

Mid-term exam, 5 March 2026

All aids are allowed, except a computer and personal assistance, as well as the use of any external information pertaining to the specific data and questions. 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 one question with five parts, or sub-questions (labeled by letters **a**) – **e**)), which should all be answered.

The mid-term exam accounts for 15% of the course mark; however, every student may choose to waive the result of the mid-term exam. The duration of the mid-term exam is 1 hour.

Generally, **statistical models and methods should be specified**, and all statistical analyses should be summarized in conclusions.

Question 1. (15 points)

As part of a study in human medicine on Hodgkin's disease (or Hodgkin's lymphoma, a cancer in the lymphomatic system), lymphocyte counts were measured in blood samples from 20 patients in remission from Hodgkin's disease. In the Minitab data listing below, the variables T4 and T8 give the numbers of T_4 and T_8 cells per mm^3 in these blood samples, respectively. Lymphocyte counts are sometimes analysed on logarithmic scale, and the listing also includes variables, **lnT4** and **lnT8**, with the values for each count on natural log scale. The researchers were, among other things, interested in comparing the levels of T_4 and T_8 cells. The listing therefore additionally contains the differences, both on original and logarithmic scales; the corresponding variables are named **diffT4T8** and **difflnT4lnT8**, respectively.

Use this information about the study as well as the output from Minitab shown on the following pages to answer the questions **a**) – **e**). The questions can be answered independently of each other, although it may be advantageous to answer them in the order they are listed.

a) (2 points)

Characterize the study type (e.g., experimental or another type) and the statistical design comprised by the T_4 and T_8 cell counts

b) (4 points) The Minitab output contains descriptive statistics and graphical representations of the distribution of the numbers of T_4 and T_8 cells, computed on both original and natural-log transformed scales. Use (some of) this information to (briefly) describe the distribution for both measurements. Concentrate on the features you would find notable before commencing a statistical analysis.

c) (3 points) Compute a 95% confidence interval for the mean T_4 cell count on *either* original or logarithmic scale, of your own choice. Interpret the confidence interval by explaining in non-technical terms what it tells you about T_4 cell counts among Hodgkin's patients. Discuss briefly whether this interval should be considered as approximate or exact.

- d) (3 points) Carry out a statistical analysis to compare the levels of T_4 and T_8 cells among Hodgkin's patients. Select the method and scale you find appropriate for the analysis based on the information provided (indicate if you think further information is needed) and justify your choice. Draw conclusions from the test and explain in non-technical terms what the analysis tells you about the levels of T_4 and T_8 cells when compared to each other. (*Hint*: You may find the following formula useful: $\ln(x_1) - \ln(x_2) = \ln(x_1/x_2)$.)
- e) (3 points) For this question, you should answer **one of the two parts i)-ii)** below, each with the same score. It is allowed (but not recommended) to answer both parts, in which case your score for **e)** will be for the best part among those two.
- i) As a supplement to the quantitative comparison between T_4 and T_8 cell levels, the researchers also wanted to assess how likely it is for a Hodgkin's patient to have higher T_4 than T_8 cell levels. Estimate this probability from the data provided, with a 95% confidence interval. In another patient group, every patient had higher T_4 than T_8 cell levels — how does that (informally) compare to Hodgkin's patients?
- ii) The study also included 20 patients with diverse, disseminated malignancies which we will simply refer to as the non-Hodgkin's group. The main focus of the study was to compare the Hodgkin's and non-Hodgkin's groups with respect to their lymphocyte counts. Outline how you would carry out statistical inference to compare T_4 cell levels between Hodgkin's and non-Hodgkin's patients. Your outline should explain decisions (or necessary steps towards decisions) about the choice of scale for analysis, the choice of statistical model and parameter of interest, and relevant details about the statistical inference (e.g., confidence interval or test) you want to use.

Minitab output for Question 1:

Data							
Row	patient	T4	T8	lnT4	lnT8	diffT4T8	difflnT4lnT8
1	1	431	311	6.066	5.740	120	0.326
2	2	397	340	5.984	5.829	57	0.155
3	3	554	670	6.317	6.507	-116	-0.190
4	4	435	446	6.075	6.100	-11	-0.025
5	5	1378	686	7.228	6.531	692	0.698
6	6	1104	1335	7.007	7.197	-231	-0.190
7	7	902	412	6.805	6.021	490	0.784
8	8	795	449	6.678	6.107	346	0.571
9	9	1621	811	7.391	6.698	810	0.693
10	10	1212	1678	7.100	7.425	-466	-0.325
11	11	288	236	5.663	5.464	52	0.199
12	12	295	262	5.687	5.568	33	0.119
13	13	1004	786	6.912	6.667	218	0.245
14	14	1283	336	7.157	5.817	947	1.340
15	15	958	286	6.865	5.656	672	1.209
16	16	171	212	5.142	5.357	-41	-0.215
17	17	2415	936	7.789	6.842	1479	0.948
18	18	396	836	5.981	6.729	-440	-0.747
19	19	257	272	5.549	5.606	-15	-0.057
20	20	568	978	6.342	6.886	-410	-0.543

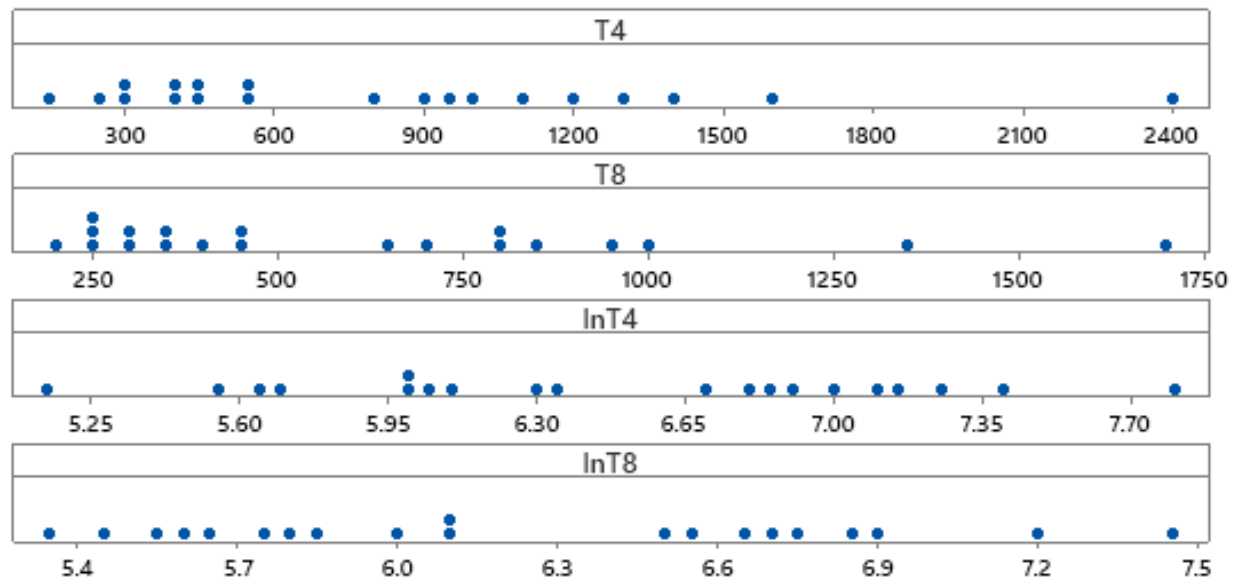
(continues on the next page)

Descriptive Statistics: T4, T8, lnT4, lnT8

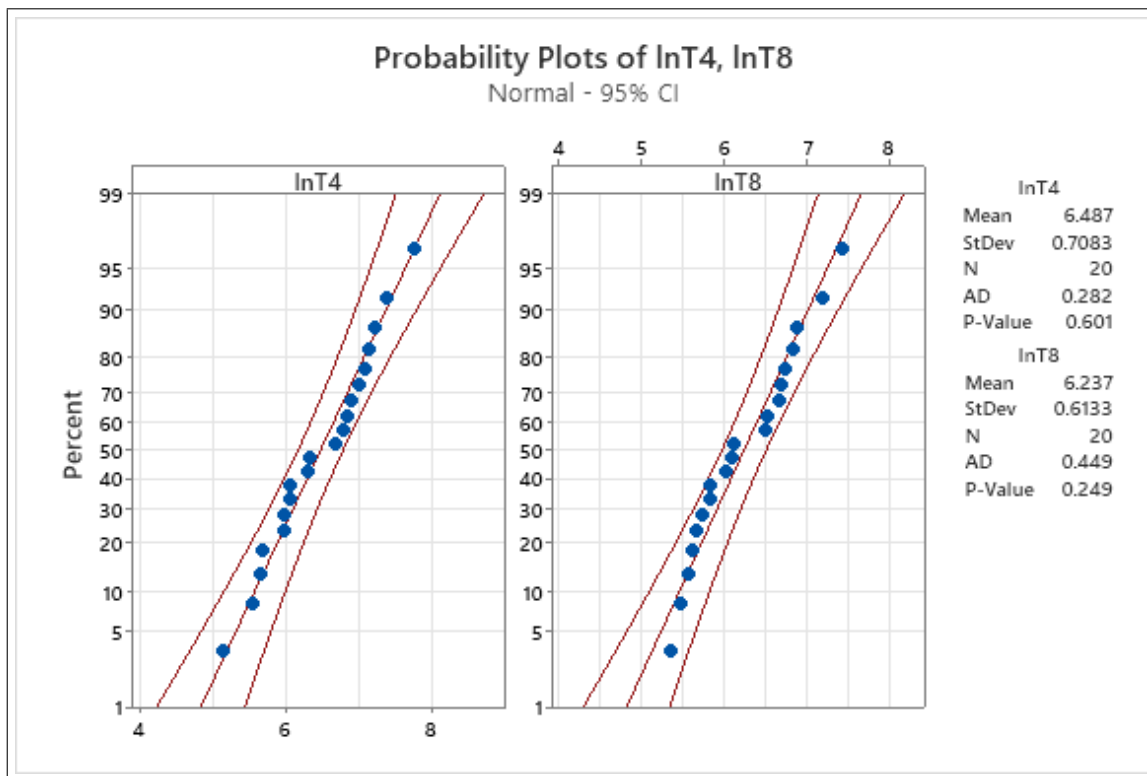
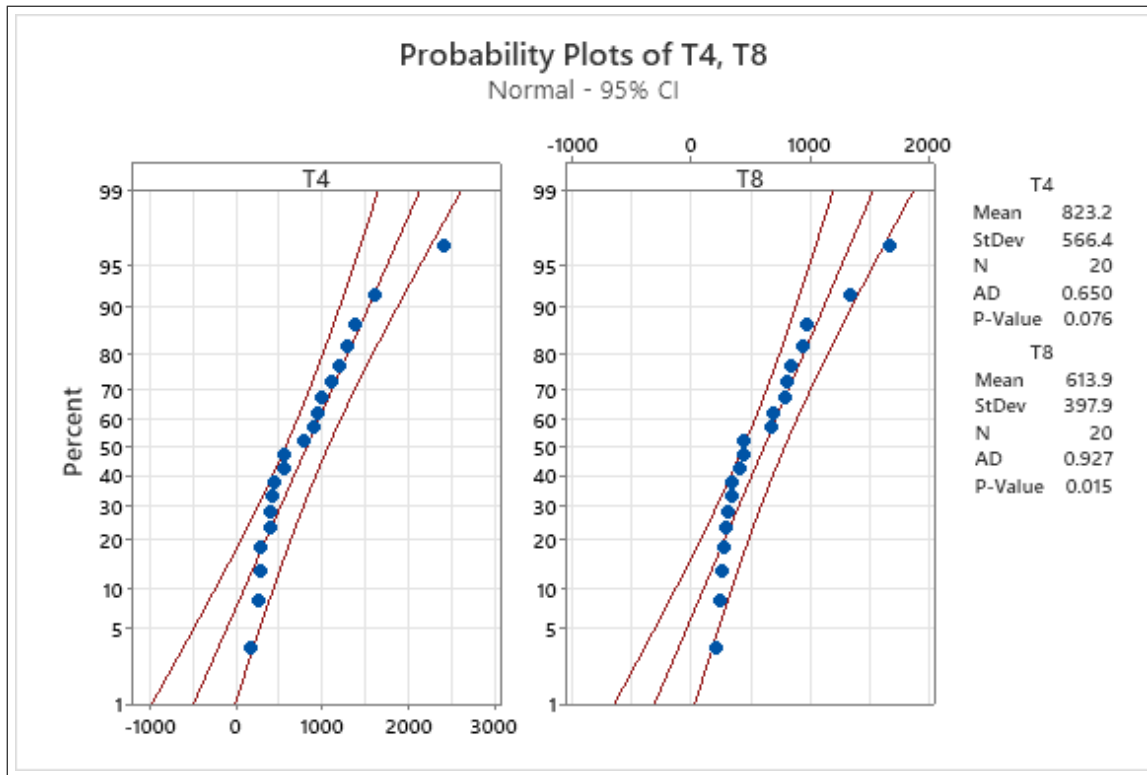
Statistics

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Skewness
T4	20	823.2	566.385	171	396.25	681.5	1185	2415	1.23
T8	20	613.9	397.854	212	292.25	447.5	829.75	1678	1.26
lnT4	20	6.48693	0.708302	5.14166	5.98204	6.51023	7.07669	7.78945	-0.09
lnT8	20	6.23730	0.613288	5.35659	5.67694	6.10367	6.72104	7.42536	0.31

Dotplots of T4, T8, lnT4, lnT8



(continues on the next page)



(continues on the next page)

WORKSHEET 1

Two-Sample T-Test and CI: T4, T8

Method

μ_1 : population mean of T4

μ_2 : population mean of T8

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
T4	20	823	566	127
T8	20	614	398	89

Estimation for Difference

95% CI for	
Difference	Difference
209	(-105, 524)

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
1.35	34	0.185

WORKSHEET 1

Two-Sample T-Test and CI: lnT4, lnT8

Method

μ_1 : population mean of lnT4

μ_2 : population mean of lnT8

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
lnT4	20	6.487	0.708	0.16
lnT8	20	6.237	0.613	0.14

Estimation for Difference

95% CI for	
Difference	Difference
0.250	(-0.175, 0.674)

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
1.19	37	0.241

WORKSHEET 1

One-Sample T: diffT4T8, difflnT4lnT8

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean	95% CI for μ
diffT4T8	20	209	506	113	(-28, 446)
difflnT4lnT8	20	0.250	0.569	0.127	(-0.016, 0.516)

μ : population mean of diffT4T8, difflnT4lnT8

Test

Null hypothesis $H_0: \mu = 0$

Alternative hypothesis $H_1: \mu \neq 0$

Sample	T-Value	P-Value
diffT4T8	1.85	0.080
difflnT4lnT8	1.96	0.064