NAME: Student #:

BME1480H1S
Experimental Design and Multivariate Analysis In Bioengineering

MIDTERM #1
October 15 2014, 1:10pm-2:10pm
HA401

Instructor: Prof Julie Audet

Open book/notes/aid sheet
Faculty Approved Calculator
No computers, No cell phones

NAME: Student #:

1)	/25
2)	/25
3)	/25
4)	/25
Total	**/100**

For all questions, use α=0.05 when appropriate. Show your calculations to get full credit, except for simple sample averages and simple sample variances (which can be done by your calculator).

Q.1) (25%) Photoresist is a light-sensitive material applied to wafers during microchip fabrication so that patterns can be imaged on to the wafer. After application, the coated wafers are baked to remove the solvent in the photoresist mixture and to harden the resist. Here are measurements of photoresist thickness (in kÅ) for eight wafers baked at two different temperatures. Assume that all of the runs were made in random order.

Table 1

95 ºC	100 ºC
11.176	5.623
7.089	6.748
8.097	7.461
11.739	7.015
11.291	8.133
10.759	7.418
6.467	3.772
8.315	8.963

The variance for the sample at 95 °C is 4.41 and the variance for the sample at 100 °C is 2.54.

(a) (15 points) Is there evidence to support the claim that the higher baking temperature results in wafers with a lower mean photoresist thickness? **Indicate clearly your null hypothesis and alternative hypothesis, the test statistic, the criteria you used for reaching a conclusion and the conclusion itself.**

This is a regular two-sample t-test (assumption of equal variances).

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(8-1)(4.41) + (8-1)(2.54)}{8+8-2} = 3.48$$
$$S_p = 1.86$$
$$t_0 = \frac{\overline{y}_1 - \overline{y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{9.37 - 6.89}{1.86\sqrt{\frac{1}{8} + \frac{1}{8}}} = 2.65$$
$$t_{0.05,14} = 1.761$$

Since $t_{0.05,14} = 1.761$, and t_o=2.65 > 1.761, or $P(t>t_o) = 0.009$.

-> Reject H_0. There appears to be a lower mean thickness at the higher temperature.

(b) (10%) As seen in Table 1, the researcher who designed the experiment decided to use n = 8 for both samples. In your opinion, what were the factors taken into account when he made this decision (i.e why not use n=4 or n=16 for instance)? Answer in 5 sentences or less.

The researcher has to consider:

1) difference in means that he/she wants to be able to detect (higher n required to detect smaller differences)
2) the variability between replicates (variance, noise) since higher variance requires higher n
3) the desired statistical power, in particular for prob for type II error (β) since higher statistical power (1-β) requires higher n
4) the amount of work required for each test since higher n is a lot of work, time, materials and, therefore, $$$ (often this needs to be minimized).

see slides at the beginning of LEC 3 (related to optional assignment).

Q.2 (25%) The diameter of a lens was measured by 12 surgeons, each using two different kinds of ophthalmic calipers. The results were (in cm):

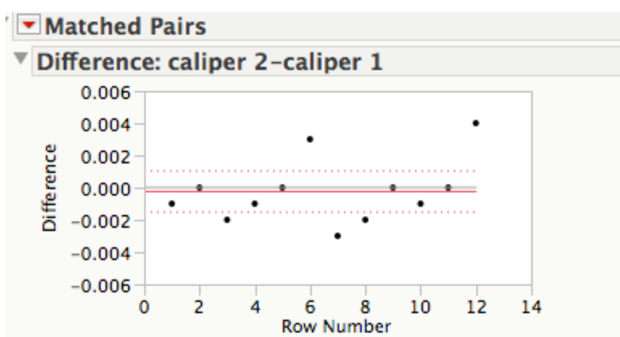
Surgeon	Caliper 1	Caliper 2	Difference (d)
1	0.265	0.264	0.001
2	0.265	0.265	0.000
3	0.266	0.264	0.002
4	0.267	0.266	0.001
5	0.267	0.267	0.000
6	0.265	0.268	-0.003
7	0.267	0.264	0.003
8	0.267	0.265	0.002
9	0.265	0.265	0.000
10	0.268	0.267	0.001
11	0.268	0.268	0.000
12	0.265	0.269	-0.004

When the data was analyzed to see if there was a significant difference between the means of the population of measurements represented by the two samples, the researcher decided to use a paired t-test, with the following hypotheses:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

The report from the statistical software used by the researcher for the analysis is presented below:



caliper 2	0.266	t-Ratio	-0.43179
caliper 1	0.26625	DF	11
Mean Difference	-0.0003	Prob > |t|	0.6742
Std Error	0.00058	Prob > t	0.6629
Upper 95%	0.00102	Prob < t	0.3371
Lower 95%	-0.0015		
N	12		
Correlation	0.12763		

Do you think, in this case, it was justified to use a paired t-test (instead of a conventional two-sample t-test) for the analysis? Answer in 5 sentences or less.

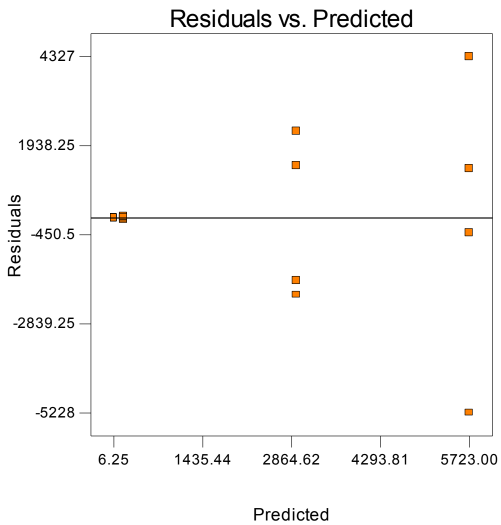
It is possible that it was not necessary to use a paired t-test since there is not a lot of variation between surgeons (blocks). This is revealed by the fact that the average measurement (average for caliper 1 and caliper 2) for each surgeon is similar (it could be advantageous in terms of statistical power to use a paired t test if there was a lot of variation from surgeons to surgeons). The reports also indicate that the correlation is low. Therefore, this also suggests that there is not a lot of surgeon-to-surgeon variability. In that context, a regular two-sample t-test may result in a similar or even higher statistical power since the degree of freedom for the test would be much higher.

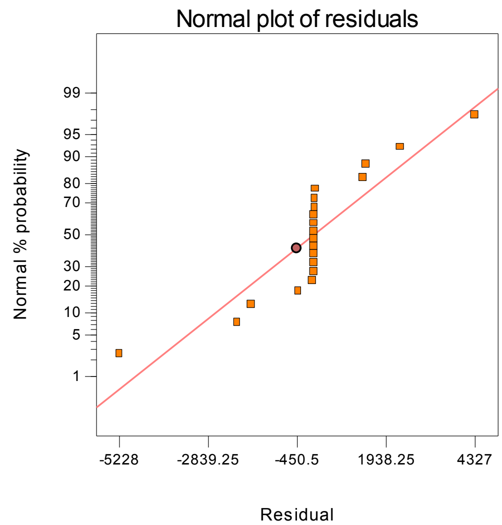
See end of LEC 3

Q.3) (25%) An experiment was performed to investigate the effectiveness of five drug delivery materials. Four samples of each material were used to encapsulate the same drug and the capsules were tested in a buffer solution at an elevated temperature to accelerate the time at which the drug is released from the capsule. The release times (in minutes) are shown below.

Material	Release Time (minutes)			
1	110	157	194	178
2	1	2	4	18
3	880	1256	5276	4355
4	495	7040	5307	10050
5	7	5	29	2

The data was analysed using ANOVA. A plot of the residuals *versus* the predicted response and a normal probability plot of the residuals are presented below. What information do these plots convey (discuss in terms of the basic assumptions of the ANOVA method). Answer in 5 sentences or less.





One of the basic assumptions for ANOVA is that the residuals are normally distributed. The plot of residuals versus predicted has a strong outward-opening funnel shape, which indicates the variance of the original observations is not constant. The normal probability plot also indicates that the normality assumption is not valid (the line represents a nomal distribution and the dots represent the residuals; the dots do not fit with the line). Therefore the ANOVA assumption cannot be validated for this dataset. A data transformation is recommended (notice large variation in the response data, several orders of magnitude).
See residual analysis in LEC 4.

Q.4) (25%) A clinical engineer is working on a brain-computer interface project where subjects equipped with brain electrodes are moving a cursor on a computer screen. The clinical engineer is interested in comparing the performance of four versions of a software (*A, B, C, D*) that interprets brain signals. She has access to only 4 subjects (1, 2, 3, 4), for only a short period of time, and she would like to measure the time each subject takes to move the cursor and reach the target. Also, each subject will need to test each of the 4 software versions. Furthermore, the engineer knows that each time a subject tests a software, it produces such fatigue that the time required to reach the target with the last software tested may be greater than the time required with the first software tested, regardless of the software version. That is, a trend develops in the time required to reach the target.

a) (12%) We have seen in class many different types of experimental designs; can you suggest the best one to meet the goal of the clinical engineer and account for the difference sources of variability? To answer this question, name the type of design and include a table (as seen in class) describing each experimental unit for the proposed experiment (i.e. with 4 subjects and 4 software versions).

Here we have two nuisance factors that we want to block (subject and order) and one factor that we want to study (software version). The only design seen in class that allows to block 2 factors during the analysis is a Latin Square (4x4).

Order	Subject			
tested	1	2	3	4
1	*C*	*D*	*A*	*B*
2	*B*	*C*	*D*	*A*
3	*A*	*B*	*C*	*D*
4	*D*	*A*	*B*	*C*

The analysis of a Randomized Complete Block (RCB) design only allows to block one nuisance factor (while randomizing the other). It is a possible alternative but will have severely decreased statistical power if the “order” has a very large effect (this was a major concern here). The reason is that the variation caused by the “order” will be pooled with the unexplained variation. Therefore the signal to noise ratio will decrease since the error term (denominator) will be larger for the F-test for the software version. In the latin square, the variation caused by the order is kept separate form the unexplained variation.

Subject			
1	2	3	4
C	*D*	*A*	*B*
B	*C*	*D*	*A*
A	*B*	*C*	*D*
D	*A*	*B*	*C*

If you have trouble understanding, I have included a spreadsheet (Clin eng_latin square) on Blackboard with some data for this experiment. This can be analysed as a latin square (fit model) or as a RCB (just ignore the order column and factor). Notice the increase in the SS_{error} whith RBC which is equal to SS_{error} + SS_{order} obtained when analysing as a Latin Square.

b) (13%) For the experimental design that you have chosen in a), what are the different components (source of variation) for the ANOVA ? To answer this question, you can list the different sum of squares that would need to be calculated to test the effect of the software version.

If you answered latin square:

Software version (SStreatment)
Order the software was tested (SS block1)
Subject-to-subject variability (SS block2)
Unexplained variation (SSerror, also referred to a SSresiduals)

If you answered RCB design:

Software version (SStreatment)
Subject-to-subject variability (SS block)
Unexplained variation (SSerror, also referred to a SSresiduals)

See end of LEC 3 and LEC 4

Scrap paper