Cassie GLMM

# Cassie Bond - Generalised Linear Mixed Models for microplastics data

library(nlme)  
library(multcomp)

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##   
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':  
##   
## geyser

library(car)

## Loading required package: carData

### within each of the catchments - run t test

to see whether there is a difference between the catchments or sites

cat("\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Comparisons by Drain (Catchment) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\n")

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Comparisons by Drain (Catchment) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

meanz <- signif(tapply(cassie$count\_corr\_across\_sizes.L, cassie$Drain, mean),3)  
sdz <- signif(tapply(cassie$count\_corr\_across\_sizes.L, cassie$Drain, sd),3)  
rsdz <- signif(100\*sdz/meanz,3); rbind(meanz,sdz, rsdz)

## B C H K SB SC  
## meanz 12.60 12.80 20.1 9.22 12.80 17.5  
## sdz 3.63 4.68 11.6 2.08 4.22 10.3  
## rsdz 28.80 36.60 57.7 22.60 33.00 58.9

with(cassie, oneway.test(count\_corr\_across\_sizes.L ~ Drain))

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: count\_corr\_across\_sizes.L and Drain  
## F = 2.276, num df = 5.000, denom df = 13.391, p-value = 0.1061

with(cassie, pairwise.t.test(count\_corr\_across\_sizes.L, Drain, p.adjust.method = "none", pool.sd = F))

##   
## Pairwise comparisons using t tests with non-pooled SD   
##   
## data: count\_corr\_across\_sizes.L and Drain   
##   
## B C H K SB   
## C 0.914 - - - -   
## H 0.179 0.199 - - -   
## K 0.086 0.128 0.070 - -   
## SB 0.915 0.995 0.195 0.102 -   
## SC 0.308 0.345 0.691 0.107 0.338  
##   
## P value adjustment method: none

with(cassie, kruskal.test(count\_corr\_across\_sizes.L ~ Drain))

##   
## Kruskal-Wallis rank sum test  
##   
## data: count\_corr\_across\_sizes.L by Drain  
## Kruskal-Wallis chi-squared = 9.6008, df = 5, p-value = 0.08737

cat("\n\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Comparisons by Site \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\n")

##   
## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Comparisons by Site \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

meanz <- signif(tapply(cassie$count\_corr\_across\_sizes.L, cassie$Site, mean),3)  
sdz <- signif(tapply(cassie$count\_corr\_across\_sizes.L, cassie$Site, sd),3)  
rsdz <- signif(100\*sdz/meanz,3); rbind(meanz,sdz, rsdz)

## B1 B2 C1 C2 H1 H2 K1 K2 SB1 SB2 SC1 SC2  
## meanz 13.90 11.20 14.70 11.00 25.1 15.20 9.67 8.77 12.90 12.70 24.7 10.30  
## sdz 3.55 3.84 5.35 3.95 16.0 2.62 2.17 2.35 1.47 6.51 10.4 2.21  
## rsdz 25.50 34.30 36.40 35.90 63.7 17.20 22.40 26.80 11.40 51.30 42.1 21.50

with(cassie, oneway.test(count\_corr\_across\_sizes.L ~ Site))

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: count\_corr\_across\_sizes.L and Site  
## F = 1.3224, num df = 11.0000, denom df = 9.3764, p-value = 0.3392

with(cassie, pairwise.t.test(count\_corr\_across\_sizes.L, Site, p.adjust.method = "none", pool.sd = F))

##   
## Pairwise comparisons using t tests with non-pooled SD   
##   
## data: count\_corr\_across\_sizes.L and Site   
##   
## B1 B2 C1 C2 H1 H2 K1 K2 SB1 SB2 SC1   
## B2 0.417 - - - - - - - - - -   
## C1 0.848 0.414 - - - - - - - - -   
## C2 0.389 0.945 0.390 - - - - - - - -   
## H1 0.352 0.270 0.382 0.264 - - - - - - -   
## H2 0.647 0.220 0.894 0.208 0.399 - - - - - -   
## K1 0.165 0.588 0.240 0.651 0.237 0.050 - - - - -   
## K2 0.114 0.412 0.185 0.464 0.219 0.035 0.652 - - - -   
## SB1 0.677 0.534 0.624 0.495 0.319 0.273 0.109 0.073 - - -   
## SB2 0.797 0.747 0.707 0.713 0.316 0.591 0.507 0.407 0.969 - -   
## SC1 0.206 0.140 0.233 0.136 0.978 0.249 0.123 0.110 0.184 0.178 -   
## SC2 0.224 0.756 0.292 0.824 0.251 0.072 0.729 0.448 0.180 0.597 0.132  
##   
## P value adjustment method: none

with(cassie, kruskal.test(count\_corr\_across\_sizes.L ~ Site))

##   
## Kruskal-Wallis rank sum test  
##   
## data: count\_corr\_across\_sizes.L by Site  
## Kruskal-Wallis chi-squared = 17.33, df = 11, p-value = 0.09848

We see here that *there is no difference in microplastic concentrations between Drains or Sites*.

We should check if the two sites in each drain have significantly different counts:

cat("Welch t-tests comparing 2 sites in each drain separately:\n\n")

## Welch t-tests comparing 2 sites in each drain separately:

for (i in 1:nlevels(cassie$Drain)){  
 tt0 <- with(droplevels(subset(cassie, cassie$Drain==levels(cassie$Drain)[i])),   
 t.test(count\_corr\_across\_sizes.L ~ Site))  
 n1 <- substr(names(tt0$estimate[1]),nchar(names(tt0$estimate)[1])-2,  
 nchar(names(tt0$estimate)[1]))  
 n2 <- substr(names(tt0$estimate[2]),nchar(names(tt0$estimate)[2])-2,  
 nchar(names(tt0$estimate)[2]))  
 cat("Sites in drain",levels(cassie$Drain)[i],  
 ": p = ",signif(tt0$p.value,3),"; mean",n1,"=",signif(tt0$estimate[1],3),  
 ", mean",n2,"=",signif(tt0$estimate[2],3),"\n")  
}

## Sites in drain B : p = 0.417 ; mean B1 = 13.9 , mean B2 = 11.2   
## Sites in drain C : p = 0.39 ; mean C1 = 14.7 , mean C2 = 11   
## Sites in drain H : p = 0.399 ; mean H1 = 25.1 , mean H2 = 15.2   
## Sites in drain K : p = 0.652 ; mean K1 = 9.67 , mean K2 = 8.77   
## Sites in drain SB : p = 0.969 ; mean SB1 = 12.9 , mean SB2 = 12.7   
## Sites in drain SC : p = 0.132 ; mean SC1 = 24.7 , mean SC2 = 10.3

rm(list = c("tt0","n1","n2"))

So all within-drain comparisons show no significant difference in mean plastic counts.

*There is no difference between Drains or Sites, or between sites within drains.*

# Variance decomposition (Type-II ANOVA model)

## one level decomposition

one.level <- lme(count\_corr\_across\_sizes.L ~ 1,   
 random = ~1|Drain, data = cassie)   
VarCorr(one.level)

## Drain = pdLogChol(1)   
## Variance StdDev   
## (Intercept) 7.207002 2.684586  
## Residual 49.874305 7.062174

# rough check  
7/57

## [1] 0.122807

Overall, drains don’t explain the variation (12%)

## Two level decomposition

A two level decomposition maybe makes more sense, so lets do that properly

two.level <- lme(count\_corr\_across\_sizes.L ~ 1, random = ~1|Drain/Site, data = cassie)   
VarCorr(two.level)

## Variance StdDev   
## Drain = pdLogChol(1)   
## (Intercept) 1.896093 1.376987  
## Site = pdLogChol(1)   
## (Intercept) 13.277125 3.643779  
## Residual 41.908058 6.473643

var.out <- VarCorr(two.level)  
drains <- as.numeric(var.out[2,1])  
site <- as.numeric(var.out[4,1])  
within.site <- as.numeric(var.out[5,1])  
tot <- drains + site + within.site; cat("\n")

drains/tot

## [1] 0.03321742

site/tot

## [1] 0.2326004

within.site/tot

## [1] 0.7341822

* 3% variation at the drain level
* 23% variation by site, unrelated to drain
* 73% variation residual:

I think this converts to a “plastic is ubiquitous” story; neither site nor drain explains the variation.

# AnOVa model treating Drain as a factor

lme.a <- lme(log(count\_corr\_across\_sizes.L) ~  
 Drain,  
 random = ~1|Site, data = cassie)   
summary(lme.a)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 55.27822 66.4878 -19.63911  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.1981225 0.3654216  
##   
## Fixed effects: log(count\_corr\_across\_sizes.L) ~ Drain   
## Value Std.Error DF t-value p-value  
## (Intercept) 2.4957541 0.2046503 24 12.195213 0.0000  
## DrainC -0.0047824 0.2894192 6 -0.016524 0.9874  
## DrainH 0.4069784 0.2894192 6 1.406190 0.2093  
## DrainK -0.2983593 0.2894192 6 -1.030890 0.3423  
## DrainSB 0.0000254 0.2894192 6 0.000088 0.9999  
## DrainSC 0.2402400 0.2894192 6 0.830076 0.4383  
## Correlation:   
## (Intr) DrainC DrainH DrainK DranSB  
## DrainC -0.707   
## DrainH -0.707 0.500   
## DrainK -0.707 0.500 0.500   
## DrainSB -0.707 0.500 0.500 0.500   
## DrainSC -0.707 0.500 0.500 0.500 0.500  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.8535336 -0.5974086 0.1574209 0.5102421 2.1293338   
##   
## Number of Observations: 36  
## Number of Groups: 12

Anova(lme.a)

## Analysis of Deviance Table (Type II tests)  
##   
## Response: log(count\_corr\_across\_sizes.L)  
## Chisq Df Pr(>Chisq)  
## Drain 6.9876 5 0.2216

# allow for unequal group variance at the drain level

lme.b <- lme(log(count\_corr\_across\_sizes.L) ~  
 Drain, weights = varIdent(form = ~1 | Drain),  
 random = ~1|Site, data = cassie)   
summary(lme.b)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 62.47006 80.68563 -18.23503  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.178673e-05 0.2919083  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.3503739 1.5372922 0.8370998 1.3188235 1.8531365   
## Fixed effects: log(count\_corr\_across\_sizes.L) ~ Drain   
## Value Std.Error DF t-value p-value  
## (Intercept) 2.4957541 0.1191711 24 20.942614 0.0000  
## DrainC -0.0047824 0.2002468 6 -0.023883 0.9817  
## DrainH 0.4069784 0.2185504 6 1.862172 0.1119  
## DrainK -0.2983593 0.1554137 6 -1.919774 0.1033  
## DrainSB 0.0000254 0.1972379 6 0.000129 0.9999  
## DrainSC 0.2402400 0.2509426 6 0.957351 0.3754  
## Correlation:   
## (Intr) DrainC DrainH DrainK DranSB  
## DrainC -0.595   
## DrainH -0.545 0.325   
## DrainK -0.767 0.456 0.418   
## DrainSB -0.604 0.360 0.329 0.463   
## DrainSC -0.475 0.283 0.259 0.364 0.287  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.82873936 -0.70641681 0.06528532 0.63108153 1.93366013   
##   
## Number of Observations: 36  
## Number of Groups: 12

anova(lme.b,lme.a)

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.b 1 13 62.47006 80.68563 -18.23503   
## lme.a 2 8 55.27822 66.48780 -19.63911 1 vs 2 2.80816 0.7295

*Allowing for heterogeneous variance is not a statistically significant improvement*

So model b is over fitting, and then with over fitting we get a difference…

… but just to see what happens

Anova(lme.b)

## Analysis of Deviance Table (Type II tests)  
##   
## Response: log(count\_corr\_across\_sizes.L)  
## Chisq Df Pr(>Chisq)   
## Drain 14.492 5 0.01277 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Anova(lme.a)

## Analysis of Deviance Table (Type II tests)  
##   
## Response: log(count\_corr\_across\_sizes.L)  
## Chisq Df Pr(>Chisq)  
## Drain 6.9876 5 0.2216

Tukey.b <- glht(lme.b, linfct = mcp(Drain = "Tukey"))  
summary(Tukey.b)

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: lme.formula(fixed = log(count\_corr\_across\_sizes.L) ~ Drain, data = cassie,   
## random = ~1 | Site, weights = varIdent(form = ~1 | Drain))  
##   
## Linear Hypotheses:  
## Estimate Std. Error z value Pr(>|z|)   
## C - B == 0 -0.0047824 0.2002468 -0.024 1.00000   
## H - B == 0 0.4069784 0.2185504 1.862 0.41451   
## K - B == 0 -0.2983593 0.1554137 -1.920 0.37896   
## SB - B == 0 0.0000254 0.1972379 0.000 1.00000   
## SC - B == 0 0.2402400 0.2509426 0.957 0.92795   
## H - C == 0 0.4117608 0.2438433 1.689 0.52807   
## K - C == 0 -0.2935769 0.1893375 -1.551 0.62073   
## SB - C == 0 0.0048078 0.2249401 0.021 1.00000   
## SC - C == 0 0.2450224 0.2732535 0.897 0.94481   
## K - H == 0 -0.7053377 0.2086006 -3.381 0.00905 \*\*  
## SB - H == 0 -0.4069530 0.2413785 -1.686 0.52979   
## SC - H == 0 -0.1667384 0.2869372 -0.581 0.99185   
## SB - K == 0 0.2983847 0.1861524 1.603 0.58556   
## SC - K == 0 0.5385993 0.2423264 2.223 0.21891   
## SC - SB == 0 0.2402146 0.2710562 0.886 0.94746   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

Tukey.a <- glht(lme.a, linfct = mcp(Drain = "Tukey"))  
summary(Tukey.a)

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: lme.formula(fixed = log(count\_corr\_across\_sizes.L) ~ Drain, data = cassie,   
## random = ~1 | Site)  
##   
## Linear Hypotheses:  
## Estimate Std. Error z value Pr(>|z|)  
## C - B == 0 -0.0047824 0.2894192 -0.017 1.000  
## H - B == 0 0.4069784 0.2894192 1.406 0.723  
## K - B == 0 -0.2983593 0.2894192 -1.031 0.908  
## SB - B == 0 0.0000254 0.2894192 0.000 1.000  
## SC - B == 0 0.2402400 0.2894192 0.830 0.962  
## H - C == 0 0.4117608 0.2894192 1.423 0.713  
## K - C == 0 -0.2935769 0.2894192 -1.014 0.913  
## SB - C == 0 0.0048078 0.2894192 0.017 1.000  
## SC - C == 0 0.2450224 0.2894192 0.847 0.959  
## K - H == 0 -0.7053377 0.2894192 -2.437 0.144  
## SB - H == 0 -0.4069530 0.2894192 -1.406 0.723  
## SC - H == 0 -0.1667384 0.2894192 -0.576 0.993  
## SB - K == 0 0.2983847 0.2894192 1.031 0.908  
## SC - K == 0 0.5385993 0.2894192 1.861 0.426  
## SC - SB == 0 0.2402146 0.2894192 0.830 0.962  
## (Adjusted p values reported -- single-step method)

# slightly different variance structure

lme.c <- lme(log(count\_corr\_across\_sizes.L) ~  
 Drain, weights = varIdent(form = ~1 | Site),  
 random = ~1|Site, data = cassie)   
summary(lme.c)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 66.48044 93.10319 -14.24022  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.111808e-05 0.2595235  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.2366108 1.5076971 1.5299736 2.3690862 0.6601576 0.8676632 1.0072926   
## SB1 SB2 SC1 SC2   
## 0.4447457 1.9163639 3.2887908 0.8805611   
## Fixed effects: log(count\_corr\_across\_sizes.L) ~ Drain   
## Value Std.Error DF t-value p-value  
## (Intercept) 2.5200699 0.1165083 24 21.629955 0.0000  
## DrainC -0.0269438 0.1986592 6 -0.135629 0.8966  
## DrainH 0.2189722 0.1505106 6 1.454862 0.1959  
## DrainK -0.3148704 0.1525676 6 -2.063809 0.0846  
## DrainSB 0.0268603 0.1333716 6 0.201395 0.8470  
## DrainSC -0.1449958 0.1726783 6 -0.839687 0.4333  
## Correlation:   
## (Intr) DrainC DrainH DrainK DranSB  
## DrainC -0.586   
## DrainH -0.774 0.454   
## DrainK -0.764 0.448 0.591   
## DrainSB -0.874 0.512 0.676 0.667   
## DrainSC -0.675 0.396 0.522 0.515 0.589  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5263439 -0.7765503 0.2762470 0.7298214 1.6775555   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.c,lme.a)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.c 1 19 66.48044 93.10319 -14.24022   
## lme.a 2 8 55.27822 66.48780 -19.63911 1 vs 2 10.79778 0.4604

Tukey.c <- glht(lme.c, linfct = mcp(Drain = "Tukey"))  
summary(Tukey.c)

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: lme.formula(fixed = log(count\_corr\_across\_sizes.L) ~ Drain, data = cassie,   
## random = ~1 | Site, weights = varIdent(form = ~1 | Site))  
##   
## Linear Hypotheses:  
## Estimate Std. Error z value Pr(>|z|)   
## C - B == 0 -0.02694 0.19866 -0.136 0.99999   
## H - B == 0 0.21897 0.15051 1.455 0.68233   
## K - B == 0 -0.31487 0.15257 -2.064 0.29579   
## SB - B == 0 0.02686 0.13337 0.201 0.99995   
## SC - B == 0 -0.14500 0.17268 -0.840 0.95785   
## H - C == 0 0.24592 0.18700 1.315 0.76777   
## K - C == 0 -0.28793 0.18866 -1.526 0.63574   
## SB - C == 0 0.05380 0.17351 0.310 0.99959   
## SC - C == 0 -0.11805 0.20527 -0.575 0.99218   
## K - H == 0 -0.53384 0.13705 -3.895 0.00135 \*\*  
## SB - H == 0 -0.19211 0.11530 -1.666 0.54196   
## SC - H == 0 -0.36397 0.15913 -2.287 0.19084   
## SB - K == 0 0.34173 0.11797 2.897 0.04144 \*   
## SC - K == 0 0.16987 0.16108 1.055 0.89411   
## SC - SB == 0 -0.17186 0.14303 -1.202 0.82905   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

Maybe there is something about drain K, but also this is over fitting

**Lets not chase p-values! that would be bad science**

There is no additional variance structure assumption needed. *No different across drains*.

*Now explore regression model options for covariates that might make sense*.

# REGRESSION on CATCHMENT POPULATION

## MP on population, ungrouped

ctrl <- lmeControl(opt='optim') # less flakey search strategy  
ctrl <- lmeControl(opt='optim', # default is 50: it will  
 maxIter=1000) # run for a while if you crank it up  
  
cassie$pop <- cassie$catchment\_pop/1e3  
  
lme.base <- lme(count\_corr\_across\_sizes.L ~ pop,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 251.1685 257.274 -121.5843  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 3.671485 6.473634  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ pop   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.345930 2.0940005 24 5.895858 0.0000  
## pop 0.108102 0.0852063 10 1.268711 0.2333  
## Correlation:   
## (Intr)  
## pop -0.692  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.1618705 -0.5420506 -0.1010006 0.2026473 3.4414169   
##   
## Number of Observations: 36  
## Number of Groups: 12

There is no overall effect but this may be because the variation differs between Drain or Site groups:

## MP on Population varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ pop,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 240.8591 254.5963 -111.4295  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.01735427 3.439066  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.2565910 3.3499556 0.6216689 1.3668834 3.5623551   
## Fixed effects: count\_corr\_across\_sizes.L ~ pop   
## Value Std.Error DF t-value p-value  
## (Intercept) 9.761266 0.7920813 24 12.323566 0.0000  
## pop 0.139552 0.0502560 10 2.776814 0.0196  
## Correlation:   
## (Intr)  
## pop -0.602  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6864565 -0.2893083 0.2163576 0.6519775 2.3630871   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 251.1685 257.2740 -121.5843   
## lme.base.het1 2 9 240.8590 254.5963 -111.4295 1 vs 2 20.30947 0.0011

The anova shows a low p-value, meaning the heterogeneous model *population by Drain* has a better fit.

## MP on Population varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ pop,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 235.5769 258.4723 -102.7885  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.02188431 3.299638  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0426062 1.3705683 1.1434073 5.0089474 0.7449622 0.5998639 0.7658523   
## SB1 SB2 SC1 SC2   
## 0.8047672 1.7429086 5.0807031 0.5783662   
## Fixed effects: count\_corr\_across\_sizes.L ~ pop   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.265677 0.6141013 24 16.716585 0.000  
## pop 0.104089 0.0290939 10 3.577683 0.005  
## Correlation:   
## (Intr)  
## pop -0.578  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6488061 -0.6680410 0.2307581 0.9226319 1.7153029   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 251.1685 257.2740 -121.5843   
## lme.base.het1 2 9 240.8590 254.5963 -111.4295 1 vs 2 20.30947 0.0011  
## lme.base.het2 3 15 235.5769 258.4723 -102.7885 2 vs 3 17.28212 0.0083

Heterogeneous 2 model (by Site) is a stat better fit that heterogeneous 1 model (by Drain): variation at site level catchment **population** is statistically significant. *This can be treated as an empirical question*.

# REGRESSION on CATCHMENT AREA

## MP on area, ungrouped

lme.base <- lme(count\_corr\_across\_sizes.L ~ catchment\_area,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 256.0241 262.1295 -124.012  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 2.6728 6.473814  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ catchment\_area   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.433184 2.1263957 24 4.906511 0.0001  
## catchment\_area 0.004009 0.0017767 10 2.256481 0.0477  
## Correlation:   
## (Intr)  
## catchment\_area -0.782  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.18805490 -0.58884398 -0.08986117 0.20779485 3.41510500   
##   
## Number of Observations: 36  
## Number of Groups: 12

There does seem to be an *effect of catchment area (p < 0.05)*… …so we test if there’s a better fit if we separate the effect of Drain:

## MP on area varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ catchment\_area,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 246.1358 259.873 -114.0679  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.08954967 3.537553  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.1915889 3.1674671 0.5903148 1.3159130 2.9931068   
## Fixed effects: count\_corr\_across\_sizes.L ~ catchment\_area   
## Value Std.Error DF t-value p-value  
## (Intercept) 9.471062 0.7992652 24 11.849712 0.0000  
## catchment\_area 0.003506 0.0010832 10 3.236379 0.0089  
## Correlation:   
## (Intr)  
## catchment\_area -0.624  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5997073 -0.3683373 0.1275379 0.6341354 2.3311896   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 256.0241 262.1295 -124.0120   
## lme.base.het1 2 9 246.1358 259.8730 -114.0679 1 vs 2 19.88832 0.0013

The anova shows a low p-value, meaning the heterogeneous model *catchment area by Drain* has a better fit.

## MP on area varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ catchment\_area,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 242.6997 265.5951 -106.3499  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.0154414 3.32745  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0279426 1.3923926 1.0453394 4.9332013 0.7415092 0.5707430 0.6744615   
## SB1 SB2 SC1 SC2   
## 0.8988398 1.7706742 4.6180756 0.6270381   
## Fixed effects: count\_corr\_across\_sizes.L ~ catchment\_area   
## Value Std.Error DF t-value p-value  
## (Intercept) 9.669730 0.6828195 24 14.161473 0.0000  
## catchment\_area 0.002502 0.0006639 10 3.768067 0.0037  
## Correlation:   
## (Intr)  
## catchment\_area -0.678  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5945441 -0.6302912 0.3850650 0.8557732 1.7154666   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 256.0241 262.1295 -124.0120   
## lme.base.het1 2 9 246.1358 259.8730 -114.0679 1 vs 2 19.88832 0.0013  
## lme.base.het2 3 15 242.6997 265.5951 -106.3499 2 vs 3 15.43605 0.0171

Heterogeneous 2 model (by Site) is a stat better fit that heterogeneous 1 model (by Drain): variation at site level catchment **area** is statistically significant.

# REGRESSION on RAINFALL

## MP on rainfall, ungrouped

lme.base <- lme(count\_corr\_across\_sizes.L ~  
 rainfall, control = ctrl,  
 random = ~1|Site, data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 252.162 258.2674 -122.081  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 4.169632 6.473319  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ rainfall   
## Value Std.Error DF t-value p-value  
## (Intercept) 13.041184 3.234771 24 4.031563 0.0005  
## rainfall 0.043873 0.107629 10 0.407626 0.6921  
## Correlation:   
## (Intr)  
## rainfall -0.866  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.07843925 -0.51701085 -0.07600452 0.12485520 3.52507235   
##   
## Number of Observations: 36  
## Number of Groups: 12

So overall preceding rainfall has no effect, but what if we split the data by Drain …

## MP on rainfall varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ rainfall,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 245.168 258.9053 -113.584  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.741964 3.437067  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.2922985 3.8332450 0.7143689 1.2479137 3.1293086   
## Fixed effects: count\_corr\_across\_sizes.L ~ rainfall   
## Value Std.Error DF t-value p-value  
## (Intercept) 11.414699 2.2298876 24 5.118957 0.0000  
## rainfall 0.018692 0.0706251 10 0.264660 0.7966  
## Correlation:   
## (Intr)  
## rainfall -0.912  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5714636 -0.3151768 0.2211803 0.4879263 2.3159502   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nANOVA to compare models\n"); anova(lme.base,lme.base.het1)

##   
## ANOVA to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 252.1619 258.2674 -122.081   
## lme.base.het1 2 9 245.1680 258.9053 -113.584 1 vs 2 16.99391 0.0045

AIC and low p-value mean that the heterogeneous model (by Drain) is a better fit, but **no significant effect of rainfall**.

## MP on rainfall varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ rainfall,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 240.7509 263.6463 -105.3755  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.598637 3.441973  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.9976691 1.5944115 1.0466855 5.4166708 0.8968568 0.6493501 0.7847766   
## SB1 SB2 SC1 SC2   
## 0.4187276 1.6405705 4.3748456 0.6362196   
## Fixed effects: count\_corr\_across\_sizes.L ~ rainfall   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.164674 1.7089203 24 7.118339 0.0000  
## rainfall -0.010945 0.0608694 10 -0.179812 0.8609  
## Correlation:   
## (Intr)  
## rainfall -0.899  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5007813 -0.5863399 0.2399159 0.6723020 1.6860357   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nANOVA to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## ANOVA to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 252.1619 258.2674 -122.0810   
## lme.base.het1 2 9 245.1680 258.9053 -113.5840 1 vs 2 16.99391 0.0045  
## lme.base.het2 3 15 240.7509 263.6463 -105.3755 2 vs 3 16.41714 0.0117

Based on AIC and p-values, heterogeneous 2 model (by Site) is a statistically better fit that heterogeneous 1 model (by Drain): *BUT* variation from **rainfall is not statistically significant**.

The variance decomposition (*at the beginning*) possibly means that, at site level, variance makes sense; this can be treated as an empirical question

Exploration of log(Y) is something you can look at. [**Note**: anova() can’t be used to compare models explaining log(Y) with models explaining Y]

lme.base.het2.log <- lme(log(count\_corr\_across\_sizes.L) ~  
 rainfall,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.log)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 67.57817 90.47358 -18.78908  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.1372887 0.2453874  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.2062018 1.5554105 1.5188723 3.0870272 0.7672516 1.0647548 1.4613090   
## SB1 SB2 SC1 SC2   
## 0.4623805 2.0705056 2.8099255 0.9110542   
## Fixed effects: log(count\_corr\_across\_sizes.L) ~ rainfall   
## Value Std.Error DF t-value p-value  
## (Intercept) 2.4976911 0.1379054 24 18.11163 0.0000  
## rainfall -0.0001922 0.0048076 10 -0.03998 0.9689  
## Correlation:   
## (Intr)  
## rainfall -0.884  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5647038 -0.7097299 0.1740455 0.5970721 1.6198760   
##   
## Number of Observations: 36  
## Number of Groups: 12

So using log(rainfall) still doesn’t let us see a significant effect of rainfall

### Looking for nonlinear effect with quadratic of rainfall

lme.base.het2.quad <- lme(count\_corr\_across\_sizes.L ~ rainfall + I(rainfall^2),  
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)  
summary(lme.base.het2.quad)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 249.8983 273.8424 -108.9491  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.458522 3.759575  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.9042141 1.4297905 0.9730885 4.8020906 0.7175661 0.5653016 0.6644362   
## SB1 SB2 SC1 SC2   
## 0.4028223 1.5174035 4.0387003 0.5814978   
## Fixed effects: count\_corr\_across\_sizes.L ~ rainfall + I(rainfall^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 15.171998 2.3770141 24 6.382797 0.0000  
## rainfall -0.302273 0.1837482 9 -1.645039 0.1344  
## I(rainfall^2) 0.005471 0.0033708 9 1.623005 0.1390  
## Correlation:   
## (Intr) ranfll  
## rainfall -0.889   
## I(rainfall^2) 0.735 -0.950  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.4877196 -0.4881392 0.1636891 0.6390189 1.6942094   
##   
## Number of Observations: 36  
## Number of Groups: 12

nothing here…

let’s scale up..

cassie$rain <-cassie$rainfall/1000  
  
lme.base.het2.quad2 <- lme(count\_corr\_across\_sizes.L ~ rain + I(rain^2),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.quad2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 208.4518 232.3959 -88.22588  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.458522 3.759575  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.9042141 1.4297905 0.9730885 4.8020906 0.7175661 0.5653016 0.6644362   
## SB1 SB2 SC1 SC2   
## 0.4028223 1.5174035 4.0387003 0.5814978   
## Fixed effects: count\_corr\_across\_sizes.L ~ rain + I(rain^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 15.172 2.377 24 6.382797 0.0000  
## rain -302.273 183.748 9 -1.645039 0.1344  
## I(rain^2) 5470.903 3370.847 9 1.623005 0.1390  
## Correlation:   
## (Intr) rain   
## rain -0.889   
## I(rain^2) 0.735 -0.950  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.4877196 -0.4881392 0.1636891 0.6390189 1.6942094   
##   
## Number of Observations: 36  
## Number of Groups: 12

Still no effect of rainfall

### maybe impose cubic

lme.base.het2.cub <-   
 lme(count\_corr\_across\_sizes.L ~ I(rainfall^3),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.cub)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 256.3052 279.2006 -113.1526  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.203787 3.478035  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.9681817 1.4959313 0.9938272 5.1701073 1.1090477 0.6831958 0.8771018   
## SB1 SB2 SC1 SC2   
## 0.4141936 1.6143629 4.4974928 0.6079855   
## Fixed effects: count\_corr\_across\_sizes.L ~ I(rainfall^3)   
## Value Std.Error DF t-value p-value  
## (Intercept) 11.373979 0.8479475 24 13.413542 0.0000  
## I(rainfall^3) 0.000015 0.0000186 10 0.808532 0.4376  
## Correlation:   
## (Intr)  
## I(rainfall^3) -0.616  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5591444 -0.5885520 0.2277523 0.7553822 1.6976295   
##   
## Number of Observations: 36  
## Number of Groups: 12

Still no effect of rainfall

summary(lme.base.het2.quad2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 208.4518 232.3959 -88.22588  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.458522 3.759575  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.9042141 1.4297905 0.9730885 4.8020906 0.7175661 0.5653016 0.6644362   
## SB1 SB2 SC1 SC2   
## 0.4028223 1.5174035 4.0387003 0.5814978   
## Fixed effects: count\_corr\_across\_sizes.L ~ rain + I(rain^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 15.172 2.377 24 6.382797 0.0000  
## rain -302.273 183.748 9 -1.645039 0.1344  
## I(rain^2) 5470.903 3370.847 9 1.623005 0.1390  
## Correlation:   
## (Intr) rain   
## rain -0.889   
## I(rain^2) 0.735 -0.950  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.4877196 -0.4881392 0.1636891 0.6390189 1.6942094   
##   
## Number of Observations: 36  
## Number of Groups: 12

### log x would capture this, approximately

lme.base.het2.xlog <- lme(count\_corr\_across\_sizes.L ~ log(rainfall),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.xlog)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 234.3702 257.2656 -102.1851  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.555974 3.534343  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.9659246 1.5793911 1.0340011 5.3615923 0.7950427 0.6186110 0.7378872   
## SB1 SB2 SC1 SC2   
## 0.4130589 1.6103315 4.2106896 0.6293135   
## Fixed effects: count\_corr\_across\_sizes.L ~ log(rainfall)   
## Value Std.Error DF t-value p-value  
## (Intercept) 14.617897 3.625619 24 4.031835 0.0005  
## log(rainfall) -0.891725 1.160911 10 -0.768126 0.4602  
## Correlation:   
## (Intr)  
## log(rainfall) -0.979  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.4971250 -0.5046830 0.2673270 0.6568381 1.6831051   
##   
## Number of Observations: 36  
## Number of Groups: 12

**None of the options show a significant effect of prior rainfall**

# REGRESSION on RESIDENTIAL PROPORTION

## MP on residential, ungrouped

lme.base <- lme(count\_corr\_across\_sizes.L ~ residential,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 252.1212 258.2267 -122.0606  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 3.720099 6.473638  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ residential   
## Value Std.Error DF t-value p-value  
## (Intercept) 18.241483 3.687885 24 4.946327 0.0000  
## residential -0.068673 0.056842 10 -1.208129 0.2548  
## Correlation:   
## (Intr)  
## residential -0.911  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.11055708 -0.57777453 -0.05994797 0.16251311 3.72669815   
##   
## Number of Observations: 36  
## Number of Groups: 12

No ungrouped effect of proportion of residential land in catchments…

## MP on Residential varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ residential,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 240.3185 254.0557 -111.1592  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.009755506 3.460297  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.3892895 3.9104260 0.6208482 1.1563355 2.9043926   
## Fixed effects: count\_corr\_across\_sizes.L ~ residential   
## Value Std.Error DF t-value p-value  
## (Intercept) 19.69285 3.0589514 24 6.437778 0.0000  
## residential -0.11677 0.0407899 10 -2.862728 0.0169  
## Correlation:   
## (Intr)  
## residential -0.979  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.7034472 -0.7032183 0.2350699 0.5524045 2.3432498   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 252.1212 258.2267 -122.0606   
## lme.base.het1 2 9 240.3185 254.0557 -111.1593 1 vs 2 21.80272 6e-04

Lower AIC and p<0.05 mean heterogeneous model (by Drain) is a better fit…

## MP on residential varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~  
 residential, weights = varIdent(form = ~1 | Site), control = ctrl,  
 random = ~1|Site, data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 242.1001 264.9955 -106.0501  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.760509 3.382191  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0240115 1.5195236 1.0543031 5.4511786 0.8766809 0.6493518 0.7635929   
## SB1 SB2 SC1 SC2   
## 0.4244132 1.6582927 4.3181188 0.7145900   
## Fixed effects: count\_corr\_across\_sizes.L ~ residential   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.634262 2.1099403 24 5.987972 0.0000  
## residential -0.010935 0.0315138 10 -0.346982 0.7358  
## Correlation:   
## (Intr)  
## residential -0.929  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.4704253 -0.6388693 0.2718548 0.6309142 1.6902983   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 252.1212 258.2267 -122.0606   
## lme.base.het1 2 9 240.3185 254.0557 -111.1593 1 vs 2 21.80272 0.0006  
## lme.base.het2 3 15 242.1001 264.9955 -106.0501 2 vs 3 10.21836 0.1158

So heterogeneous 1 model (by Drain) is a statistically better fit than heterogeneous 2 model (by Site) = variation at *catchment* level.

**Residential proportion** is statistically significant - *negative* relationship.

# REGRESSION on AGRICULTURAL PROPORTION

## Ungrouped model – MP by agricultural

lme.base <- lme(count\_corr\_across\_sizes.L ~ agricultural,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 253.1385 259.244 -122.5693  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 3.894599 6.473644  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ agricultural   
## Value Std.Error DF t-value p-value  
## (Intercept) 13.508807 1.7069621 24 7.913947 0.0000  
## agricultural 0.042768 0.0441838 10 0.967949 0.3559  
## Correlation:   
## (Intr)  
## agricultural -0.408  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.09754580 -0.53278813 -0.09646189 0.16974403 3.68796147   
##   
## Number of Observations: 36  
## Number of Groups: 12

So in an ungrouped model the proportion of agricultural land is not significant.

## MP on agric and Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ agricultural,  
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 244.2358 257.9731 -113.1179  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.403883 3.468708  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.2772536 3.8861718 0.7142936 1.2085186 2.9011696   
## Fixed effects: count\_corr\_across\_sizes.L ~ agricultural   
## Value Std.Error DF t-value p-value  
## (Intercept) 11.60651 0.8592022 24 13.508480 0.0000  
## agricultural 0.06287 0.0455768 10 1.379422 0.1978  
## Correlation:   
## (Intr)  
## agricultural -0.199  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6197616 -0.5687206 0.1753759 0.4846682 2.3271010   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 253.1386 259.2440 -122.5693   
## lme.base.het1 2 9 244.2358 257.9731 -113.1179 1 vs 2 18.90273 0.002

Low AIC and p-value say heterogeneous model a better fit, but agricultural proportion still shows no significant effect.

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ agricultural,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 242.479 265.3744 -106.2395  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.533351 3.373494  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0217319 1.4956852 1.0567440 5.4332531 0.9002261 0.6880135 0.8589729   
## SB1 SB2 SC1 SC2   
## 0.4231467 1.6470524 4.7545886 0.6594777   
## Fixed effects: count\_corr\_across\_sizes.L ~ agricultural   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.092197 0.7898942 24 15.308630 0.0000  
## agricultural -0.011971 0.0223589 10 -0.535403 0.6041  
## Correlation:   
## (Intr)  
## agricultural -0.374  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5152815 -0.6693329 0.2145542 0.7520637 1.6935414   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 253.1386 259.2440 -122.5693   
## lme.base.het1 2 9 244.2358 257.9731 -113.1179 1 vs 2 18.90273 0.0020  
## lme.base.het2 3 15 242.4790 265.3744 -106.2395 2 vs 3 13.75679 0.0325

heterogeneous 2 model is a stat better fit that heterogeneous 1 model: variation at site level

**Agricultural proportion is still not stat significant**.

## exploration of log(agricultural)

lme.base.het2.log <- lme(log(count\_corr\_across\_sizes.L) ~ agricultural,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.log)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 68.89105 91.78646 -19.44552  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.002195614 0.2285678  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.3683167 1.4996268 1.6968824 3.2450166 0.9815759 1.5517320 2.0253918   
## SB1 SB2 SC1 SC2   
## 0.4616302 2.1607567 3.7374164 1.0001570   
## Fixed effects: log(count\_corr\_across\_sizes.L) ~ agricultural   
## Value Std.Error DF t-value p-value  
## (Intercept) 2.5414716 0.04394433 24 57.83389 0.0000  
## agricultural -0.0017601 0.00142916 10 -1.23159 0.2463  
## Correlation:   
## (Intr)  
## agricultural -0.326  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.68719456 -0.88716461 0.08546522 0.66418229 1.65695389   
##   
## Number of Observations: 36  
## Number of Groups: 12

### Looking for nonlinear effect

lme.base.het2.quad <-   
 lme(count\_corr\_across\_sizes.L ~ agricultural + I(agricultural^2),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.quad)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 243.3496 267.2938 -105.6748  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.461638 3.37525  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0186957 1.4950348 1.0541964 5.4332764 0.9141990 0.6971729 0.8802878   
## SB1 SB2 SC1 SC2   
## 0.4224549 1.6425623 3.0701626 0.6553970   
## Fixed effects: count\_corr\_across\_sizes.L ~ agricultural + I(agricultural^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.095601 0.774818 24 15.610900 0.0000  
## agricultural -24.050682 10.780799 9 -2.230881 0.0526  
## I(agricultural^2) 0.254756 0.114218 9 2.230441 0.0527  
## Correlation:   
## (Intr) agrclt  
## agricultural -0.002   
## I(agricultural^2) 0.001 -1.000  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5266857 -0.7031061 0.1001569 0.7781843 1.6937885   
##   
## Number of Observations: 36  
## Number of Groups: 12

maybe something here: hard to see

### so lets scale so we can see what is happening

cassie$ag <-cassie$agricultural/1000  
  
lme.base.het2.quad2 <-   
 lme(count\_corr\_across\_sizes.L ~ ag + I(ag^2),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.quad2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 201.9031 225.8472 -84.95155  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.461638 3.37525  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0186957 1.4950348 1.0541964 5.4332764 0.9141990 0.6971729 0.8802878   
## SB1 SB2 SC1 SC2   
## 0.4224549 1.6425623 3.0701626 0.6553970   
## Fixed effects: count\_corr\_across\_sizes.L ~ ag + I(ag^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.10 0.77 24 15.610900 0.0000  
## ag -24050.68 10780.80 9 -2.230881 0.0526  
## I(ag^2) 254756.28 114217.89 9 2.230441 0.0527  
## Correlation:   
## (Intr) ag   
## ag -0.002   
## I(ag^2) 0.001 -1.000  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5266857 -0.7031061 0.1001569 0.7781843 1.6937885   
##   
## Number of Observations: 36  
## Number of Groups: 12

### maybe impose cubic

lme.base.het2.cub <-   
 lme(count\_corr\_across\_sizes.L ~ I(agricultural^3),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.cub)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 260.672 283.5674 -115.336  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.534462 3.373712  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0218348 1.4953444 1.0565702 5.4330091 0.8999646 0.6876309 0.8586613   
## SB1 SB2 SC1 SC2   
## 0.4231467 1.6472348 4.7567442 0.6594997   
## Fixed effects: count\_corr\_across\_sizes.L ~ I(agricultural^3)   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.090867 0.7901254 24 15.302466 0.000  
## I(agricultural^3) -0.000001 0.0000025 10 -0.530953 0.607  
## Correlation:   
## (Intr)  
## I(agricultural^3) -0.374  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5151542 -0.6689895 0.2151243 0.7520106 1.6935589   
##   
## Number of Observations: 36  
## Number of Groups: 12

Not cubic so look closely at the quadratic maybe runs the other way

summary(lme.base.het2.quad2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 201.9031 225.8472 -84.95155  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.461638 3.37525  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0186957 1.4950348 1.0541964 5.4332764 0.9141990 0.6971729 0.8802878   
## SB1 SB2 SC1 SC2   
## 0.4224549 1.6425623 3.0701626 0.6553970   
## Fixed effects: count\_corr\_across\_sizes.L ~ ag + I(ag^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 12.10 0.77 24 15.610900 0.0000  
## ag -24050.68 10780.80 9 -2.230881 0.0526  
## I(ag^2) 254756.28 114217.89 9 2.230441 0.0527  
## Correlation:   
## (Intr) ag   
## ag -0.002   
## I(ag^2) 0.001 -1.000  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5266857 -0.7031061 0.1001569 0.7781843 1.6937885   
##   
## Number of Observations: 36  
## Number of Groups: 12

**None of the options show a significant effect of the proportion of agricultural land in each catchment**

# REGRESSION on INDUSTRIAL PROPORTION

## MP on proportion industrial – ungrouped

lme.base <- lme(count\_corr\_across\_sizes.L ~ industrial,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 251.788 257.8934 -121.894  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 4.229718 6.473407  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ industrial   
## Value Std.Error DF t-value p-value  
## (Intercept) 14.083799 2.2170519 24 6.352490 0.0000  
## industrial 0.009383 0.1417252 10 0.066203 0.9485  
## Correlation:   
## (Intr)  
## industrial -0.678  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.1419595 -0.5455609 -0.0491773 0.1671380 3.5677407   
##   
## Number of Observations: 36  
## Number of Groups: 12

So, no ungrouped effect of industrial land proportion.

## MP on Proportion industrial varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ industrial,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 243.92 257.6573 -112.96  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.04404927 3.638908  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.3677138 3.7455951 0.6419287 1.1088647 3.2719610   
## Fixed effects: count\_corr\_across\_sizes.L ~ industrial   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.290192 0.8336058 24 12.344194 0.0000  
## industrial 0.096459 0.0530853 10 1.817064 0.0993  
## Correlation:   
## (Intr)  
## industrial -0.597  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.7503038 -0.2187003 0.2182537 0.5739280 2.3375295   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 251.788 257.8934 -121.894   
## lme.base.het1 2 9 243.920 257.6573 -112.960 1 vs 2 17.86795 0.0031

So, no effect of industrial land proportion grouped by Drain, but low AIC and p-value < 0.05 mean heterogeneous model is a better fit.

## MP on Proportion industrial varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ industrial,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,  
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 235.8785 258.774 -102.9393  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.009901528 3.134079  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.4845622 1.9452442 1.0703649 5.8935965 1.1202902 0.6339473 0.8085304   
## SB1 SB2 SC1 SC2   
## 0.4293923 1.7066131 5.3239052 0.6087797   
## Fixed effects: count\_corr\_across\_sizes.L ~ industrial   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.301967 0.6282550 24 16.397748 0.0000  
## industrial 0.135242 0.0432021 10 3.130455 0.0107  
## Correlation:   
## (Intr)  
## industrial -0.692  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6187450 -0.5493009 0.2182551 0.8525389 1.6970734   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 251.7880 257.8934 -121.8940   
## lme.base.het1 2 9 243.9200 257.6573 -112.9600 1 vs 2 17.86795 0.0031  
## lme.base.het2 3 15 235.8785 258.7740 -102.9393 2 vs 3 20.04149 0.0027

So heterogeneous 2 model (by Site) is a statistically better fit than heterogeneous 1 model (by Drain) = variation at *Site* level.

**Industrial proportion** is statistically significant – *positive* relationship.

# REGRESSION on NATURAL PROPORTION

(sum of water environment and natural environment)

## basic prediction of MP using proportion of natural land

lme.base <- lme(count\_corr\_across\_sizes.L ~ natural,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 246.3155 252.421 -119.1578  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 3.746106 6.47364  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ natural   
## Value Std.Error DF t-value p-value  
## (Intercept) 11.999869 2.405831 24 4.987827 0.0000  
## natural 1.264555 1.076421 10 1.174777 0.2673  
## Correlation:   
## (Intr)  
## natural -0.773  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.3568146 -0.5011484 -0.1480583 0.1785412 3.6216179   
##   
## Number of Observations: 36  
## Number of Groups: 12

So ungrouped proportion natural has no effect on MP count.

## Prediction of MP by proportion natural varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ natural,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 238.6173 252.3545 -110.3086  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.032652 3.782558  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.1658541 3.4088269 0.6052465 1.1758657 2.7228157   
## Fixed effects: count\_corr\_across\_sizes.L ~ natural   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.617305 0.9341645 24 11.365563 0.0000  
## natural 0.938852 0.5291607 10 1.774229 0.1064  
## Correlation:   
## (Intr)  
## natural -0.56   
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.80025129 -0.28771247 0.04223199 0.59328072 2.36199759   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 246.3155 252.4210 -119.1578   
## lme.base.het1 2 9 238.6173 252.3545 -110.3086 1 vs 2 17.69825 0.0033

Lower AIC and low p-value say heterogeneous model a better fit

## Prediction of MP by proportion natural, varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ natural,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 234.1207 257.0161 -102.0604  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.02566547 4.864267  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.6946028 1.1740476 0.6723941 3.8519554 0.8290803 0.4040321 0.5019910   
## SB1 SB2 SC1 SC2   
## 0.2901727 1.1021618 3.1418614 0.4976385   
## Fixed effects: count\_corr\_across\_sizes.L ~ natural   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.095044 0.7279272 24 13.868204 0.0000  
## natural 0.709022 0.2688420 10 2.637318 0.0248  
## Correlation:   
## (Intr)  
## natural -0.754  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5975405 -0.3339090 0.2535897 0.7988213 1.6931135   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 246.3155 252.4210 -119.1578   
## lme.base.het1 2 9 238.6173 252.3545 -110.3086 1 vs 2 17.69825 0.0033  
## lme.base.het2 3 15 234.1207 257.0161 -102.0604 2 vs 3 16.49656 0.0113

So heterogeneous 2 model (by Site) is a statistically better fit than heterogeneous 1 model (by Drain) = variation at *Site* level.

**Natural land-use proportion** has statistically significant effect on MP counts – *positive* relationship.

# REGRESSION on SERVICES PROPORTION

## basic prediction of MP using proportion of services land

lme.base <- lme(count\_corr\_across\_sizes.L ~ services,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 248.9492 255.0547 -120.4746  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 3.780377 6.473641  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ services   
## Value Std.Error DF t-value p-value  
## (Intercept) 16.059783 2.2614227 24 7.101628 0.000  
## services -0.342731 0.3033824 10 -1.129698 0.285  
## Correlation:   
## (Intr)  
## services -0.735  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.16902020 -0.54971794 -0.08626217 0.17216998 3.69506840   
##   
## Number of Observations: 36  
## Number of Groups: 12

So ungrouped proportion services has no effect on MP count.

## Prediction of MP by proportion Services varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ services,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 237.3411 251.0783 -109.6705  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.01092281 3.566104  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.2409353 3.7418190 0.5834449 1.1053946 2.8382872   
## Fixed effects: count\_corr\_across\_sizes.L ~ services   
## Value Std.Error DF t-value p-value  
## (Intercept) 14.07972 1.1422051 24 12.326790 0.000  
## services -0.30369 0.0975524 10 -3.113101 0.011  
## Correlation:   
## (Intr)  
## services -0.843  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6814218 -0.4080084 0.1585657 0.5919091 2.3497936   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 248.9492 255.0547 -120.4746   
## lme.base.het1 2 9 237.3411 251.0783 -109.6706 1 vs 2 21.60811 6e-04

There is a *significant effect of service land-use proportion*. Lower AIC and low p-value mean the heterogeneous model (by Drain) is a better fit.

## Prediction of MP by proportion services, varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ services,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,  
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 235.6954 258.5908 -102.8477  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.01309108 3.02982  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.2815774 1.5710276 1.3030561 6.0822055 1.2296194 0.6477070 0.7432028   
## SB1 SB2 SC1 SC2   
## 0.4493206 1.7645235 4.5284285 1.3658902   
## Fixed effects: count\_corr\_across\_sizes.L ~ services   
## Value Std.Error DF t-value p-value  
## (Intercept) 13.964154 0.8250062 24 16.926121 0.0000  
## services -0.276061 0.0859174 10 -3.213097 0.0093  
## Correlation:   
## (Intr)  
## services -0.814  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6589817 -0.8426074 0.1834111 0.6035024 1.6977128   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 248.9492 255.0547 -120.4746   
## lme.base.het1 2 9 237.3411 251.0783 -109.6706 1 vs 2 21.60811 0.0006  
## lme.base.het2 3 15 235.6954 258.5908 -102.8477 2 vs 3 13.64571 0.0339

So based on lower AIC and p<0.05, heterogeneous 2 model (by Site) is a statistically better fit than heterogeneous 1 model (by Drain) = variation at *Site* level.

**Services land-use proportion** has statistically significant effect on MP counts – *positive* relationship.

# REGRESSION on PUBLIC OPEN SPACE PROPORTION

## basic prediction of MP using proportion of public open space land

lme.base <- lme(count\_corr\_across\_sizes.L ~ public\_open,   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 250.4745 256.58 -121.2373  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 4.227121 6.473403  
##   
## Fixed effects: count\_corr\_across\_sizes.L ~ public\_open   
## Value Std.Error DF t-value p-value  
## (Intercept) 13.97123 2.5748591 24 5.426015 0.0000  
## public\_open 0.02896 0.2722765 10 0.106363 0.9174  
## Correlation:   
## (Intr)  
## public\_open -0.774  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.14001161 -0.53651527 -0.06127124 0.15115073 3.56527170   
##   
## Number of Observations: 36  
## Number of Groups: 12

So ungrouped proportion public\_open has no effect on MP count.

## Prediction of MP by proportion public open space, varying by Drain

lme.base.het1 <- lme(count\_corr\_across\_sizes.L ~ public\_open,   
 weights = varIdent(form = ~1 | Drain),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het1)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 240.5486 254.2858 -111.2743  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.08267427 3.468826  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.3907152 3.8080729 0.6189398 1.1566308 3.5319148   
## Fixed effects: count\_corr\_across\_sizes.L ~ public\_open   
## Value Std.Error DF t-value p-value  
## (Intercept) 9.783033 0.8287201 24 11.804990 0.0000  
## public\_open 0.262754 0.1086697 10 2.417912 0.0362  
## Correlation:   
## (Intr)  
## public\_open -0.646  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6667635 -0.2837071 0.2463968 0.5461714 2.3538308   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 250.4745 256.5800 -121.2373   
## lme.base.het1 2 9 240.5486 254.2858 -111.2743 1 vs 2 19.92596 0.0013

There is a *significant effect of public open space land-use proportion*. Lower AIC and low p-value mean the heterogeneous model (by Drain) is a better fit.

## Prediction of MP by proportion services, varying by Site

lme.base.het2 <- lme(count\_corr\_across\_sizes.L ~ public\_open,   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 233.3105 256.2059 -101.6552  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.006887345 3.195163  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.1324685 1.4208211 1.8337814 5.5654093 0.9896338 0.6008445 0.7402829   
## SB1 SB2 SC1 SC2   
## 0.4138627 1.6723836 5.3242994 0.6112170   
## Fixed effects: count\_corr\_across\_sizes.L ~ public\_open   
## Value Std.Error DF t-value p-value  
## (Intercept) 9.988675 0.6583043 24 15.173342 0.0000  
## public\_open 0.319303 0.0894456 10 3.569804 0.0051  
## Correlation:   
## (Intr)  
## public\_open -0.736  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6017441 -0.7756191 0.2778688 0.6979100 1.7065204   
##   
## Number of Observations: 36  
## Number of Groups: 12

cat("\nAnOVa to compare models\n"); anova(lme.base,lme.base.het1,lme.base.het2)

##   
## AnOVa to compare models

## Model df AIC BIC logLik Test L.Ratio p-value  
## lme.base 1 4 250.4745 256.5800 -121.2373   
## lme.base.het1 2 9 240.5486 254.2858 -111.2743 1 vs 2 19.92596 0.0013  
## lme.base.het2 3 15 233.3105 256.2059 -101.6552 2 vs 3 19.23810 0.0038

So based on lower AIC and p<0.05, heterogeneous 2 model (by Site) is a statistically better fit than heterogeneous 1 model (by Drain) = variation at *Site* level.

**Public open space land-use proportion** has statistically significant effect on MP counts – *positive* relationship.

=-=-=-=-=-=-

# What if linear relationship not found

use this section of code if you need to explore other options

=-=-=-=-=-=-

*exploration of log(Y) is something you can look at. Note anova() cant be used to compare log(Y)~ models and Y~ models*

lme.base.het2.log <- lme(log(count\_corr\_across\_sizes.L) ~  
 catchment\_pop, weights = varIdent(form = ~1 | Site), control = ctrl,  
 random = ~1|Site, data = cassie)   
summary(lme.base.het2.log)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 76.94974 99.84515 -23.47487  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.00134332 0.2395254  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.2526241 1.3848470 1.6133277 2.6378447 0.6817708 0.9603527 1.3302524   
## SB1 SB2 SC1 SC2   
## 0.8797571 2.0387166 3.6369106 0.8070128   
## Fixed effects: log(count\_corr\_across\_sizes.L) ~ catchment\_pop   
## Value Std.Error DF t-value p-value  
## (Intercept) 2.3484540 0.06140024 24 38.24829 0.0000  
## catchment\_pop 0.0000078 0.00000226 10 3.44307 0.0063  
## Correlation:   
## (Intr)  
## catchment\_pop -0.688  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6875844 -0.7521101 0.2766510 0.8250737 1.6852995   
##   
## Number of Observations: 36  
## Number of Groups: 12

Looking for nonlinear effect

lme.base.het2.quad <- lme(count\_corr\_across\_sizes.L ~  
 catchment\_pop + I(catchment\_pop^2), weights = varIdent(form = ~1 | Site), control = ctrl,  
 random = ~1|Site, data = cassie)   
summary(lme.base.het2.quad)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 289.3201 313.2642 -128.66  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.009975611 3.360063  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0824718 1.3525714 1.1844786 4.9223153 0.7732357 0.5904193 0.7510431   
## SB1 SB2 SC1 SC2   
## 0.7833858 1.7088157 4.9950080 0.5722494   
## Fixed effects: count\_corr\_across\_sizes.L ~ catchment\_pop + I(catchment\_pop^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.241366 0.6570034 24 15.587995 0.0000  
## catchment\_pop 0.000122 0.0001087 9 1.120804 0.2914  
## I(catchment\_pop^2) 0.000000 0.0000000 9 -0.161754 0.8751  
## Correlation:   
## (Intr) ctchm\_  
## catchment\_pop -0.478   
## I(catchment\_pop^2) 0.345 -0.961  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6168158 -0.7017130 0.2302874 0.8478853 1.7146908   
##   
## Number of Observations: 36  
## Number of Groups: 12

so maybe something here: hard to see

so lets scale so we can see what is happening

cassie$pop <-cassie$catchment\_pop/1000 # scale up by 1E3  
  
lme.base.het2.quad2 <-   
 lme(count\_corr\_across\_sizes.L ~ pop + I(pop^2),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.quad2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 247.8736 271.8177 -107.9368  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.009975601 3.360063  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0824718 1.3525714 1.1844786 4.9223152 0.7732357 0.5904193 0.7510431   
## SB1 SB2 SC1 SC2   
## 0.7833858 1.7088157 4.9950079 0.5722494   
## Fixed effects: count\_corr\_across\_sizes.L ~ pop + I(pop^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.241366 0.6570034 24 15.587996 0.0000  
## pop 0.121825 0.1086946 9 1.120804 0.2914  
## I(pop^2) -0.000379 0.0023429 9 -0.161754 0.8751  
## Correlation:   
## (Intr) pop   
## pop -0.478   
## I(pop^2) 0.345 -0.961  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6168158 -0.7017130 0.2302874 0.8478853 1.7146908   
##   
## Number of Observations: 36  
## Number of Groups: 12

can solve for the turning point (quadratic and see)

#maybe impose cubic

lme.base.het2.cub <-   
 lme(count\_corr\_across\_sizes.L ~ I(catchment\_pop^3),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,  
 data = cassie)   
summary(lme.base.het2.cub)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 293.9065 316.8019 -131.9533  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 1.012912 3.6553  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 0.9052590 1.3235544 0.9407145 4.5409037 0.7215728 0.5962918 0.7481385   
## SB1 SB2 SC1 SC2   
## 0.4482377 1.5391747 4.3518970 0.5627205   
## Fixed effects: count\_corr\_across\_sizes.L ~ I(catchment\_pop^3)   
## Value Std.Error DF t-value p-value  
## (Intercept) 11.15169 0.6715225 24 16.606574 0.0000  
## I(catchment\_pop^3) 0.00000 0.0000000 10 2.384688 0.0383  
## Correlation:   
## (Intr)  
## I(catchment\_pop^3) -0.429  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.5564137 -0.5077142 0.1586697 0.7794884 1.7097909   
##   
## Number of Observations: 36  
## Number of Groups: 12

#Hmmmmmmmm so look closely at the quadratic maybe runs the other way

summary(lme.base.het2.quad2)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 247.8736 271.8177 -107.9368  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.009975601 3.360063  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.0824718 1.3525714 1.1844786 4.9223152 0.7732357 0.5904193 0.7510431   
## SB1 SB2 SC1 SC2   
## 0.7833858 1.7088157 4.9950079 0.5722494   
## Fixed effects: count\_corr\_across\_sizes.L ~ pop + I(pop^2)   
## Value Std.Error DF t-value p-value  
## (Intercept) 10.241366 0.6570034 24 15.587996 0.0000  
## pop 0.121825 0.1086946 9 1.120804 0.2914  
## I(pop^2) -0.000379 0.0023429 9 -0.161754 0.8751  
## Correlation:   
## (Intr) pop   
## pop -0.478   
## I(pop^2) 0.345 -0.961  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6168158 -0.7017130 0.2302874 0.8478853 1.7146908   
##   
## Number of Observations: 36  
## Number of Groups: 12

a log on x would capture this, approximately

lme.base.het2.xlog <-   
 lme(count\_corr\_across\_sizes.L ~ log(catchment\_pop),   
 weights = varIdent(form = ~1 | Site),   
 control = ctrl,  
 random = ~1|Site,   
 data = cassie)   
summary(lme.base.het2.xlog)

## Linear mixed-effects model fit by REML  
## Data: cassie   
## AIC BIC logLik  
## 230.7116 253.607 -100.3558  
##   
## Random effects:  
## Formula: ~1 | Site  
## (Intercept) Residual  
## StdDev: 0.02141546 2.976445  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.3926242 1.4976531 1.4907442 5.6780373 0.8075713 0.6910454 0.9032710   
## SB1 SB2 SC1 SC2   
## 0.4715166 1.7973789 5.8393447 0.7593973   
## Fixed effects: count\_corr\_across\_sizes.L ~ log(catchment\_pop)   
## Value Std.Error DF t-value p-value  
## (Intercept) 6.062041 1.7055564 24 3.554289 0.0016  
## log(catchment\_pop) 0.778437 0.2139992 10 3.637572 0.0046  
## Correlation:   
## (Intr)  
## log(catchment\_pop) -0.963  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -1.6521037 -0.7092796 0.1144710 0.6213612 1.7148103   
##   
## Number of Observations: 36  
## Number of Groups: 12

so maybe there is a population effect … this is showing whether variance is constant - HINT it is not, therefore use heterogeneous

summary(lme.base.het2$modelStruct$varStruct)

## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Site   
## Parameter estimates:  
## B1 B2 C1 C2 H1 H2 K1 K2   
## 1.0000000 1.1324685 1.4208211 1.8337814 5.5654093 0.9896338 0.6008445 0.7402829   
## SB1 SB2 SC1 SC2   
## 0.4138627 1.6723836 5.3242994 0.6112170

summary(lme.base.het1$modelStruct$varStruct)

## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Drain   
## Parameter estimates:  
## B C H K SB SC   
## 1.0000000 1.3907152 3.8080729 0.6189398 1.1566308 3.5319148

*lots to explore*