UNIVERSITY OF TORONTO

Faculty of Arts and Science

August 2018 EXAMINATIONS

STA303H1S

Duration - 3 Hours

Aids Allowed: Non-programmable calculator

First Name:____

Last Name:____

Student Number:____

This exam booklet contains 32 pages, and 100 multiple choice questions. Write all answers on the attached scantron sheet, in pencil. Nothing written in the exam booklet will be marked. You must hand in your exam booklet. Aids permitted: non-programmable calculator.

Each question is worth 1 mark; the exam is out of 100 marks.

1. Suppose Y follows an exponential family distribution with density/mass function $f(y;\theta) = \exp\left(\frac{y\theta - b(\theta)}{\phi} - c(y;\phi)\right)$. Which is true regarding the relationship between the mean and variance of Y?
   a. $E(Y) = Var(Y)$
   b. $Var(Y) = \phi \times E(Y)$
   c. $Var(Y) = dE(Y)/d\theta \times \phi$
   d. It is not possible to say in generality; different exponential family distributions have different mean/variance relationships
2. Suppose specifically that $Y \sim Poisson(\lambda)$, with mass function $P(Y = y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$. This is an exponential family- what is the canonical parameter?
   a. $\theta = \log \lambda$
   b. $\theta = \log y$
   c. $\theta = \exp \lambda$
   d. $\theta = \exp y$
3. If $Y \sim Binomial(n,p)$ instead, with mass function $P(Y = y; p) = \binom{n}{y} p^y (1-p)^{n-y}$, and we write this in exponential family form, what is the dispersion parameter ϕ?
   a. $\phi = 1$
   b. $\phi = \binom{n}{y}$
   c. $\phi = n$
   d. It is not possible to say; this depends on the choice of canonical parameter
4. Consider a random variable Y_i with mean $E(Y_i)$ that we wish to link to a linear predictor $\eta_i = x_i'\beta$, with $x_i, \beta \in \mathbb{R}^p$, using some link function $g(\cdot)$ such that $g\left(E(Y_i)\right) = \eta_i$. Which of the following represent the mathematical conditions necessary to impose on $g(\cdot)$ for it to be a valid link function in a generalized linear model?
   a. Monotone and differentiable
   b. Strictly monotone and differentiable
   c. Monotone and continuously differentiable
   d. Strictly monotone and continuously differentiable
5. In the above scenario, why bother linking the mean $E(Y)$ to the linear predictor η_i- why not just transform the response $Y \to g(Y)$? Which of the following is one good reason not to do this?
   a. Data transformations have to be invertible, and $g(\cdot)$ might be anything
   b. $g(\cdot)$ might not be defined for individual datapoints y_i
   c. Applying $g(\cdot)$ directly to Y violates the Box-Cox principle of data trasnformations
   d. $g(Y)$ will always have higher variance than Y itself, as applying a function to observed data results in a more complicated mean-variance relationship

6. What is meant by a "canonical" link function?
   a. The link function that gives the same transformation of the mean as would be obtained from a Box-Cox transformation
   b. The link function with the lowest variance out of all possible link functions
   c. The link function that is the canonical parameter in the exponential family-representation of the probability distribution of the data
   d. Any link function that is strictly monotone on the parameter space
7. We have data $y = (y_1, \ldots, y_n)$, which are realizations from a random variable Y with $E(Y) = \mu$. We wish to fit a Generalized Linear Model with linear predictor η_i, link function $g(\cdot)$, and corresponding variance function $V(\mu)$. The Iteratively ReWeighted Least Squares (IRWLS) algorithm is used to fit Generalized Linear Models. At iteration $t+1$ of this algorithm, the first step is to:
   a. Compute the transformed response $z^t = g(y_i)$
   b. Compute the transformed mean, $\bar{z}^t = g(\bar{y})$
   c. Compute the transformed response $z^t = g^{-1}(y_i)$
   d. Compute the transformed response, $z^t = \hat{\eta}^t + (y - \hat{\mu}^t) \times \frac{\partial \eta^t}{\partial \mu^t}|_{\hat{\eta}^t}$
8. The next step is to then:
   a. Compute weights $w^t = V(\hat{\mu}^t)$
   b. Compute weights $w^t = V(\hat{\mu}^t)^{-1}$
   c. Compute weights $w^t = \left(\left(\frac{\partial \eta^t}{\partial \mu^t}|_{\hat{\eta}^t}\right)^2\right)^{-1}$
   d. Compute weights $w^t = \left(\left(\frac{\partial \eta^t}{\partial \mu^t}|_{\hat{\eta}^t}\right)^2 V(\mu^t)\right)^{-1}$
9. We compute the updated estimate $\hat{\beta}^{t+1}$ by:
   a. Running a weighted linear regression on the above-computed weights and the transformed response, with $\hat{\beta}^{t+1} = \left(X'W^tX\right)^{-1} X'W^tz^t$
   b. Running a simple linear regression on the transformed response, with $\hat{\beta}^{t+1} = \left(X'X\right)^{-1} X'z^t$. The weights are used if we wish to estimate $Var(\hat{\beta}^{t+1})$.
   c. Running a weighted linear regression on the above-computed weights and the original response, with $\hat{\beta}^{t+1} = \left(X'W^tX\right)^{-1} X'W^ty$. The transformed response z^t is used to monitor the convergence of the algorithm
   d. Running a simple linear regression on the original response, with $\hat{\beta}^{t+1} = \left(X'X\right)^{-1} X'y$. The weights are used if we wish to estimate $Var(\hat{\beta}^{t+1})$, and the transformed response z^t is used to monitor the convergence of the algorithm.
10. In Generalized Linear Model terminology and using the notation above, a saturated model is one where
    a. There are as many estimated parameters as there are data points
    b. We include all the potential covariates x_i
    c. We include all the potential covariates x_i and all possible interactions
    d. The fitted model contains the true model as a special case, but potentially includes unnecessary covariates

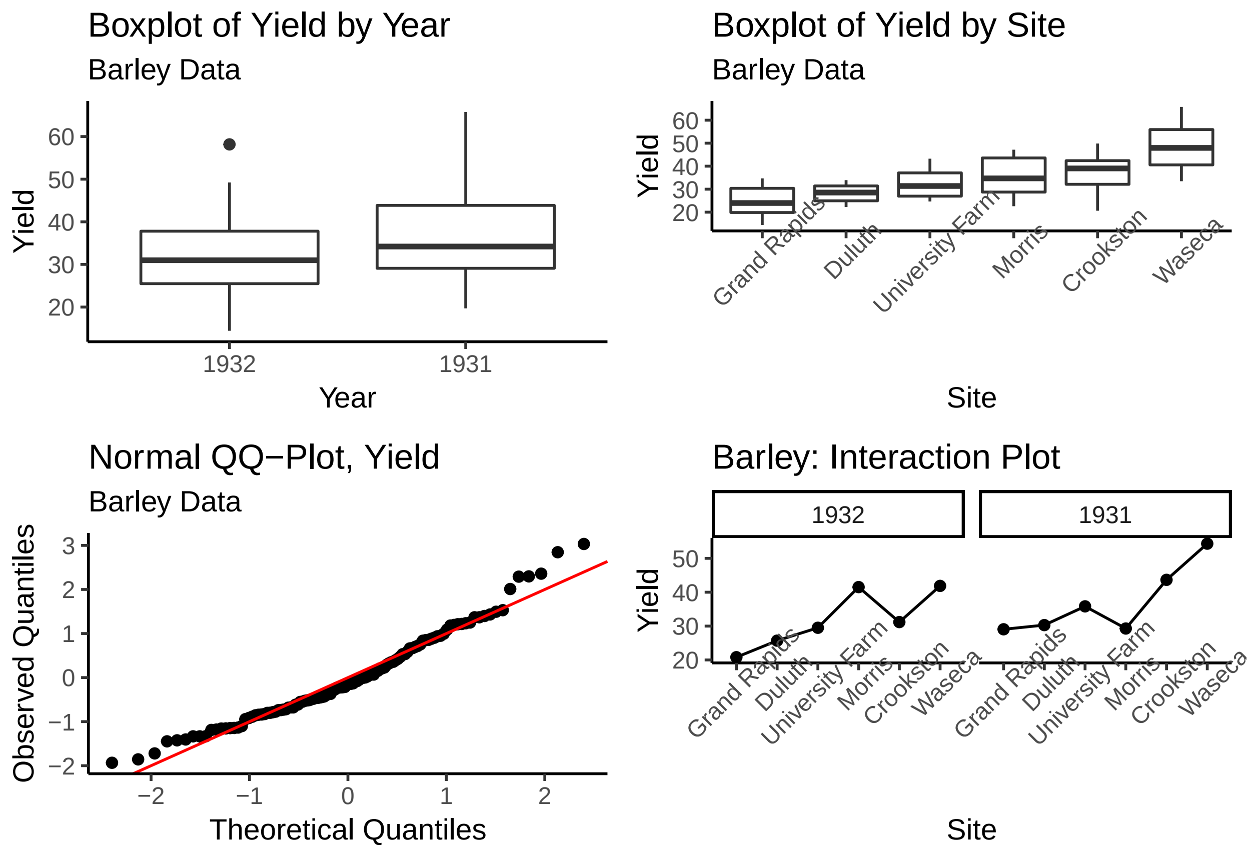
11. Which of the following best describes how well the saturated model fits the data?
    a. The saturated model always has perfect fit, meaning any measure of model fit will attain its best possible value on the saturated model, out of all models that could be fit to the observed data
    b. The saturated model fits the observed data as best as is possible for the measured covariates; if we had the true model, though, this would fit the observed data better
    c. We can't say how well the saturated model fits the data; it depends on the variance of the data, $\hat{\sigma}_y^2$. If the variance is too high then no model will fit well
    d. We can't say how well the saturated model fits the data; it depends on which covariates we measured. If we missed some covariates that are part of the true model, then the saturated model may fit the data worse than some other simpler model we choose
12. Fill in the blanks: the deviance of a model measures ___ and is computed as ___:
    a. How well the model fits the observed data; twice the log-likelihood of the fitted model
    b. How well the model fits the observed data; twice the difference in log-likelihood of the saturated model and the fitted model
    c. How well the model generalizes to new data; twice the log-likelihood of the fitted model
    d. How well the model generalizes to new data; twice the difference in log-likelihood of the saturated model and the fitted model
13. Which of the following best describes how we use the residual deviance of a single model? n is the number of datapoints and p is the number of estimated parameters. Assume n is "large".
    a. Compare it to a χ^2_n distribution; large values indicate poor fit
    b. Compare it to a χ^2_p distribution; large values indicate poor fit
    c. Compare it to a χ^2_{n-p} distribution; large values indicate poor fit
    d. Different GLMs may give deviances which have different, or unknown, asymptotic distributions, so it's not possible to interpret the deviance in general as a measure of model fit
14. Which of the following model assumptions does a plot of deviance residuals vs the linear predictor NOT test?
    a. Validity of the mean-variance relationship/correctness of the variance function
    b. Nonlinearity/misspecification of the structural part of the model
    c. The link function is correct
15. Which of the following aspects of the model and data can NOT be identified from a partial residual plot?
    a. Validity of the mean-variance relationship/correctness of the variance function
    b. Nonlinear relationships between individual covariates and the link-transformed mean response
    c. Outliers and/or influential points

16. Recall the single random effect model from lecture: $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$, $i = 1 \ldots K$, $j = 1 \ldots n$, with $\alpha_i \overset{iid}{\sim} N(0, \sigma_\alpha^2)$ and $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, independently. This could represent, for example, n repeated measurements on K experimental units. We try an ANOVA, with sum of squares decomposition $SST = SSA + SSE$, with associated mean squares $MSA = SSA/(K-1)$ and $MSE = SSE/(K(n-1))$. The associated variance estimates are $\hat{\sigma^2} = MSE$ and $\hat{\sigma}_\alpha^2 = \frac{MSA - MSE}{n}$. What is one problem with this?
    a. $\hat{\sigma}_\alpha^2$ is not unbiased
    b. $\hat{\sigma}_\alpha^2$ can be negative
    c. α_i is a random variable, which we don't observe, so we can't compute SSA and SSE
    d. α_i is a random variable, which we don't observe, so the usual sum of squares decomposition SST = SSA + SSE doesn't apply
17. Consider an experiment in which some continuous measurement is made on each of 30 rats for 6 consecutive weeks. This produces a dataframe with three columns: the continuous response variable `response`, an id for each `rat`, and a label for each `week`. In a regression model `response ~ rat + week`, `week` may most reasonably be considered to be a
    a. Fixed effect
    b. Random effect
    c. Mixed effect
    d. Ordinal effect
18. Continuing with the above question, `rat` may most reasonably be considered to be a
    a. Fixed effect
    b. Random effect
    c. Mixed effect
    d. Ordinal effect
19. Suppose we have a dataset containing customers' monthly credit card spending, measured in dollars, and some other potentially predictive variables. We wish to build a predictive model, such that we can predict any of these customers' spending in any given month of the year. We have multiple years of data, so multiple measurements on each of the 12 months, and we only want to predict spend for these customers, not other customers. We build a mixed effects model containing `customer` as a random effect. Which is the clearest and most reasonable interpretation for the random effect of `customer`?
    a. The amount by which a customer's monthly spend can be expected to fluctuate around that customer's average monthly spend, for a given customer in a given month
    b. Baseline spend for a given customer, holding all other variables in the model at zero
    c. Baseline spend for a given customer, holding all other variables in the model at their means
    d. The amount by which a customer's monthly spend can be expected to deviate from the average spend of a customer with the same values for all fixed-effect covariates in the model

20. In the above scenario, including **`customer`** as a random effect effectively models each customer's monthly spending as a linear function with a different intercept, and the same slope, for each customer. Suppose we wanted to let each customer have a different slope for **`month`**. How could we achieve this?
    a. We can't; we don't have enough data
    b. We could add in an interaction between the fixed effect of **`month`** and the random effect of **`customer`**
    c. We could add an interaction between the fixed effect of **`month`** and the effect of **`customer`**, however we would have to consider **`customer`** to be a fixed effect
    d. We could build a separate model of monthly spend for each **`customer`**
21. Let y_{ij} be the spend of the i^{th} customer in the j^{th} month, x_{ij} be the corresponding fixed-effect covariates, and b_i be the corresponding random effect, we may write the model $y_{ij} = x'_{ij}\beta + b_i + \epsilon_{ij}$, $b_i \overset{iid}{\sim} N(0, \sigma_b^2)$, $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, $b_i \perp \epsilon_{ij}$. What is the most appropriate interpretation of the ϵ_{ij}?
    a. The measurement error in recording customers' spends
    b. The random deviation of the spend of the i^{th} customer in the j^{th} month from the mean of customers with the same fixed effect covariates
    c. The random deviation of the spend of the i^{th} customer in the j^{th} month from the mean of the i^{th} customer's monthly spend
    d. The variability that is unexplained by the fixed effect term $x'_{ij}\beta$
22. What is $Cov(y_{ij}, y_{mk})$ for $i \neq m$, $j \neq k$?
    a. 0
    b. 1
    c. σ_b^2
    d. $\sigma^2 + \sigma_b^2$
23. What is $Cov(y_{ij}, y_{ik})$ for $j \neq k$?
    a. 0
    b. 1
    c. σ^2
    d. $\sigma^2 + \sigma_b^2$
24. In reference to the above question, we can let $y_i = (y_{i1}, \ldots, y_{in})'$ be the vector of monthly spend values for the i^{th} customer, $i = 1 \ldots K$, and write the model as $y_i = X_i\beta + Z_i b + \epsilon_i$. If $b = (b_1, \ldots, b_K)'$ is the vector of all the random effects terms b_i from the scalar form of the model above, what is Z_i?
    a. A vector of ones
    b. A matrix of ones
    c. A vector with a one in the i^{th} position
    d. A matrix with a column of n ones in the i^{th} column

25. We can further write $y = (y_1', \dots, y_K')' = (y_{11}, \dots, y_{1n}, \dots\dots, y_{Kn})'$ be the entire sample stacked together in one (column) vector and write $y = X\beta + Zb + \epsilon$. What is the purpose of the Z matrix?
    a. It is a "grouping" matrix- it picks out the random effect corresponding to the appropriate entry in y
    b. It is a "grouping" matrix- it picks out the random effect with the lowest variance
    c. It is a regularization term, which prevents the Restricted Maximum Likelihood Estimate from being degenerate
    d. It is a regularization term, which ensures that the Restricted Maximum Likelihood Estimate is unique
26. For a mixed effects model fit by maximum likelihood, what is one undesirable property of the resulting estimates of the variance parameters σ^2, σ_b^2?
    a. They are biased
    b. They are inefficient
    c. They are not unique
    d. They can be negative
27. We wish to compare using a likelihood ratio two mixed effects models with the same random effects terms, but different fixed effects terms. Can we do this if we fit the models using Restricted Maximum Likelihood (REML)- why, or why not?
    a. Yes, as long as the models have the same random effects, it doesn't matter how they were fit
    b. No, REML fits models that are not well approximated by a normal distribution, and hence the likelihood ratio is not asymptotically χ^2 distributed
    c. Yes, normally mixed effects models are biased so aren't comparable using a likelihood ratio, but REML corrects this bias
    d. No, the likelihood that REML optimizes depends on an estimate of the fixed effects parameters, which will be different for models with different fixed effects structures
28. We aren't sure whether adding a random effect into our model is a reasonable thing to do for a given dataset. Hence we're interested in comparing a model without a random effect to the same model with the random effect. Specifically, we want to compare $y_{ij} = x_{ij}'\beta + \epsilon_{ij}$ to $y_{ij} = x_{ij}'\beta + b_i + \epsilon_{ij}$, with $b_i \overset{iid}{\sim} N(0, \sigma_b^2)$ using a formal statistical procedure. What is the null hypothesis?
    a. $H_0 : b_i = 0$ for all i
    b. $H_0 : \sum_i b_i = 0$
    c. $H_0 : E(b_i) = 0$
    d. $H_0 : \sigma_b^2 = 0$

29. Why is this a difficult test to perform?
    a. The null hypothesis is on the boundary of the parameter space, so usual asymptotic likelihood theory does not apply
    b. The null hypothesis involves quantities which are unobservable
    c. The null hypothesis involves quantities which are random
    d. We require a lot of data for the asymptotic distribution of the test statistic to be valid
30. Suppose we fit a mixed effects model, and we wish to predict a datapoint y_{ij} for some i, j. Denote the fitted fixed effects parameter by $\hat{\beta}$ and the predicted random effects parameter by $\tilde{b}$. Suppose i is a group that was present in the dataset used to fit the model. What is the predicted value $\hat{y}_{ij}$?
    a. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + \tilde{b}_i$
    b. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + E_{\tilde{b}}(b_i)$
    c. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + E_{\tilde{b}|y}(b_i)$
    d. $\hat{y}_{ij} = E_{\tilde{b}|y}\left(x'_{ij}\hat{\beta} + b_i\right)$
31. Suppose i is a group that was not present in the dataset used to fit the model. What is the predicted value $\hat{y}_{ij}$?
    a. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + \tilde{b}_i$
    b. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + E_{\tilde{b}}(b_i)$
    c. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + E_{\tilde{b}|y}(b_i)$
    d. $\hat{y}_{ij} = E_{\tilde{b}|y}\left(x'_{ij}\hat{\beta} + b_i\right)$
32. For response **`y`**, fixed effect **`x`** and random effects **`g1`** and **`g2`**, which of the following **`lmer`** formulas will fit a model of the form $y_{ij} = x'_{ij}\beta + \epsilon_{ij}$?
    a. `y ~ x`
    b. `y ~ (1|x)`
    c. `y ~ x - g1`
    d. `y ~ x - g1 - g2`
33. As in the previous question, which of the following **`lmer`** formulas will fit a model of the form $y_{ij} = x'_{ij}\beta + g_{1i} + \epsilon_{ij}$?
    a. `y ~ x`
    b. `y ~ g1`
    c. `y ~ (1+x|g1)`
    d. `y ~ x + (1|g1)`
34. As in the previous question, which of the following **`lmer`** formulas will fit a model of the form $y_{ij} = x'_{ij}\beta + g_{1i} + g_{1i} \times x_{ij} + \epsilon_{ij}$?
    a. `y ~ x`
    b. `y ~ g1`
    c. `y ~ (1+x|g1)`
    d. `y ~ x + (1|g1)`



The following questions concern the now-familiar **`barley`** data from lecture and the midterm: **`yield`** of 10 varieties of Barley was measured across 6 sites in each of 2 years, with all combinations appearing once in the data. Here we average across **`variety`** and want to know: do the different **`sites`** have different barley **`yields`**, on average, and do any such differences change across **`year`**? **`yield`** is a continuous variable measured in bushels per acre, **`year`** and **`site`** are categorical.

35. What type of statistical technique might be appropriate for addressing these questions using these data?
    a. A t-test
    b. A simple linear regression
    c. An ANOVA
    d. A logistic regression

36. Consider the array of plots pertaining to the Barley dataset. Which plot would be used to visually assess whether **`yield`** could be modelled as being normally distributed?
    a. Boxplot of Yield by Year
    b. Boxplot of Yield by Site
    c. Normal QQ-Plot, Yield
    d. Barley: Interaction Plot
37. Consider the array of plots pertaining to the Barley dataset. Which plot would be used to visually assess a potential difference in means across values of **`year`**?
    a. Boxplot of Yield by Year
    b. Boxplot of Yield by Site
    c. Normal QQ-Plot, Yield
    d. Barley: Interaction Plot
38. Consider the array of plots pertaining to the Barley dataset. Which plot would be used to visually assess whether the shape of the distribution of **`yield`** is similar across values of **`site`**?
    a. Boxplot of Yield by Year
    b. Boxplot of Yield by Site
    c. Normal QQ-Plot, Yield
    d. Barley: Interaction Plot
39. Consider the array of plots pertaining to the Barley dataset. Which plot would be used to visually assess whether any **`site`** contained outlying values of **`yield`**?
    a. Boxplot of Yield by Year
    b. Boxplot of Yield by Site
    c. Normal QQ-Plot, Yield
    d. Barley: Interaction Plot
40. Consider the array of plots pertaining to the Barley dataset. Which plot would be used to visually assess whether any difference in mean **`yield`** between **`sites`** was changing over **`years`**?
    a. Boxplot of Yield by Year
    b. Boxplot of Yield by Site
    c. Normal QQ-Plot, Yield
    d. Barley: Interaction Plot

I fit a particular type of model to the data, obtaining the below table:

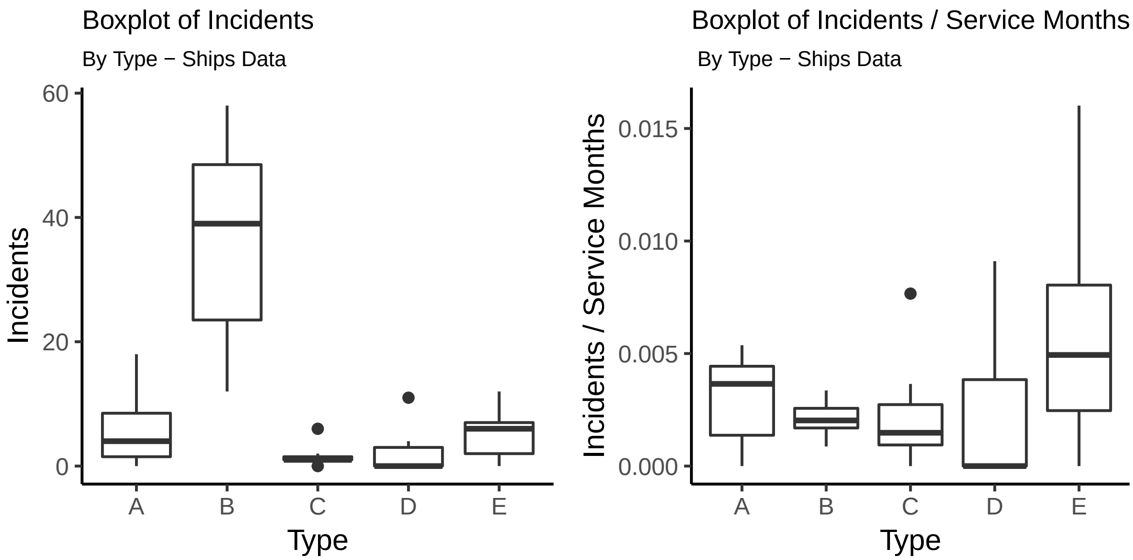
	Df	Sum of Squares	Mean Square	F	p
site	(1)	6634	(2)	(3)	0
year	(4)	847	(5)	(6)	0
Residuals	(7)	5229	(8)		

41. What `R` function did I use to obtain the above table? You can assume I passed the results of calling this function to `summary`:
   a. `aov`
   b. `lm`
   c. `anova`
   d. `anova.table`
42. What formula did I pass this function?
   a. `yield ~ site + year`
   b. `yield ~ site * year`
   c. `yield ~ site : year`
   d. `yield ~ year : site`
43. What value should go in the cell labelled (1)?
   a. 6
   b. 5
   c. 114
   d. 115
44. What value should go in the cell labelled (2)?
   a. 1,327
   b. 1,106
   c. 7.83
   d. 1.27
45. What value should go in the cell labelled (3)?
   a. 23.89
   b. 11.74
   c. 26.54
   d. 28.67
46. What value should go in the cell labelled (4)?
   a. 113
   b. 5
   c. 1
   d. 2
47. What value should go in the cell labelled (5)?
   a. 7.50
   b. 847.00
   c. 423.50
   d. 0.162

48. What value should go in the cell labelled (6)?
    a. 18.31
    b. 9.15
    c. 7.50
    d. 14.63
49. What value should go in the cell labelled (7)?
    a. 113
    b. 119
    c. 114
    d. 120
50. What value should go in the cell labelled (8)?
    a. 43.94
    b. 45.89
    c. 43.58
    d. 46.27
51. Suppose we want to add an interaction term between **`site`** and **`year`**. Which of the following **`R`** formulas would achieve this?
    a. **`yield ~ site + year`**
    b. **`yield ~ site * year`**
    c. **`yield ~ site : year`**
    d. **`yield ~ year : site`**
52. What would the sum of sqaures for **`site`** be in this new ANOVA table?
    a. It is not possible to say
    b. 6,634
    c. 847
    d. 3,740.5
53. What would the sum of sqaures for **`year`** be in this new ANOVA table?
    a. It is not possible to say
    b. 6,634
    c. 847
    d. 3,740.5
54. Would you expect the F-value for **`site`** in the ANOVA table with the interaction to be higher, lower, or the same as the F-value for **`site`** in the ANOVA table without the interaction?
    a. Higher
    b. Lower
    c. Same
    d. It is not possible to say

55. We fit this interaction model and obtain a p-value for the `site:year` interaction that is pretty much 0 (3.95×10^{-5}). What do we conclude, assuming we're satisfied with all model assumptions?
    a. There is evidence that the mean `yield` differs across `site`
    b. There is evidence that the difference in mean `yield` across `site` is different across `year`
    c. There is not evidence that the mean `yield` differs across `site`
    d. There is not evidence that the difference in mean `yield` across `site` is different across `year`

The `ships` data, familiar from the midterm, contains observations on 34 ships, their aggregate months of `service`, and the number of damage `incidents`, both integers. Also included are the `type` of ship (labelled `A` through `E`), the `year` of construction (a factor with levels `60`, `65`, `70`, `75`), and the `period` of operation, a factor with levels `60` and `75`. Can we bulid a model to predict `incidents` using the available variables?



56. Consider the boxplots of `incidents` and `incidents / service` by ship `type`. Which of the following conclusions from these plots is the most reasonable?
    a. `type == C` ships have the least damage incidents in the dataset
    b. `type == B` ships have normally distributed damage incidents
    c. `type == B` ships are in service for longer than other types in these data
    d. `type == E` ships have a higher variance of incidents than other types in these data
57. What is one good reason not to fit a normal linear regression to these data?
    a. The data represents counts, hence we know it is poisson-distributed and a normal linear model is not appropriate
    b. The variance is not equal across levels of `type`
    c. The data contains a mix of continuous and categorical variables, and a normal linear regression can only handle one type of variable at a time
    d. The normal approximation might not be reasonable since the data contains some very low counts

58. Suppose I really want to fit a normal linear regression to these data, so I try a log transformation of **incidents**. Is this a reasonable thing to do? Choose the most correct statement:
    a. Yes, **log** transformations are often useful when the data is strictly positive and/or exhibits right-skew
    b. No, **log** transformations are only warrented if the Box-Cox λ value is within one standard error of zero
    c. No, there are zero counts in the data, so **log** isn't well-defined
    d. It is not possible to judge this without seeing a scatterplot of **incidents**

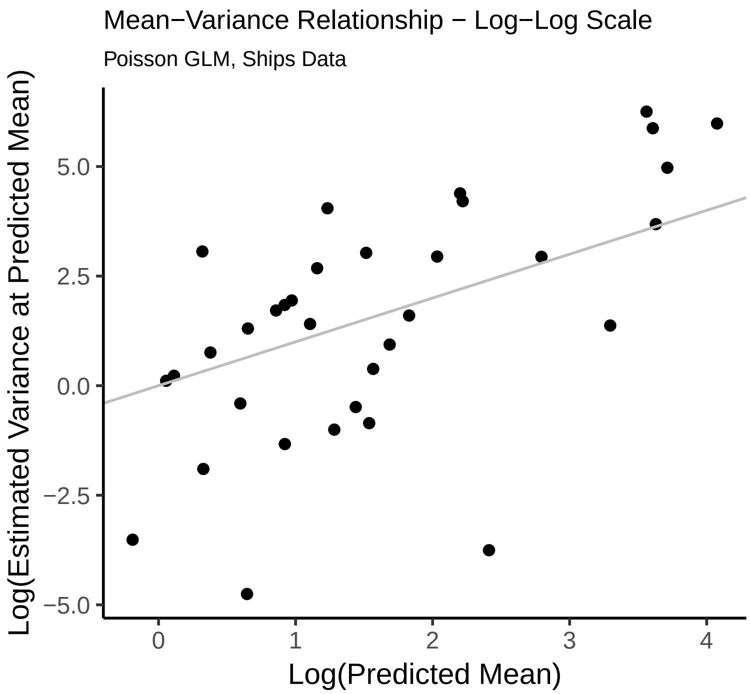
Consider the following model fit to these data:

```
## ======================SHIPS MODEL 1======================

##
## Call:
## glm(formula = incidents ~ ., family = poisson, data = ships)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.4598  -1.6558  -0.3323   0.6261   3.5104
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.387e+00  1.151e+00  -3.811 0.000138 ***
## typeB        9.622e-01  2.058e-01   4.675 2.94e-06 ***
## typeC       -1.212e+00  3.274e-01  -3.701 0.000215 ***
## typeD       -8.652e-01  2.875e-01  -3.009 0.002619 **
## typeE       -1.105e-01  2.350e-01  -0.470 0.638160
## year         5.280e-02  1.378e-02   3.831 0.000127 ***
## period       3.642e-02  9.245e-03   3.939 8.19e-05 ***
## service      4.785e-05  7.050e-06   6.787 1.15e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 614.54  on 33  degrees of freedom
## Residual deviance: 120.56  on 26  degrees of freedom
## AIC: 234.43
##
```

```
## Number of Fisher Scoring iterations: 5

## Estimated dispersion parameter:  4.385883
```



59. Denoting y_i as the observed counts, $\hat{y}_i$ as the predicted counts, $\hat{\eta}_i$ as the predicted linear predictor and $\hat{\mu}_i$ as the predicted mean, how was the dispersion parameter at the bottom of the output estimated?
    a. $\hat{\phi} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{y_i}$
    b. $\hat{\phi} = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}$
    c. $\hat{\phi} = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$
    d. Both B and C are correct
60. If we refit the model using $\hat{\phi}$ as the dispersion parameter, what would change?
    a. Estimated coefficients would change, and their standard errors would change
    b. Estimated coefficients would change, and their standard errors would not change
    c. Estimated coefficients would not change, and their standard errors would change
    d. Estimated coefficients would not change, and their standard errors would not change
61. What model assumption can be investigated using the plot entitled "Mean-Variance Relationship - Log-Log Scale"? The grey line is set to have slope = 1.
    a. $E(Y_i) = Var(Y_i)$
    b. $E(Y_i) > Var(Y_i)$
    c. $E(Y_i) = \mu$
    d. $Var(Y_i) = \sigma^2$

62. When you would be satisfied with the assumption?
    a. When the points fall exactly on the grey line
    b. When the points fall uniformly about the grey line and aren't too spread out
    c. When the points mostly fall above the grey line
    d. When the points mostly fall below the grey line
63. Holding other variables constant, what is the effect of changing ship type from **`A`** to **`B`** on the predicted mean number of incidents?
    a. Add 0.962
    b. Multiply by 0.962
    c. Add exp(0.962)
    d. Multiply by exp(0.962)
64. Based on deviance, does this model fit the data well?
    a. Yes, the residual deviance is less than the null deviance, indicating improved fit over the null model
    b. No, the residual deviance is much higher than its degrees of freedom, indicating poor fit
    c. No, the residual deviance is lower than the AIC, indicating poor fit
    d. It is not possible to say without referring to specific χ^2 values
65. The ship-building company is concerned that you haven't properly accounted for the obvious fact that ships that have been in service for longer are more likely to have higher counts. What is your response?
    a. We have accounted for this by including **`service`** in the model
    b. The poisson distribution assumption accounts for this because of the mean-variance relationship
    c. We can account for this by putting **`log(service)`** in the model as an offset and comparing the fit
    d. We can account for this by dividing **`incidents`** by **`service`**, like we did in the above boxplot

The **`smoking`** dataset from assignment 2 contains data on a survey of $1,314$ from Whickham, in the North of England, and measures their smoking habits, age, and mortality after a 20 year followup. This means that these women were contacted once, and their age and smoking status at contact was recorded; then 20 years later it was recorded whether they had died or not. Here is the data:

```
## =================Smoking: Raw Data=================

## # A tibble: 28 x 4
##    age   smoker death count
##    <fct>  <dbl> <chr> <dbl>
##  1 18-24      1 alive    53
##  2 18-24      0 alive    61
##  3 25-34      1 alive   121
##  4 25-34      0 alive   152
##  5 35-44      1 alive    95
##  6 35-44      0 alive   114
```

```
##  7 45-54      1 alive   103
##  8 45-54      0 alive    66
##  9 55-64      1 alive    64
## 10 55-64      0 alive    81
## # ... with 18 more rows
```

```
## =================Smoking: Summary Tables=================
```

```
## , , smoker = 0
##
##        death
## age     alive dead
##   18-24    61    1
##   25-34   152    5
##   35-44   114    7
##   45-54    66   12
##   55-64    81   40
##   65-74    28  101
##   75+       0   64
##
## , , smoker = 1
##
##        death
## age     alive dead
##   18-24    53    2
##   25-34   121    3
##   35-44    95   14
##   45-54   103   27
##   55-64    64   51
##   65-74     7   29
##   75+       0   13
```

```
## =================Smoking: Smoking and Mortality, Averaged over age=================
```

```
## # A tibble: 2 x 3
## # Groups:   smoker [2]
##   smoker alive  dead
##    <dbl> <dbl> <dbl>
## 1      0   502   230
## 2      1   443   139
```

```
## =================MODEL 1 OUTPUT=================

##
## Call:
## glm(formula = num ~ smoker + death, family = poisson, data = .)
##
## Deviance Residuals:
##      1       2       3       4
## -1.074   1.672   1.183  -1.962
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  6.26613    0.04078 153.644  < 2e-16 ***
## smoker      -0.22931    0.05554  -4.129 3.64e-05 ***
## deathdead   -0.94039    0.06139 -15.319  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 287.6348  on 3  degrees of freedom
## Residual deviance:   9.2003  on 1  degrees of freedom
## AIC: 45.239
##
## Number of Fisher Scoring iterations: 4
```

66. We discussed a number of sampling schemes for cross-classified data such as this. Suppose the data were collected by sampling 1,314 women, and then cross-classifying them by age, mortality and smoking status. Which would be an appropriate distribution to model the counts in the table?
    a. Poisson
    b. Multinomial
    c. Binomial
    d. Something else/not possible to say
67. Suppose we had put out a survey in the study area for a fixed length of time, received 1,314 responses, and then cross-classified them by age, mortality and smoking status. Which would be an appropriate distribution to model the counts in the table?
    a. Poisson
    b. Multinomial
    c. Binomial
    d. Something else/not possible to say

68. Suppose we sampled 582 smokers and 732 non-smokers, and then cross-classified them based on age and mortality. Which would be an appropriate distribution to model the counts in the table?
    a. Poisson
    b. Multinomial
    c. Binomial
    d. Something else/not possible to say
69. Ignore age for a moment and focus only on smoking status and mortality, and assume that the response of interest is mortality. What is the sampling scheme called in a study where we go out and sample a certain number of smokers and a certain number of non-smokers, and then measure their mortality (record whether they die during the study period)?
    a. Prospective sampling
    b. Retrospective sampling
    c. Stratified sampling
    d. Clustered sampling
70. What is the sampling scheme called in a study where we go out and sample a certain number of people who died and a certain number of people who remained alive during the study period, and then record whether or not they smoked during the study period?
    a. Prospective sampling
    b. Retrospective sampling
    c. Stratified sampling
    d. Clustered sampling
71. In the first type of study (where we sample smokers/non-smokers and record mortality), which of the following quantities can be estimated?
    a. Population proportion of smokers
    b. Population mortality rate
    c. Joint probability of smoking and mortality
    d. Odds-ratio of mortality for smokers vs non-smokers
72. In the second type of study (where we sample based on mortality and then record smoking status), which of the following quantities cannot be estimated?
    a. Population proportion of smokers
    b. Population mortality rate
    c. Joint probability of smoking and mortality
    d. Odds-ratio of mortality for smokers vs non-smokers
73. Which type of study was performed here?
    a. Prospective
    b. Retrospective
    c. Neither
    d. Both

74. Based on the output above, what is the mortality rate for smokers and non-smokers in this study?
    a. Smokers: 38%. Non-smokers: 62%.
    b. Smokers: 53%. Non-smokers: 62%.
    c. Smokers: 24%. Non-smokers: 31%.
    d. Not enough information given
75. Which of the following is the most reasonable explanation for the observation that smokers have a lower mortality rate than non-smokers?
    a. There are more smokers than non-smokers in the dataset, so the denominator of the mortality rate is larger and hence the rate is smaller
    b. There are less smokers than non-smokers in the dataset, so the mortality rate is more variable; hence it is not unlikely that we would see a statistical artifact such as this
    c. The smokers in the dataset are younger on average, and younger people have a lower mortality rate irrespective of whether or not they smoke
    d. An explanation isn't needed; intuition is sometimes wrong, and these data merely illustrate this
76. We wish to base inference on the mortality rate for smokers and non-smokers using a linear model of the form $\log \mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, where $i = 1, 2$ indicates smoking status (smoker/non-smoker) and $j = 1, 2$ indicates mortality (dead/alive). Is this model acceptable for this goal?
    a. No, the model has 4 parameters and the data has 4 data points, so it is saturated
    b. Yes, the log-link is the appropriate link for contingency table data
    c. No, the log-link is only acceptable for poisson-distributed data, and we don't know whether these data were collected according to a possion or a multinomial sampling scheme
    d. Yes, the model is set up in such a way that allows us to separate the effects of smoking and mortality in accordance with our research question
77. Supposing we proceed with this model, a hypothesis test of which term equalling zero would correspond to the scientific hypothesis of smoking and mortality being statistically independent?
    a. $H_0 : \mu = 0$
    b. $H_0 : \alpha_i = 0$
    c. $H_0 : \beta_j = 0$
    d. $H_0 : (\alpha\beta)_{ij} = 0$
78. Refer to the output entitled "MODEL 1 OUTPUT". What do you conclude about the relationship between smoking and mortality?
    a. The fitted model does not contain enough information to answer this question
    b. The p-values of both `smoker` and `deathdead` (representing the indicator for `death` = "dead") are very low, so there is evidence to suggest that these variables are not statistically independent
    c. The residual deviance is much higher than its degrees of freedom, hence we conclude that this model does not fit the data as well as the saturated model, and hence there is evidence to suggest that these variables are not statistically independent
    d. We cannot make a conclusion unless we also fit the model with the interaction term

79. Consider now the model with four terms: μ, α_i, β_j and γ_k, with the first three the same as before, and γ_k representing age group, $k = 1 \dots 7$. Which linear model below represents smoking and mortality being jointly independent of age?
    a. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$
    b. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$
    c. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
    d. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
80. Which linear model below represents smoking and mortality being independent, conditonal on age?
    a. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$
    b. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$
    c. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
    d. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
81. Which linear model below represents smoking and mortality and age being mutually independent?
    a. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$
    b. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$
    c. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk}$
    d. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
82. Which type of independence corresponds to the research question of whether smoking and mortality are related to age?
    a. Mutual
    b. Joint
    c. Conditional
    d. Marginal
83. Which type of independence corresponds to the research question of whether the dependence between smoking and mortality can be entirely attributed to these variables' individual dependence on age?
    a. Mutual
    b. Joint
    c. Conditional
    d. Marginal
84. If you wanted to test for mutual independence between all three variables, which model would be your base model for comparison?
    a. $\log \mu_{ijk} = \mu$
    b. $\log \mu_{ijk} = \mu + \alpha_i$
    c. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j$
    d. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$

85. Which of the following models would be the most appropriate to compare to this base model?
   a. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$
   b. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$
   c. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk}$
   d. $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$
86. Which pair of models would be most appropriate to compare if you wished to test whether the relationship between smoking status and mortality is independent of age?
   a. Model 1: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$ against Model 2: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$
   b. Model 1: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$ against Model 2: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk}$
   c. Model 1: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij}$ against Model 2: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik}$
   d. Model 1: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik}$ against Model 2: $\log \mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + (\alpha\gamma)_{ik}$

Recall from lecture the Rat Growth data: data on the weights of each of 30 rats was measured for 5 consecutive weeks. We are interested in modelling growth as a function of week, accounting for the fact that weight measurements on the same rat will be correlated. We model these data using a linear mixed effects model. Here is the data; `y` is the response (weight, units not specified), `rat` is an ID for which rat it is, and `week` is an ID for which week it is.

```
## # A tibble: 150 x 3
##    rat    week     y
##   * <fct> <int> <dbl>
##   1 1         0   151
##   2 1         1   199
##   3 1         2   246
##   4 1         3   283
##   5 1         4   320
##   6 2         0   145
##   7 2         1   199
##   8 2         2   249
##   9 2         3   293
##  10 2         4   354
## # ... with 140 more rows
```

```
## ===============RAT MODEL 1===============
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ 1 + (1 | rat) + week
##    Data: rat
##
```

```
## REML criterion at convergence: 1127.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7919 -0.4897  0.1287  0.5794  2.4702
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  rat      (Intercept) 191.86   13.851
##  Residual              64.29    8.018
## Number of obs: 150, groups:  rat, 30
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 156.0533     2.7715   56.31
## week         43.2667     0.4629   93.46
##
## Correlation of Fixed Effects:
##      (Intr)
## week -0.334

## ===============RAT MODEL 2===============

## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ 1 + (week | rat) + week
##    Data: rat
##
## REML criterion at convergence: 1084.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7275 -0.5670  0.1213  0.5562  2.3688
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  rat      (Intercept) 119.53   10.933
##           week         12.49    3.535   0.18
##  Residual              33.84    5.817
## Number of obs: 150, groups:  rat, 30
##
```

```
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 156.0533     2.1590   72.28
## week         43.2667     0.7275   59.47
##
## Correlation of Fixed Effects:
##      (Intr)
## week 0.007
```

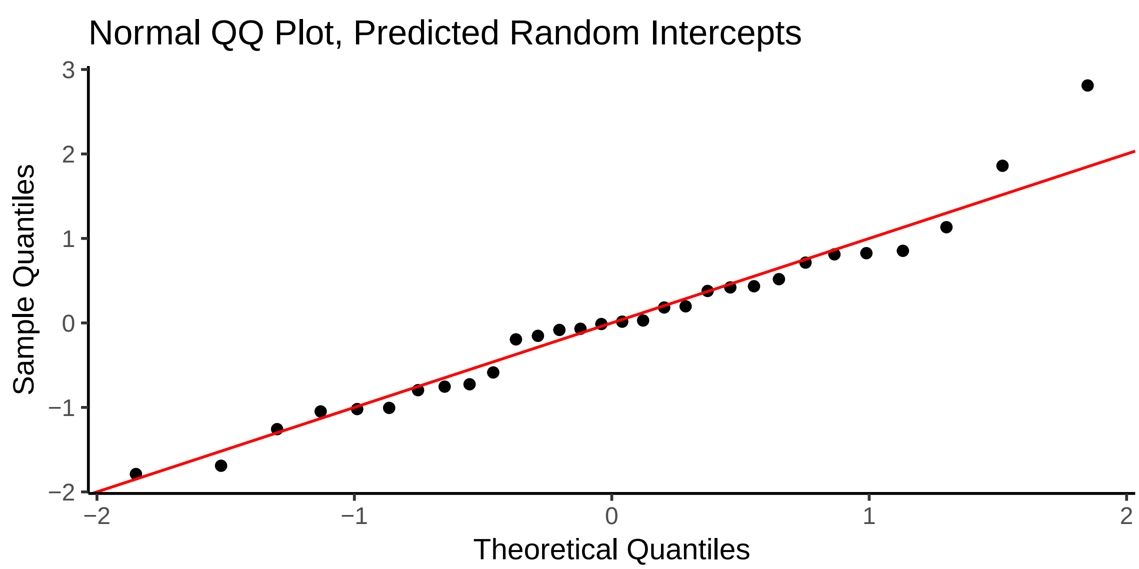
```
## ===============RAT MODEL 3===============
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: y ~ 1 + (1 | rat) + week
##    Data: rat
##
##      AIC      BIC   logLik deviance df.resid
##   1139.2   1151.2   -565.6   1131.2      146
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.8057 -0.4932  0.1276  0.5779  2.4817
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  rat      (Intercept) 185.14   13.607
##  Residual              63.76    7.985
## Number of obs: 150, groups:  rat, 30
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  156.053      2.729   57.19
## week          43.267      0.461   93.85
##
## Correlation of Fixed Effects:
##      (Intr)
## week -0.338
```

```
## ===============RAT MODEL 4===============
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: y ~ 1 + (week | rat) + week
```

```
##    Data: rat
##
##      AIC      BIC   logLik deviance df.resid
##   1101.1   1119.2   -544.6   1089.1      144
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7219 -0.5575  0.1208  0.5631  2.3722
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  rat      (Intercept) 114.87   10.718
##           week         11.97    3.459   0.19
##  Residual              33.84    5.817
## Number of obs: 150, groups:  rat, 30
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 156.0533     2.1227   73.52
## week         43.2667     0.7153   60.49
##
## Correlation of Fixed Effects:
##      (Intr)
## week 0.007
```



87. Which is/are good reason(s) for treating `rat` as a random effect, as opposed to a fixed effect with 30 levels?
    a. Treating it as a random effect lets us generalize our inference to new rats; treating as a fixed effect conditions on the rats actually used in this study, and does not generalize
    b. Treating it as a random effect allows us to separate the variability in weight that is due to `rat` from haphazard/unspecific variability
    c. Both A and B
    d. Neither A nor B
88. What is/are good reason(s) for treating `week` as a fixed effect, as opposed to a random effect?
    a. It is not reasonable to treat `week` as being a random sample from the population of possible weeks in which weight could be observed
    b. `week` represents a positive random variable and hence cannot be normally distributed
    c. Both A and B
    d. Neither A nor B
89. Consider the model "RAT MODEL 1". Which `R` function generated this model? You can assume the results were passed to the `summary()` function:
    a. `lmer(y ~ 1+(1|rat)+week,data = rat)`
    b. `lmer(y ~ 1+(week|rat)+week,data = rat)`
    c. `lmer(y ~ 1+(1|rat)+week,data = rat,REML = FALSE)`
    d. `lmer(y ~ 1+(week|rat)+week,data = rat,REML = FALSE)`
90. Consider the model "RAT MODEL 2". Which `R` function generated this model? You can assume the results were passed to the `summary()` function:
    a. `lmer(y ~ 1+(1|rat)+week,data = rat)`
    b. `lmer(y ~ 1+(week|rat)+week,data = rat)`
    c. `lmer(y ~ 1+(1|rat)+week,data = rat,REML = FALSE)`
    d. `lmer(y ~ 1+(week|rat)+week,data = rat,REML = FALSE)`
91. Consider the model "RAT MODEL 3". Which `R` function generated this model? You can assume the results were passed to the `summary()` function:
    a. `lmer(y ~ 1+(1|rat)+week,data = rat)`
    b. `lmer(y ~ 1+(week|rat)+week,data = rat)`
    c. `lmer(y ~ 1+(1|rat)+week,data = rat,REML = FALSE)`
    d. `lmer(y ~ 1+(week|rat)+week,data = rat,REML = FALSE)`
92. Consider the model "RAT MODEL 4". Which `R` function generated this model? You can assume the results were passed to the `summary()` function:
    a. `lmer(y ~ 1+(1|rat)+week,data = rat)`
    b. `lmer(y ~ 1+(week|rat)+week,data = rat)`
    c. `lmer(y ~ 1+(1|rat)+week,data = rat,REML = FALSE)`
    d. `lmer(y ~ 1+(week|rat)+week,data = rat,REML = FALSE)`

93. For the model "RAT MODEL 1", what is the proportion of total variance explained by `rat`?
    a. 75%
    b. 25%
    c. 63%
    d. 37%
94. Consider the models "RAT MODEL 1" and "RAT MODEL 2". The latter contains one more term than the former; hence we wish to compare them using a likelihood ratio test. Can we do this?
    a. Yes, the models are nested (model 1 is a special case of model 2). LRT is appropriate for nested model comparison
    b. Yes, the data are normally distributed so the LRT has an exact distribution under the null hypothesis of the simple model explaining the data just as well as the more complicated model
    c. No, random effects cannot be compared using an unadjusted LRT, as the null hypothesized value of the parameter lies on the boundary of the parameter space
    d. No, the models were fit with REML, and hence their likelihoods are uncomparable
95. What statistical hypothesis would correspond to the nested model comparison above? The more complicated model may be written $y_{ij} = x'_{ij}\beta + b_{i1} + b_{i2}week_j + \epsilon_{ij}$, where $x_{ij} = (1, week_j)$ is the vector of fixed effect covariates, $week_j$ is a vector of length 4 containing an indicator for the j^{th} week, and $b_{i1}, b_{i2} \perp \epsilon_{ij}$ for all i, j with (b_{i1}, b_{i2}) jointly normal with mean 0 and $Var(b_{i1}) = \sigma^2_{b1}$, $Var(b_{i2}) = \sigma^2_{b2}$, $Cov(b_{i1}, b_{i2}) = \rho\sigma^2_{b1}\sigma^2_{b2}$.
    a. $H_0 : \beta = 0$
    b. $H_0 : b_{i2} = 0$
    c. $H_0 : E(b_{i2}) = 0$
    d. $H_0 : \sigma^2_{b2} = 0$
96. Consider the models "RAT MODEL 1" and "RAT MODEL 3". Why are the estimated variances different?
    a. ML estimation is exact, where as REML uses numerical approximations, and hence the resulting estimates are less accurate
    b. ML estimation results in variance estimators that are downwardly biased
    c. REML estimation results in variance estimators that are upwardly biased
    d. Both REML and ML estimate only the variance ratio; the individual variances are not interpretable

97. Consider the models "RAT MODEL 3" and "RAT MODEL 4". The latter contains one more term than the former; hence we wish to compare them using a likelihood ratio test. Can we do this?
    a. Yes, the models are nested (model 1 is a special case of model 2). LRT is appropriate for nested model comparison
    b. Yes, the data are normally distributed so the LRT has an exact distribution under the null hypothesis of the simple model explaining the data just as well as the more complicated model
    c. No, random effects cannot be compared using an unadjusted LRT, as the null hypothesized value of the parameter lies on the boundary of the parameter space
    d. No, the models were fit with ML, and hence their likelihoods are uncomparable
98. Consider the plot titled "Normal QQ Plot, Predicted Random Intercepts", which is a normal QQ-plot of the predicted random intercepts from RAT MODEL 1. What model assumption is being tested here?
    a. Normality of the residuals
    b. Equal variance of the residuals
    c. Normality of the random effects
    d. Equal variance of the random effects
99. If we wanted to predict the weight in a given week after birth for a given **rat** from the 30 included in the dataset using RAT MODEL 1, what would be the prediction equation? $\tilde{b}_{i1}, \tilde{b}_{i2}$ are the predicted random effects for the i^{th} rat.
    a. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + \tilde{b}_{i1}$
    b. $\hat{y}_{ij} = x'_{ij}\hat{\beta}$
    c. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + \tilde{b}_{i1} + \tilde{b}_{i2}$
    d. $\hat{y}_{ij} = \tilde{b}_{i1}$
100. If we wanted to predict the weight in a given week after birth for a new **rat** that was not one of the 30 included in the dataset using RAT MODEL 1, what would be the prediction equation?
    a. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + \tilde{b}_{i1}$
    b. $\hat{y}_{ij} = x'_{ij}\hat{\beta}$
    c. $\hat{y}_{ij} = x'_{ij}\hat{\beta} + \tilde{b}_{i1} + \tilde{b}_{i2}$
    d. $\hat{y}_{ij} = \tilde{b}_{i1}$

THIS PAGE IS FOR ROUGH WORK. NOTHING ON THIS PAGE WILL BE MARKED

THIS PAGE IS FOR ROUGH WORK. NOTHING ON THIS PAGE WILL BE MARKED

THIS PAGE IS FOR ROUGH WORK. NOTHING ON THIS PAGE WILL BE MARKED

THIS PAGE IS FOR ROUGH WORK. NOTHING ON THIS PAGE WILL BE MARKED