Name: Student #

BME1480H1F
Experimental Design and Multivariate Analysis In Bioengineering

MIDTERM
October 22 2015, 1:10pm-3pm
MSB 3278

Instructor: Prof Julie Audet

Open book/notes/aid sheet
Calculator, Computer
No wi-fi, no cell phones

Name: Student #

1)	/13
2)	/12
Total	**/25**

Show your calculations (including degrees of freedom) and clearly present your conclusions to get full credit. Use α=0.05 unless specified otherwise.

Q.1) (15 points) A group of researchers who studied the stress of air traffic controllers proposed some key modifications to the software used in the controller's work station. Two software modifications were selected as having the best potential for reducing controller stress; they were tested along with the standard software currently in use. The key question was: To what extent do the software alternatives differ in terms of their effect on controller stress?

To provide the necessary data to answer the question, the three softwares (Software A, Software B, Software C) were installed at the Montreal Area Control Center. System A is the standard software currently in use. Six controllers in total were selected and they were assigned to operate each of the three softwares in a random order. Following each trial, each controller was immediately subjected to a cortisol blood test and the cortisol level (in μg/dL) was used to measure his/her stress level. It is expected that controllers differ substantially in their ability to handle stressful situations. The results are presented in Table I.

Table I
(Air traffic controller stress test , cortisol in μg/dL)

	Software A	Software B	Software C
Controller 1	g	a	m
Controller 2	n	h	b
Controller 3	i	c	o
Controller 4	j	d	p
Controller 5	q	k	e
Controller 6	f	r	l

* numerical values for the raw data will not be provided for this question.

Name: Student #

After the data for the experiments had been collected, the researchers realized that the group had yet to agree on the best method to analyse the data. Three different analysis methods were considered (listed below):

METHOD 1) Randomized Complete Block Design (RCBD) ANOVA, for which the sum of squares for Total, Controllers and Error are provided below:

$$SS_{Total} = 70$$
$$SS_{Controllers} = 30$$
$$SS_{Error} = 19$$

METHOD 2) Completely Randomized Design (CRD) ANOVA, also referred to in class as simple one-way ANOVA

METHOD 3) Completely Randomized Design (CRD) ANOVA with Software B and Software C after normalizing the cortisol levels to the corresponding levels obtained with Software A (see Table II).

Table II
(Air traffic controller stress test with normalized values i.e. fold change in cortisol level relative to that obtained with software A)

	Software B	Software C
Controller 1	a/g	m/g
Controller 2	h/n	b/n
Controller 3	c/i	o/i
Controller 4	d/j	p/j
Controller 5	k/q	e/q
Controller 6	r/f	l/f

a) (5 points) Do you find evidence that the softwares differ in terms of their effect on controller stress when you use Method 1 (RCBD) to analyse the data ? Test appropriate hypotheses (specify H_o and H_1), complete the ANOVA Table for the RCBD, and state your conclusion clearly (i.e. the softwares differ in terms of their effect on controller stress or the softwares do not differ).

Name: Student #

ANOVA Table for the RCBD

Source	d.f.	SS	MS	F ratio	p-value (or range)

* the number of rows in the table is more than you need

Response = Stress
Treatments = Software A, B, or C = (Factor of interest)
Controllers = Block = Nuisance factor

H_o: $\tau_A = \tau_B = \tau_C = 0$
H_1: $\tau_i \neq 0$ for at least one i

$SS_{treatment} = SS_{Total} - SS_{controllers} - SS_{error} = 70-30-19=21$
d.f. treatment = 3-1=2
d.f. error = d.f. total-d.f. controllers-d.f. treatment= (18-1)-(6-1)-2=10

$MS_{treatment} = 21/2 = 10.5$

$MS_{Error} = 19/10=1.9$

$F=10.5/1.9=5.53$

$F_{0.05,2,10} = 4.10 < 5.53$ and $F_{0.025,2,10}=5.46 < 5.53$

Reject H_o. Softwares make a difference on controller stress ($p<0.025$).

ANOVA Table for RCBE:

Source	**d.f.**	**SS**	**MS**	**F**	**p**
Treatments (Softwares)	2	21	10.5	5.53	<0.025
Controllers (Block)	5	30			
Error	10	19	1.9		
Total	17	70			

b) (4 points) Do you find evidence that the softwares differ in terms of their effect on controller stress when you use Method 2 (CRD) to analyse the data ? Test appropriate hypotheses (specify H_o and H_1), complete the ANOVA Table for the CRD, and state your conclusion clearly (i.e. the softwares differ in terms of their effect on controller stress or the softwares do not differ).

ANOVA Table for the CRD

Source	d.f.	SS	MS	F ratio	p-value (or range)

* the number of rows in the table is more than you need

Completely Randomized Experiment (CRD):

H_o: $\tau_A = \tau_B = \tau_C = 0$
H_1: $\tau_i \neq 0$ for at least one i

Here you have to redistribute the SS and d.f. obtained for the RCBD to get the SS and d.f. of the CRD. In a CRD, the SS and d.f. associated with Block are pooled into the SS and d.f. for Error

$SS_{Error, CRD} = SS_{Error,RCBD} + SS_{Controllers} = 30+19=49$

d.f. Error CRD = d.f. Error RCBD + d.f. Controllers = 10 + 5 = 15

$MS_{Error, CRD} = 49/15 = 3.27$

$MS_{treatment}$ does not change = 10.5 but now $MS_{Error, CRD} = 3.27$

$F_{CRD} = 10.5/3.27 = 3.21$

$F_{0.05,2,15} = 3.68 > 3.21$, $p>0.05$, fail to reject H_o. Softwares do not make a difference on controller stress.

ANOVA Table for CRD:

Source	**d.f.**	**SS**	**MS**	**F**	**p**
Treatments (Softwares)	2	21	10.5	3.21	>0.05
Error	15	49	3.27		
Total	17	70			

c) (2 points) Are the results of the test in b) with METHOD 2 different than what you found in a) with METHOD 1? Compare your conclusions in a) and b) and briefly discuss the reasons for the difference and/or similarity.

We have different conclusions based on F test with a RCBD versus CRD because when the data is analysed as CRE, you get a much larger error term since the controller-to-controller variation (Block) is pooled into the error. You get a larger denominator for the F ratio for the Treatment test. This decreases the sensitivity of

the Treatment (Software) test (Software is the factor of interest). Note that it can be beneficial in terms of test sensitivity to increase d.f. error but only if the block variation is negligible.

d) (2 points) The results of the analysis in JMP using METHOD 3 i.e. ANOVA with normalized responses (see Table II) are presented below. JMP automatically provides the results of a two-sample t-test (top panel) with the ANOVA (bottom panel).

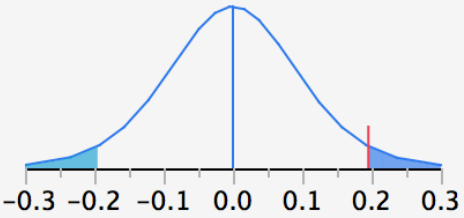
JMP Report

t Test

C–B

Assuming equal variances

Difference	0.19542	t Ratio	2.090705
Std Err Dif	0.09347	DF	10
Upper CL Dif	0.40368	Prob > |t|	0.0631
Lower CL Dif	−0.01285	Prob > t	0.0315*
Confidence	0.95	Prob < t	0.9685



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Software	1	0.11456302	0.114563	4.3710	0.0631
Error	10	0.26209521	0.026210		
C. Total	11	0.37665823			

Compare Methods 1 (RCBD ANOVA) *versus* Method 3 (CRD ANOVA with normalized values) in terms of

i) the hypotheses tested
ii) p-values obtained (explain the main reason(s) for the differences or similarities)
iii) conclusions that can be made (explain the main reason(s) for the differences or similarities)

i)
Method 1:
H_o: $\tau_A = \tau_B = \tau_C = 0$
H_1: $\tau_i \neq 0$ for at least one i

Method 3:
H_o: $\tau_B = \tau_C = 0$
H_1: $\tau_i \neq 0$ for at least one i

Thus, the two methods are not really testing the same thing.

ii)

The p-values are different mainly because we have lost one degree of freedom in the ANOVA for treatment with Method 3 compared with Method A, less total d.f. , so decreased statistical power (i.e. sensitivity).

The ratios may have large error due to error propagation (error cortical levels for old and new software).

iii) CRD ANOVA with normalized (Method 3) values compares B and C and not differences between A, B and C which is what we are mostly interested in. If both B and C decrease stress in a similar manner (relative to A), we will still fail to reject Ho since their effects is no longer compared with A. Method 3 still takes into account the variability between subjects but it tend to be less sensitive than a RCBD since it has on less d.f. because treatment A was used to normalize.

Q2) (12) A researcher is investigating the recovery times (days) of patients on a new postsurgical treatment; he would like to see if these recovery time are shorter than when a standard treatment is used. Seven subjects were selected for the study and n=4 are randomly assigned to be on the new treatment while the remaining m=3 received the standard treatment.

Table III

Recovery (days)	
New treatment (n=4)	Standard treatment (m=3)
19, 22, 25, 26	23, 33, 40
Average=23	Average=32
Standard dev=3.16	Standard dev=8.54

a) (1 point) Write appropriate hypotheses (specify H_o and H_1) for a relevant statistical test

1=new and 2=standard

H_o: $\mu_1-\mu_2$ = mean recovery time with standard – mean recovery time with new= 0
H_1: $\mu_1-\mu_2 < 0$

b) (4 points) The researcher decides to perform a permutation test to determine if there is evidence that the new treatment significantly decreases the recovery times. Using the data in Table III and Table IV (below), test the hypotheses that you listed in a). Calculate (show your calculations) a p-value for the test (or a p-value range) and state your conclusion clearly.

In Table III, the difference between the average recovery time with the new treatment (avg Y_1) and the standard treatment (avg Y_2) is -9.

From Table IV, p value = P(avg Y_1-avg $Y_2 \leq -9$) = $3/35 \approx 0.0857$

No strong evidence of differences in recovery times.

Name: Student #

Table IV

No.	Randomization: New treatment				Randomization: Standard treatment			Difference in means	Sum of new	Sum of standard	Difference in medians
▲ 1	19	22	25	26	23	33	40	−9.00	92	96	−9.5
2	22	23	25	26	19	33	40	−6.67	96	92	−9.0
3	22	33	25	26	19	23	40	−0.83	106	82	2.5
4	22	25	26	40	19	23	33	3.25	113	75	2.5
5	19	23	25	26	22	33	40	−8.42	93	95	−9.0
6	19	25	26	33	22	23	40	−2.58	103	85	2.5
7	19	25	26	40	22	23	33	1.50	110	78	2.5
8	19	22	23	26	25	33	40	−10.17	90	98	−10.5
9	19	22	26	33	23	25	40	−4.33	100	88	−1.0
10	19	22	26	40	23	25	33	−0.25	107	81	−1.0
11	19	22	23	25	26	33	40	−10.75	89	99	−10.5
12	19	22	25	33	23	26	40	−4.92	99	89	−2.5
13	19	22	25	40	23	26	33	−0.83	106	82	−2.5
14	23	25	26	33	19	22	40	−0.25	107	81	3.5
15	22	23	26	33	19	25	40	−2.00	104	84	−0.5
16	22	23	25	33	19	26	40	−2.58	103	85	−2.0
17	19	23	26	33	22	25	40	−3.75	101	87	−0.5
18	19	23	25	33	22	26	40	−4.33	100	88	−2.0
19	19	22	23	33	25	26	40	−6.08	97	91	−3.5
20	23	25	26	40	19	22	33	3.83	114	74	3.5
21	22	23	26	40	19	25	33	2.08	111	77	−0.5
22	22	23	25	40	19	26	33	1.50	110	78	−2.0
23	19	23	26	40	22	25	33	0.33	108	80	−0.5
24	19	23	25	40	22	26	33	−0.25	107	81	−2.0
25	19	22	23	40	25	26	33	−2.00	104	84	−3.5
26	25	26	33	40	19	22	23	9.67	124	64	7.5
27	22	26	33	40	19	23	25	7.92	121	67	6.5
28	22	25	33	40	19	23	26	7.33	120	68	6.0
29	19	26	33	40	22	23	25	6.17	118	70	6.5
30	19	25	33	40	22	23	26	5.58	117	71	6.0
31	19	22	33	40	23	25	26	3.83	114	74	2.5
32	23	26	33	40	19	22	25	8.50	122	66	7.5
33	23	25	33	40	19	22	26	7.92	121	67	7.0
34	22	23	33	40	19	25	26	6.17	118	70	3.0
35	19	23	33	40	22	25	26	4.42	115	73	3.0

From M.D. Earnst. 2004.

c) (2 points) Why are there 35 rows in Table IV ? Show how to calculate this number.

Name: Student #

There are 35 possible permutations of 7 recovery times to 2 treatment groups of sizes n=4 and m=3.

Use the binomial coefficient with n=7 and k=4.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

= 7!/4!(7-4)! = 1x2x3x4x5x6x7/(1x2x3x4)(1x2x3)=5x6x7/1x2x3=210/6=35

d) (5 points) Test the hypothesis using the t-distribution assuming equal variances (and see if you obtain a p-value similar to that obtained in b).

For full credit, provide an answer for the seven steps listed below (make sure you show the calculation of the t ratio, specify the degrees of freedom and state clearly your conclusion).

1. Parameter of interest:
2. Null hypothesis (H_o)
3. Alternative hypothesis (H_1)
4. Test statistic chosen:
5. Reject H_o if:
6. Computations (make sure you specify the degrees of freedom):
7. Conclusions:

1. Parameter of interest is the mean recovery time with the new and with the standard treatments

2. and 3.
H_o: $\mu_1-\mu_2$ = mean recovery time with new – mean recovery time with standard= 0
H_1: $\mu_1-\mu_2 < 0$

New = 1 and standard=2

4. Test statistics chosen:

Two-sample t-test method

Assume equal variance (use pooled variance S^2_p)
$n=4=n_1$ and $m=3=n_2$

S_1=3.16 S_2=8.54
Average standard=32
Average new= 23

$$t_0 = \frac{avg\ Y1 - avg\ Y2}{Sp\sqrt{\frac{1}{n1} + \frac{1}{n2}}}$$

Pooled variance:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

5. Reject H_o if $|t_0| > |t_{0.05,v}|$

6. Computation

$S_p^2 = [(3\text{-}1)\ (8.54)^2 + (4\text{-}1)(3.16)^2]/(3+4-2)=(2*72.93 + 3*9.99)/5=(146+29.97)/5=35.2$

S_p= 5.93

$$t = \frac{23 - 32}{5.93\sqrt{\frac{1}{4} + \frac{1}{3}}} = \frac{-9}{5.93\sqrt{0.58}} = \frac{-9}{4.5} = -2.0$$

d.f. = 3+4-2 = 5
$|t_{0.05,5}| = 2.02 > |-2.0|$, therefore cannot reject H_o (borderline, p value for one tail test slightly above 5%). $|t_{0.10,5}| = 1.5$, so p-value in the range 0.05 to 0.10.

7. Conclusion: Not enough evidence to claim that the recovery time with the new treatment is shorter than with the standard treatment.

Name: Student #

Name: Student #

SCRAP PAPER (TEAR OFF)