Final exam, 13 December 2021

All aids are allowed, except a computer and personal assistance. Restricted use of some computer-like devices (including 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 was carried out (some time ago) to investigate the impact of early breeding of heifers. Monozygotic twin pairs of calves were split onto two treatment groups, one where calves were bred early after having received feed supplementation, and another group where calves were bred and fed as usual. In each twin pair, one calf was randomly assigned to the early breeding group and the other calf to the normal breeding group. Early breeding is generally considered to be more stressful for the animals, and may therefore lead to lower and/or more variable performance in the subsequent lactation. Here we consider data on milk yield through the first lactation (up to 250 days, and measured in kg) for 23 such twin pairs. The table on the next page shows all the data values, including also the difference between yields of the two heifers in each pair, computed as "Normal minus Early". Use the information in the table and in the subsequent Minitab listings to answer the questions **a)-f)** below.

a) (*2 points*)
Characterize the study type (e.g., experimental or another type) and statistical design (e.g., one-sample or another design). Discuss briefly, e.g. by using specific examples, which observations may be considered as independent of each other, and which observations (if any) should not be considered as independent of each other.

b) (*3 points*)
Carry out a brief comparative, descriptive analysis of the milk yield measurements for the two breeding groups; contrast the distributions using standard descriptors such as the center, spread, distributional shape and extremes.

c) (*2 points*)
The study included more heifer pairs than presented here. One particular pair had yield values $(3100, 4700)$ for the pair of (normal, early) bred heifers. Does this pair in your view constitute an outlier when compared to the data shown? Support your judgement by relevant statistical calculation(s).

Data table for Question 1:

Milk yield (kg)	Breeding group		
Pair	Normal	Early	Difference
1	4339	3042	1297
2	3883	4354	-471
3	4728	4758	-30
4	5626	4902	724
5	4283	3145	1138
6	3832	1887	1945
7	3672	3265	407
8	4556	4269	287
9	5049	4894	155
10	4171	2282	1889
11	3326	2601	725
12	3080	3127	-47
13	4950	5121	-171
14	4744	2720	2024
15	4658	3323	1335
16	4769	3150	1619
17	3734	4043	-309
18	3031	1173	1858
19	3881	3599	282
20	5098	3378	1720
21	5152	3192	1960
22	4071	2523	1548
23	2365	2528	-163
mean	4217	3355	862
stand. dev.	794	1013	857

d) (*4 points*)
The primary purpose of the study was to compare the two breeding groups. With this purpose in mind, formulate an appropriate statistical model and analyse the data to assess whether a difference seems to exist – and if so, seems to align with the expectations prior to the study. Formulate your conclusion both statistically and in a less technical version that could be used as a valid summary for a non-statistical audience.

e) (*2 points*)
The researchers also discussed whether a nonparametric analysis would be preferable to assess the research question. For the two questions below, select your preferred answers among those listed, as their relate to a nonparametrical analysis for *these data*. Note that each question may have **more than one correct answer**.

(A) Which nonparametric analyses would be applicable to statistically compare the two breeding groups?

i) the Mann-Whitney-Wilcoxon rank (sum) test,
ii) the Wilcoxon signed rank test,
iii) the Kruskal-Wallis test,
iv) the sign test,
v) another nonparametric method (give details): ......

(B) Which of the following statements are valid and relevant to describe the rationale for choosing between a parametric and a nonparametric method *for these data*?

i) a nonparametric method is attractive because the data show non-normal distribution(s) for the outcome(s) of interest,
ii) nonparametric methods (i.e., those listed above) offer statistical inference about the population mean without assuming any specific distribution type,
iii) the t-test is not robust enough to give valid inference about the population mean,
iv) the conclusion for the comparison of yields between the two groups is so strong, that it is unlikely to differ between parametric and nonparametric analyses.

f) (*2 points*)
In addition to the mean yield, the researchers also wanted to quantify the difference between the two breeding groups in terms of the probability of first lactation milk yields to exceed 4000 kg. Make relevant distributional assumptions and estimate this probability for a heifer in the normal breeding group. Additionally, use this value to estimate the probability that two normally bred and randomly selected heifers from the population both have yields above 4000 kg. Finally, discuss whether you could use the same approach to estimate the probability that in a twin pair with one normally bred and one early bred heifer, both of them have milk yields above 4000 kg (without carrying out any calculations).

Minitab listings and plots for Question 1:

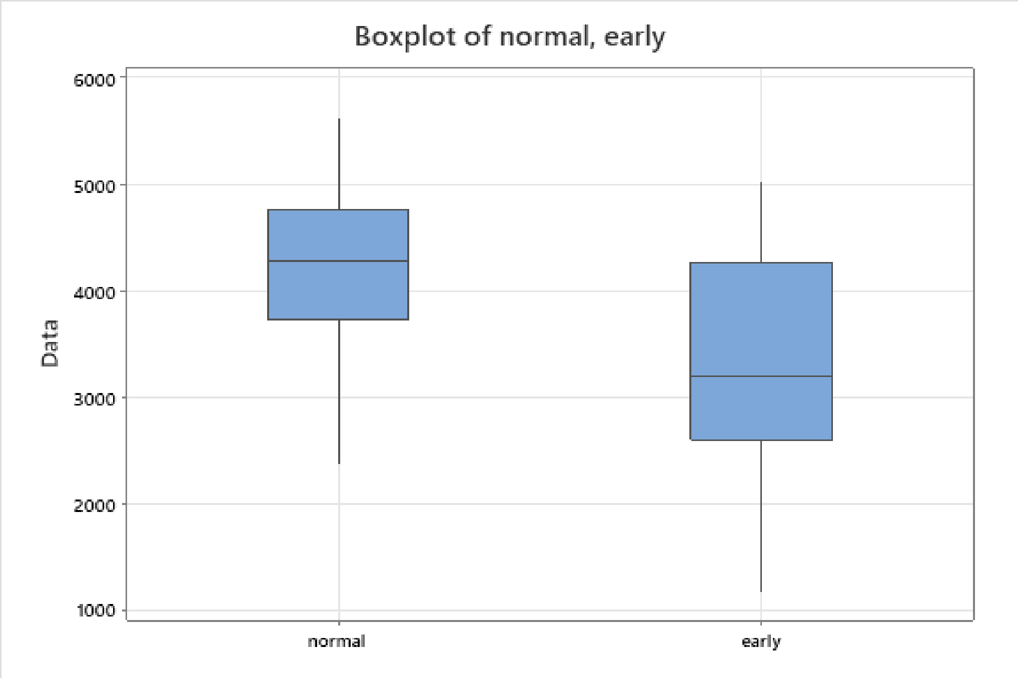
HEIFER

Descriptive Statistics: normal, early, diff

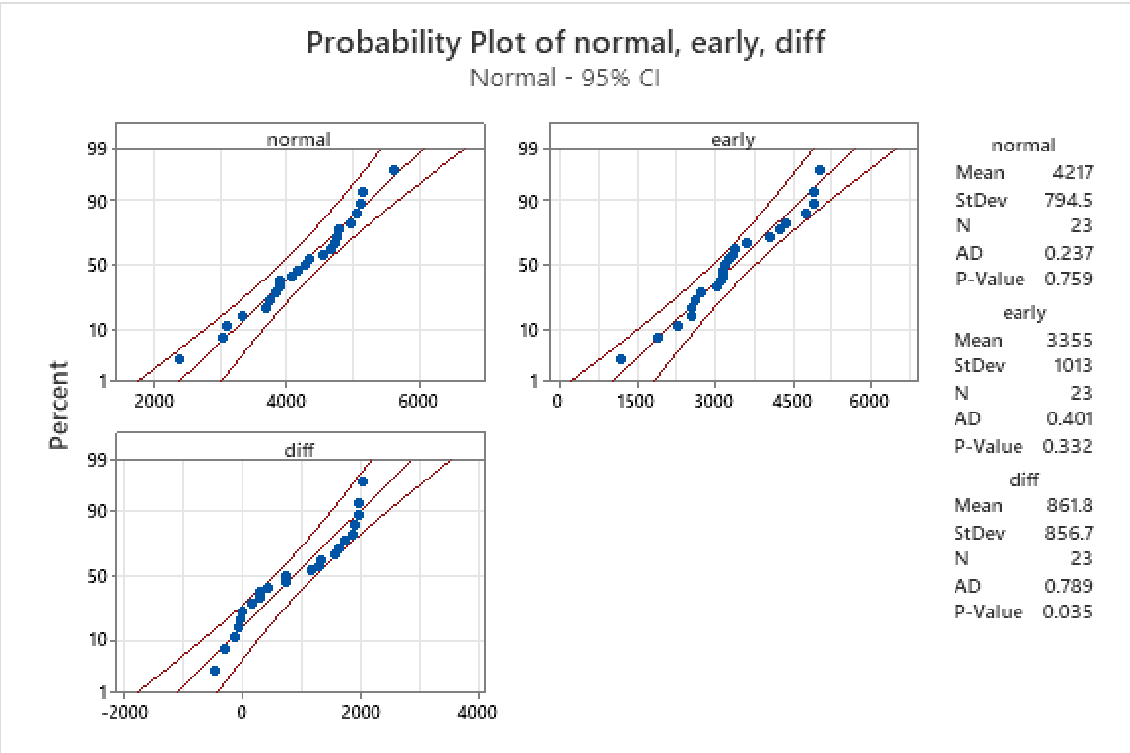
Statistics

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Skewness
normal	23	4217	794	2365	3734	4283	4769	5626	-0.47
early	23	3355	1013	1173	2601	3192	4269	5021	0.01
diff	23	862	857	-471	-30	725	1720	2024	-0.03

(continues on next page)



Boxplot of normal, early
Data
6000
5000
4000
3000
2000
1000
normal
early



Probability Plot of normal, early, diff
Normal - 95% CI
Percent
normal
early
diff
99
90
50
10
1
2000
4000
6000
0
1500
3000
4500
6000
-2000
0
2000
4000
normal
Mean 4217
StDev 794.5
N 23
AD 0.237
P-Value 0.759
early
Mean 3355
StDev 1013
N 23
AD 0.401
P-Value 0.332
diff
Mean 861.8
StDev 856.7
N 23
AD 0.789
P-Value 0.035

Question 2. (*17.5 points*)

When (human) patients undergo surgery, the operating room is kept cool (for various reasons). The exposure to cold, in addition to impairment of temperature regulation caused by anesthesia and altered distribution of body heat, may result in mild hypothermia (approximately 2°C below the normal core body temperature). As a result of the hypothermia, patients may have an increased susceptibility to postoperative wound infections or heart conditions. We discuss two studies from the 1990s to investigate whether keeping the patients' temperature within a normal range using a blanket has beneficial effects on outcomes of an operation.

An Austrian study focused on the risk of infection and included patients undergoing colon or rectal surgeries, which are typically associated with a high risk of infection. Random assignment of patients to the two temperature management groups was done during the induction of anesthesia. In the *normothermic* group, patients' core temperatures were maintained at near normal 36.5°C, while in the *hypothermic* group, patients' core temperatures were allowed to decrease to about 34.5°C. Both groups used a forced-air cover on the upper bodies of patients during the operation, but it delivered heated air in the normothermic group only (it delivered surrounding air in the hypothermic group). In order to keep the surgeons and operating room personnel from detecting which patient was in which group, shields and drapes were placed over all devices which would indicate group assignment. There was no postoperative control of patient temperatures, and the patients did not know their group assignments. Determinations of when to begin postoperative feeding, suture removal, and hospital discharge were made by attending surgeons who were unaware of the patients' group assignments and core temperatures during surgery.

a) (*2 points*)
Based on the description above, characterize the study type (e.g., experimental vs. observational) and design (e.g., one-sample or two-sample). Note also any additional features of the study design of importance for the validity of its conclusions.

A few results of the study are shown in the table below.

Variable	Group normothermic	Group hypothermic
Total number of patients	104	96
Patients with postoperative infections: number (%)	6 (5.8%)	18 (18.8%)
Days of hospitalization: mean ± stand. dev.	12.1 ± 4.4	14.7 ± 6.5

b) (*4 points*)
Use the information from the table, as well as the attached Minitab listings, to construct a 99% confidence interval for the difference between the mean length of stay in the hospital for the two groups. What conclusion may be drawn from this confidence interval about the P-value for a significance test to compare the groups?

c) (*4 points*)
Use the information from the table, as well as the attached Minitab listings, to give a statistical assessment of whether the treatment affects the proportion of postoperative infections, and draw conclusions about the impact of the treatment.

A North American study focused on patients with high risk of coronary conditions/diseases. It used the same two treatment groups, generally implemented in a similar way (though not identical in all details). A few results from this study are shown in the table below; "cardiac event" includes a range of different events of abnormal heart function.

Variable	Group normothermic	Group hypothermic
Total number of patients	142	158
Minor cardiac event during operation: number (%)	13 (9.2%)	15 (9.5%)
Major event up to 24h after operation: number (%)	2 (1.4%)	10 (6.3%)
Duration of surgery: mean ± stand. error	3.6 ± 0.9	3.4 ± 1.1
Estimated blood loss (ml): mean ± stand. error	390 ± 70	520 ± 60

d) (*4.5 points*)
Outline the approach you would use for a statistical analysis to study each of the parameters:

i) the effect of treatment on occurrence of major cardiac events after operation,

ii) the association between duration of surgery and blood loss during surgery,

iii) the association between occurrences of cardiac events during and after operation.

Indicate if the information presented would allow you to carry out the analysis (without statistical software), or whether additional information is needed (specify what you need). In all cases i)-iii), a discussion of the statistical approach and any particular issues to be aware of in the analysis is considered sufficient for a full score, without necessarily carrying out any extra analysis.

e) (*3 points*)
In a discussion about the results of the table for the duration of surgery and the estimated blood loss during surgery, the following three views were put forward.

1) Approximate 95% confidence intervals for the mean blood loss during surgery are: $390 \pm 2 \cdot (70/\sqrt{142}) = (378, 402)$ for normothermic patients, and correspondingly $520 \pm 2 \cdot (60/\sqrt{158}) = (510, 530)$ for hypothermic patients.

2) About 95% of the surgeries for patients in the normothermic group lasted between 1.8 and 5.4 hours, whereas in the hypothermic group about 95% of the surgeries lasted between 1.2 and 5.6 hours.

3) The standard error of the mean for duration of surgery (or blood loss) in either group seems unusually large, since the standard deviation for the individual observations is equal to the standard error of the mean multiplied by $\sqrt{n}$. Thus, the mean ± one standard deviation would go far outside the range of possible values for the duration of surgery (or blood loss).

For each of these statements, indicate whether it is correct or incorrect. It is *not* required to give explanations to your assessments. However, added explanations can be useful because partly correct arguments behind an incorrect answer may still be counted positively towards your score.

Minitab listings and plots for Question 2, parts b)- c):

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	104	12.10	4.40	0.43
Sample 2	96	14.70	6.50	0.66

Estimation for Difference

Difference	95% CI for Difference
-2.600	(-4.163, -1.037)

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
-3.29	165	0.001

(continues on next page)

WORKSHEET 1

Tabulated Statistics: Worksheet rows, Worksheet columns

Rows: Worksheet rows   Columns: Worksheet columns

	C1	C2	All
1	6	18	24
	11.79	12.21	
2	104	96	200
	98.21	101.79	
All	110	114	224

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	6.251	1	0.012
Likelihood Ratio	6.528	1	0.011

WORKSHEET 1

Tabulated Statistics: Worksheet rows, Worksheet columns

Rows: Worksheet rows   Columns: Worksheet columns

	C3	C4	All
1	6	18	24
	12.48	11.52	
2	98	78	176
	91.52	84.48	
All	104	96	200

Cell Contents
Count
Expected count

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	7.965	1	0.005
Likelihood Ratio	8.237	1	0.004

Question 3. (*17.5 points*)

A study was carried out to investigate the earth vibrations caused by waterfalls. The paper published on the study (Rinehart JS (1969), *Science* **164**, 1513 – 1514) described the findings as follows:

> "Typically a waterfall produces continuous earth vibrations with one frequency predominating; this frequency is inversely proportional[1] to the height of the waterfall."

In this question, we will consider the dataset these findings were based on. It includes a total of nine waterfalls in the United States, Canada and Iceland, where for each of them the height was recorded together with the frequency, measured by a suitable instrument at the brink or base of the waterfall. The resulting nine values of frequency (Hz) and height (m) are shown in the Minitab listing, together with the values of two additional variables: the inverse height (m^{-1}), calculated as 1/`height`, and the inverse frequency (Hz^{-1}), calculated as 1/`freq`.

a) (*5 points*)
Consider the four analyses presented in the Minitab listings following the list of data values. Choose among those four analyses the one that in your view best matches the description of the biological mechanisms and the purpose of analysis reflected in the above quote. Specify the statistical model your chosen analysis is based on, and discuss how well the model describes the data, in particular whether the model assumptions appear to be met to a reasonable degree. Include a brief statement about the predictive ability of the model.

b) (*4 points*)
Estimate all the parameters of the model, and explain their interpretation. Give also 95% confidence intervals for the parameters (where feasible).

c) (*4 points*)
In the paper's description of the relation between the height and frequency, it is implicitly understood that the relation determined from the data could not be attributed to chance. Use a statistical test to quantify the evidence contained in the data of a relation (of the type investigated) between height and frequency, and draw conclusions.

In the description of the relation between the height and frequency (and in the graph shown in the paper), it is also implicitly understood that when extrapolating the relation it will pass through the origin (that is, the point where both $x = 0$ and $y = 0$). Give a statistical assessment (using a test or a confidence interval) of the evidence contained in the data in favour of or against the fact that the relation passes through the origin.

d) (*4.5 points*)
Assume that another researcher was not quite convinced about the correctness or accuracy of the relation, and wanted to collect information (height and frequency) about one additional waterfall in order to validate the relation. What height should this researcher

[1] Two variables are inversely proportional (or varying inversely, or in inverse variation, or in inverse proportion or reciprocal proportion) if one of the variables is directly proportional with the multiplicative inverse (reciprocal) of the other, or equivalently if their product is a constant. (Wikipedia)

approximately target for the additional waterfall, in order to get maximal "value" out of the extra observation? Explain your recommended value (*Note:* It is beyond the methods of VHM 801 to determine a single, definitively "best" choice of the height.) Compute the predicted frequency for your recommended height. Explain how you would assess statistically whether the observed frequency for this additional waterfall is in compliance with the model, and use the information provided to roughly determine a range where the observed frequency for the additional waterfall would not considered to be in disagreement with the current data and model.

Minitab listings and graphical displays for Question 3:

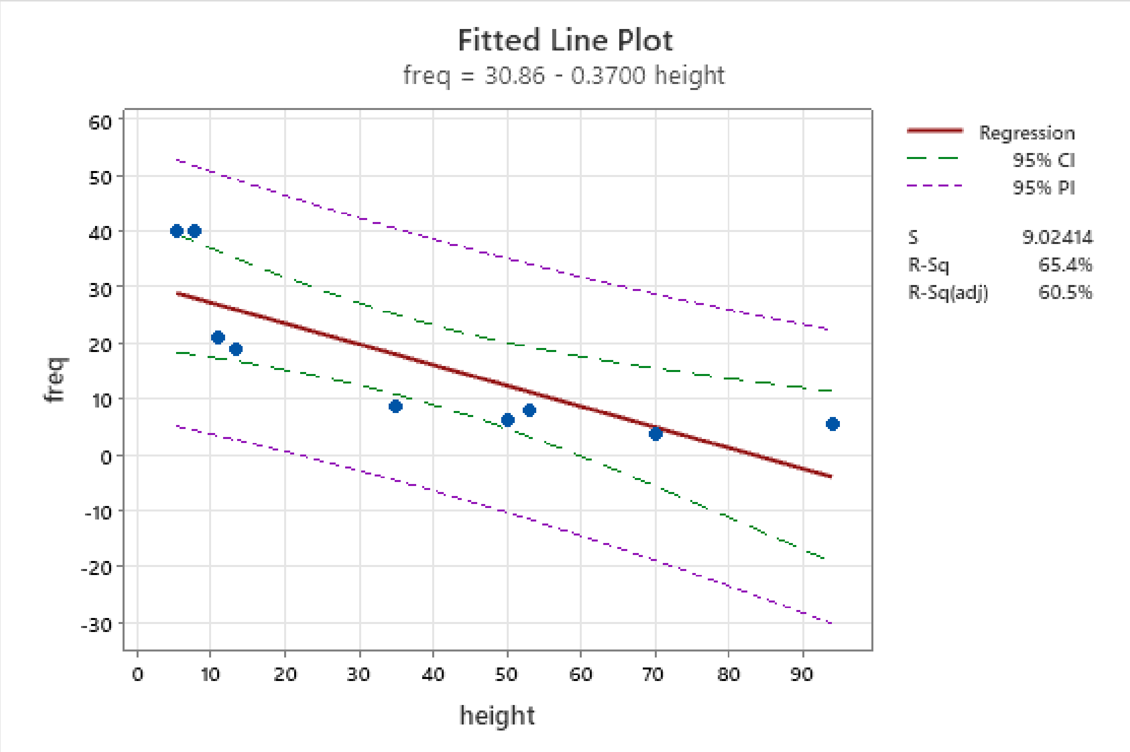
WATERFALL

Data Display

Data

Row	freq	height	invheight	invfreq
1	5.5	94.0	0.010638	0.181818
2	3.9	70.0	0.014286	0.256410
3	6.2	50.0	0.020000	0.161290
4	8.0	53.0	0.018868	0.125000
5	8.7	35.0	0.028571	0.114943
6	19.0	13.4	0.074627	0.052632
7	21.0	10.9	0.091743	0.047619
8	40.0	7.6	0.131579	0.025000
9	40.0	5.2	0.192308	0.025000

(continues on the next four pages)



WATERFALL

Regression Analysis: freq versus height

Regression Equation

freq = 30.86 - 0.370 height

Coefficients

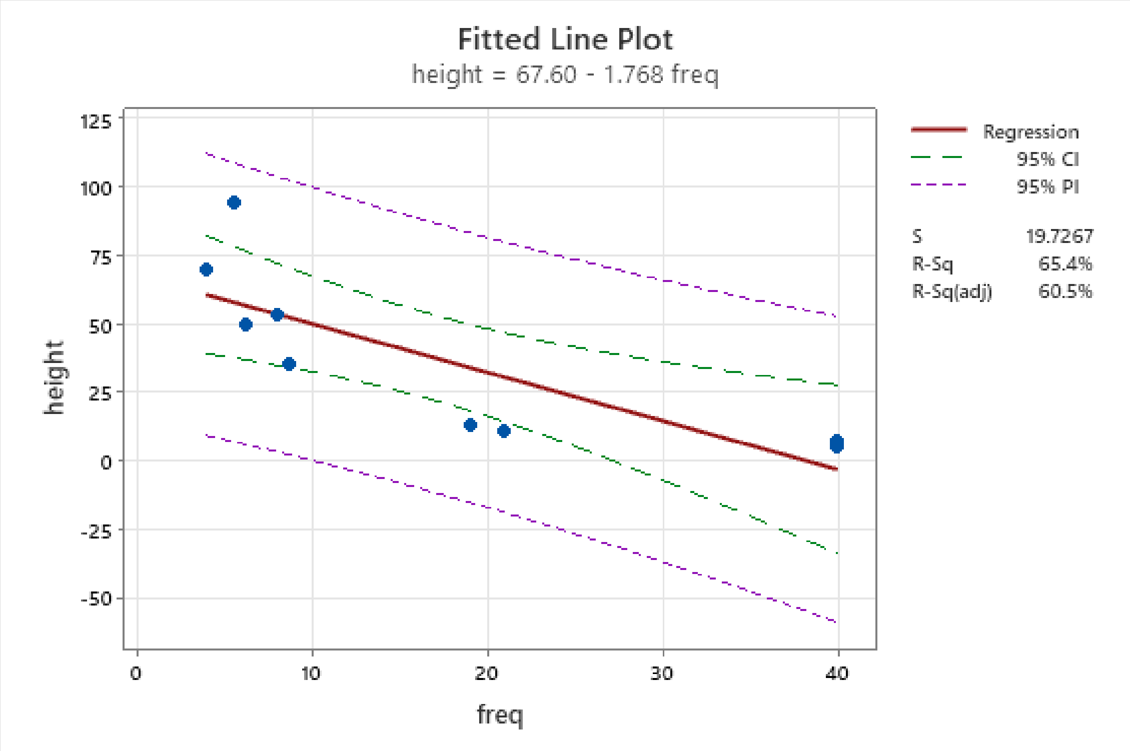
Term	Coef	SE Coef	T-Value	P-Value
Constant	30.86	4.87	6.34	0.000
height	-0.370	0.102	-3.64	0.008

Model Summary

S	R-sq	R-sq(adj)
9.02414	65.42%	60.48%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1078.3	1078.29	13.24	0.008
Error	7	570.0	81.44		
Total	8	1648.3			



WATERFALL

Regression Analysis: height versus freq

Regression Equation

height = 67.6 - 1.768 freq

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	67.6	10.5	6.42	0.000
freq	-1.768	0.486	-3.64	0.008

Model Summary

S	R-sq	R-sq(adj)
19.7267	65.42%	60.48%

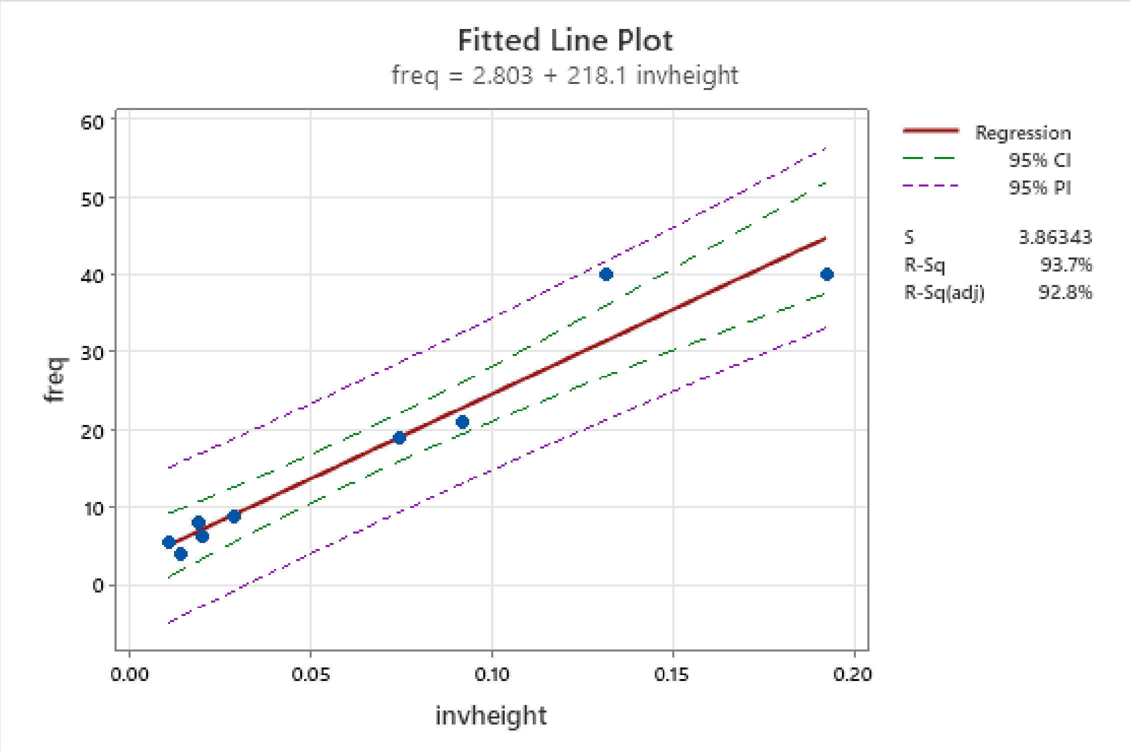
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	5153	5152.6	13.24	0.008
Error	7	2724	389.1		
Total	8	7877			

Fits and Diagnostics for Unusual Observations

Obs	height	Fit	Resid	Std Resid	
1	94.00	57.87	36.13	2.04	R

R Large residual



WATERFALL

Regression Analysis: freq versus invheight

Regression Equation

freq = 2.80 + 218.1 invheight

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	2.80	1.89	1.48	0.182
invheight	218.1	21.4	10.17	0.000

Model Summary

S	R-sq	R-sq(adj)
3.86343	93.66%	92.76%

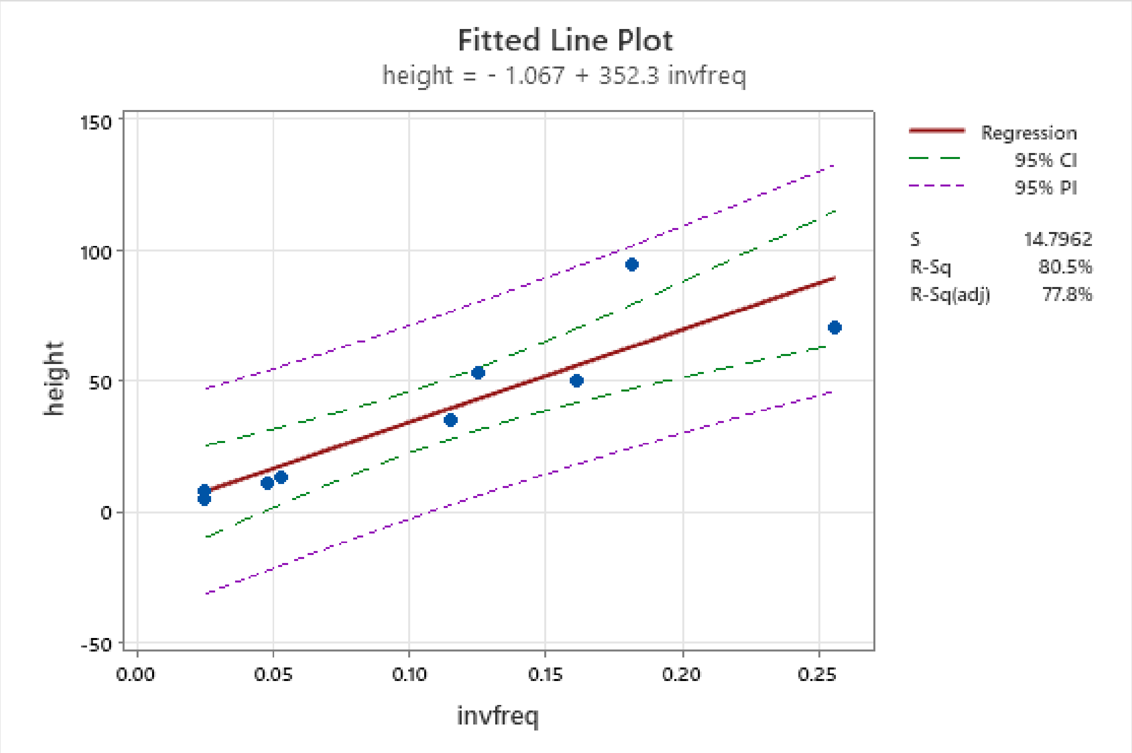
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1543.9	1543.85	103.43	0.000
Error	7	104.5	14.93		
Total	8	1648.3			

Fits and Diagnostics for Unusual Observations

Obs	freq	Fit	Resid	Std Resid	
8	40.00	31.50	8.50	2.54	R

R Large residual



WATERFALL

Regression Analysis: height versus invfreq

Regression Equation

height = -1.07 + 352.3 invfreq

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant	-1.07	8.73	-0.12	0.906
invfreq	352.3	65.5	5.38	0.001

Model Summary

S	R-sq	R-sq(adj)
14.7962	80.54%	77.76%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	6344	6344.1	28.98	0.001
Error	7	1532	218.9		
Total	8	7877			

Fits and Diagnostics for Unusual Observations

Obs	height	Fit	Resid	Std Resid
1	94.00	62.99	31.01	2.36 R

R Large residual