

STOR 557: Fall 2023
Final Exam: December 9, 2023

Open book in-class exam: time limit 3 hours.

You are allowed to consult course notes and text (printed or e-read), homework assignments and any personal notes you have made during the course. The exam includes examples that you are expected to work in R-Studio or R-Markdown; personal computers or ipads are allowed in the exam and you are expected to use them. You are also allowed to use your computer for routine internet queries (e.g. looking up the syntax of an R command) but they must not be used to communicate with any person outside the exam. Answers may be prepared in any communicable format (e.g. R-Markdown, Word, Latex) but should be uploaded into gradescope before you leave the exam room. If for some reason you cannot access gradescope, you may email your solution to the instructor, but only during the time period allowed for the exam. You may consult the instructor if the wording is unclear, or if you think there might be an error, or if you encounter a technical error involving either R or gradescope, but the instructor will not give general advice about R, or hints how to solve the exam. The university Honor Code is in effect at all times.

To facilitate copying text from the exam questions (which is allowed), this entire exam has been placed online at

https://rls.sites.oasis.unc.edu/Final_Dec-09-2023.pdf

There are three questions. Question 1 is worth 30 points each and Questions 2 and 3 are worth 35 points each. Statistical tables will not be required.

1. The `carrots` dataset concerns samples of carrots exposed to an insecticide. The experiment consisted of 8 different doses of the insecticide conducted in 3 blocks (24 experiments in total). For each experiment, the numbers of undamaged (`success`) and damaged (`failure`) carrots were noted. The data may be downloaded using the following R code:

```
carrots=read.csv('https://rls.sites.oasis.unc.edu/carrots.csv')
```

- (a) Perform a binomial glm analysis with `logdose` as a numerical covariate and `block` as a factor covariate. Based on the deviance of the fitted model, would you say this model fits the data well? **[8 points]**
 - (b) Repeat the analysis of part (a) including an overdispersion parameter. Briefly summary how your conclusions differ from those in (a). **[4 points]**
 - (c) Now perform some diagnostics on the model in (a). One of the observations appears to be very different from the others. Which one, and what features lead you to conclude that this observation may be an outlier? **[10 points]**
 - (d) Repeat the analysis of part (a) removing the outlier. Does the model now fit the data? Briefly summarize your overall conclusions. **[8 points]**
2. The `oatvar` dataset concerns a randomized block design comparing the yields of 8 varieties of oats. You can download the data with `library(faraway)` followed by `data(oatvar)`. The growing area was divided into 5 blocks, each planted with a single plot of each variety.
 - (a) Plot the data and comment on any differences among either the varieties or the blocks. **[6 points]**

- (b) Fit the data using a standard (fixed effects) analysis of variance, treating both block and variety as factor variable, and assuming the data have normal responses. What do you conclude about the differences among both the block and variety effects? **[6 points]**
 - (c) Using *either* STAN or INLA (do not attempt to do both!), carry out a Bayesian analysis of the data, treating variety as a fixed effect and block as a random effect, and make plots of the SD densities of the error and the block. **[9 points]**
 - (d) Compute a numerical summary of the posteriors. Which varieties show a significant difference from the first variety? **[7 points]**
 - (e) Plot the densities of the variety effects, and summarize your conclusions about the differences among the 8 varieties. **[7 points]**
3. The `salamanders` dataset is based on an experiment conducted by a team of ecologists in the Appalachian mountains. The intent of the study was to understand the extent to which salamanders from different populations will interbreed. The data may be downloaded using the following R code:

```
salamanders=read.csv('https://rls.sites.oasis.unc.edu/salamanders.csv')
```

Description of the dataset. There were three experiments, conducted at different times with what we are assuming to be different sets of animals. The salamanders were drawn from two species, Rough Butt (labelled R in the descriptions that follow) and Whiteside (labelled W). The POP variable shows the population of both the female and male salamander used in a particular observation, for example W/R means that the female was a Whiteside salamander and the male was a Rough Butt salamander. The variables FEMALE and MALE indicate which specific female and male animal was used in a particular experiment. For example FR1, FR2 through FR10 indicate females numbered 1 through 10 from the Rough Butt population, with similar notations for MR1, . . . MR10 and FW1, . . . FW10 and MW1, . . . MW10. Note that the same notation is used for animals in all three experiments, but each label (for example, FR1) is referring to a *different* animal in each of the experiments 1, 2 and 3. Each of the three experiments had the same design, whereby each of the 20 female salamanders was paired 6 times with a different male salamander, but because of time limitations, the experiments did not attempt to pair all possible combinations of male and female salamanders. Finally, the response variable y was a binary response where y=1 means that mating occurred and y=0 means mating did not occur.

- (a) Briefly summarize the data, focusing on the proportion of successful matings among each of the four possible combinations of species (R/R, R/W, W/R, W/W). What do you conclude based on this? (There is no need for any formal testing at this stage.) **[5 points]**
- (b) Now conduct an analysis of variance in which R/R, R/W, W/R, W/W are considered treatments and EXPT is a block variable, but ignoring the details of which female salamander is paired with which male salamander (equivalently, we assume all 480 responses are independent). Are the differences among species combinations statistically significant? Summarize your conclusions (for example, say which species combinations have the highest and lowest probabilities of mating and to what extent you can be certain of your conclusions based on the standard errors of the differences). **[7 points]**

- (c) To take account of the fact that some individual animals are likely to mate than others, it makes sense to treat both **FEMALE** and **MALE** as random effects. In other words, assume there is a different random effect associated with each female animal and a different random effect associated with each male animal, and of course, the effects associated with Rough Butt and Whiteside salamanders are independent, as are the random effects for females and males in different experiments (because we are assuming the animals used are different for the three experiments). However, to simplify the analysis, we shall assume that there is only one standard deviation of the random effect for females and one (presumably different) standard deviation of the random effect for males. Use the Laplace method of approximation to fit this random effects model, with the species combinations (R/R, R/W, W/R, W/W) and EXPT treated as fixed effects. Summarize your conclusions. [**9 points**]
- (d) Repeat the analysis of part (c) with (i) the species combination effect omitted, (ii) the EXPT effect omitted. By comparing the fitted models with the one you fitted in part (c), comment on whether the the species combination effect and the EXPT effect are statistically significant. [**7 points**]
- (e) Using any plots or summary statistics you consider appropriate, perform diagnostics on the model in part (c). Is there anything unusual about the residuals? [**7 points**]

Sample Solutions

Note: There are no unique “right answers” to computational problems of this nature. The answers given here are intended as examples of responses that are adequate to answer the questions. Alternative answers will be accepted so long as they are appropriate and fully explained.

1. (a) Consider the following R code:

```
carrots=read.csv('https://rls.sites.oasis.unc.edu/carrots.csv')
m1=glm(cbind(success,failure)~logdose+block,family=binomial,carrots)
summary(m1)
```

The results include:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.0226	0.6501	3.111	0.00186	**
logdose	-1.8174	0.3439	-5.285	1.26e-07	***
blockB2	0.3009	0.1991	1.511	0.13073	
blockB3	-0.5424	0.2318	-2.340	0.01929	*

...

Residual deviance: 39.976 on 20 degrees of freedom

The result shows a very highly significant coefficient for `logdose` and likely a block effect as well (e.g. the p-value of `Be` is 0.019) but the residual deviance is too high for the degrees of freedom (`pchisq(39.976,20,lower.tail=F)` gives the value 0.005). This suggests the model does not fit the data.

- (b) The corresponding code with `family=quasibinomial` yields an overdispersion parameter of about 2.06 and correspondingly increased standard errors:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.0226	0.9328	2.168	0.04237	*
logdose	-1.8174	0.4934	-3.683	0.00147	**
blockB2	0.3009	0.2857	1.053	0.30479	
blockB3	-0.5424	0.3326	-1.631	0.11858	

It still looks as though `logdose` is statistically significant but probably the block effect is now. However, this analysis doesn't resolve the issue of whether the overdispersion model fits the data.

- (c) `plot(m1)` leads to the plots in Figure 1. Observation 14 has a residual almost twice as large as any other, and it also has a large Cook's distance (0.818). Therefore, it is both an outlier and an influential value.

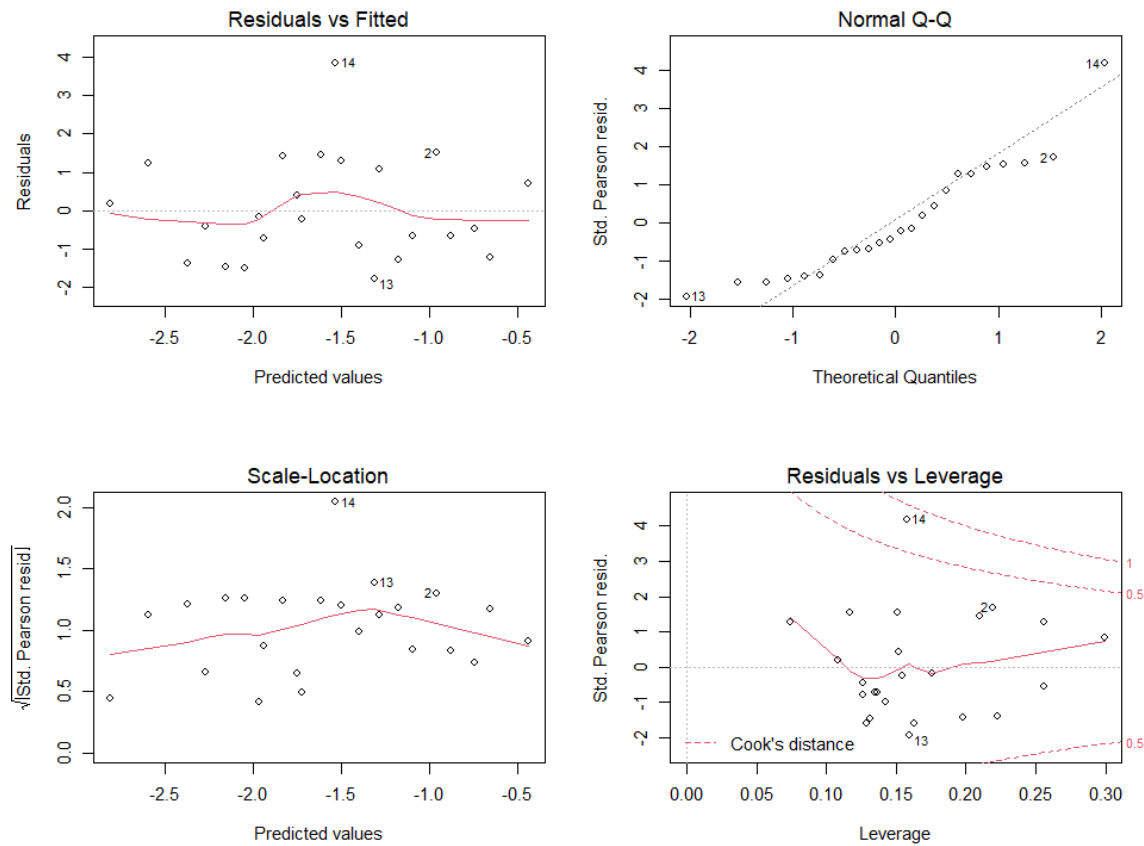


Figure 1: Diagnostic plots for model m1

(d) Repeat the analysis with observation 14 deleted:

```
carrot2=carrots[-14,]
m3=glm(cbind(success,failure)~logdose+block,family=binomial,carrot2)
summary(m3)
```

The result once again shows a very highly significant coefficient for `logdose` and it looks as though the block effect B3 is significant as well:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.68659	0.69482	3.867	0.00011	***
logdose	-2.17902	0.37124	-5.870	4.37e-09	***
blockB2	0.08988	0.21149	0.425	0.67084	
blockB3	-0.54581	0.23312	-2.341	0.01921	*
Residual deviance: 25.289 on 19 degrees of freedom					

The residual plots (not shown here) show no evidence of further outliers though there is still some evidence that the residuals are not normally distributed.

Summary: Observation 14 is unusual because of the high number of successes in comparison with the other samples at the same or similar doses. However, the model appears to fit well to the rest of the data. There is clearly a very highly significant effect due to

the insecticide, and there is also evidence that block B3 has a lower success rate than the others.

Comment on student solutions. The most common error was some version of the following analysis: first define `rate=carrots$success/(carrots$success+carrots$failure)`, and then do a binomial glm with `rate` as the response. However, glm with `family=binomial` only works correctly if the response variable is either binary (0 or 1), or a two-column matrix of successes and failures as in this example. The model fitted in this case is meaningless, *but*, if you did it this way, you should still have been able to identify observation 14 as an outlier. So students who followed this route still got some credit.

2. (a) See Figure 2.

```
library(faraway)
data(oatvar)
par(mfrow=c(1,2))
plot(oatvar$yield~factor(oatvar$variety))
plot(oatvar$yield~factor(oatvar$block))
```

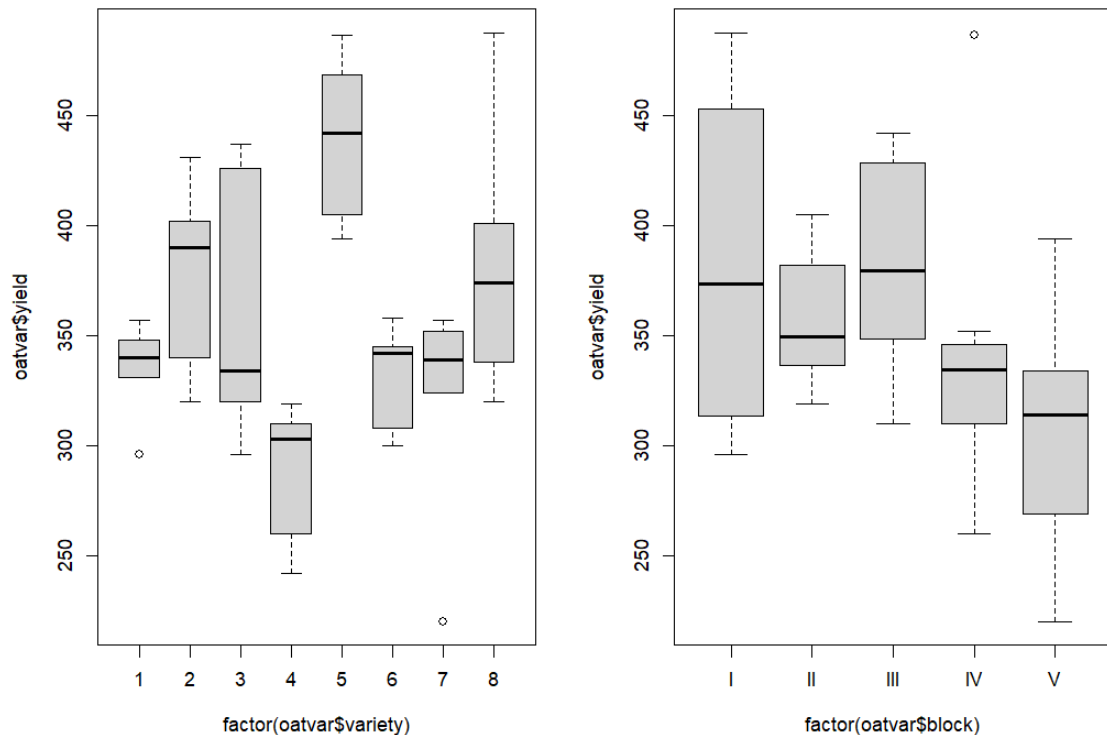


Figure 2: Initial plots for oatvar data

(b) See the code:

```
mod1=lm(yield~factor(variety)+factor(block),oatvar)
mod2=lm(yield~factor(variety),oatvar)
```

```

mod3=lm(yield~factor(block),oatvar)
anova(mod1,mod2)
anova(mod1,mod3)

```

The p-values for mod2 against mod1 and for mod3 against mod1 are 0.001 and 1.8×10^{-5} respectively. Clearly both the variety effect and the block effect are significant, but especially the variety effect.

INLA:

(c) Possible code:

```

library(INLA)
library(ggplot2)
data(oatvar, package="faraway")
formula <- yield ~ variety + f(block, model="iid")
result <- inla(formula, family="gaussian", data=oatvar)
summary(result)
# remark: we could change the prior as in some of the class examples, but
# in this case, the results seem worse so go back to the original model
sigmaalpha <- inla.tmarginal(function(x) 1/sqrt(exp(x)),
result$internal.marginals.hyperpar[[2]])
sigmaepsilon <- inla.tmarginal(function(x) 1/sqrt(exp(x)),
result$internal.marginals.hyperpar[[1]])
restab <- sapply(result$marginals.fixed,
function(x) inla.zmarginal(x,silent <- TRUE))
restab <- cbind(restab, inla.zmarginal(sigmaalpha,silent=TRUE))
restab <- cbind(restab, inla.zmarginal(sigmaepsilon,silent=TRUE))
colnames(restab) <- c("mu","V2-V1","V3-V1","V4-V1","V5-V1","V6-V1","V7-V1","V8-V1",
"alpha","epsilon")
c(inla.mmarginal(sigmaalpha), inla.mmarginal(sigmaepsilon))
ddf <- data.frame(rbind(sigmaalpha,sigmaepsilon),errterm=gl(2,2048,
labels = c("alpha","epsilon")))
library(ggplot2)
ggplot(ddf, aes(x,y, linetype=errterm))+geom_line()+
xlab("sigma")+ylab("density")+xlim(10,60)

```

The result is the top plot of Figure 3.

(d) The table may be produced by

```

restab[,2:10]=round(as.numeric(restab[,2:10]),2)
data.frame(restab)

```

with the result

	mu	V2.V1	V3.V1	V4.V1	V5.V1	V6.V1	V7.V1	V8.V1	alpha	epsilon
mean	348.1278	22.03	11.22	-47.29	70.5	-13.48	-22.9	27.89	23.43	36.75
sd	14.99764	17.31	17.24	17.29	17.9	17.17	17.18	17.35	3.42	5.28
quant0.025	318.3771	-12.54	-23.06	-80.87	34.13	-47.3	-56.6	-6.84	17.38	27.43
quant0.25	338.32	10.47	-0.32	-59.01	58.67	-25.04	-34.48	16.32	21.01	33.01

quant0.5	348.0589	22.13	11.27	-47.52	70.84	-13.54	-23.01	28.02	23.2	36.39
quant0.75	357.7976	33.63	22.75	-35.88	82.64	-2.06	-11.49	39.54	25.59	40.09
quant0.975	377.6885	55.47	44.68	-12.9	104.46	20.2	10.93	61.35	30.78	48.14

Varieties 4 and 5 are significantly different (negative and positive, respectively). The others do not appear significant when the posterior means are compared with the posterior standard deviations.

(e) The bottom plot of Figure 3 is produced by the following code.

```
x <- seq(-100,100,length.out = 100)
rden <- sapply(result$marginals.fixed,function(y) inla.dmarginal(x, y))[, -1]
ddf <- data.frame(yield=rep(x,7), density=as.vector(rden), treat=gl(7,100,
labels=c("V2-V1","V3-V1","V4-V1","V5-V1","V6-V1","V7-V1","V8-V1")))
ggplot(ddf, aes(x=yield, y=density, linetype=treat))+geom_line()
```

STAN:

(c) Code:

```
library(faraway)
library(rstan)
library(reshape2)
data(oatvar)
oatdat <- list(N=40, Nt=8, Nb=5, y=oatvar$yield, treat=as.numeric(oatvar$variety),
blk=as.numeric(oatvar$block))
fit <- stan(file = "C:/Users/rls/aug20/UNC/STOR590-F20/R-Code/rbd.stan",
data = oatdat)
traceplot(fit, pars="sigmabl", inc_warmup = FALSE)
traceplot(fit, pars="sigmaepsilon", inc_warmup = FALSE)
postsig <- rstan::extract(fit, pars=c("sigmabl","sigmaepsilon"))
ref <- melt(postsig,value.name="sigma")
ggplot(data=ref,aes(x=sigma, linetype=L1))+geom_density()+scale_linetype(name="SD",
labels=c("block","error"))+xlim(0,100)
```

(d) Code:

```
print(fit, pars=c("trt","sigmabl","sigmaepsilon","bld"), probs=c(0.025,0.5,0.975))
...
Inference for Stan model: rbd.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.
```

	mean	se_mean	sd	2.5%	50%	97.5%	n_eff	Rhat
trt[1]	334.48	1.19	26.72	283.23	334.32	390.90	506	1.00
trt[2]	376.76	1.14	26.31	325.65	376.22	428.83	534	1.00
trt[3]	362.80	1.11	26.78	309.19	362.97	416.02	584	1.00
trt[4]	286.60	1.05	26.29	233.01	286.96	338.19	624	1.00
trt[5]	439.72	1.15	26.69	388.65	438.95	497.44	538	1.00
trt[6]	330.67	1.10	26.12	279.26	330.72	384.11	561	1.00

trt[7]	318.62	1.16	26.79	263.85	318.53	372.74	537	1.00
trt[8]	384.11	1.11	26.88	330.46	384.24	440.46	582	1.00
sigmablκ	42.79	1.29	24.91	13.19	36.54	106.18	375	1.01
sigmaepsilon	38.38	0.13	5.50	29.40	37.72	51.00	1782	1.00
bld[1]	23.99	1.04	23.29	-24.08	23.15	75.11	506	1.00
bld[2]	2.90	1.04	23.03	-43.91	2.66	50.92	493	1.00
bld[3]	24.43	1.06	23.61	-22.74	23.75	73.79	500	1.00
bld[4]	-11.03	1.13	23.19	-63.31	-10.39	37.21	425	1.01
bld[5]	-40.86	1.15	24.38	-97.50	-39.59	1.54	447	1.00

Samples were drawn using NUTS(diag_e) at Sat Dec 09 18:00:43 2023. For each parameter, `n_eff` is a crude measure of effective sample size, and `Rhat` is the potential scale reduction factor on split chains (at convergence, `Rhat=1`).

(e) Code:

```
trteff <- rstan::extract(fit, pars="trt")
rdf <- data.frame(yield=unlist(trteff), treat=gl(8,4000, labels=1:8))
ggplot(data=rdf,aes(x=yield, linetype=treat))+geom_density()
```

The plots from parts (c) and (e) are shown in the top and bottom plots of Figure 4.

Summary of Conclusions (from both STAN and INLA). Both Bayesian analyses show posterior densities on a similar scale for the block and residual effects (which implies, among other things, that the block effect needs to be taken into account). Although it is a little hard from the variety plots to interpret which variety is which, the table of model results makes the order clear: variety (treatment) 4 has the smallest posterior mean and variety 5 has the largest. The others do not appear significantly different from variety 1 when taking the posterior standard deviations into account.

Comment on student solutions. Numerous students who used INLA found the same problem as with the examples in class, that estimates of the precision parameter for the random effect were unrealistically large, implying that `sigmaalpha` (the SD of the random effect) is very close to 0, much smaller than `sigmaepsilon`. That issue didn't affect my own analysis; however, it's possible that different versions of INLA are producing slightly different results. For some students, this issue persisted even after trying the alternative prior density. The main consequence was with the posterior density plots for `sigmaalpha` and `sigmaepsilon`: if the scales are totally different, there's no sensible way to plot both densities on the same graph (but you should still have been able to produce separate plots for each density). This issue will also have affected the posterior densities of the block effects if you tried to graph those, but because the variety effects are treated as fixed effects, the glitch I've just described does not affect the variety effects much at all. In fact, most of the numerical results for variety effects were correct: the difficulty here comes in drawing the plots (which, I have to confess, took a few attempts myself before I got it right). Far fewer students used STAN, but the ones that did seemed to fare better: certainly, the post-model-fitting analyses (such as drawing posterior density plots) are easier to handle in STAN than in INLA.

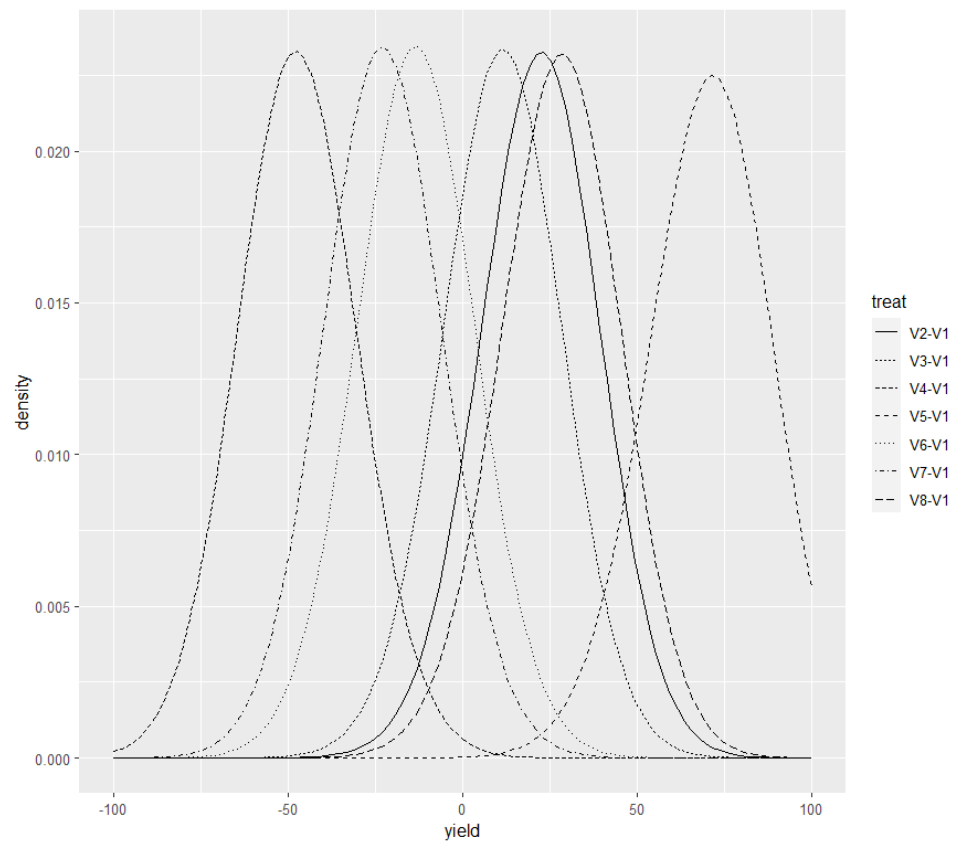
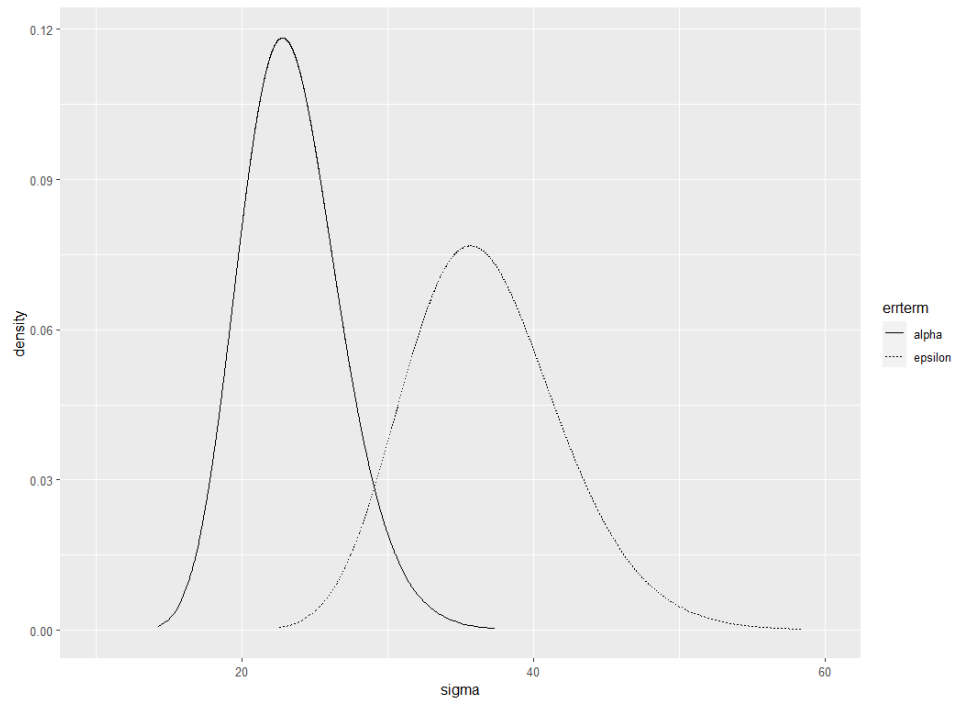


Figure 3: INLA plots

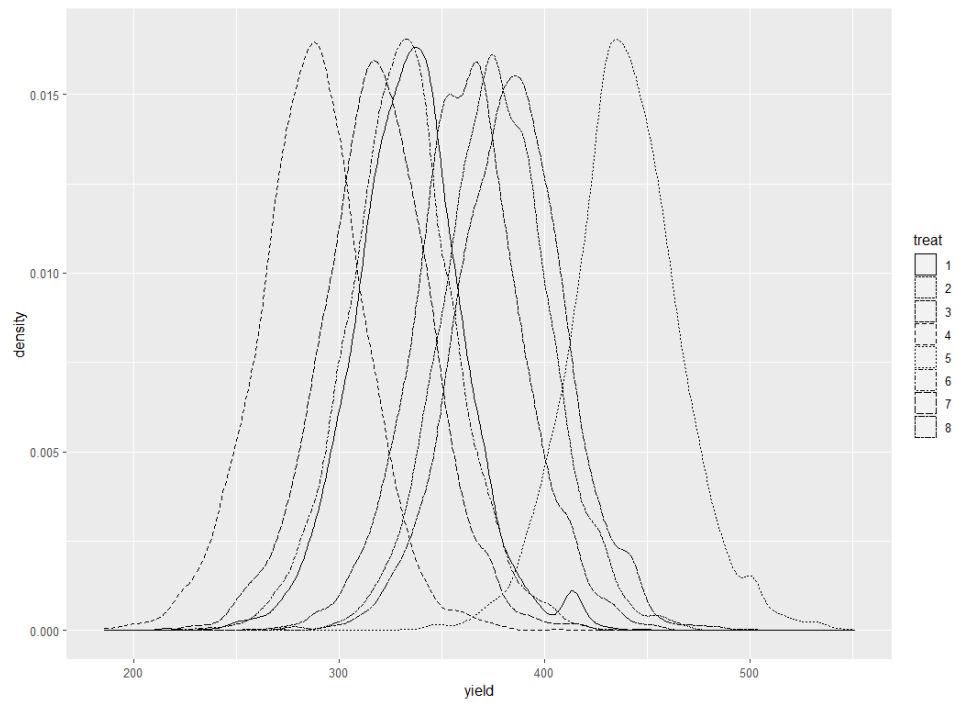
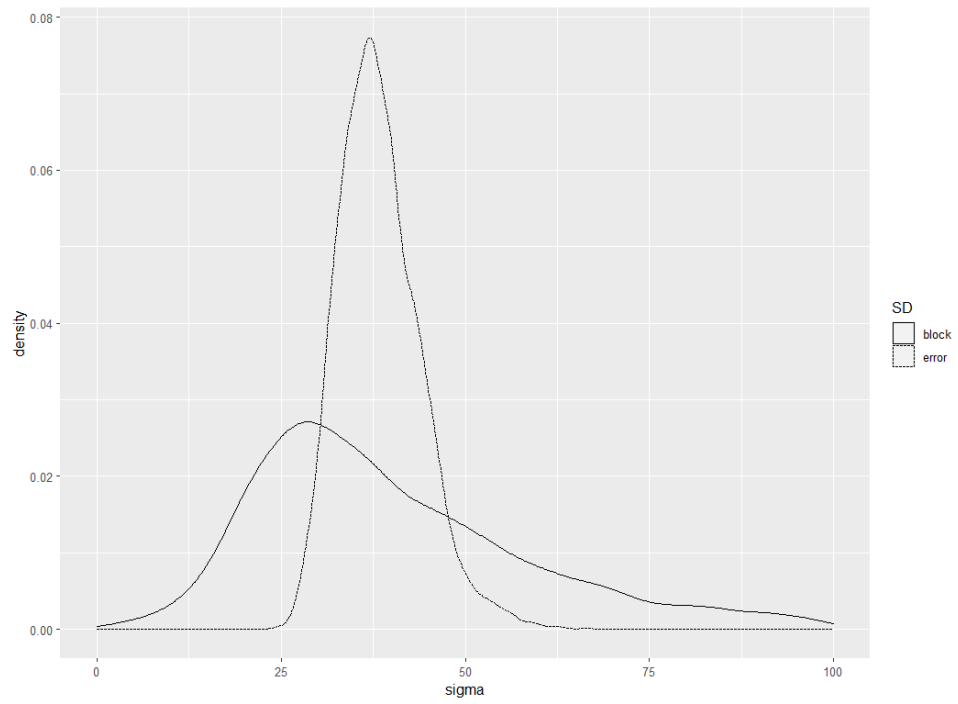


Figure 4: STAN plots

3. (a) Loading the dataset and then

```
mean(salamanders$y[salamanders$POP=='R/R'])
mean(salamanders$y[salamanders$POP=='R/W'])
mean(salamanders$y[salamanders$POP=='W/R'])
mean(salamanders$y[salamanders$POP=='W/W'])
```

shows mating rates of 0.67, 0.56, 0.21, 0.67 in the four populations. The observed mating rate in the W/R population is substantially lower than in the other three.

- (b) Code such as

```
g0=glm(y~factor(EXPT),family=binomial,salamanders)
g1=glm(y~factor(EXPT)+POP,family=binomial,salamanders)
anova(g0,g1,test='LRT')
```

leads to a very highly significant effect for POP ($p \approx 2 \times 10^{-11}$) and the following (edited) table of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	9.750e-01	2.824e-01	3.453	0.000555	***
factor(EXPT)2	-4.321e-01	2.816e-01	-1.535	0.124886	
factor(EXPT)3	-3.935e-01	2.816e-01	-1.397	0.162300	
POP R/W	-4.741e-01	3.096e-01	-1.531	0.125664	
POP W/R	-2.029e+00	3.435e-01	-5.906	3.5e-09	***
POP W/W	5.287e-16	3.175e-01	0.000	1.000000	

Residual deviance: 442.68 on 354 degrees of freedom

This shows that the different mating rate for W/R is very highly significant, but the contrast between R/W and the baseline R/R is not. Also, the effect for EXPT is not significant. However, the residual deviance (442.68) is too large for the degrees of freedom (354), implying the model is not a good fit.

- (c) One way for fit these models is

```
library(lme4)
salamanders$EXPT=as.factor(salamanders$EXPT)
g2=glmer(y~EXPT+POP+(1|FEMALE:EXPT)+(1|MALE:EXPT),nAGQ=1,family=binomial,
g22=glmer(y~POP+factor(EXPT)+(1|FEMALE)+(1|MALE),family=binomial,
g23=glmer(y~POP+as.numeric(EXPT)+(1|FEMALE)+(1|MALE),family=binomial,
data=salamanders)
```

Note the format of the random effects nested within EXPT, which reflects that different animals were used for the different experiments. Part of `summary(g2)` reads as follows:

	AIC	BIC	logLik	deviance	df.resid
	433.2	464.3	-208.6	417.2	352

Random effects:

Groups	Name	Variance	Std.Dev.
FEMALE:EXPT	(Intercept)	1.139	1.067
MALE:EXPT	(Intercept)	1.009	1.004

Number of obs: 360, groups: FEMALE:EXPT, 60; MALE:EXPT, 60

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.39247	0.52457	2.655	0.00794	**
EXPT2	-0.61598	0.57726	-1.067	0.28594	
EXPT3	-0.53255	0.57689	-0.923	0.35593	
POPR/W	-0.70031	0.45910	-1.525	0.12715	
POPW/R	-2.90058	0.55869	-5.192	2.08e-07	***
POPW/W	-0.02153	0.53882	-0.040	0.96813	

We may note that there still seems to be some discrepancy between the deviance and the `df.resid`, implying that the model is not a good fit. However, the coefficients of the fixed effects are similar to those earlier, with somewhat larger standard errors. The biggest different is in POPW/R, which has gone from -2.029 to -2.9 . So it looks as though taking account of the random effects has enhanced the difference between W/R and the other three populations.

(d) The code

```
g3=glmer(y~EXPT+(1|FEMALE:EXPT)+(1|MALE:EXPT),nAGQ=1,family=binomial,
data=salamanders)
g4=glmer(y~POP+(1|FEMALE:EXPT)+(1|MALE:EXPT),nAGQ=1,family=binomial,
data=salamanders)
anova(g2,g3,test='LRT')
anova(g2,g4,test='LRT')
```

shows that the p-value for g3 against g2 is $\approx 2.5 \times 10^{-12}$, while the p-value for g4 against g2 is ≈ 0.51 . This confirms the results from part (b), that the POP effect is clearly significant but the EXPT effect is not.

(e) Various things you could do here for instance, but a starting point would be to plot `residuals(g2)` against `fitted(g2)` — see the left hand plot in Figure 5. One possible way to go beyond this is by adapting some of the code from Chapter 2, for example

```
library(dplyr)
salamanders=mutate(salamanders,residuals=residuals(g2),linpred=fitted(g2))
gdf=group_by(salamanders,cut(linpred,breaks=unique(quantile(linpred,(1:20)/21))))
diagdf=summarise(gdf,residuals=mean(residuals),linpred=mean(linpred))
plot(residuals~linpred,diagdf,xlab='linear predictor',ylab='mean residual')
```

which leads to the right hand plot of Figure 5.

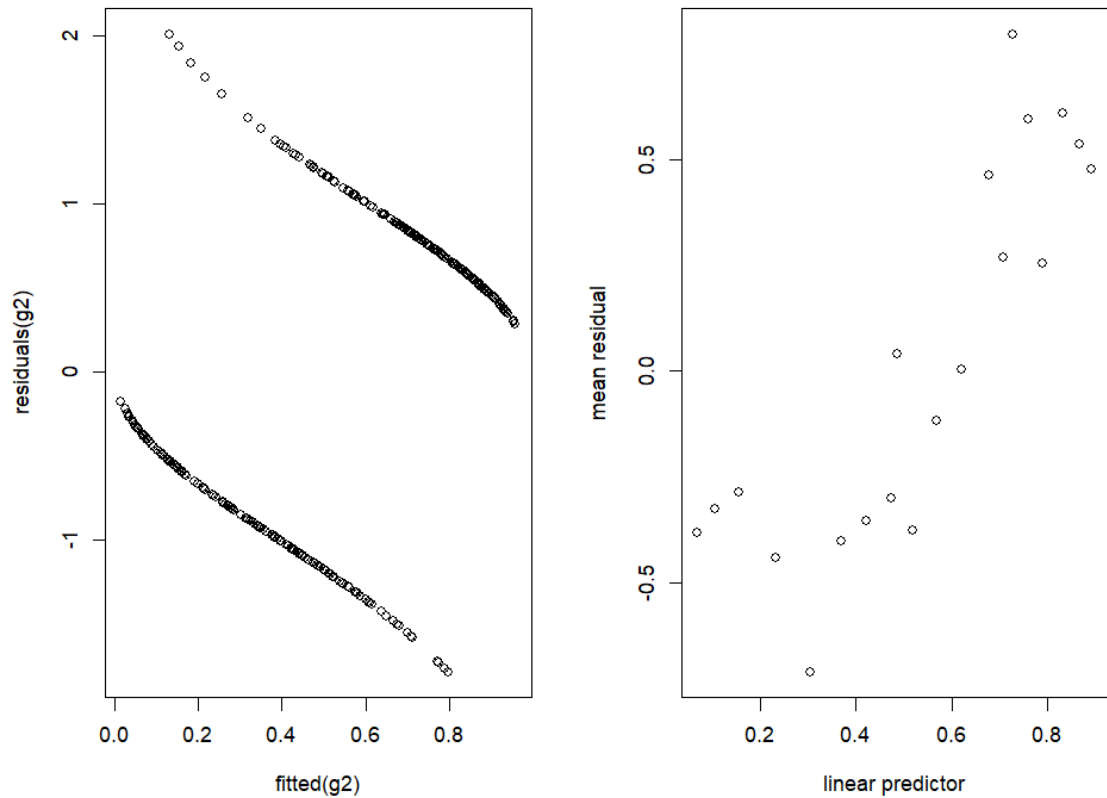


Figure 5: Two versions of the residuals v. fitted values plot, the second with grouping

It seems there is a problem with the pattern of residuals not looking random, which makes me wonder whether the `glmer` is actually working correctly in this instance. Another calculation is

```
mean(residuals(g2)[salamanders$POP=='R/R'])
mean(residuals(g2)[salamanders$POP=='R/W'])
mean(residuals(g2)[salamanders$POP=='W/R'])
mean(residuals(g2)[salamanders$POP=='W/W'])
```

which leads to the four means 0.057, 0.050, -0.093, 0.057, which further exemplifies the issue that the residuals do not appear to be random.

Another plot you could do is

```
d <- fortify.merMod(g2)
ggplot(d, aes(sample=.resid))+stat_qq()+facet_grid(EXPT~POP)
```

which leads to the plot in Figure 6. If you look at this plot carefully, you can see the same problem, that the residuals in the W/R column appear to be larger than those in the other three.

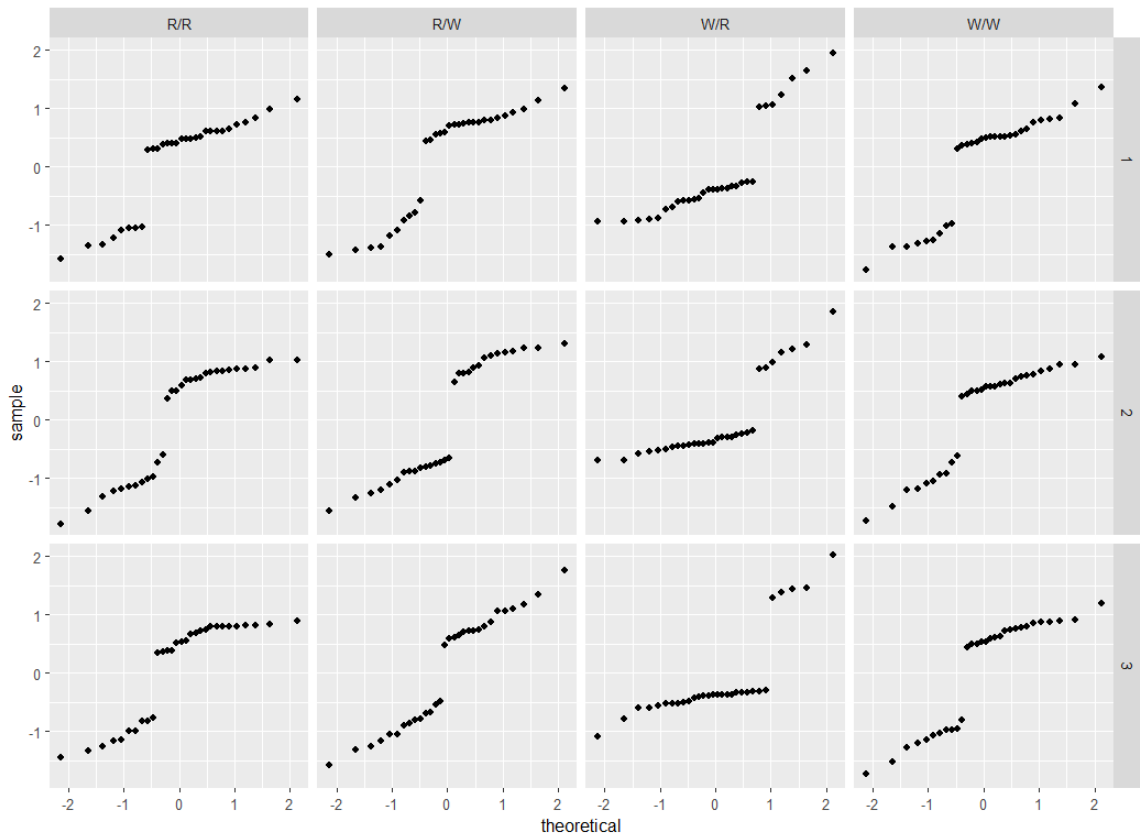


Figure 6: QQ plots subsetted by experiment and population

Comment on student solutions. Two issues came up repeatedly in grading this question. First, most students missed the point about nesting the random effects within EXPT, so they wrote $(1|FEMALE)+(1|MALE)$ instead of $(1|FEMALE:EXPT)+(1|MALE:EXPT)$. However, I tried to emphasize that the analysis assumed different animals were used in different experiments: the point was the same as the technicians-within-labs analysis which was one of the Chapter 10 examples. The second point was simply not identifying the issue about the unequal distribution of residuals among the four populations: however, that was a hard one to spot if you didn't know what to look for. Kudos to the few students who got this point. (In fact, the assumption about using different animals in different experiments was not quite correct; that was the original intention, but the experiments ended up using some of the same animals twice. I learned about this dataset from Peter McCullagh, who discussed it in some detail in the book *Generalized Linear Models* by McCullagh and Nelder; the dataset was also the subject of the research paper *Generalized Linear Models with Random Effects; Salamander Mating Revisited* by M. Rezaul Karim and Scott L. Zeger, *Biometrics*, Jun., 1992, Vol. 48, No. 2 (Jun., 1992), pp. 631-644, <https://www.jstor.org/stable/2532317>. The analysis here is a simplified version of what was given in those original references.)