ANOVA)發現中山大學海工系水文學之出席率與期中考(顯著性 0.982)、期末考(顯著性 0.970)並沒有相關性，反而只有總成績(顯著性 0.004)與調整後總成績(顯著性 0.004)具有相關性。研究發現出席率與期中、期末考

無關的可能原因為課堂學生均有學長姐提供歷屆講義、考古題，於是學生自行學習的狀態下，與親臨課堂學習者在學習表現上並無太大差異，不過因為出席率會納入部分總成績，導致出席較低者總成績的表現也較差。

因為官校生的出席率皆為 100% 所以改以平時成績作為平時課堂表現，平時成績分為 70~80 分、80~90 分、90 分以上三組進行分析，分析結果發現顯著性為 0.20(>0.05)，表示這三群平時在總成績的表現並沒有顯著差異。接續使用線性回歸分析發覺總成績與平時成績顯著 0.03(<0.05) (表 11)，兩者有達到線性相關，相關係數為 0.52，表示總成績與平時成績達到中度相關，關係式為：總成績=平時成績*0.52+38.37。另外同樣一組數據從 excel 所繪製而成的圖，所求得的方程式 $y = 0.525x + 38.37$ ，也與 SPSS 線性迴歸式相同。

參考文獻

- [1] Chow V. T., D. R. Maidment, L. W. Mays, *Applied Hydrology*, McGraw-Hill Pub. Co., New York, 1988.
- [2] Linsley R. K., J. B. Franzini, D. L. Freyberg, G. Tchobanoglous, *Water Resources Engineering*, 4 ed., New York, 1995.
- [3] Ross M. S., *Introduction to probability and statistics for engineering and scientists*, John Wiley & Sons, New York, 1987.
- [4] Duncan, D. B. (1955). "Multiple range and multiple F tests". *Biometrics*. 11: 1–42. doi:10.2307/3001478.
- [5] 李玉玲，國立中山大學開放式課程，統計生物學 NSYSU Open course ware.
- [6] Welch, B. L. (1947). "The generalization of "Student's" problem when several different population variances are involved". *Biometrika*. 34 (1–2): 28–35. doi:10.1093/biomet/34.1-2.28. MR 19277.
- [7] Brown, Morton B.; Forsythe, Alan B. (1974). "Robust tests for equality of variances". *Journal of the American Statistical Association*. 69: 364–367. doi:10.1080/01621459.1974.10482955. JSTOR2285659.
- [8] Allen & Unwin, *SPSS Survival Mannul, A step by step guide to data analysis using SPSS for Windows (Version 12)*, Julie Pallant, 2nd edn., 2005, www.allenandunwin.com/spss.htm.
- [9] Fisher, Ronald. *Studies in Crop Variation*. I. An examination of the yield of dressed grain from Broadbalk (PDF). *Journal of Agricultural Science*. 1918, 11: 107–135.
- [10] Scheffé, Henry. *The Analysis of Variance*. New York: Wiley. 1959.

The influence of usual performance on the grades of midterm and final exams

Yo-Inn Shen¹ , Yu-Shi Hu² , Weihao Chung³

¹Department of Marine Environment and Engineering, Sun Yet-sen University

²Department of Tourism Management, Tajen University

³Department of Civil Engineering, Chinese Military Academy

Abstract

This paper aims at discussing the influence of students' usual performance on the midterm and final exams. The usual performance includes homework, appearance rate, and usual tests. The midterm and final exams are chosen to be closed-book. In order to understand whether the degree of freedom or sample size will change the confidence interval size, and meanwhile to predict the accuracy of mean of the actual population, analysis is performed respectively on the grades of hydrology in Sun Yet-sen University and the grades of engineering mathematics in Chinese Military Academy. Analysis methods adopt t-test, single factor analysis variance (ANOVA) and Duncan's multiple range test. The analysis software employed is SPSS.

Keywords: degree of freedom, *t*-test, ANOVA, Duncan's multiple range test