# ST2304 Exam 2020, marks scheme and comments on answers

Bob O'Hara 4/30/2021

Extra details from the run through session 11.05.2021

**All example answers are only a single example – many possible answers would get full marks.**

## Problem 1: Lazy Mole Rats

Question 1
1 mark. Correct answer: The value of the parameter that makes the data most likely

Question 2
1 mark. Correct answer: -0.10 (about! I allow for some error)

Question 3
1 mark. Correct answer: 0.249 (not a well-calibrated question!) To find this you need to approximate the area under the curve ABOVE 0. If you do this, the answer is actually between 0.312 and 0.249 so really, it would be hard to distinguish between these. You only need the graph to do this, it is approximate despite the options being to 3 decimal places.

Question 4
Up to 4 marks. Correct answer: No (2 marks) 2 marks for "insignificant p-value" e.g. No, the p-value calculated was greater than 0.05 (0.249 for one sided, 0.5 for two sided). This question asked about a TWO-SIDED question, it was is there ANY difference, not which direction it was in.

Question 5
1 mark. Correct answer: 0.39

Question 6
4 marks. Correct answer: 0.10 to 0.69 You calculate this by doing: estimate ± t * SE e.g. 0.39 ± (-2.03*0.14611)

Question 7
1 mark. Correct answer: 41%, make sure to round correctly

Question 8
1 mark. Correct answer: Confirmatory

Question 9
4 marks. Correct answer: 2 for "yes, there's an effect" 2 for quoting p-value and other statistics e.g. Yes. P-value is lower than 0.05, it is 0.0111984, so, there is less than 5% chance we would see an effect of caste this strong if the null hypothesis was true.

Question 10
2 marks. Correct answer: energy~mass*caste

Question 11
4 marks. Correct answer:1 for saying the effect is there + 1 for some judgement of size (at worst, it is statistically significant), bettergiving the effect, and some sort of context.2 marks for justifying their answer, e.g. by comparison to mass effect, or (possibly) with R^2
e.g. I would interpret that there is a statistically significant effect of caste on energy expenditure. The p-value for the effect of caste was <0.05 (0.0111). Figure 2 shows as well that workers have higher expenditure than the lazy caste (and the estimate from an lm for this was +0.39 for the worker caste). The confidence intervals for this effect did not cross 0 (0.1-0.67). There is also an affect of body mass, with lazy individuals being heavier and log mass increasing energy expenditure but not enough to remove the effect of caste. So, lazy caste individuals do use less energy than workers despite being larger in body mass (around 0.39 log kj/day less energy for the lazy caste).

# Greek Sheep

## Question 12
1 mark. Correct answer: Exploratory

## Question 13
1 mark. Correct answer: AIC

## Question 14
2 marks. Correct answer: Area + PropPasture + Gods

## Question 15
4 marks. Correct answer: 1 mark for "it has the lowest AIC." (or equivalent), 1 mark for citing it
2 marks for comparison to R2 (e.g. although model Y has a higher R2, it also has more parameters), and/orsaying how close it is to others (e.g. "model Z has and AIC that is only 1 higher, but is a bit simpler, so wecould consider that).
e.g. I chose this model because it had the lowest AIC (-612.2), which should mean it is the best for prediction. However, the next 'best' model (Area + PropPasture) had an AIC of <2 different (-610.7). So either is probably ok. But the R2 (57.9 vs 56.9) is higher for the model with Gods, so this is explaining more variance than the next 'best' model based on the AIC.

## Question 16
2 marks. Correct answers:Error has equal variance along line The relationship is linear There are no outliers

## Question 17
2 marks. Correct answer: Normally distributed error (residuals) There are no outliers

## Question 18
4 marks. Correct answer: Not great (2 marks):1. one outlier  - we couldn't actually find this! (2 marks)2. heteroscedasticity (2 marks)3. Linearity OK (2 marks)4. Might be one influential point? (2 marks)
e.g. Assumptions are not met fully. There is unequal variance in Figure 3, the variance increases as the fitted values increase. The assumption of normality of residuals is also not met. Figure 4 shows strong deviation from normality at the tails. I could not see any outliers, all extreme points seem part of the whole dataset. But, the relationship does look mostly linear.

## Question 19
4 marks. Correct answer:1. remove outlier? (2 marks)2. Box-Cox transformation (2 marks)3. Possibly remove largest point, if it's influential (2 marks) You need to match this bit to your answer in Qu18! E.g. You could try transforming the data. You could check the best transformation using Boxcox transformations. This would be to equalise the variance and improve normality. You could look at removing points with high influence or collecting more data.

## Question 20
4 marks. Correct answer: 1. 2 marks: plot against predictors, to check linearity of response to that predictor(&outliers?) 1. 2 marks: plot Cook's D to check if there are influential points 1. 2 marks for other sensiblesuggestions

## Question 21
1 mark. Correct answer: cloglog

## Question 22
4 marks. Correct answer: 0.14 0.68

You can get to this in R or with a calculator. Need to predict using linear predictor equation
(alpha + beta1*x1 + beta2*x2) this gives the prediction on the link scale. Then use the inverse of the cloglog link to get the prediction as a probability:

$$1 - e^{-e^{PREDICTION\ ON\ LINK\ SCALE}}$$

R code for this is:

# a + Bx

-1.833 + (0.71*(log(5))) + (1.66*0.5) = 0.14

# cloglog (INVERSE = 1-exp(-exp()))

1-exp(-exp(0.1397009)) = 0.68

## Question 23
4 marks. Correct answer: 0.13 0.6789119


## Question 24
4 marks.Correct answer: 1 mark: island 2, because it has the smallest probability + 2 if "but the difference is small(<1%)" 1 mark: doesn't matter, + 2 because the difference is small (<1%) 1 marks: the largest island, onlyif +2 because the probabilities are about the same, but on a larger island there are more caves to hide in (orsomething equivalent!) Also, 1 mark for noting that the difference is smaller than the standard error, so anydifferences could just be noise (or some equivalent statement).Or, if a different answer is given, but it makes some sense and relies on the predictions, up to 4 markNote: One student suggested that they would prefer Island 1 because the probabilities were almost the same,and on Island 1 there was more pasture, so it would be easier to see cyclops from a distance, and be able tothrow sheep at them an run away. This was an excellent (if unexpected answer): they had assessed the dataand used it to justify their answer.3