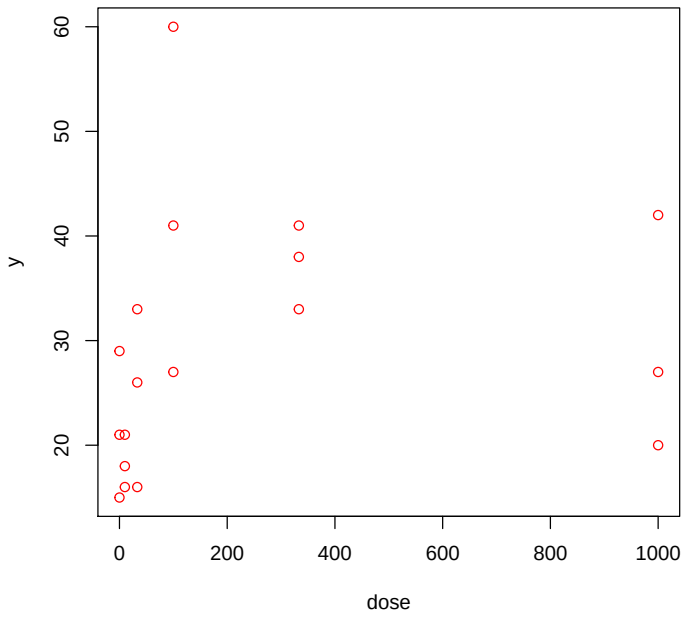
Chap 5, 19
(a) It's easier to see the distribution of the points when the x-axis of these points are evenly distributed

(b) The p-values are 0.1, <1e-5, <1e-5 for the base, log10, and factor models, so the first is not sifgnificant, but significant for the other two.

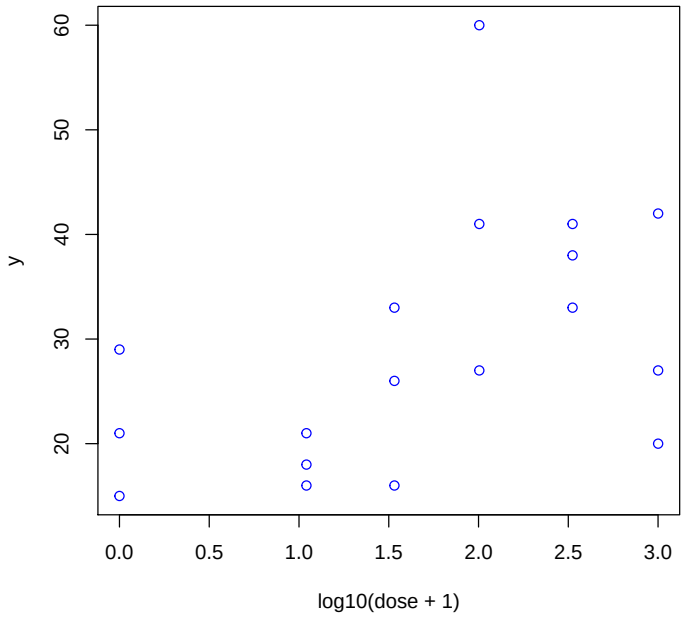
(c) Compare Dev/DF with 2 (like the critical value 1.96), the "factor" model (Dev/DF = 2.8) is less lack of fit. The "dose" model (Dev/DF = 4.7) is most lack of fit.

Model	$-2\log\Lambda$	DF	p-value	Residual Dev	Residual DF	Dev/DF
dose	2.55	1	0.11	75.8	16	4.7
log10(dose+1)	18.73	1	1.5e-5	59.6	16	3.7
factor(dose)	44.86	5	1.5e-8	33.4	12	2.8

```
library(aod)
data("salmonella")
dose = salmonella$dose
y = salmonella$y
plot(dose, y, col="red")
```



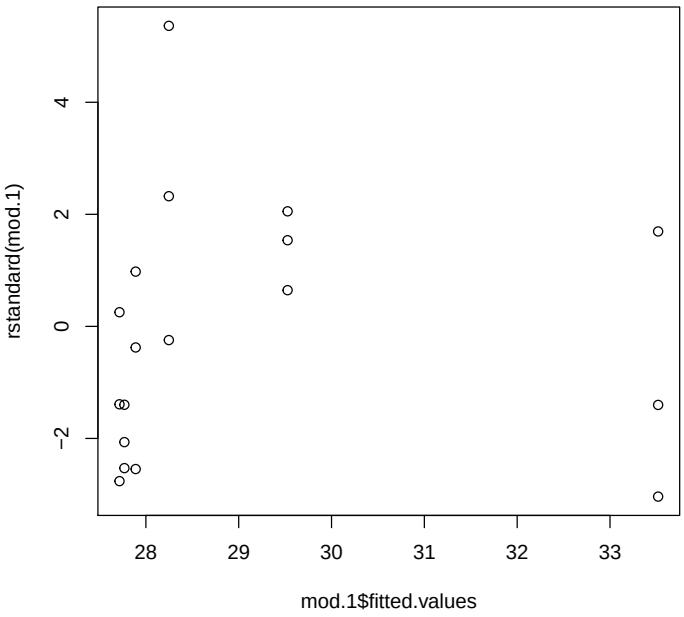
```
plot(log10(dose+1), y, col="blue")
```



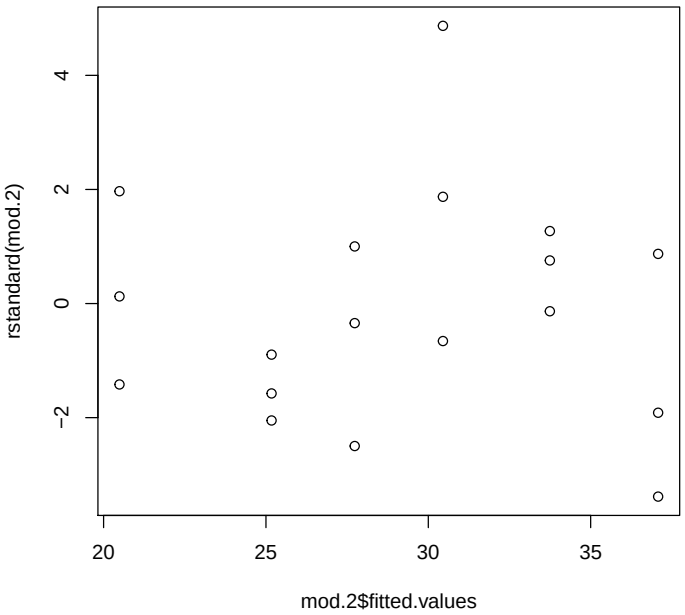

```
mod.1 <- glm( y ~ dose, family = poisson())
mod.2 <- glm( y ~ log10(dose+1), family = poisson())
mod.3 <- glm( y ~ as.factor(dose), family = poisson())
A3 = anova(mod.3, test = "Chisq")
```

(d)

```
plot(mod.1$fitted.values, rstandard(mod.1))
```

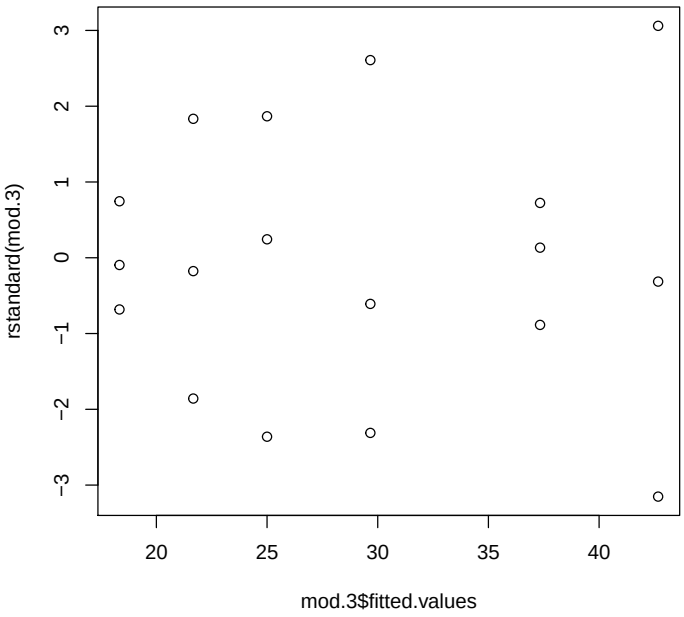



```
plot(mod.2$fitted.values, rstandard(mod.2))
```



(e) The mean residual is now 0 everywhere, and so the extreme residuals are not so large. There are still two that lie beyond ± 3 and three more beyond ± 2, which is too many for a sample of size 18. There is some kind of overdispersion relative to a Poisson model.

```
plot(mod.3$fitted.values, rstandard(mod.3))
```

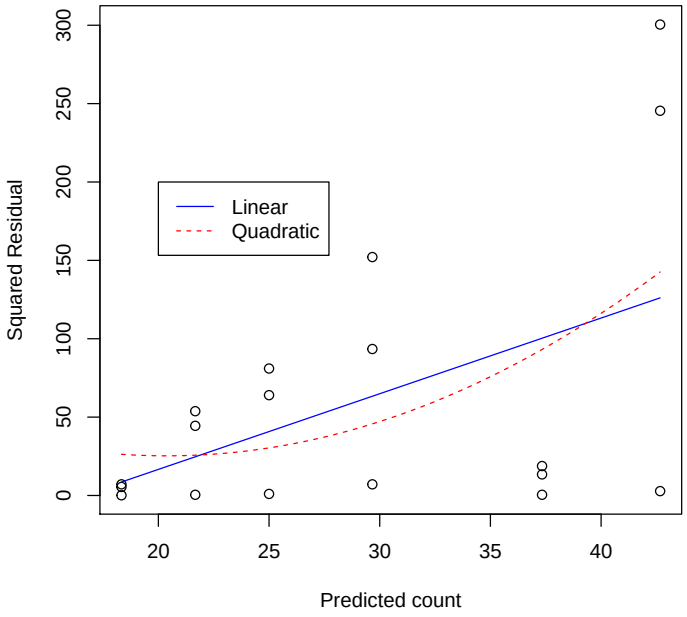


Chap 5, 35

(a) Negative-binomial has smaller devience residual, so it is better than Quasi-possion.

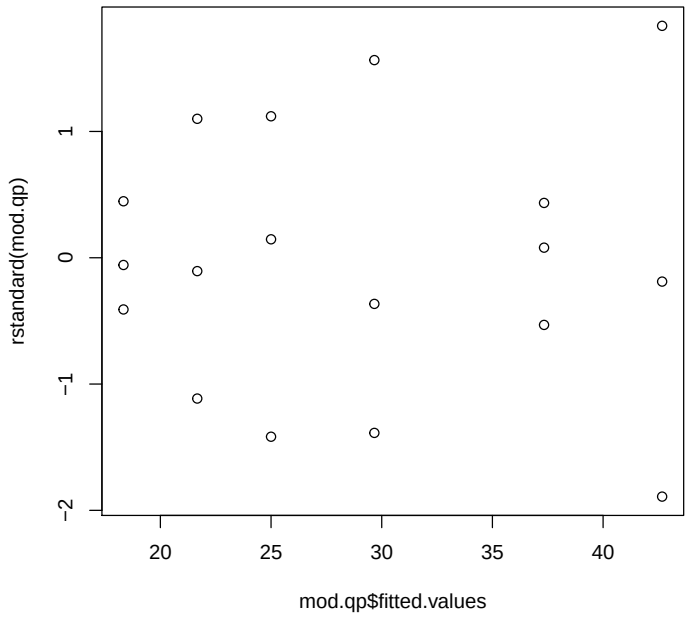
```
library(MASS)
library(car)
mod.nb <- glm.nb(y ~ as.factor(dose))
mod.qp <- glm(y ~ as.factor(dose), family = quasipoisson(link = "log"))
print(as.matrix(anova(mod.qp, mod.nb)),quote = "FALSE")
 Resid. Df Resid. Dev Df Deviance
1        12   33.49596 NA       NA
2        12   17.27234  0 16.22362
res.sq <- residuals ( object = mod.3 , type = "response")^2
set1 <- data.frame ( res.sq , mu.hat = mod.3$fitted.values )
fit.lin <- lm( formula = res.sq ~ mu.hat , data = set1 )
fit.quad <- lm( formula = res.sq ~ mu.hat + I(mu.hat ^2), data = set1 )
plot ( x = set1$mu.hat , y = set1$res.sq , xlab = "Predicted count" , ylab = "Squared R
curve ( expr = predict ( object = fit.lin , newdata =
               data.frame (mu.hat = x ) , type = "response") , col = "blue" , add
               = TRUE , lty = "solid")
curve (expr = predict ( object = fit.quad , newdata =
               data.frame ( mu.hat = x ) , type = "response") , col = "red" , add =
               TRUE , lty = "dashed")
```

```
 legend ( x = 20 , y = 200 , legend = c ("Linear" , "Quadratic") , col
                = c ("blue", "red") , lty = c ("solid" , "dashed") , bty = "n ")
 print(as.matrix(anova(mod.qp, test="Chisq")), quote = "FALSE")
                  Df Deviance Resid. Df Resid. Dev    Pr(>Chi)
NULL              NA       NA        17   78.35758          NA
as.factor(dose)   5 44.86162        12   33.49596 0.006431595
```



(b) The p-value is 0.006432, which indicates the dose has an significant effect.
(c)

```
plot(mod.qp$fitted.values, rstandard(mod.qp))
```



Chap 6, Q4

(a)

i. Each explanatory variable is assumed to have the same effect on each subject, but their overall (log) means (or intercepts; the"baselin" number of drinks when all explanatory variables are zero) are assumed to vary randomly.

ii. We tried 1, 3, 5, 7, 9, 12, 15 quadrature points. The variance component estimate is 0.208 for 1 and 3, and 0.209 thereafter. Fixed effect parameter estimates are bouncing around seemingly randomly but not by more than the 3rd decimal place, so we conclude that the additional quadrature points beyond about 5 are not useful.

```
library(lme4)
data <- read.csv(
                "http://www.chrisbilder.com/categorical/Chapter4/DeHartSimplified.csv")
data[,3] = as.factor(data[,3])
mod <- glmer(formula = numall ~   nrel+ prel+ negevent+ posevent
                              + gender+ rosn+ age+ desired+ state + dayweek + (1|id),
                              data = data,  family = poisson(), nAGQ = 5)
mod.glm <- glm(formula = numall ~ dayweek  + nrel+ prel + negevent
                                +posevent+ gender+ rosn+ age+ desired+ state
                                + dayweek , data = data,  family = poisson())
```

iii

$T = 43.744$ with p-value $3.17e-10$

```
m1 <- lmer(numall ~ dayweek + nrel+ prel
                          + negevent+ posevent+ gender+ rosn
                          + age+ desired+ state + (1|id), data = data,REML=F)
m0 <- lm(numall ~ dayweek + prel+ negevent
                         + posevent+ gender+ rosn+ age
                         + desired+ state, data = data,REML=F)
anova(m1,m0) ## two sequential tests
```

```
Data: data
Models:
m0: numall ~ dayweek + prel + negevent + posevent + gender + rosn +
m0:     age + desired + state
m1: numall ~ dayweek + nrel + prel + negevent + posevent + gender +
m1:     rosn + age + desired + state + (1 | id)
   Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
m0 16 2828.5 2899.3 -1398.3   2796.5
m1 18 2788.8 2868.4 -1376.4   2752.8 43.744      2   3.17e-10 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

iiii

	Df	AIC	LRT	Pr(Chi)
<none>		1034.85		
nrel	1.00	1040.49	7.65	0.01
prel	1.00	1033.24	0.39	0.53
negevent	1.00	1037.15	4.31	0.04
posevent	1.00	1032.85	0.00	0.97
gender	1.00	1033.41	0.56	0.45
rosn	1.00	1034.64	1.79	0.18
age	1.00	1033.19	0.34	0.56

```
scope = names(data)[-c(1,2,4)]
ltr1 = drop1(mod, scope = c("nrel","prel", "negevent", "posevent"),
             data = na.omit(data),  test ="Chisq")
xtable(as.matrix(ltr2))
ltr2 = drop1(mod, scope = c( "gender", "rosn", "age"),
             data = na.omit(data),  test ="Chisq")
```

(b)
(i) Independence means y_{ij} and y_{ik} are independent; while exchangible means that exchanging a par of $\mathbf{y_i} := \{y_{i1}, y_{i2}, \cdots y_{iJ}\}$, the joint distribution of $\mathbf{y_i}$ doesn't change.

(ii) There are significant effects due to prel, negevent, desired, and dayweek. The significance of prel is different from the GLMM, as is the nonsignificance of nrel.

```
library(geepack)
#data = data[sort(data[,1]),]
data = na.omit(data)
mod.gee <- geeglm(numall ~ nrel+ prel+ negevent+ posevent
                                + gender+ rosn+ age+ desired+ state + dayweek,
                                data = data, id = id, corstr = "exchangeable",
                                family = poisson("log"))
print(coef(summary(mod.gee)))
```

```
               Estimate    Std.err        Wald     Pr(>|W|)
(Intercept) -0.326354009 0.74788112  0.19042009 6.625672e-01
nrel         0.090566890 0.04490711  4.06732368 4.372049e-02
prel         0.016480588 0.02780300  0.35136790 5.533398e-01
negevent    -0.291707311 0.13591679  4.60625943 3.185545e-02
posevent     0.013180937 0.06479952  0.04137605 8.388138e-01
gender      -0.119160707 0.11732462  1.03154412 3.097963e-01
rosn         0.129808682 0.12798412  1.02871563 3.104605e-01
age         -0.004888138 0.01110577  0.19372666 6.598329e-01
desired      0.278878955 0.03947359 49.91351724 1.606715e-12
state       -0.078026059 0.08945524  0.76079522 3.830798e-01
dayweek2    -0.177600673 0.13547804  1.71850742 1.898853e-01
dayweek3    -0.115100268 0.11314645  1.03483425 3.090259e-01
dayweek4     0.094241360 0.10979824  0.73670288 3.907191e-01
dayweek5     0.228328975 0.11068685  4.25529854 3.912807e-02
dayweek6     0.534209881 0.12147080 19.34105104 1.093304e-05
dayweek7     0.242976747 0.14411676  2.84249981 9.180129e-02
```

(iii) The test results change because the tests are performed assuming that variables are entered into the model in the order in which they are listed in the formula. Now negevent and prel are no longer significant, but nrel is.
(iiii) From the Anova table below, dayweek is significant.

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
mod.wo.dayweek	11.00	1103.47	1152.16	-540.73	1081.47			
mod	17.00	1034.85	1110.10	-500.42	1000.85	80.62	6.00	0.00

```
mod <- glmer(formula = numall~ nrel+ prel+ negevent+ posevent+ gender
+ rosn+ age+ desired+ state + dayweek + (1|id),
data = data,  family = poisson(), nAGQ = 5)

mod.wo.dayweek <- glmer(formula = numall~nrel+ prel+ negevent+ poseven+ gender
+ rosn+ age+ desired+ state + (1|id),
data = data,  family = poisson(), nAGQ = 5)
```

```
a = anova(mod,mod.wo.dayweek,test = "Chisq")
```