Cover Page

Examination paper for ST2304 Statistical Modelling for Biologists/Biotechnologists

Date: May 5th 2026
Time: 15:00-19:00

Course contact: Bob O'Hara
Present at the exam location: NO

Permitted examination support material: C. An A4 sheet of paper, with material hand-written or printed on both sides. A specific basic calculator is allowed, and mathematical tables (Tabeller og formler i statistikk (Tapir forlag, Fagbokforlaget), Matematiske formelsamling (K. Rottmann)

OTHER INFORMATION

Read the questions carefully and make your own assumptions. In your answers, explain clearly what assumptions you have made and how you have understood or limited the assignment

If there are direct errors or omissions in the assignment set and you cannot make your own assumptions, please refer to the information about complaints regarding formal errors on the NTNU website “Explanation of grades and appeals”.

SPECIFIC INFORMATION FOR YOUR COURSE

The questions are in English but can be answered in English or Norwegian.

No paper drawings:
This exam does not include hand drawings. If you receive hand drawing sheets, this is by mistake. **You will not be able to submit the sheets, and they will not be graded.**

Weighting: The maximum number of marks is given for each question.

Change the language of the question set:
If the question set is given in both English and Norwegian, you can change the language by going to the “hamburger menu” in the top hand-right corner and click “Language”

Withdrawing from the exam:
If you wish to submit a blank test/withdraw from the exam for another reason, go to the “hamburger menu” in the top right-hand corner and click “Submit blank”. This cannot be undone, even if the test is still open.

Access to your answers:
After the exam, you can find your answers under previous tests in Inspera. Be aware that it may take a working day until any paper drawings material is available in “previous tests”.

1 Confidence Interval Definition

Which of these statements about a 95% confidence interval is correct?

Select one alternative:

- If we repeated the data collection, there is a 95% probability that the new estimate would be true
- I am 95% certain to get this answer wrong
- If we repeated the data collection, there is a 95% probability of getting an estimate of the parameter within the limits
- If we repeated the data collection, there is a 95% probability that the true value of the parameter will be within the new confidence interval.

Maximum marks: 1

2 Linear model Assumptions

Which of the following is **NOT** a standard assumption of a linear regression model?

Select one or more alternatives:

- [ ] The independent variables are normally distributed
- [ ] The residuals have constant variance (homoscedasticity).
- [ ] The residuals are zero-mean normally distributed.
- [ ] The relationship between the independent and dependent variable is linear.
- [ ] The sample size must be greater than 100.

Maximum marks: 2

3 GLM Assumptions

What are two assumptions a GLM makes compared to a linear model that make a GLM more general?

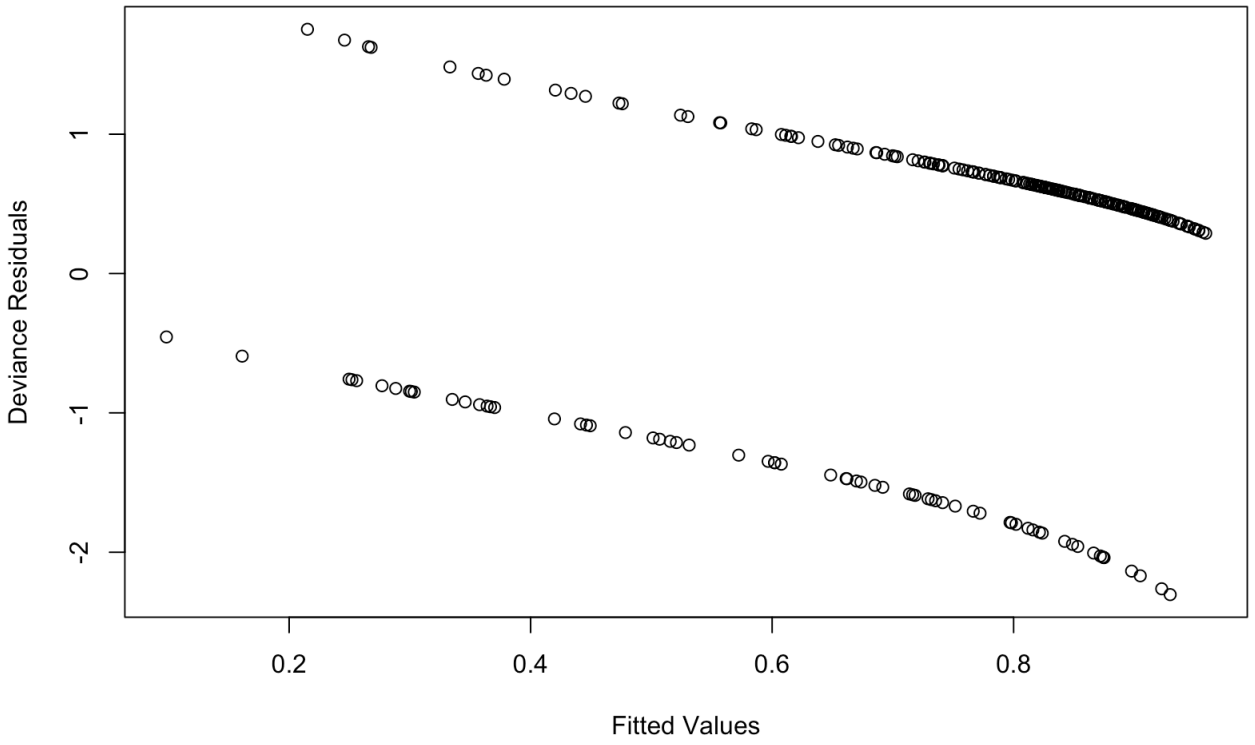
Select one or more alternatives:

- [ ] Constant variance
- [ ] The right link function
- [ ] Correct distribution
- [ ] No outliers
- [ ] Normality
- [ ] No multicollinearity
- [ ] Dependence of errors
- [ ] Just fewer assumptions

Maximum marks: 2

4 What is the model?

Here is a residual plot from a model fitted to some data. What distribution could the data have come from?



Select one alternative

- Poisson
- Something I haven't seen
- Zero truncated one inflated negative binomial distribution
- Binomial
- Normal

Maximum marks: 2

It is well known that trolls live under bridges. This creates obvious problems for traffic, so you have been tasked with tackling this problem. As a first step, you have decided to look at the sort of bridge trolls prefer, as this means you (or - preferably - someone bigger) can concentrate on clearing these bridges. Your team (all of whom are bigger than you) have surveyed 263 bridges in the region, looking at the following variables:

- the size of the bridge (length, in metres)
- what the bridge is made of (straw, wood, or stone)
- the landscape around the bridge: whether it is forest, fields or a town
- whether the bridge is regularly used by goats
- how old the bridge is (in years)
- whether there is a troll under the bridge (this will be your response)

Your aim is to say whether any particular bridge is likely to have a troll under it, so you can concentrate your troll eradication efforts on the bridges likely to be troll infested. The first thing you have to do is decide is which variables are important, and should be kept in your model.

The table below shows some model comparison statistics you could use.

Model Selection Statistics

Model	k	Deviance	AIC	BIC
1	1	306.8	308.8	312.3
Length	2	251.9	255.9	263.1
Habitat	3	305.9	311.9	322.6
Material	3	247.0	253.0	263.7
Material + Habitat	5	242.5	252.5	270.4
Material + Length	4	244.9	252.9	267.2
Habitat + Length	4	247.6	255.6	269.9
Material + Habitat + Length	6	240.3	252.3	273.7

5 What Type of Question?

What sort of model is this?

Select one alternative:

- Exploratory
- Predictive
- Explanatory
- Something Else

Maximum marks: 1

6 Model Comparison Statistic

For this problem, which is the best statistic to use to decide between the models?

Select one alternative:

- Deviance
- AIC
- Degrees of Freedom (k)
- BIC

Maximum marks: 1

7 Model Selection

Which is the best model?

Select one alternative:

- 1
- Length
- Habitat
- Material
- Material + Habitat
- Material + Length
- Habitat + Length
- Material + Habitat + Length

Maximum marks: 1

Because model selection is too complicated to explain to the goblins who are paying you, you decide to use the following model:

```
> mod <- glm(Troll ~ Material + Habitat + Length, data=Bridge, family = binomial())
> summary(mod)

Call:
glm(formula = Troll ~ Material + Habitat + Length, family = binomial(),
    data = Bridge)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.39694    1.42877   0.278  0.78115
MaterialStraw -2.10237    1.02941  -2.042  0.04112 *
MaterialWood  -1.40289    0.53594  -2.618  0.00885 **
HabitatForest  0.43146    0.34498   1.251  0.21106
HabitatTown   -0.71618    0.53976  -1.327  0.18456
Length         0.11128    0.07469   1.490  0.13626
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 306.77  on 262  degrees of freedom
Residual deviance: 240.27  on 257  degrees of freedom
AIC: 252.27

Number of Fisher Scoring iterations: 5
```

8 Links

Which link function was used in this binomial model?

Select one alternative:

- logit
- cloglog
- log
- identity
- probit

Maximum marks: 1

9 Best Bridges for Trolls?

What type of bridge are you most likely to find a troll under (everything else being equal?)

Select one alternative:

- Wood bridge in fields
- Stone bridge in a forest
- Stone bridge in a town
- Straw bridge in a town
- Wood bridge in a town
- Straw bridge in fields
- Straw bridge in a forest
- Wood bridge in a forest
- Stone bridge in fields

Maximum marks: 2

10 Choose a bridge. Find a troll.

On the basis of this analysis, what sorts of bridge should you focus your troll removing efforts on? Are there any additional factors you should think about beyond this analysis?

Write your answer in the box below. Changes are saved automatically.

Maximum marks: 5

Whilst returning from a day's bridge surveying, you are stopped by a Helpful Knight at a fork in the road. He tells you that you can go left or right. If you go left, you will reach a bridge in a forest that is that is 3 m long and made of wood. If you go right, you will reach a bridge in a forest that is that is 10 m long and made of stone.

You are obviously worried about meeting a troll (because you are not traveling with anyone bigger than you), so you need to decide which way to go.

11 Left? Or right?

Should you go left or right if you want to avoid a troll?

Select one alternative:

- Right
- Left
- Doesn't matter
- I have no idea but perhaps the Helpful Knight has taken ST2304?

Maximum marks: 1

12 Why Left? Or right?

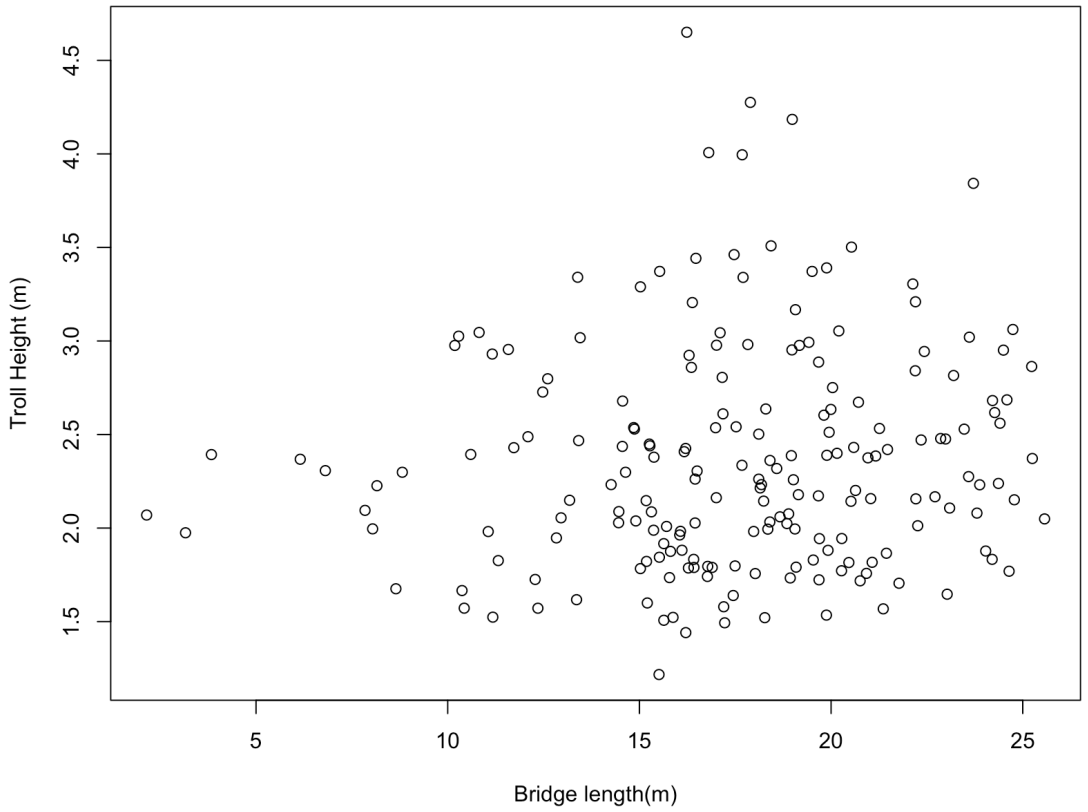
Explain you choice. Give any numbers you calculated to reach this decision

Write your answer in the box below. Changes are saved automatically.

Maximum marks: 4

Whilst you were surveying bridges, you and your team have measured the sizes of any trolls you have found: obviously you need more and bigger people to deal with bigger trolls.

In your conversation with the Helpful Knight, he suggested that bigger trolls might prefer larger bridges, so you decide to test this theory. Below is a plot of troll size against bridge size.


Troll Height (m)
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5
10
15
20
25
Bridge length(m)

13 Bigger bridge, bigger troll?

A model is fitted which gives the following ANOVA table.

```
> HeightMod <- lm(Height ~ Length, data=Troll)
> anova(HeightMod)
Analysis of Variance Table

Response: Height
           Df Sum Sq Mean Sq F value Pr(>F)
Length      1   0.79 0.78999  2.2112 0.1386
Residuals 194  69.31 0.35727
```

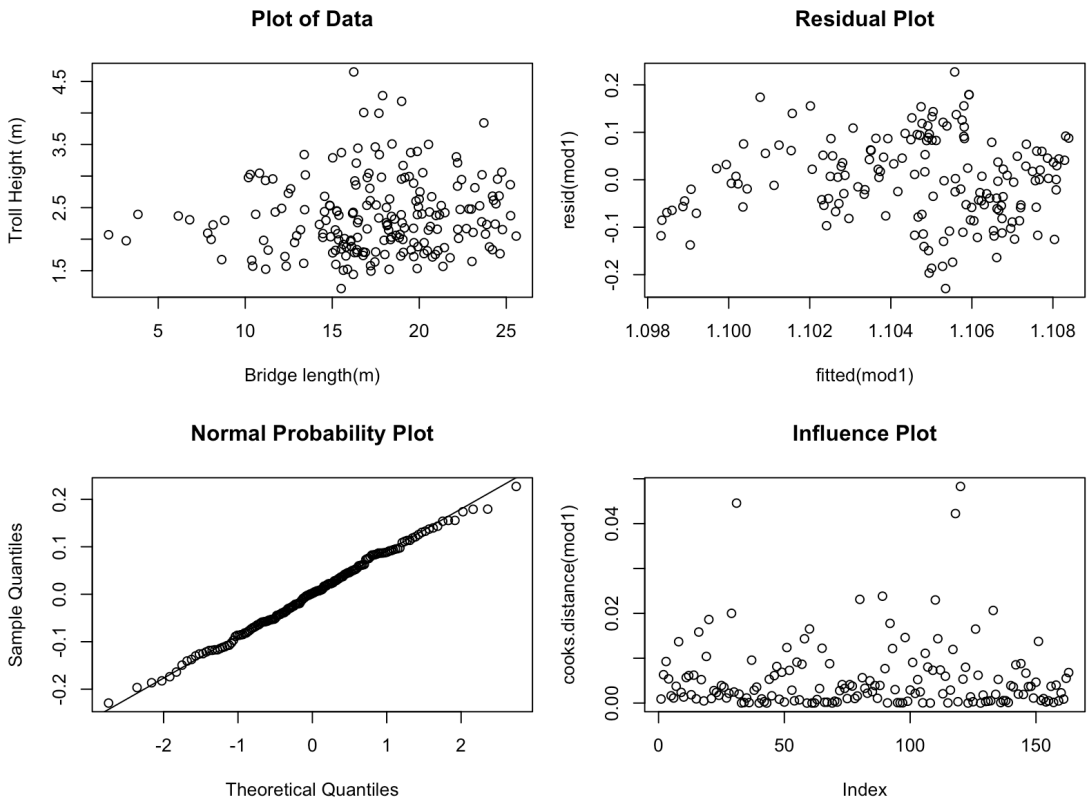
Would you conclude from this table that there is an effect of bridge length on troll height?

Select one alternative:

- No
- Really, it would be better if the Helpful Knight was here
- Not clear to me
- Yes, but I don't know what direction the effect is in
- Yes, longer bridges mean shorter trolls
- Yes, longer bridges mean taller trolls

Maximum marks: 1

You receive a scroll from the Helpful Knight suggesting you check how well the model fits the data. So you send back a scroll noting that you had already made the plots shown below.



14 But Which Assumptions Hold?

From these plots, which assumptions are you concerned about?

Select one or more alternatives:

- [ ] If the relationship is linear
- [ ] If there are some strangely influential points
- [ ] None - the model looks fine
- [ ] Are there are any outliers?
- [ ] Do the residuals look normally distributed?
- [ ] Heteroscedsticity: do the residuals have equal variance along the line?

Maximum marks: 2

15 What to Do?

The helpful knight responds with some wise comments. What does he suggest?

Select one alternative:

- You should remove some of the data
- You should add another term into the model
- You should transform troll height
- You should transform bridge length

Maximum marks: 1

The Helpful Knight gets back to you saying that the results look strange. He is convinced that, at least on stone bridges bigger trolls are on longer bridges. So you fit a model with that, and get the following summary:

```
> modH2 <- lm(Height ~ Length + Material, data=Bridge)
> summary(modH2)

Call:
lm(formula = Height ~ Length + Material, data = Bridge)

Residuals:
     Min       1Q   Median       3Q      Max
-1.06897 -0.39550 -0.07312  0.31039  2.03676

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.95458    0.23864   4.000 8.27e-05 ***
Length        0.06210    0.01163   5.342 2.02e-07 ***
MaterialStraw 0.91271    0.17599   5.186 4.34e-07 ***
MaterialWood  0.65094    0.09110   7.145 9.12e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5429 on 259 degrees of freedom
Multiple R-squared:  0.1692,    Adjusted R-squared:  0.1596
F-statistic: 17.58 on 3 and 259 DF,  p-value: 2.027e-10
```

16 R squared

What percentage of the variation is explained by the model? Answer to the nearest integer.

[ ].

Maximum marks: 1

17 Changing Views

How does this new model change your view on whether bridge size affects troll size?

Write your answer in the box below. Changes are saved automatically.

Maximum marks: 4

18 Why?

Why does the effect of bridge size change? What could you with the data do to investigate the change?

(hint: think about how bridge type might affect bridge size. Also, this might require some thought)

Write your answer in the box below. Changes are saved automatically.

Maximum marks: 6

19 Recommendations

What would you recommend as a strategy of troll control? In particular

- what bridges would you concentrate on when looking for trolls?
- which bridges would you send your biggest co-workers to?

Write your answer in the box below. Changes are saved automatically.

Maximum marks: 6