grouping\_effect.R

bacsi

Sun Oct 28 11:25:21 2018

library(tidyverse)

## -- Attaching packages ---------------------------------------------------------- tidyverse 1.2.1 --

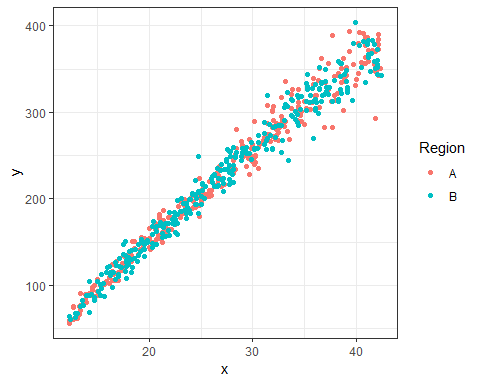
## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts ------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

data(abdom,package="gamlss.data")  
  
str(abdom)

## 'data.frame': 610 obs. of 2 variables:  
## $ y: int 59 64 56 61 74 60 75 63 62 67 ...  
## $ x: num 12.3 12.3 12.3 12.4 12.7 ...

set.seed(123)  
idsplit=caret::createDataPartition(y=abdom$x, p=0.5,list=FALSE)  
regA=abdom[idsplit,]%>%mutate(Region="A")  
regB=abdom[-idsplit,]%>%mutate(Region="B")  
  
df = rbind(regA,regB)  
  
df%>%ggplot(aes(x,y,col=Region))+  
 geom\_point()+theme\_bw()



glm(y~x\*Region,data=df)%>%summary()

##   
## Call:  
## glm(formula = y ~ x \* Region, data = df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -84.672 -8.112 -0.209 8.108 54.262   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -55.39030 2.79111 -19.845 <2e-16 \*\*\*  
## x 10.34549 0.09710 106.550 <2e-16 \*\*\*  
## RegionB 0.43704 4.01211 0.109 0.913   
## x:RegionB -0.01518 0.14060 -0.108 0.914   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 214.7392)  
##   
## Null deviance: 4785899 on 609 degrees of freedom  
## Residual deviance: 130132 on 606 degrees of freedom  
## AIC: 5012.4  
##   
## Number of Fisher Scoring iterations: 2

glm(y~x:Region,data=df)%>%summary()

##   
## Call:  
## glm(formula = y ~ x:Region, data = df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -84.590 -8.105 -0.185 8.058 54.315   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -55.17880 2.00339 -27.54 <2e-16 \*\*\*  
## x:RegionA 10.33847 0.07258 142.45 <2e-16 \*\*\*  
## x:RegionB 10.33793 0.07382 140.05 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 214.3896)  
##   
## Null deviance: 4785899 on 609 degrees of freedom  
## Residual deviance: 130135 on 607 degrees of freedom  
## AIC: 5010.5  
##   
## Number of Fisher Scoring iterations: 2

glm(y~x+x:Region,data=df)%>%summary()

##   
## Call:  
## glm(formula = y ~ x + x:Region, data = df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -84.590 -8.105 -0.185 8.058 54.315   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.518e+01 2.003e+00 -27.543 <2e-16 \*\*\*  
## x 1.034e+01 7.258e-02 142.449 <2e-16 \*\*\*  
## x:RegionB -5.477e-04 4.156e-02 -0.013 0.989   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 214.3896)  
##   
## Null deviance: 4785899 on 609 degrees of freedom  
## Residual deviance: 130135 on 607 degrees of freedom  
## AIC: 5010.5  
##   
## Number of Fisher Scoring iterations: 2

library(broom.mixed)  
  
library(brms)

## Loading required package: Rcpp

## Loading 'brms' package (version 2.4.0). Useful instructions  
## can be found by typing help('brms'). A more detailed introduction  
## to the package is available through vignette('brms\_overview').  
## Run theme\_set(theme\_default()) to use the default bayesplot theme.

rsmod <- brm(y ~ x + (x|Region), data = df, iter = 1500, warmup = 500, chains = 1)

## Compiling the C++ model

## Start sampling

##   
## SAMPLING FOR MODEL 'gaussian brms-model' NOW (CHAIN 1).  
##   
## Gradient evaluation took 0 seconds  
## 1000 transitions using 10 leapfrog steps per transition would take 0 seconds.  
## Adjust your expectations accordingly!  
##   
##   
## Iteration: 1 / 1500 [ 0%] (Warmup)  
## Iteration: 150 / 1500 [ 10%] (Warmup)  
## Iteration: 300 / 1500 [ 20%] (Warmup)  
## Iteration: 450 / 1500 [ 30%] (Warmup)  
## Iteration: 501 / 1500 [ 33%] (Sampling)  
## Iteration: 650 / 1500 [ 43%] (Sampling)  
## Iteration: 800 / 1500 [ 53%] (Sampling)  
## Iteration: 950 / 1500 [ 63%] (Sampling)  
## Iteration: 1100 / 1500 [ 73%] (Sampling)  
## Iteration: 1250 / 1500 [ 83%] (Sampling)  
## Iteration: 1400 / 1500 [ 93%] (Sampling)  
## Iteration: 1500 / 1500 [100%] (Sampling)  
##   
## Elapsed Time: 21.389 seconds (Warm-up)  
## 46.935 seconds (Sampling)  
## 68.324 seconds (Total)

## Warning: There were 79 divergent transitions after warmup. Increasing adapt\_delta above 0.8 may help. See  
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup

## Warning: There were 846 transitions after warmup that exceeded the maximum treedepth. Increase max\_treedepth above 10. See  
## http://mc-stan.org/misc/warnings.html#maximum-treedepth-exceeded

## Warning: Examine the pairs() plot to diagnose sampling problems

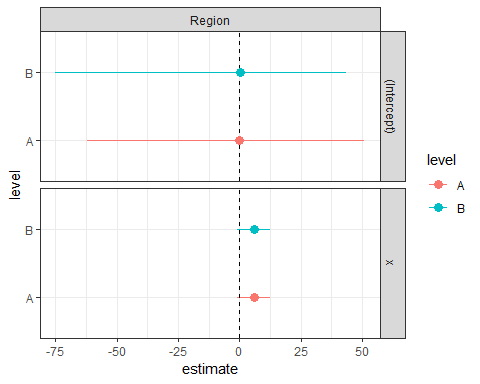
tidy(rsmod,  
 conf.method="HPDinterval",   
 conf.level = 0.975)

## # A tibble: 6 x 8  
## effect component group term estimate std.error conf.low conf.high  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 fixed cond <NA> (Interce~ -5.49e+1 18.4 -1.09e+2 6.23   
## 2 fixed cond <NA> x 4.90e+0 4.57 -2.12e+0 11.2   
## 3 ran\_pa~ cond Region sd\_\_(Int~ 2.47e+1 36.2 7.83e-3 113.   
## 4 ran\_pa~ cond Region sd\_\_x 1.03e+1 9.18 2.39e-3 31.0   
## 5 ran\_pa~ cond Region cor\_\_(In~ 2.71e-3 0.669 -9.93e-1 0.992  
## 6 ran\_pa~ cond Resid~ sd\_\_Obse~ 1.47e+1 0.426 1.39e+1 15.7

tidy(rsmod, effects = "ran\_pars",   
 robust = T,  
 conf.method="HPDinterval",   
 conf.level = 0.975)

## # A tibble: 4 x 8  
## effect component group term estimate std.error conf.low conf.high  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 ran\_pa~ cond Region sd\_\_(Int~ 11.7 14.6 7.83e-3 113.   
## 2 ran\_pa~ cond Region sd\_\_x 9.21 11.8 2.39e-3 31.0   
## 3 ran\_pa~ cond Region cor\_\_(In~ -0.0129 0.964 -9.93e-1 0.992  
## 4 ran\_pa~ cond Resid~ sd\_\_Obse~ 14.7 0.441 1.39e+1 15.7

tidy(rsmod, effects = "ran\_vals",   
 robust = T,  
 conf.method="HPDinterval",   
 conf.level = 0.975) %>%  
 ggplot(aes(estimate,level,xmin=conf.low,xmax=conf.high,col=level))+  
 geom\_errorbarh(height=0)+  
 geom\_vline(xintercept=0,lty=2,col="black")+  
 facet\_grid(term~group,scale="free\_x")+  
 geom\_point(size=3)+  
 theme\_bw()



rsmod\_2 <- brm(y ~ x\*Region, data = df, iter = 1500, warmup = 500, chains = 1)

## Compiling the C++ model  
## Start sampling

##   
## SAMPLING FOR MODEL 'gaussian brms-model' NOW (CHAIN 1).  
##   
## Gradient evaluation took 0 seconds  
## 1000 transitions using 10 leapfrog steps per transition would take 0 seconds.  
## Adjust your expectations accordingly!  
##   
##   
## Iteration: 1 / 1500 [ 0%] (Warmup)  
## Iteration: 150 / 1500 [ 10%] (Warmup)  
## Iteration: 300 / 1500 [ 20%] (Warmup)  
## Iteration: 450 / 1500 [ 30%] (Warmup)  
## Iteration: 501 / 1500 [ 33%] (Sampling)  
## Iteration: 650 / 1500 [ 43%] (Sampling)  
## Iteration: 800 / 1500 [ 53%] (Sampling)  
## Iteration: 950 / 1500 [ 63%] (Sampling)  
## Iteration: 1100 / 1500 [ 73%] (Sampling)  
## Iteration: 1250 / 1500 [ 83%] (Sampling)  
## Iteration: 1400 / 1500 [ 93%] (Sampling)  
## Iteration: 1500 / 1500 [100%] (Sampling)  
##   
## Elapsed Time: 0.296 seconds (Warm-up)  
## 0.356 seconds (Sampling)  
## 0.652 seconds (Total)

tidy(rsmod\_2,  
 conf.method="HPDinterval",   
 conf.level = 0.975)

## # A tibble: 5 x 8  
## effect component group term estimate std.error conf.low conf.high  
## <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 fixed cond <NA> (Interc~ -5.52e+1 2.73 -61.2 -49.4   
## 2 fixed cond <NA> x 1.03e+1 0.0970 10.1 10.6   
## 3 fixed cond <NA> RegionB 2.08e-1 3.87 -8.11 8.40   
## 4 fixed cond <NA> x:Regio~ -7.95e-3 0.135 -0.270 0.302  
## 5 ran\_pa~ cond Residu~ sd\_\_Obs~ 1.47e+1 0.406 13.9 15.6