Suggested Exercises from M&M Chapter 7 *Homegrown exercises begin on page 3*

These pages were updated on Oct 12

To start with, do some of the odd-numbered exercises. answers to all odd-numbered exercises are given on textbook pages S-1 onwards.

Do some or all of the following even-numbered exercises. You are asked to hand in answers to designated ones.. see the list, and the deadline, on the main course page. Some of these will be discussed in tutorials or answers to them posted on the course web page

§ 7.1	§ 7.2
7.2	7.48
7.4	7.49
7.9	7.50
7.10	7.51
7.20	7.52
7.21	7.57
7.22	7.61
7.24	7.62
7.26	7.63
7.30	7.72
7.32	7.89 (use table from p 11 of notes to get an idea of the
7.33	statistical power) or use spreadsheet provided
7.34	under Resources for Ch 7)
7.38	7.98
7.40	7.100
	7.101
	7.102
	7.103
	7.104(a)
	7.118 (could use TINV function in Excel)

Suggested Exercises from M&M Chapter 7 *Homegrown exercises begin on page 3*

Options for data analysis by EXCEL (see e.g.'s on Resources page)

1-sample (or paired-sample) data

Under your control...

• Calculate mean & SD (s) by AVERAGE & STDEV formulae
• Calculate SEM, i.e., s/SQRT[n]

100(1 – α)% CI
• Determine t* based on α and df (TINV function)
• Note that TINV function is 2-sided

P-Value (Evidence against H_0 and in favour of H_{alt})
• Calculate ratio (test statistic)
• P-Value = Prob (Ratio this -or more - favourable to H_{alt} | H_0)
• Can use the TDIST(ratio, df, # tails) function.

Formal TEST of H_0 vs H_{alt}, with pre-set α
Compare P-value with pre-set "Threshold of positivity" / "level of extremeness" (α).

If you let Microsoft do it all for you...

• Use TTEST function (if 1 sample you will need to have a corresponding array containing the test value μ_0; if paired samples, data arrays must be "paired")

• Use T-test from Toolpak under Data Analysis under Tools menu
• Reconstruct CI from the output

2-sample (unpaired) data

3 options

By yourself...

use AVERAGE & STDEV formulae for each sample
calculate SEM's and SE(difference of two means)
CI: diff in means ± t* SE[diff in means] ... t* by TINV fn.
Test: Calculate ratio and use TDIST function to get P-value

Use inbuilt TTEST function (P-value only .. no CI)

Use t-test in Toolpak; & output to construct CI (cf Excel e.g. in Resources)

Use modern approach of representing a difference of means as a regression coefficient of an indicator (0/1) variable representing the two groups (cf same Excel e.g. in Resources) Can do so using Regression in Toolpak

Options for data analysis by SAS

1-sample (or paired-sample) data

INSIGHT

Analysis Menu

-> Distribution

Highlight response variable and indicate that it is the "Y" variable

Request/turn off extra output via "Output" options

or, after you have said OK, via Tables menu

(In my version 6, I cannot specify a "μ_0" via output, whereas I can via Tables)

Program Editor

PROC MEANS if single column or column of differences
(type HELP MEANS in command line, or cf PROC MEANS in HELP)

See example in Resources in Ch 7

2-sample (unpaired) data

INSIGHT

See Example in Ch 7 Resources, which uses the modern approach of representing a difference of means as a regression coefficient of an indicator (0/1) variable representing the two groups

Program Editor

PROC TTEST
(type HELP TTEST in command line, or see TTEST under help menu)

See example in Resources in Ch 7

Can also use regression approach in Program Editor -- see example

"Homegrown" Exercises around M&M Chapter 7

-1- Respiratory health of adults living near sour gas wells in Southern Alberta [Spitzer et al 1985]

Suppose that the study of the respiratory health of adults living near sour gas wells in Southern Alberta had not included a 'control' area but that, instead, the authors had used for comparison some published data from a VERY LARGE study of adults somewhere else in Canada or the US, where the mean FEV_1 in a similar age group was found to be 3.9 litres (SD: 0.95 litres) for males and 3.00 litres (SD: 0.70 litres) for females.

a Compare the FEV_1's of the adults in the index area with those found in the external "control" population (female students analyze females; male students analyze males)

b Is the difference statistically significant?

c If it were, how would you interpret it to the residents of the index area?

d If it were not, what would you infer and what other calculations might you do?

e Repeat a-d using the internal comparison population and using the computer printout (excerpted from SAS) below:

sex	area	n	Mean	SD	Var
m	control	52	3.97	0.92	0.8401
m	index	48	3.70	1.01	1.0350
f	control	45	2.95	0.65	0.4242
f	index	46	3.07	0.67	0.4469

-2- How big a sample?

A dental epidemiologist wishes to estimate the mean weekly consumption of sweets (candy) among children of a given age in his area. After devising a scheme which enables him to determine the sweet consumption of a child during a week, he does a pilot survey and finds that the standard deviation of the weekly sweet consumption is about 3 ounces. He considers taking a random sample of either 25 children, or 100 children, or 625 children for the main survey. Estimate the standard error of the sample mean for each of these three sample sizes. How do these standard errors explain why large samples are more reliable than small samples?

-3- Age at Menarche

In a study of the age of menarche in women in the U.S.A., the following distributions were observed for 2 samples of women born in the 1930's and the 1950's respectively. The ages at menarche quoted are to the nearest whole number of years. Test the hypothesis that there is no difference in the average age of menarche between the two groups of women. Calculate a 95% confidence interval for the true difference.

Age at menarche	Women born in 1930's	Women born in 1950's
10	0	3
11	2	11
12	8	28
13	14	23
14	27	12
15	5	1
16	8	0
17	1	0
18	1	0
Total	66	78

See Resources for Ch 7 (from main web page) for how to do this efficiently in SAS, or SAS/INSIGHT.

-4- THE INFANT SEAT AS TREATMENT FOR GASTROESOPHAGEAL REFLUX

Susan R. Orenstein, MD, P F. Whitington, MD, and D M. Orenstein, MD
(N Engl J Med 1983; 309:760-3.)

Abstract:

Positioning in the infant seat ("chalasia chair") for treatment of infants with gastroesophageal reflux is presumed to have a beneficial effect. We undertook a controlled, prospective study of such positioning to evaluate this purported benefit. Nine infants with documented gastroesophageal reflux participated in 18 paired two-hour postprandial trials in an infant seat and in the horizontal prone position.

Distal esophageal pH monitoring demonstrated longer exposure to gastroesophageal reflux while infants were in the seat than when they were prone (28.2 ± 6.4 per cent vs. 12.8 ± 3.7 per cent of total time with pH < 4.0, P = 0.023), a difference due largely to more episodes (16.0 ± 2.4 vs. 10.1 ± 2.3 per two-hour postprandial period, P = 0.002).

We conclude that the infant seat, rather than being therapeutic in gastroesophageal reflux in children under six months of age, is actually detrimental, when compared with simply placing an infant prone.

See table in next column

a From the four measures on which the 'seat' and 'prone' method were compared, what is your impression of the benefit of the seat?

b Perform a formal statistical test on the 'GER time' to see if your initial impressions are borne out.

c What do the S.E.M.'s given at the bottom of each column represent? Are they of any use in this trial?

Table 1: Clinical data and results of paired trials comparing the effects of the infant seat (S) and the prone position (P) on GastroEsophageal Reflux (GER)

		Prelim studies		---Seat trial----				---Prone trial---				Order
		E	M	GER Time	GER Time	epi >5	epi >5	max	GER Time	epi >5	epi >5	max	
		gr	mmHg	%	%	#	#	min	%	#	#	min	
1	VS	0	11.7	19.3	14	18	0	3	17	11	0	4	S->P
2	S	4	31.8	62.5	62	28	5	11	16	19	1	5	S->P
3	VSI		12.7	15.5	18	16	0	3	12	14	0	3	P->S
4	VPI	0	15	13.5	9	7	0	3	3	4	0	1	S->P
5	VPFIA	4	11.5	16.5	49	25	3	13	16	15	0	4	S->P
6	VPF	4	19	21.5	18	8	0	9	4	2	1	5	S->P
7	VPI	4	12.7	25.3	10	15	1	2	8	7	0	4	S->P
8	S		15.3	24.5	30	11	3	8	1	1	0	1	S->P
9	VFPI	0	9.5	66	44	16	3	8	38	18	3	9	P->S
	Mean		15.5	29.4	28	16	1.7	6.7	12.8	10	0.6	4	
	S.E.M.		2.2	6.7	6.4	2.4	0.6	1.3	3.7	2.3	0.3	0.8	

Sx: Vomiting, Failure to thrive, Spells, Pulmonary sx, Irritability Anorexia E: Esophagogram; gr: grade 0 (no GER) to 4 GER to thoracic inlet M: Manometry; During the two trials, GER Time was the percentage of time with pH < 4 during one 2-hour postprandial period, and number of episodes was the number of episodes with pH < 4 during a two-hour postprandial period. epi:episodes; epi>5: episodes > 5 min; max: longest episode

-5- Mr. W.P. [From text by Ingelfinger et al]

The Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure included the following recommendations concerning the goals of anti-hypertensive therapy:

"The initial goal of anti-hypertensive therapy is to achieve and maintain diastolic pressure at less than 90 mm Hg.

i Mr. W.P. is started on treatment. He has the following blood pressures at his next 4 visits:

86, 92, 82, 84

- a Assuming that the standard deviation of his blood pressure is 5, about average, compute the 80% and 95% confidence intervals for his mean blood pressure. What is your confidence that his mean blood pressure is below 90mm Hg?
- b Use the measurements to estimate Mr. W.P.'s standard deviation (compute s).
- c Compute the 80% and 95% confidence limits for his mean blood pressure using s, n, and the t distribution.

ii Mr. W.P. is followed and his average blood pressure over many visits is 85 mm Hg. Suppose that his true standard deviation for individual measurements is 6 mm Hg.

- a How often would you expect a reading of 95 or higher? 100 or higher?
- b On the next visit, his blood pressure is 95. How could you settle whether his average pressure is no longer below the goal of 90 mm Hg?

iii (continuation) After measuring Mr. W.P.s blood pressure on several visits, you find his new average to be 95.

- a How many measurements must you have made to be 90% confident that his new mean in 90 mm Hg or greater?
- b How many observations would be required if the new observed mean were 91?

iv You follow Mr. W.P. and his blood pressure is consistently above 90 mm Hg. His pulse on 3 visits is 80, 85, and 75. You prescribe propranolol (an antihypertensive agent which also slows the pulse). On the next 5 visits, his blood pressure is unchanged but his pulse is 70, 65, 75, 60, and 65.

- a Compute the 55% confidence limits for the change in Mr. W.P.'s pulse.
- b Do you think the reason his blood pressure has not responded is that he has not taken the propranolol, or that the dose prescribed was not effective? Why?

-6- Which SE to use?

Why would you get contradictory answers if you tested the differences in the means using (a) the SE's in columns 1 and 2, and (b) the SE in column 3? Which is correct/ Why?

Postoperative Effect on Plasma Ascorbic Acid for 105 Cases; Readings on the Same Individuals

	(1)	(2)	(3)
All Cases No. = 105	PreopValue mg/100 ml	PostopValue mg/100 ml	Difference (postop -preop) mg/100 ml
Mean	0.43	0.36	-0.07
SE	0.036	0.028	0.015
95% CI	0.36 to 0.50	0.30 to 0.41	-0.10 to -0.04
			t = |4.93|
			P ≤ 0.01(sig)

-7- From SEM's to test-statistic and P-Value

Whiting, R. B., et al., in an article titled "Idiopathic Hypertrophic Subaortic Stenosis in the Elderly", (New Eng. J. Med. 285:196-200, 1971)

The authors compare certain parameters of 30 patients 60 years of age and under and 14 patients over 60 years of age with this disorder. The following is from the abstract:

> "Older patients had a mean resting left ventricular outflow tract gradient of 53.3 ± 8.8 mm (mean ± S.E.M.) of mercury, and younger patients a mean resting gradient of 32.2 ± 5.2 mm (mean ± S.E. M.) of mercury (P less than 0.05)."

a From the information presented, how would you verify the P value?

b Which group shows greater subject to subject variation? Why?

-8- Paired and Unpaired

In a study of interviewer effects in psychiatric epidemiology, a sample of six males and 15 females was interviewed twice, first by a physician (MD) who identified himself as such and later by another physician (MR) who pretended he was a lay person. The following data emerged for the scores of the subjects on a Physical Condition System Scale

		MD	MR
Males	$\bar{x}$	3.50	0.50
	s	2.43	0.84
	n	6	6
Females	$\bar{x}$	1.33	4.00
	s	1.59	2.17
	n	15	15

a Comment - in a very descriptive way - on the findings.

b Suppose you wanted to formally test the null hypothesis that males and females give the same amount of information <u>when the interviewer is an MD</u>.

- What test of significance would you use?
- Do you have enough information to do the test?
- Give the relevant formula for doing the test.
- Beyond what value of the test statistic you would reject H_0?

c Suppose now you want to test the null hypothesis that there is no difference between the amount of information females give to an MD and how much they give to a (supposedly) lay person.

- What test of significance would you use?
- Do you have enough information to do the test?
- Give the relevant formula.

-9- EARLY DISCHARGE AT 2000g OF PRETERM INFANTS

Summary A study was conducted at l'Hôpital Sainte-Justine 10/79 à 6/80 to see if the mean discharge weight of preterm infants born at ≤2000g could be safely reduced.

A study group* (21 infants) was discharged "early" at a mean weight of 2010g (1890-2190g) when 5 preset criteria were met: no medical problem; adequate weight gain; stable temperature control in room air; all feedings by nipple; mother ready to have the baby home. [* born on on odd days] a

A control group+ (17 infants) was discharged at a mean weight of 2261g (2200-2400g). [+ born on even days] a

The duration of hospitalization for the "early" group was reduced by 11.6 days. At expected date of delivery, the weight was similar for both groups (3095 ± 403g vs 3146 ± 453g) as well as length, head circumference and hemoglobin concentration. Follow up until expected data of delivery, showed no morbidity or mortality in either group. Early discharge did not affect mothering confidence. This study demonstrated that discharge to an adequate home environment of low birth weight infants at 2000 ± 100g is safe provided appropriate criteria are met and adequate follow up is available.

a Does this method qualify as "random allocation"?

b Suppose you were the investigator and that you favoured one of the two policies. Explain how you could subvert the allocation method to make the policy "look good".

c Would a formal statistical test of the baseline differences automatically 'catch' this subversion?
[For more on significance tests on baseline data see item 15 of Table 2 of `http://www.consort-statement.org/`]

d The authors performed a statistical test on the discharge weights, and found a significant difference between the two groups ($p < 0.001$). What was the purpose of doing this test?

e What statistical test, if any, is most suitable for comparing the lengths of hospital stay? What summary statistic would you use to describe the reduction?

f What test is suitable for comparing weights at the expected date of delivery?

g The followup until the expected date of delivery showed no morbidity or mortality in either group. Do you take this as adequate statistical evidence that early discharge is as "safe" as the more conservative discharge policy? If not, why not?

Reported Data	mean ± 1 SD (range) *"Early Discharge" Group*	mean ± 1 SD (range) *Control" Group*
Neonatal		
n	21	17
male:female	10:11	8:9
birthweight (g)	1655 ± 214 (1000-1960)	1533 ± 293 (900-1940)
gestation (weeks)	32.1 ± 2 (27 - 36)	31.2 ± 2 (28 - 34)
Discharge		
weight (g)	2010 ± 84 (1890-2190)	2261 ± 59 (2200-2400)
length of stay (days)	26.3 ± 15.2 (8 - 74)	37.9 ± 14.5 (15 - 61)
At Expected Date of Delivery		
(i.e. at 40 wks. gestation)		
weight (g)	3095 ± 403 (2440-3910)	3146 ± 453 (2440-4195)

-10- Inhibition of oxidation of low-density lipoprotein with red wine Lancet 344 page 1152 October 22, 1994

Sir: Maxwell and colleagues (July 16, p 193) reported that ingestion of red wine was associated with increased antioxidant activity in human serum by use of an enhanced chemoluminescence assay.[1] They also showed that serum antioxidant values rose during the 4 h after red wine intake. Frankel and colleagues[1] reported d the inhibition of oxidation of human low-density lipoproteins (LDL) by phenolic substances in red wine in vitro. Serafini et al[] also showed that the consumption of polyphenol-rich beverages such as red wine and tea was associated with an increase in plasma antioxidant potential. Antioxidant substances inhibit LDL oxidation, thus reducing the development of atherosclerosis, thought to be provoked by atherogenic modified LDL[]' However, it is important to distinguish between the antioxidant properties of red wine compared with other beverages and foods, a consideration that formed the basis of the protocol in our study.

10 male volunteers (33-37 years old) drank vodka (40% ethanol) (wash-out period) and then red wine(Chateau Lagrange, 1989) (experimental period), corresponding to a dose of 0.8 g/kg ethanol per day for l4 days. All subjects received a standard diet (supplied by Taihei Co, Tokyo) to control their caloric and nutritional intake, and they abstained from drinking tea, coffee, and other such substances to minimize the intake of phenolic substances other than those derived from red wine throughout the study period. Fasting venous blood samples were taken at day -14, 0, and +14. Plasma LDL was prepared by ultracentrifugation (d 1.006-1-063 g/mL) and oxidation of LDL was investigated by measuring conjugated dienes formed with 2,2'-azobis (4-methoxy-2,4dimethylvaleronitrile; V-70).

Oxidation of LDL was longer at day 14 (lag time 54.7 [SE 2.6] min) than at day 0 (49.1 [2.2] min). There was no difference in oxidation of LDL between day -14 and day 0 (Figure).

Our results provide direct evidence that regular and long-term consumption of red wine, but not ethanol, inhibited LDL oxidation in vivo. It is suggested that red wine intake may reduce atherosclerosis and morbidity and mortality from coronary heart disease. In this context, our study provides a plausible explanation for the "French paradox" (the apparent incompatibility of a high-fat diet with a low incidence of coronary heart disease).

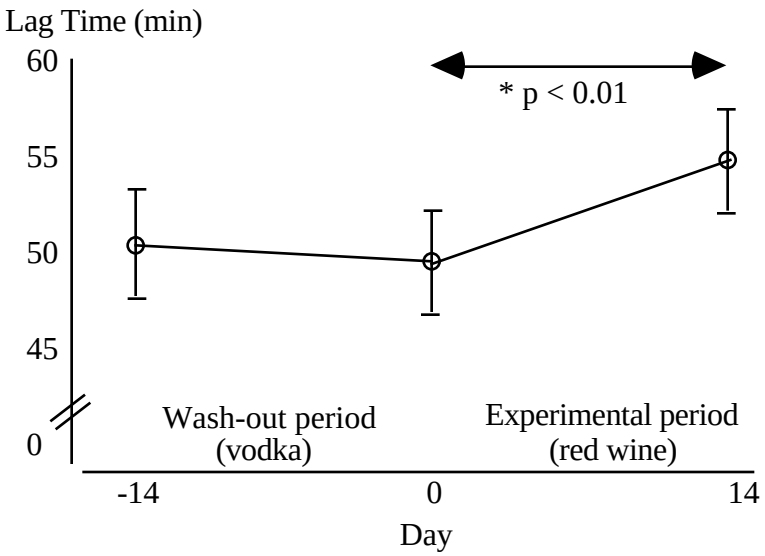

Fig 1: LDL oxidation in volunteers;
values are means(SE).
* paired t-test

- Calculate the SD and variance of the 10 lag times at Day 0 and at Day 14.
- The authors used 'error bars' of ± **1** SE rather than ± some larger multiple of the SE, presumably so as to ensure that the two CI's for day 0 and day 14 did not overlap. If you were going to put a 95% CI at each point, what multiple would you use?
- If the authors calculated a CI for the mean difference as

  $$54.7 - 49.1 \pm m\sqrt{2.6^2 + 2.2^2},$$

  they would find that for any multiple m bigger than about 1.65, the corresponding confidence interval included a mean difference of zero. Does this mean that the difference is not statistically significant at conventional levels of significance? How does one reconcile this with the p<0.01 reported by the authors?

  [Calculating the SE based as $\sqrt{s^2[1/10 + 1/10]}$ where s^2 is the weighted average of the two variances, would give the same SE in this example; so the issue is not one of separate versus pooled variances! Nor is it an issue of 1- versus 2-sided tests!]