Cover Page

Department of mathematical Sciences

Examination paper for ST2304 Statistical modelling for biologists and biotechologists

Examination date: 6th August 2020

Examination time (from-to): 09:00 – 13:00

Permitted examination support material: All support material is allowed

Academic contact during examination: Bob O'Hara
Phone: 915 54 416

Technical support during examination: Orakel support services
Phone: 73 59 16 00

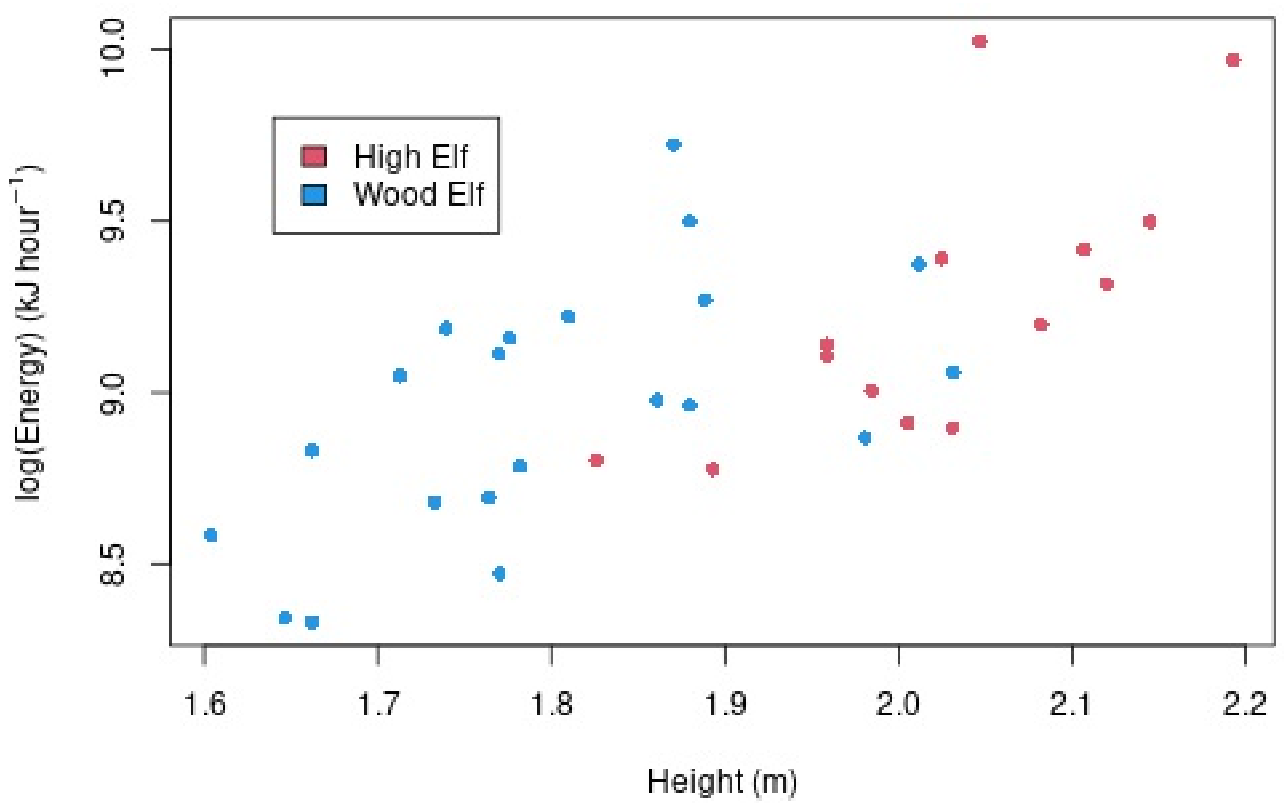
OTHER INFORMATION

- If a question is unclear/vague – make your own assumptions and specify in your answer the premises you have made. Only contact academic contact in case of errors or insufficiencies in the question set.
- **Saving:** Answers written in Inspera are automatically saved every 15 seconds. If you are working in another program remember to save your answer regularly.
- **Cheating/Plagiarism:** The exam is an individual, independent work. Examination aids are permitted. All submitted answers will be subject to plagiarism control. *Read more about cheating and plagiarism here.*
- **Notifications:** If there is a need to send a message to the candidates during the exam (e.g. if there is an error in the question set), this will be done by sending a notification in Inspera. A dialogue box will appear. You can re-read the notification by clicking the bell icon in the top right-hand corner of the screen. All candidates will also receive an SMS to ensure that nobody misses out on important information. Please keep your phone available during the exam.
- **Weighting:** Weighting of the questions is given for each question.

ABOUT SUBMISSION

- **Your answer will be submitted automatically when the examination time expires and the test closes**, if you have answered at least one question. This will happen even if you do not click “Submit and return to dashboard” on the last page of the question set. You can reopen and edit your answer as long as the test is open. If no questions are answered by the time the examination time expires, your answer will not be submitted.
- **Withdrawing from the exam:** If you wish to submit a blank test/withdraw from the exam, go to the menu in the top right-hand corner and click “Submit blank”. This can not be undone, even if the test is still open.
- **Accessing your answer post-submission:** You will find your answer in Archive when the examination time has expired.

Elven Service



There are different races of elf, in particular High Elves and Wood Elves. Anthropologists have been studying their behaviours, and in particular how much energy they put in to helping their group, carrying out tasks such as guarding against dwarves or tending their trees.

The researchers collected data for 35 elves, noting their race, and measuring the amount of energy individuals used. They also measured their heights, which we will use later.
The data are all analysed with energy use log-transformed, using a natural log.

First, we can compare the two races, to test if they use the same amount of energy. This was done by calculating the maximum likelihood estimate of the difference in means, assuming the data were normally distributed, with the same variance for all observations.

1 Elf MLE definition

Which of these is the best description of a maximum likelihood estimate?
Select one alternative:

☒ The value of the parameter that makes the data most likely

○ The most likely value of the parameter

○ The estimate most likely to to be true

○ The estimate of the data that makes the parameter most likely

Maximum marks: 1

2 Elf t-test MLE

This is the likelihood curve for the difference in means between High and Wood elves.

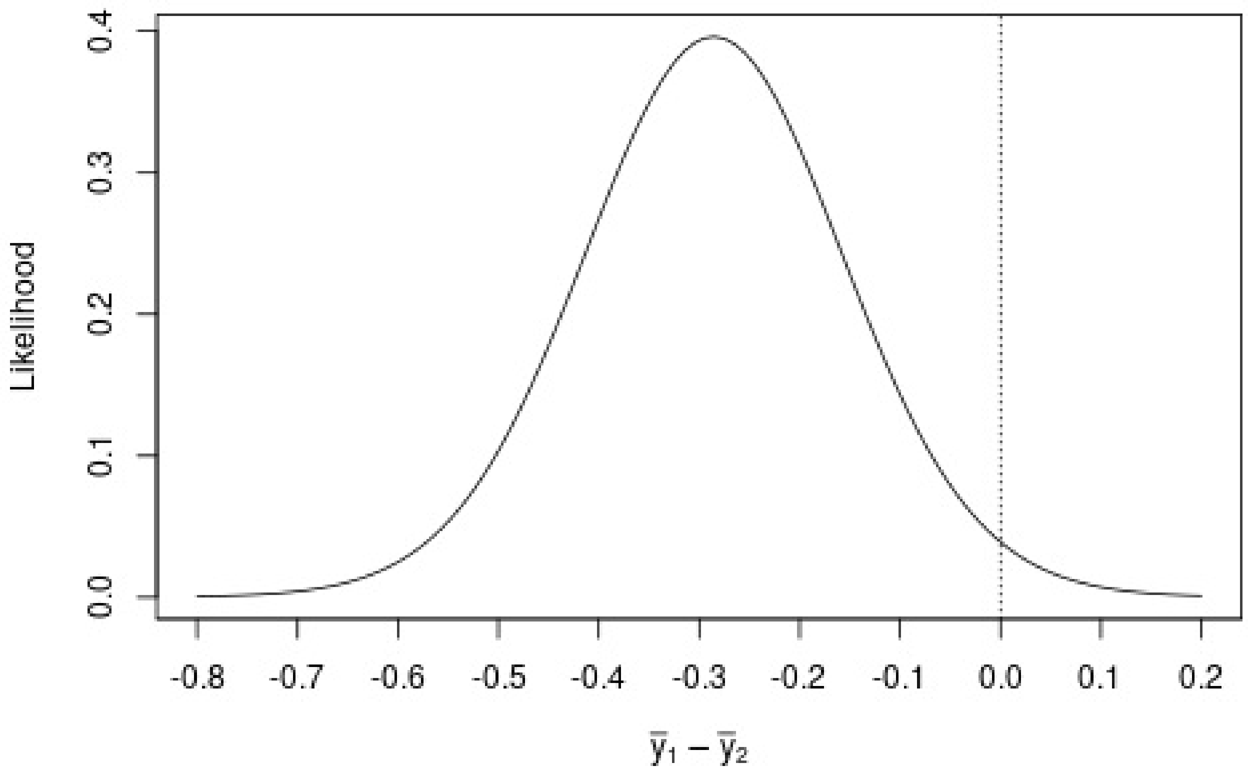


Figure 1: Likelihood for difference in log(energy use) between Wood ($\bar{y}_1$) and High ($\bar{y}_2$) elf races.

What is (approximately) the maximum likelihood estimate? □ (to no more than 2 decimal places).

0.29 (about!)

Maximum marks: 1

3 Elf t-test pvalue

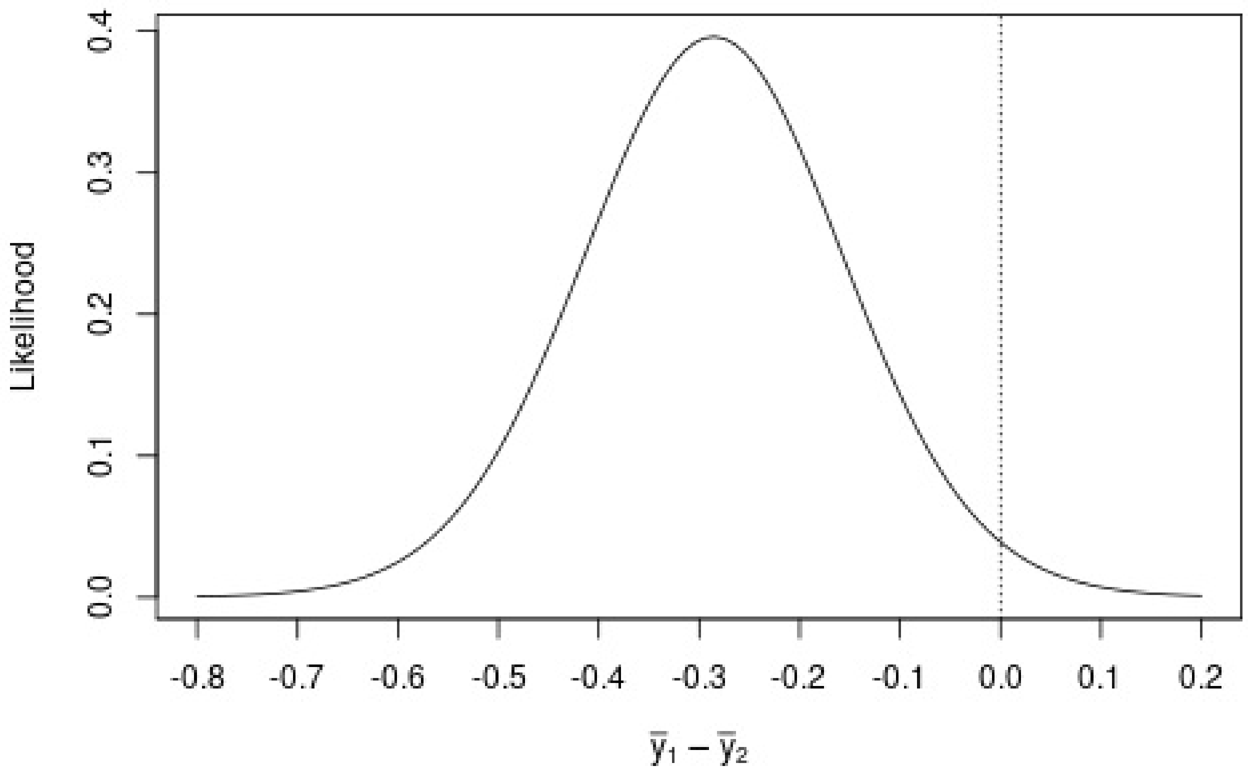


Figure 1: Likelihood for difference in log(energy use) between Wood ($\bar{y}_1$) and High ($\bar{y}_2$) elf races.

What is the probability that the difference would be great than 0?
Select one alternative:

X 0.035

X 0.04

(another bad question!)

○ 0.39

○ 0.18

Maximum marks: 1

4 Elf t-test Conclusion

From this analysis, would you conclude that there is a difference in energy use between the races?

Fill in your answer here

Yes, there does seem to be a difference, with wood elves using less energy than high elves

Maximum marks: 4

Elven Service

As noted, the anthropologists also measured the heights of the elves, because they expected that larger elves would use more energy (as they are also heavier), and this might obscure any relationship between races. The data are plotted in Figure 2, with energy use transformed with a natural log.

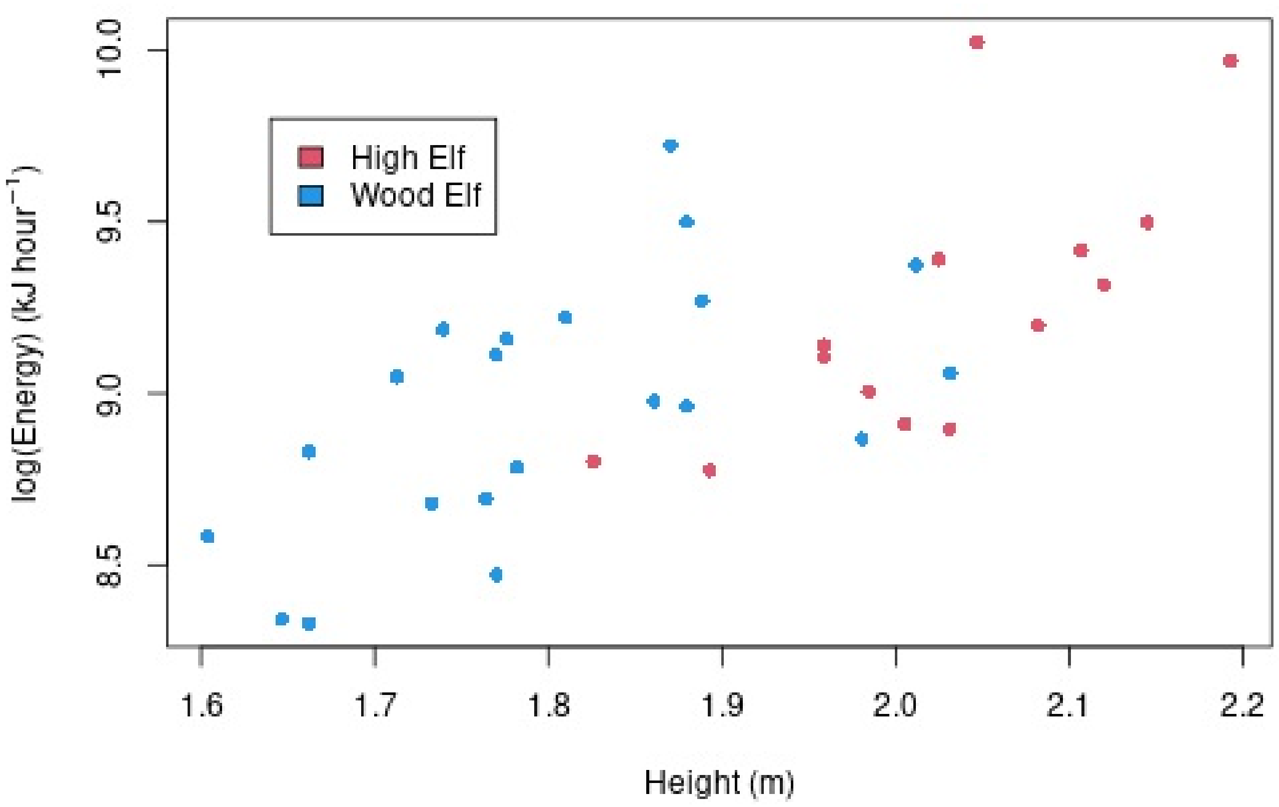


Figure 2: Energy use (on natural log scale), height and race of elves.

A linear model was fitted with Height and Race as explanatory variables. it gave the following summary:

```
Call:
lm(formula = energy ~ height + race, data = Data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.47673 -0.17578 -0.01317  0.19371  0.73388 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.09687    0.94230  -0.103    0.919    
height       0.89282    0.19303   4.625 5.89e-05 ***
raceHigh     0.19667    0.14611   1.346    0.188    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2966 on 32 degrees of freedom
Multiple R-squared:  0.477,	Adjusted R-squared:  0.4443 
F-statistic: 14.59 on 2 and 32 DF,  p-value: 3.132e-05
```

5 Elf ANCOVA Effect

What is the estimated effect of race, to 2 decimal places? **0.20** .

Maximum marks: 1

6 Elf Confidence Interval

What is the 95% confidence interval for the effect of race, to 2 decimal places?

Lower value: **-0.10** upper value: **0.49** **Calculations: 0.20 +/- 2.03*0.15**

(hint: the 2.5% quantile for a t-distribution with 33 degrees of freedom is -2.03)

Maximum marks: 4

7 Elf ANCOVA R2

How much of the variation is explained by the model (as a percentage, to the nearest whole number)? 0.48 .

Maximum marks: 1

8 Elf ANCOVA test type

An analysis of variance was conducted, and gave the following readout.

```
Analysis of Variance Table
Response: energy
          Df  Sum Sq Mean Sq F value    Pr(>F)    
height     1 2.40774 2.40774 27.3752 1.011e-05 ***
race       1 0.15937 0.15937  1.8119    0.1877    
Residuals 32 2.81450 0.08795                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Would you consider the test to be exploratory, confirmatory or something else?

Select one alternative:

✗ Confirmatory

○ Exploratory

○ Something else

Maximum marks: 1

9 Elf ANCOVA conclusion

An analysis of variance was conducted, and gave the following readout.

```
Analysis of Variance Table
Response: energy
          Df  Sum Sq Mean Sq F value    Pr(>F)    
height     1 2.40774 2.40774 27.3752 1.011e-05 ***
race       1 0.15937 0.15937  1.8119    0.1877    
Residuals 32 2.81450 0.08795                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Does it suggest a significant effect of race? How do you come to that conclusion?

Fill in your answer here

Format

This suggests that once height is controlled for there is no effect of race. Strictly, the p-value (0.19) suggests that it is not that unlikely to get an estimated race effect as large or larger than the observed estimate.

Words: 0

Maximum marks: 4

10 Elf ANCOVA conclusions

Based on the model of the data in Figure 2, and the ANOVA in Question 8, interpret the effect of race on log(energy) use.

Fill in your answer here

The analyses and figures suggest that there is no effect of race: the difference seen earlier is because high elves are taller than wood elves, and height is correlated with energy use.

Words: 0

Maximum marks: 4

A research project has been looking at flying speed in swallows. The following variables were measured:

- **Mass**: the bird's mass (g)
- **Sex**: Male or Female
- **HWI**: Hand Wing Index, which measures wing shape, and may be related to flight speed. Unitless.
- **Continent**: whether it is African or European

They want to see whether any of these variables explain the (unladen) flying speed, so they can then look at their ability to carry heavy weights (specifically coconuts).

Swallow Problem

Is this exploratory or confirmatory?

Select one alternative:

☒ Exploratory

○ Confirmatory

Maximum marks: 1

All combinations of models were fitted, and different statistics calculated to compare the models. The summaries are below: the models with R^2 > 10% are not presented.

Table 1: AIC, BIC and R^2 for models with R^2 > 10%.

Model	**AIC**	**BIC**	**R^2 (%)**
Mass	-1813.5	-1804.6	40.2
Mass + Sex	-1835.0	-1823.1	49.3
Mass + Continent	-1819.5	-1807.6	43.5
Mass + Sex + Continent	-1840.9	-1826.1	52.0
Mass + HWI	-1811.7	-1799.8	40.3
Mass + Sex + HWI	-1833.2	-1818.4	49.4
Mass + Continent + HWI	-1817.9	-1803.0	43.6
Mass + Sex + Continent + HWI	-1839.4	-1821.6	52.2

12 Swallow Which Statistic?

Which statistic is best to use to compare these models, if we want to predict the ~~population size~~ flying speed

Select one alternative:

○ BIC

○ R^2

✗ AIC

Maximum marks: 1

13 Swallow Which Model best

Which is the best model?

Select one alternative:

✗ Mass + Sex + Continent

○ Mass + Continent + HWI

○ Mass + Continent

○ Mass + HWI

○ Mass + Sex

○ Mass + Sex + Continent + HWI

○ Mass + Sex + HWI

○ Mass

Maximum marks: 2

14 Why Best Swallow Model

Why do you think this is the best model? Which statistics did you use to come to this conclusion and why?

Fill in your answer here

Format

This is the best model because it has the lowest value of AIC. An alternative might be Mass + Sex + Continent + HWI, which has and AIC that is 1.5 higher, but it is also more complicated.

Words: 0

Maximum marks: 4

The residuals for the best model are plotted below.

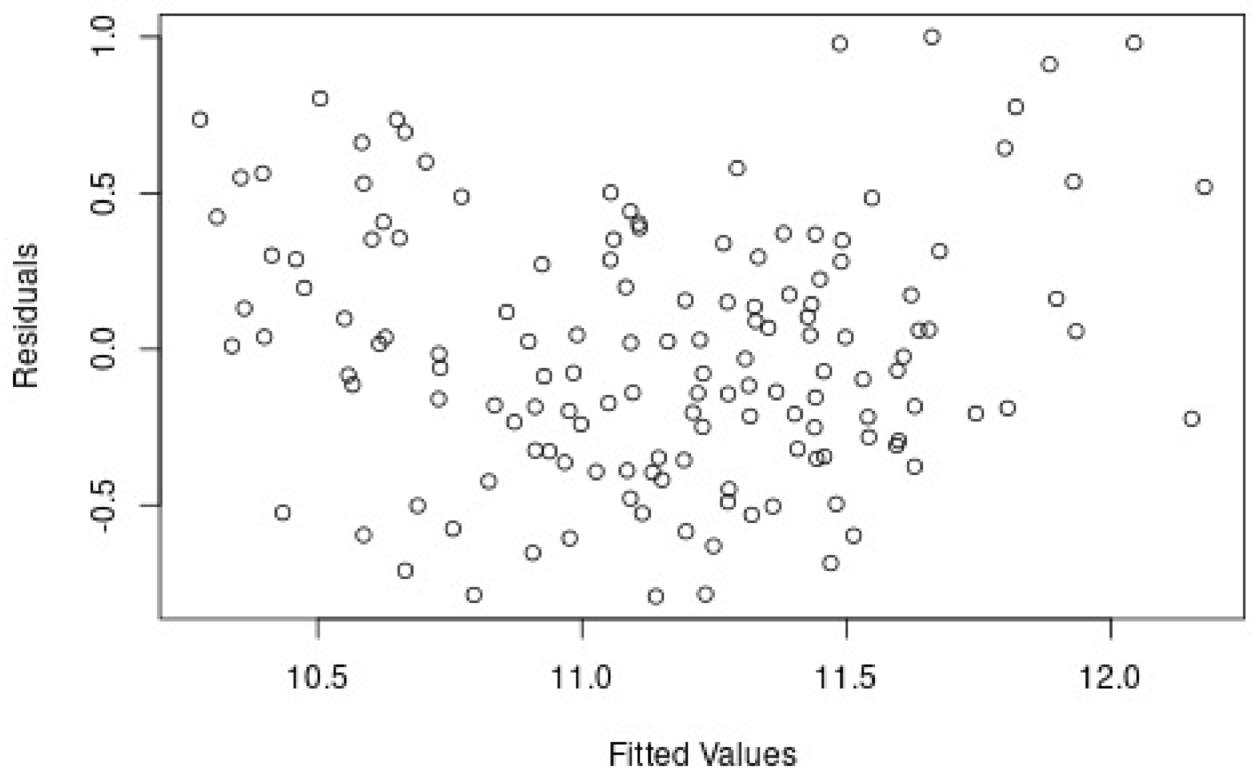


Figure 3: Residual Plot for model of flight speed of swallows

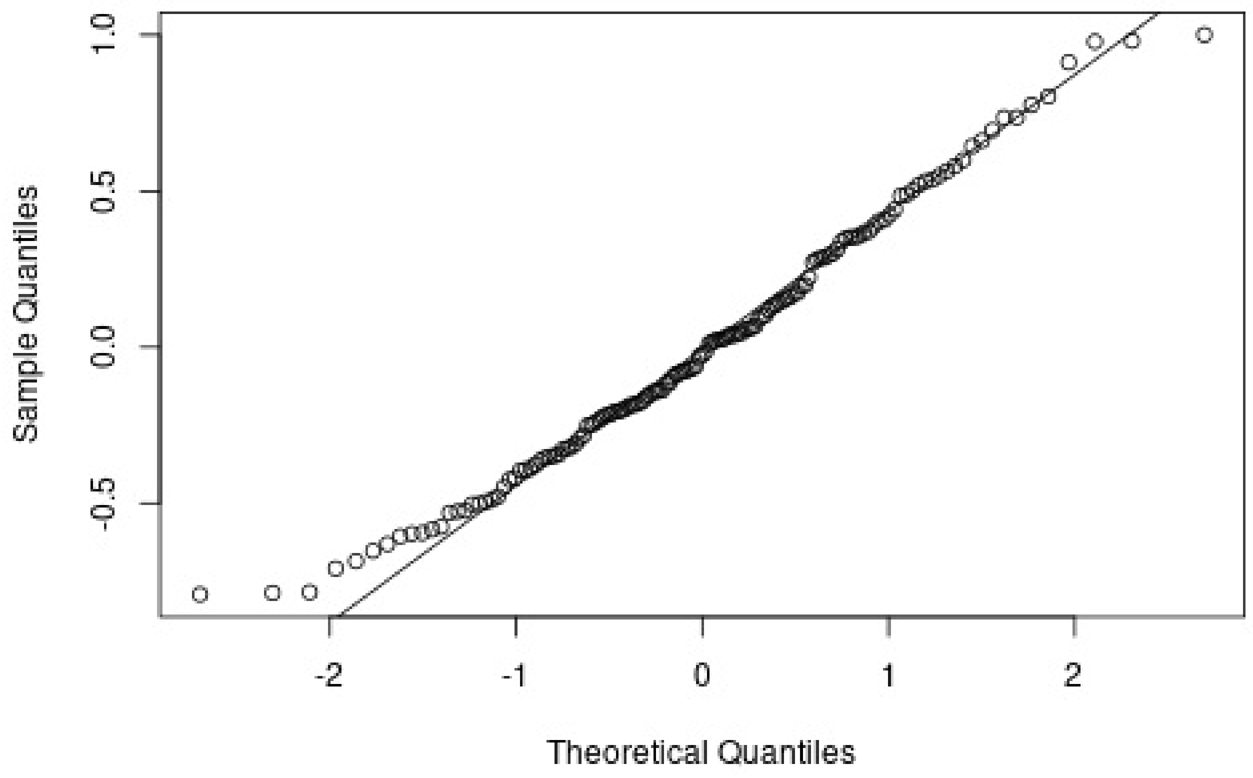


Figure 4: Normal probability plot for residuals from a model of flight speed of swallows

15 Swallow Residual Plot Assumptions Check

Which model assumptions can you use the plot in Figure 3 to check for?
Select one or more alternatives:

X The relationship is linear

X There are no outliers

□ Normally distributed error (residuals)

X Error has equal variance along line

Maximum marks: 2

16 Swallow Normal Probability Plot Assumptions Check

Which model assumptions can you use the plot in Figure 4 to check for?
Select one or more alternatives:

□ The relationship is linear

□ Error has equal variance along line

☒ There are no outliers

☒ Normally distributed error (residuals)

Maximum marks: 2

17 Swallow Residuals

Based on Figures 3 and 4, do you think the assumptions of this model are met? If not, why?

Format

Most of the assumptions seem to be met, but not the linearity assumption. The residual plot suggests that there is positive curvature. Aside from that, there do not seem to have any outliers, and the residuals seem to be normally distributed.

If a model was fitted with curvature, this might change.

Words: 0

Maximum marks: 4

18 Swallow Model Improvment

Based on your answer to question 17, how could you try to improve the model?

Fill in your answer here

Format

I would try adding quadratic terms for Mass and HWI: one or both might have a non-linear effect. The alternative would be to try a Box-Cox transformation, with a power <1.

Words: 0

Maximum marks: 4

19 Swallow Other Plots

What other plots could be used in addition to those in Fig 3 and Fig 4 to check if the model assumptions are met? What assumption would these plots check?

Fill in your answer here

Format

The obvious plots would be the residuals against HWI and Mass, to see if there is a non-linear effect of one or both.

We could also draw an influence plot, to look at whether any data points have large effects on the model.

Words: 0

Maximum marks: 4

The researchers were also interested in whether swallows were able to carry coconuts, and thus be a vector in their dispersal.

They observed the number of coconuts being carried by swallows during their migrations. They then fitted a model to this data assuming a Poisson distribution. They used mass (measured in grams), sex, and the continent the swallow came from as predictors. They got the following summary from the model:

(somehow I messed this up)

```
Call:
glm(formula = Coconuts ~ ~~log(~~Mass~~)~~ + Continent, family = poisson("log"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9755  -1.2303  -0.1411   0.5287   2.5724

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1.41332    0.79990  -1.767  0.07725 .
Mass               0.10279    0.04248   2.420  0.01552 *
ContinentEuropean -0.40279    0.15238  -2.643  0.00821 **
SexMale           -0.06592    0.15023  -0.439  0.66084
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Dispersion parameter for poisson family taken to be 1)

    Null deviance: 173.32  on 142  degrees of freedom
Residual deviance: 157.11  on 139  degrees of freedom
AIC: 409.1

Number of Fisher Scoring iterations: 5
```

20 Link Function

Which link function was used in this model for the number of coconuts being carried?

Select one alternative:

○ cloglog

○ logit

☒ log

○ probit

○ identity

Maximum marks: 1

21 Better Coconut Carrier

Based on this analysis, what type of swallow is able to carry more coconuts?

Fill in your answer here

Format

Large African swallows. African swallow can carry exp(0.4)= 1.5 times heavier coconuts than European swallows can. The effect of mass is so that an increase in 1g increases the size that can be carried by about 1.1 times (i.e. about 10%).

Words: 0

Maximum marks: 4

As part of a conservation programme, the UK is considering using swallows to help with introductions of coconuts. A conservation manager has to decide which swallows to use.

- Swallow 1: African, Male, weighs 15 g
- Swallow 2: Europena, Male, weighs 20 g

22

Coconut Prediction Swallow 1

Link scale: -1.41 + 0.10*15 + -0.40*0 - 0.07*1

What is the prediction on the link scale for swallow 1 from this model? **0.06** (answer to 2 decimal places).

What is the corresponding for the expected number of coconuts swallow 1 can carry? **1.06** (answer to 2 decimal places)

exp(0.06) = 1.06

Note: if you use the rounded numbers in the calculations, you get an answer of 0.03, and exp(0.03) = 1.03. This will still be marked right

Maximum marks: 4

23

Coconut Prediction, Swallow 2

Link scale: -1.41 + 0.10*20 + -0.40*1 - 0.07*1

What is the prediction on the link scale for swallow 2 from this model? **0.12** (answer to 2 decimal places)

What is the corresponding for the expected number of coconuts swallow 2 can carry? **1.13**

(answer to 2 decimal places)

Maximum marks: 4

24

Coconut Prediction Which Swallow

Which swallow would be a better choice to act as a vector of coconut dispersal, and why?
(to help you: the standard errors for the predictions are about 0.2)

Fill in your answer here

There is not a large difference: Swallow 2 can probably carry more coconuts, but the difference is about 0.1 coconuts, so it is not large and within one standard error. The best strategy might be to use Swallow 1 after it has put on more weight.

Maximum marks: 4