Final exam, 16 December 2024

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*)

The one-hole test is used to test the manipulative skill of job applications. This test requires subjects to grasp a pin, move it to a hole, insert it, and return for another pin. The score on the test is the number of pins inserted in a fixed time interval. In one study, male college students were compared with experienced female industrial workers. The table below summarizes scores (giving the sample size, mean, variance and standard deviation for each group) for the first minute of the test.

Group	n	$\bar{X}$	s^2	s
workers	412	37.32	14.7	3.83
students	750	35.12	18.6	4.31

a) (*3 points*)
The interest is in comparing the mean scores between workers and students, with the expectation that the experienced workers would outperform the students, at least during the first minute before learning occurs. Describe the statistical design and formulate an appropriate statistical model to study this question. Carry out a suitable statistical test to answer the question and draw conclusions; you may use any of the enclosed Minitab listings to support your analysis.

b) (*4 points*)
The five-number summary of the data for the workers was:

$$23 \; - \; 33.5 \; - \; 37 \; - \; 40.5 \; - \; 46.$$

Use this information, as well as the information already provided, to describe the distribution of scores among the workers. Discuss the data's compliance with model assumptions, and assess whether your findings in **a)** could be seriously compromised by a non-compliance with model assumptions. In case you would like to carry out additional analyses to support your conclusions, state what information you need, how you would obtain it from statistical software, and indicate how it could affect your conclusions.

c) (*3 points*)
One purpose of the study was to develop performance norms for job applicants. Based on the data above, compute a range (interval) within which 95% of the scores for workers would be expected to fall (assuming the model's assumptions were met).

d) (*3 points*)
The study also included scores for the 15th minute of the one-hole test, and the five-number summary of these scores for the workers was:

$$32 \; - \; 39 \; - \; 44 \; - \; 49 \; - \; 59.$$

Use these numbers and the numbers given in **b)** to descriptively compare the distributions of scores in the 1st and 15th minute of the test. Your comparison should include both the location, spread and shape of the distributions. In addition, describe a statistical model and analysis to assess any differences between the scores in the 1st and 15th minute of the test. 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 it would be used in the analysis.

e) (*2 points*)
Returning to the data for the first minute scores studied in **a)-c)**, a way was sought to intuitively express any difference in performance between workers and students. For this purpose, it was suggested to calculate the probability (say p) of a worker scoring higher than a student, when workers and students are selected as representative of their respective populations. The task is to assess the range for this probability by selecting which of the following four ranges it falls into:

$$\begin{array}{ll} \textit{i)}: & p < 0.5, \\ \textit{ii)}: & 0.5 \leq p < 0.75, \\ \textit{iii)}: & 0.75 \leq p < 0.95, \\ \textit{iv)}: & 0.95 \leq p, \end{array}$$

where, as described above, p is the probability of a worker's score being higher than a student's score. Select one of four options *i)-iv)* as your assessment of p. (*Note:* It is possible to calculate p from the information provided, but you are not required to do so.)

(continues on the next page)

Minitab listings for Question 1

WORKSHEET 1

Two-Sample T-Test and CI

Method

μ_1: population mean of Sample 1
μ_2: population mean of Sample 2
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	750	35.12	4.31	0.16
Sample 2	412	37.32	3.83	0.19

Estimation for Difference

Difference	95% CI for Difference
-2.200	(-2.682, -1.718)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-8.95	933	0.000

WORKSHEET 1

Two-Sample T-Test and CI

Method

μ_1: population mean of Sample 1
μ_2: population mean of Sample 2
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	750	35.1	18.6	0.68
Sample 2	412	37.3	14.7	0.72

Estimation for Difference

Difference	95% CI for Difference
-2.200	(-4.148, -0.252)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$
Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
-2.22	1019	0.027

Question 2. (*17.5 points*)

American cranberry has long been used for treatment of bladder and kidney ailments. A randomized controlled trial was designed to test whether regular drinking of cranberry juice can prevent the recurrence of urinary tract infections (UTIs) in women. A total of 150 women with a UTI were treated with antibiotics and then randomly assigned to one of three groups. One group drank cranberry juice concentrate daily for six months; another group took a lactobacillus drink daily for six months (a lactose-fermenting bacteria thought to help inhibit the growth of UTI-causing bacteria); the last group served as the control group and drank neither cranberry juice nor lactobacillus drinks for 6 months. After 6 months, the number of women in each group with recurring symptomatic UTI was recorded. The results were (one of the women did not complete the trial):

Group	Women with recurring UTI	Women total
Cranberry juice	8	50
Lactobacillus drink	19	49
Control	18	50

a) (*4 points*)
In this study, the main interest was in the cranberry group. Compute a 95% confidence interval for the proportion of women with recurring UTIs in the cranberry group. Compute also a 95% confidence interval for the same proportion in the control group. Do these two confidence intervals allow you to conclude about any significant difference between the two groups? Explain your answer (briefly, no further calculations are required)

b) (*4 points*)
A full statistical analysis of the data comprises all 3 groups. Formulate a statistical model for the data, carry out a statistical test to assess whether the occurrence of UTIs differs between the three groups, and draw conclusions for the test. Make sure to include estimates from relevant conditional or marginal distributions in your analysis. You may use any of the Minitab listings included for parts **b)-c)** to support your analysis. Note that further discussion of the analysis follows in the next question.

c) (*4 points*)
Use your model and analysis from **b)** to choose one of the options given for each of the sections *i)–iv)* below (there is only one correct choice in each section). No additional explanation needs to be given unless explicitly indicated, but make sure it is clear which model and analysis your answers are based on.

i) The condition for use of the chi-square test is met for these data because:
(a) the total sample size (number of women) is large; (b) all of the cell counts are greater than 5; (c) all of the expected cell counts are greater than 5.

ii) The contributions to the chi-square statistic for the different cells:
(a) are approximately equal for all cells; (b) include one cell with substantially larger contribution than all other cells (indicate which cell); (c) include two cells with substantially larger contributions than all other cells (indicate which two cells); (d) another pattern (describe the pattern).

iii) The alternative hypothesis for the test was:
(a) all treatment groups have different UTI occurrences; (b) a difference in UTI occurrence exists between at least two treatment groups; (c) women with and without recurring UTIs have different distributions on treatment groups; (d) there is a dependence between treatment groups and UTI occurrence.

iv) If the researchers wanted to compare all pairs of treatment groups, a Bonferroni adjustment for multiple comparisons:
(a) is appropriate and involves 3 comparisons; (b) is appropriate and involves 2 comparisons; (c) is inappropriate because Bonferroni adjustments only apply to ANOVA.

d) (*3.5 points*)
In the journal article describing the study, the authors chose to focus the reporting of the results on the comparison of the cranberry and control groups only. Use a statistical test to assess whether the study provided statistical evidence of a different occurrence of UTIs in these two groups. Draw conclusions about any effects demonstrated by the data/study.

e) (*2 points*)
The journal article describing the study included the following paragraph on sample size.

> "We calculated the sample size based on the assumption that at least 30% of women will experience a recurrence of urinary tract infection within a year. We considered a reduction in recurrences to 10% as clinically important. To detect such a reduction with a two tailed α of 0.05 and a power of 80%, we needed 70 women in each group."

Discuss, based on the Minitab listing below, the above sample size calculation (e.g., its validity) and how it matches the study (i.e., how well the actual study matched the sample size calculation).

CRANBERRY.MWX

Power and Sample Size

Test for Two Proportions
Testing comparison p = baseline p (versus ≠)
Calculating power for baseline p = 0.1
α = 0.05

Results

Comparison p	Sample Size	Target Power	Actual Power
0.3	62	0.8	0.802599

The sample size is for each group.

Minitab listings for Question 2, parts b)-c)

Rows: group Columns: UTI

	yes	total	All
cranberry	8	50	58
	13.45	44.55	
lactobac	19	49	68
	15.77	52.23	
control	18	50	68
	15.77	52.23	
All	45	149	194

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Valu
Pearson	4.147	2	0.126
Likelihood Ratio	4.452	2	0.108

Rows: group Columns: UTI

	yes	no	total	All
cranberry	8	42	50	100
	15.10	34.90	50.00	
lactobac	19	30	49	98
	14.80	34.20	49.00	
control	18	32	50	100
	15.10	34.90	50.00	
total	45	104	149	298
	45.00	104.00	149.00	
All	90	208	298	596

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Valu
Pearson	7.290	6	0.295
Likelihood Ratio	7.796	6	0.253

Rows: group Columns: UTI

	yes	no	All
cranberry	8	42	50
	15.10	34.90	
lactobac	19	30	49
	14.80	34.20	
control	18	32	50
	15.10	34.90	
All	45	104	149

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Valu
Pearson	7.290	2	0.026
Likelihood Ratio	7.796	2	0.020

Question 3. (*17.5 points*)
An experiment was carried out to compare two different methods for producing insulin with regard to the insulin's reduction of serum glucose levels. Each insulin type (A or B) was injected, for each of three different doses (2.25, 4.00, or 5.75 units), into four different rats. Thus, the experiment involved a total of 24 rats. The percentage reduction of the serum glucose level after the insulin injection was recorded for each rat. The table below displays the reductions recorded for individual rats as well as, in parenthesis, the mean and standard deviation for each group of 4 rats as well as overall.

Glucose reduction (%)	Dose			
Insulin type	2.25	4.00	5.75	Total
A	17 21 49 54 (35.25 / 18.95)	34 49 63 64 (52.50 / 14.11)	61 62 72 91 (71.50 / 13.92)	(53.08 / 21.07)
B	16 33 37 40 (31.50 / 10.72)	34 41 64 64 (50.75 / 15.56)	56 57 62 72 (61.75 / 7.32)	(48.00 / 16.61)
Total	(33.38 / 14.39)	(51.63 / 13.78)	(66.63 / 11.54)	(50.54 / 18.82)

For example, for insulin type A injected at dose 2.25, the reductions (in %) for the 4 rats were: 17, 21, 49, and 54, corresponding to a mean of 35.25 and a standard deviation of 18.95.

a) (*3 points*)
Characterize the statistical design (including descriptors such as: experimental unit, factor, design type, replication), and describe briefly how you would carry out a proper randomization for the experiment.

b) (*4 points*)
Suggest an appropriate statistical model for the data, including the model assumptions. Use the information provided above and in the Minitab listings for parts **b)-d)** to discuss whether the model assumptions seem to be met to a reasonable degree. If you need extra information to discuss this, specify what information you need, how you would obtain it from statistical software, and how you would use it. If you identify concerns with the model assumptions, describe briefly how these could be dealt with.

Your answers to the next two questions should be based on the analysis shown in the Minitab listings, without consideration of any concerns you might have had about the model assumptions or of any suggestions for alternative modeling.

c) (*4 points*)
The question of primary interest to the investigators was whether the two insulin types had different effects on the reduction of glucose levels. Such differences could either be related to specific insulin doses or could be general across doses. Explore these questions by statistical inference (including statistical significance) and interpret the graphical displays of relevance for the questions. In addition, compute a 95% confidence interval for the difference in mean glucose level reduction between the two insulin types.

d) (*3.5 points*)
Although of secondary interest, the investigators still wanted to explore effects of the insulin dose on the reduction of glucose levels. Determine and state the statistical significance of such effects, overall and (if appropriate) between specific doses, and draw conclusions. Also here, interpret the relevant graph(s) and address whether dose effects (if existent) appear to have a systematic direction.

e) (*3 points*)
In addition to the analysis in **d)**, it was also of interest to establish a functional relationship with the actual insulin doses. For simplicity, we consider an approch that does not account for insulin type, although strictly speaking not quite correct. Explain the statistical model of the Minitab listing shown below and interpret the results (briefly). Finally, use the model and results to predict the expected reduction of the serum glucose level for a rat given a dose of 5 units, with a suitable, approximate 95% interval. (*Hint:* It is allowed to determine the interval from the figure.)

INSULIN.MTW

Regression Analysis: reduc versus dose

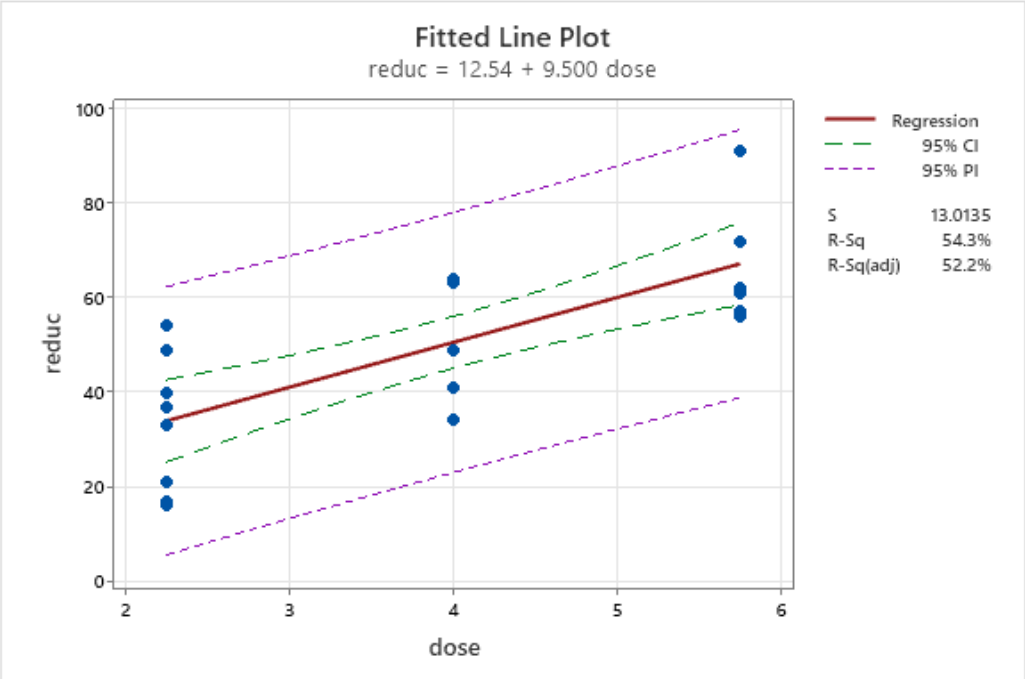
The regression equation is
reduc = 12.54 + 9.500 dose

Model Summary

S	R-sq	R-sq(adj)
13.0135	54.27%	52.20%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	4422.25	4422.25	26.11	0.000
Error	22	3725.71	169.35		
Total	23	8147.96			



Minitab listings for Question 3, parts b)-d)

INSULIN.MTW

ANOVA: reduc versus dose, type

Factor Information

Factor	Type	Levels	Values
dose	Fixed	3	2.25, 4.00, 5.75
type	Fixed	2	A, B

Analysis of Variance for reduc

Source	DF	SS	MS	F	P
dose	2	4436.33	2218.17	11.45	0.001
type	1	155.04	155.04	0.80	0.383
dose*type	2	69.33	34.67	0.18	0.838
Error	18	3487.25	193.74		
Total	23	8147.96			

Model Summary

S	R-sq	R-sq(adj)
13.9189	57.20%	45.31%