Final exam, 17 April 2026

All aids are allowed, except a fully operational computer and personal assistance. Restricted use of some computer-like devices (including laptops, tablets and smartphones) is permitted under the rules described at the VHM 801 course homepage. The exam consists of three questions that should all be answered. The weights for each of the three questions and also for each subquestion within a question are indicated; these weights total *50 points*. Note that questions, and often also subquestions, can be answered independently of each other. The duration of the exam is 3 hours.

Generally, **statistical models and methods should be specified**, and every statistical analysis should be summarized in a conclusion. Throughout, if you realize that you need more information than is provided to carry out an analysis, specify what information you need, how you would obtain it using statistical software, and how you would use it in the analysis.

Question 1. (*15 points*)

A study published close to 20 years ago on the efficiency of radiotherapy on so-called "pituitary macrotumors" (the exact meaning of this is not critical) in dogs involved 11 dogs. The size of the tumors was evaluated using computed tomography scans (a digital imaging technique). The dogs were scanned before the radiotherapy and when the dogs returned for checkup at least one month afterwards. The methods and results were reported as follows in the article:

> "Differences between heights of tumors before and after radiotherapy were compared by using the Wilcoxon test." [..]
> "Mean tumor heights before and after radiotherapy were significantly different (mean, standard deviation, respectively: 14.1, $s = 3.6$ mm and 8.8, $s = 4.1$ mm, $P < 0.02$)."

a) (*2 points*)
Describe the statistical design. Do you think it is valid to use a two-sample t-test to compare the mean tumor heights before and after radiotherapy? Justify your answer (briefly).

b) (*4 points*)
Alternatively, a parametric analysis could be based on the differences between tumor heights before and after radiotherapy, say $D = \mathit{before} - \mathit{after}$. This would require the standard deviation (s_D) of the differences, which was not reported in the paper; here we will work with a value of $s_D = 5$ mm (based on some further assumptions). Carry out a parametric analysis to give a 95% confidence interval for the difference in tumor heights, as well as a test of whether there is evidence of a reduction in tumor heights by the radiotherapy.

c) (*3 points*)
In a recent study, with a much larger sample of dogs, on variation in the size of imaging-diagnosed pituitary macrotumors in dogs, the proportion of tumors of height less than 10 mm was 21.3%. Assuming that tumor heights can be approximated well by a normal distribution (as supported by the most recent study), use the information provided from the older study to estimate the proportion of tumors of height less than 10 mm, *before radiotherapy*, and compare with the value from the recent, more comprehensive study.

d) (*4 points*)
The article also includes the following summary of the results:

> "A decrease of more than 30% in pituitary tumor height was observed in 10 of the 11 dogs reexamined."

Use this information to statistically assess the evidence in favor of a reduction in tumor height from the radiotherapy by a non-parametric test. (*Hint:* Complete first the analysis assuming that no change in tumor height was observed for the 11^{th} dog. Discuss next how different outcomes for the 11^{th} dog would affect your results; if time allows, carry out any extra analyses to substantiate your conclusion from the initial analysis.)

e) (*2 points*)
Discuss critically how the statistical analysis was carried out and reported in the article (based on the excerpts given above, which include all the information provided about this outcome); your discussion should include any suggestions you might have for improvements.

Question 2. (*15 points*)

In this question, we will consider an opinion poll on issues related to drinking and driving. In a report published by the Traffic Injury Research Foundation, the source and methodology was described as follows: "The Road Safety Monitor is an annual public opinion survey developed and conducted by the Traffic Injury Research Foundation to take the pulse of the nation on key road safety issues. The survey was administered by telephone to a random sample of Canadian drivers, and was carried out in September, 2006. A total of 1201 drivers completed the interview."

a) (*5 points*)
In one of the questions, the respondents were asked to indicate their level of concern about drinking and driving on a six-point scale from 1 (not a problem at all) to 6 (extremely serious problem). Eighty-eight percent (88.0%) of the answers fell in the categories 5 and 6 (very serious or extremely serious problem). Give a statistical model for the number of answers obtained to the question falling in either of categories 5 and 6. Motivate briefly your chosen model, and use it to compute 95% and 99% confidence intervals for the proportion of Canadian drivers who considered drinking and driving a very serious or extremely serious problem.

b) (*4.5 points*)
In all of the questions of the survey, the respondents were offered a choice between two or more answer categories, in the same way as for the question of **a)**. The publication reporting the findings of the many different questions in the survey stated that

> "[...] the results can be considered accurate within 2.9%, 19 times out of 20."

Explain the meaning of this statement in statistical terms, and describe the model and calculation you think lies behind the statement. In addition, based on your statistical model (for either **a)** or **b)**) determine the sample size required to achieve results that are "accurate within 2.9%, 19 times out of 20" (or with an accuracy of 2.9% interpreted in an alternative way, which then should be explained). Alternatively, for a slightly lower score, fill in the relevant fields in the printed Minitab menus for Question 2.

c) (*2.5 points*)
Another question of the survey was whether the respondents within the last 12 months at least once had been driving when they thought they were probably over the legal limit. Among the 88% of respondents, who considered drinking and driving a very serious problem, 7.6% reported driving at least once when above the legal limit. Conversely, among those respondents, who did not consider drinking and driving a very serious problem, 21.2% reported driving when above the legal limit. Use this information to approximately estimate the overall proportion of Canadian drivers who admit to driving while above the legal limit within the last 12 months. (*Hint:* Compute first approximately the number of respondents reporting drinking and driving in each of the two groups. The numbers do not quite add up, perhaps due to missing reponses; use round-off values in your calculation whenever necessary.)

The report states the overall proportion of respondents reporting driving while above the legal limit to be 7.7%. Compare this value with your estimate, and explain why it seems likely that at least one of the values stated in the report is wrong.

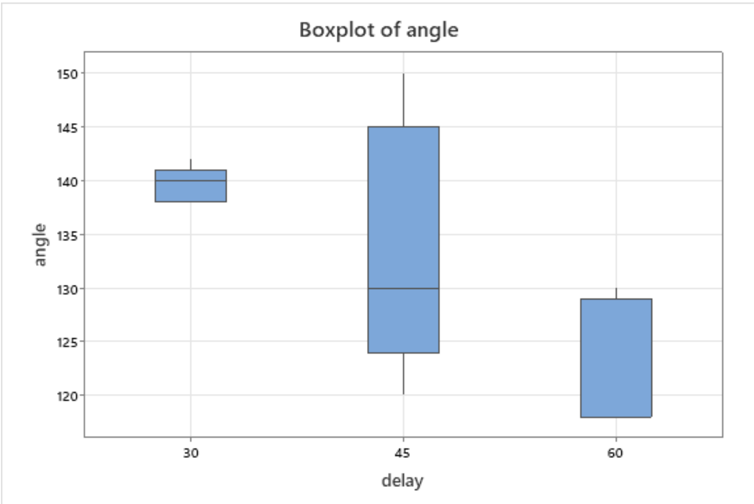
d) (*3 points*)
It is of particular interest to compare the proportions of respondents reporting driving while above the legal limit between those respondents, who consider drinking and driving a very serious problem, and those who do not. Irrespective of your previous calculations and conclusions, we will here assume the following numbers (matching roughly the conclusions of the report):

- 7 out of 33 respondents who do not consider drinking and driving a very serious problem, reported driving while above the legal limit,
- 19 out of 250 respondents who do consider drinking and driving a very serious problem, reported driving while above the legal limit.

Use this information, and any information in the attached Minitab listings you find useful, to carry out a statistical analysis to compare between the two groups the proportions of respondents reporting driving while above the legal limit. If you need additional information to carry out the analysis, specify what information you need, how you would obtain it from statistical software (or otherwise), and how you would use it.

Question 3. (*20 points*)

The leaves of certain plants in the genus *Albizia* (tropical small trees and shrubs) will fold and unfold in various light conditions. Fifteen different leaves were taken and subjected to red light for 3 minutes. The leaves were divided into three groups of five at random. The leaflet angles were then measured 30, 45 and 60 minutes after light exposure in the three groups, the values being shown in the plot below.



a) (*5 points*)
Explain the study type (e.g., observational, etc.) and the statistical design (e.g., completely randomized, etc.). Review the attached Minitab listing; determine the statistical model used for the analysis, and carry out a statistical analysis to compare the treatments. *Note:* Defer all discussion of model choice and model validity to subquestion **b)**.

b) (*5 points*)
Use the information already provided to evaluate the following statements about model choice for these data, by indicating for each statement whether it is correct or incorrect. (*Note:* The list contains more than one correct statement.) Further comments may be added to explain the evaluations, but it is not required to do so.

(b1) The first step in the analysis should be to compute the correlation between delays and angles.

(b2) The three delay groups should be compared by pairwise 2-sample t-tests.

(b3) The values within each delay group are approximated well by a normal distribution.

(b4) The validity of assuming a normal distribution within each group is difficult to assess from these data.

(b5) The data do not contain any outliers.

(b6) The textbook/course guideline for equal variances across groups is violated.

(b7) The spread in the different treatment groups is so variable that the data need to be log-transformed before analysis can be carried out.

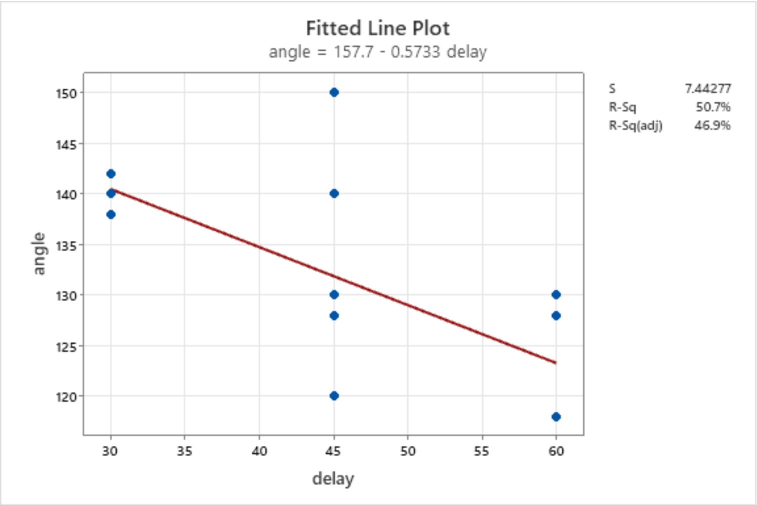
(b8) The assumptions behind the (above) analysis carried out on such a small dataset are best checked by looking at the residuals.

(b9) Because the dataset is small, a non-parametric analysis is the only valid approach to analyze these data.

(b9) Because the dataset is small, a non-parametric analysis will have too little power for these data.

(b10) If doubts exist about the validity of assumptions, it is helpful to analyze the data by different methods and compare the results.

c) (*5 points*)
The researchers expected to see a decline of the leaflet angles with time since the light treatment (delay). Describe the statistical model used for the graph below, and carry out a statistical test to determine whether the data show evidence of such a decline. (*Note:* It is possible to compute a test from the information provided, but if you are unable to do so, a partial answer can consist in explaining how you would use statistical software to request calculations to answer the question; make sure you specify exactly what information you need and how you would use it.)



d) (*5 points*)
Two different models have been studied for these data (in **a)** and **c)**). Generally, statistical analysis aims at finding the "best" model, even though that is not always achievable. One very simple way of evaluating the model from **c)** relative to the model from **a)** is by comparing predicted values for each of the three delays represented in the data to 95% confidence intervals from the latter model. Carry out this calculation, in the following steps:

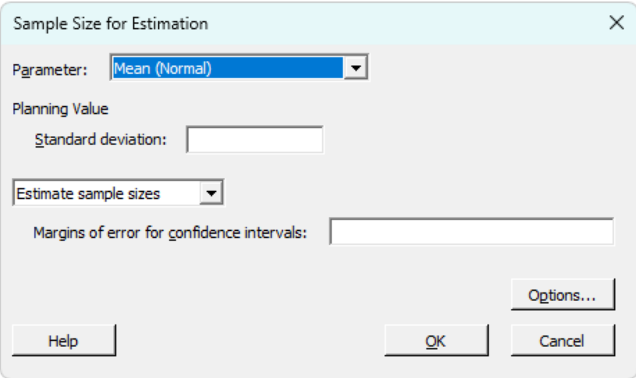
i) compute predicted values for delays 30, 45 and 60 minutes from the model in **c)**,

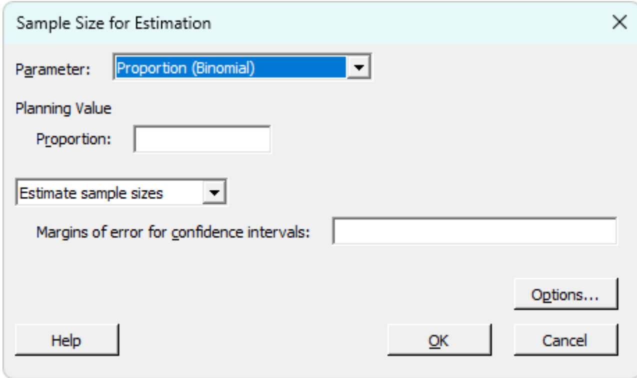
ii) compute 95% confidence intervals for the means for each delay based on the model in **a)**,

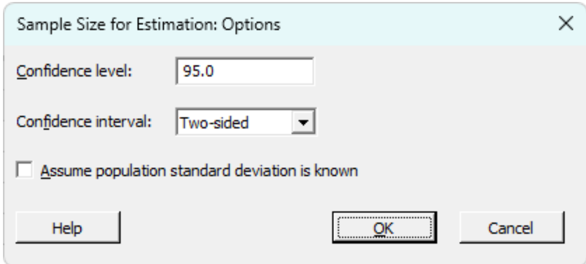
iii) compare the predictions with the intervals, and draw conclusions; use also any other information you find relevant to compare the two models.

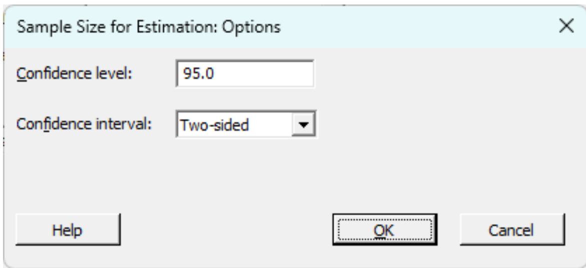
Minitab menus and output for Question 2

Menus (optional) for part b):









Output (slightly reduced) for part d):

WORKSHEET 1

Test and CI for One Proportion

Method

p Event proportion
Method Adjusted Blaker's exact method

Descriptive Statistics

N	Event	Sample p	95% CI for p
33	7	0.212121	(0.096441, 0.377875)

WORKSHEET 1

Test and CI for One Proportion

Method

p Event proportion
Method Adjusted Blaker's exact method

Descriptive Statistics

N	Event	Sample p	95% CI for p
250	19	0.076000	(0.047219, 0.114670)

WORKSHEET 1

Test and CI for Two Proportions

Method

p_1: proportion where Sample 1 = Event
p_2: proportion where Sample 2 = Event
Difference: $p_1 - p_2$

Descriptive Statistics

Sample	N	Event	Sample p
Sample 1	33	7	0.212121
Sample 2	250	19	0.076000

Estimation for Difference

Difference	95% CI for Difference
0.136121	(-0.007175, 0.279417)

CI based on normal approximation

Test

Null hypothesis H_0: $p_1 - p_2 = 0$
Alternative hypothesis H_1: $p_1 - p_2 \neq 0$

Method	Z-Value	P-Value
Normal approximation	2.54	0.011

(continues on the next page)

WORKSHEET 1

Tabulated Statistics: Worksheet rows, Worksheet columns

Rows: Worksheet rows   Columns: Worksheet columns

	C1	C2	All
1	7	19	26
	3.37	22.63	
2	33	250	283
	36.63	246.37	
All	40	269	309

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	4.922	1	0.027

WORKSHEET 1

Tabulated Statistics: Worksheet rows, Worksheet columns

Rows: Worksheet rows   Columns: Worksheet columns

	C3	C4	All
1	7	19	26
	3.03	22.97	
2	26	231	257
	29.97	227.03	
All	33	250	283

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	6.474	1	0.011

(Minitab output continues on the next page)

Minitab output (slightly reduced) for Question 3

ANGLE.MTW

Descriptive Statistics: angle

Statistics

Variable	delay	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Skewness
angle	30	5	139.6	1.67332	138	138	140	141	142	0.51
	45	5	133.6	11.6103	120	124	130	145	150	0.52
	60	5	122.4	6.06630	118	118	118	129	130	0.67

ANGLE.MTW

Data Display

Data

Row	delay	angle
1	30	140
2	30	138
3	30	140
4	30	138
5	30	142
6	45	140
7	45	150
8	45	120
9	45	128
10	45	130
11	60	118
12	60	130
13	60	128
14	60	118
15	60	118

ANGLE.MTW

One-way ANOVA: angle versus delay

Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	$\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
delay	3	30, 45, 60

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
delay	2	762.1	381.07	6.56	0.012
Error	12	697.6	58.13		
Total	14	1459.7			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
7.62452	52.21%	44.25%	25.33%