Análisis de las llamadas de A. pertinax (variable Tcall dur)

José R. Ferrer Paris

2019-11-18

## Exploración de los datos

Leemos el archivo de datos de Aratinga pertinax, versión de 2019, reclassificamos la variable Region para facilitar interpretación, y descartamos grupos con pocas observaciones:

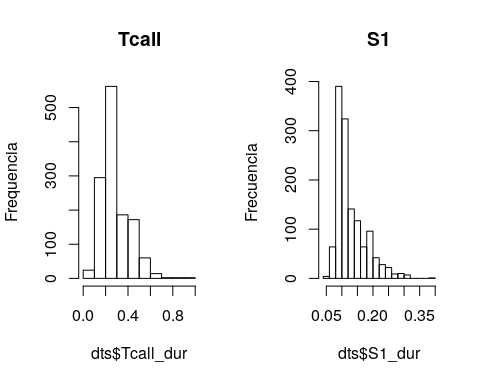
dts <- read.csv(sprintf("%s/data/mdf\_JR\_15viii19.csv",script.dir))  
str(dts)

## 'data.frame': 1351 obs. of 9 variables:  
## $ IndivGroup: Factor w/ 97 levels "AUA01","AUA02",..: 1 1 1 1 1 1 1 1 2 2 ...  
## $ soundfile : Factor w/ 1351 levels "0211327a","0211344a",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ S1\_dur : num 0.156 0.133 0.14 0.136 0.149 ...  
## $ Tcall\_dur : num 0.154 0.137 0.144 0.137 0.144 ...  
## $ RecSite : Factor w/ 37 levels "A6","A7","B1",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Lat : num 12.5 12.5 12.5 12.5 12.5 ...  
## $ Long : num -69.9 -69.9 -69.9 -69.9 -69.9 ...  
## $ LocCode : Factor w/ 14 levels "AUA","BON","CUR",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Region : Factor w/ 2 levels "isl","main": 1 1 1 1 1 1 1 1 1 1 ...

dts$Region <- factor(dts$Region,levels=c("main","isl"))  
dts <- subset(dts,!LocCode %in% c("VEN3","VEN7","VEN8","VEN9") )  
dts$LocCode <- droplevels(dts$LocCode)  
dts$IndivGroup <- droplevels(dts$IndivGroup)

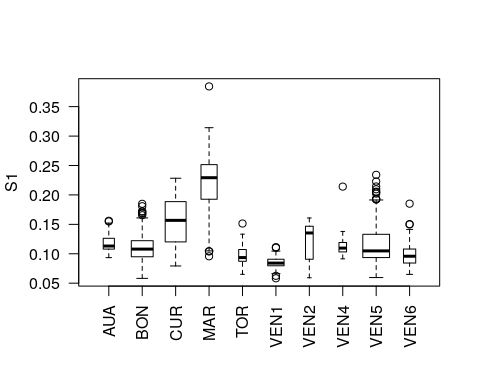
Verificamos el histograma de las dos variables (duración de la primera sílaba y duración total). Se observa asimetría marcada e indicios de multimodalidad

layout(matrix(1:2,ncol=2))  
hist(dts$Tcall\_dur,main="Tcall",ylab="Frequencia")  
hist(dts$S1\_dur,main="S1",ylab="Frecuencia")



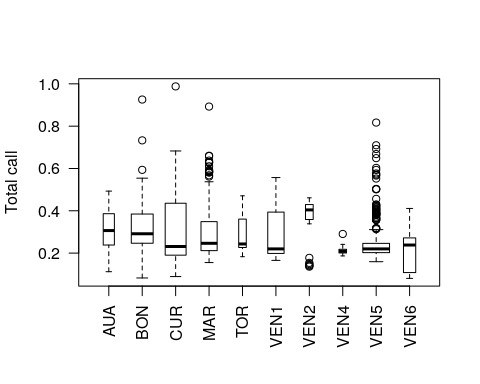
Verificamos la duración total y duración de la primera silaba por localidad (LocCode). El ancho es proporcional al número de observaciones.

boxplot(S1\_dur~LocCode,dts,las=2,varwidth=T,xlab="",ylab="S1",main="")



Para algunos grupos se observan conjuntos de valores extremos muy diferentes al resto.

boxplot(Tcall\_dur~LocCode,dts,las=2,varwidth=T,xlab="",ylab="Total call",main="")



## Ajuste de modelos

Vamos a considerar varios modelos para las variables Tcall\_dur y S1 usando transformación logarítmica. Utilizamos el valor de AIC marginal calculado por la función para escoger el mejor modelo. Debemos obviar el valor de las pruebas de t y F en las tablas, y basar la selección en las diferencias en AIC.

A partir de la exploración de los datos usamos un modelo con efecto fijo constante y un modelo alternativo para representar el efecto geográfico de la variable Region

La nomenclatura de los modelos es la siguiente

|  |  |  |
| --- | --- | --- |
| MODELO | componente | valor |
| S\_\_\_ | variable respuesta | log (S1) |
| T\_\_\_ | variable respuesta | log (Tcall) |
| \_0\_\_ | efecto fijo | constante |
| \_1\_\_ | efecto fijo | Region |
| \_\_0\_ | Efecto aleatorio | IndivGroup |
| \_\_1\_ | Efecto aleatorio | LocCode/IndivGroup |
| \_\_2\_ | Efecto aleatorio | Region/LocCode/IndivGroup |
| \_\_\_0 | Heterocedasticidad | No |
| \_\_\_1 | Heterocedasticidad | Si |

A continuación el código para ajustar todos los modelos

S000 <- lme(log(S1\_dur)~1,dts,random=~1|IndivGroup, method="ML")  
S100 <- lme(log(S1\_dur)~Region,dts,random=~1|IndivGroup, method="ML")  
  
S010 <- update(S000,random=~1|LocCode/IndivGroup)  
S110 <- update(S100,random=~1|LocCode/IndivGroup)  
  
S020 <- update(S000,random=~1|Region/LocCode/IndivGroup)  
S120 <- update(S100,random=~1|Region/LocCode/IndivGroup)  
  
S001 <- update(S000,weights=varIdent(form=~1|Region))  
S101 <- update(S100,weights=varIdent(form=~1|Region))  
S011 <- update(S010,weights=varIdent(form=~1|Region))  
S111 <- update(S110,weights=varIdent(form=~1|Region))  
S021 <- update(S020,weights=varIdent(form=~1|Region))  
S121 <- update(S120,weights=varIdent(form=~1|Region))  
  
dt2 <- subset(dts,subset=Tcall\_dur<0.6)  
T000 <- lme(log(Tcall\_dur)~1,dt2,random=~1|IndivGroup, method="ML",weights=varPower(form=~Tcall\_dur))  
T100 <- lme(log(Tcall\_dur)~Region,dt2,random=~1|IndivGroup, method="ML",weights=varPower(form=~Tcall\_dur))  
  
T010 <- update(T000,random=~1|LocCode/IndivGroup)  
T110 <- update(T100,random=~1|LocCode/IndivGroup)  
  
T020 <- update(T000,random=~1|Region/LocCode/IndivGroup)  
T120 <- update(T100,random=~1|Region/LocCode/IndivGroup)  
  
T001 <- update(T000,weights=varComb(varIdent(form=~1|Region),varPower(.1,~Tcall\_dur)))  
T101 <- update(T100,weights=varComb(varIdent(form=~1|Region),varPower(.1,~Tcall\_dur)))  
T011 <- update(T010,weights=varComb(varIdent(form=~1|Region),varPower(.1,~Tcall\_dur)))  
T111 <- update(T110,weights=varComb(varIdent(form=~1|Region),varPower(.1,~Tcall\_dur)))  
T021 <- update(T020,weights=varComb(varIdent(form=~1|Region),varPower(.1,~Tcall\_dur)))  
T121 <- update(T120,weights=varComb(varIdent(form=~1|Region),varPower(.1,~Tcall\_dur)))  
  
  
#P101 <- update(T100,weights=varPower(form=~log(Tcall\_dur)|Region))  
#E101 <- update(T100,weights=varExp(form=~log(Tcall\_dur)|Region))

Comparamos el valor de AIC para cada variable respuesta

S.aics <- AIC(S000,S100,S010,S110,S020,S120,  
 S001,S101,S011,S111,S021,S121)  
T.aics <- AIC(T000,T100,T010,T110,T020,T120,  
 T001,T101,T011,T111,T021,T121)

Reordenamos los modelos según el AIC

aic.tab.S <- cbind(S.aics,delta.AIC=S.aics[,2]-min(S.aics[,2]))  
aic.tab.T <- cbind(T.aics,delta.AIC=T.aics[,2]-min(T.aics[,2]))

### Resultados para S1

Los modelos con heterocedasticidad y dos niveles de efectos aleatorios tienen el mayor soporte y son casi equivalentes (), por lo que no se puede decir que haya evidencia definitiva sobre la importancia del efecto fijo en esta variable. Los modelos con tres niveles de efectos aleatorios tienen parametros nulos ( y el estimado del efecto de region es cero).

aic.tab.S[order(aic.tab.S$AIC),]

## df AIC delta.AIC  
## S111 6 -752.3977 0.0000000  
## S011 5 -752.1884 0.2093672  
## S121 7 -750.3977 2.0000006  
## S021 6 -750.1884 2.2093677  
## S110 5 -729.8144 22.5833546  
## S010 4 -729.6094 22.7882708  
## S120 6 -727.8144 24.5833550  
## S020 5 -727.6094 24.7882714  
## S101 5 -697.5357 54.8620382  
## S001 4 -677.9722 74.4255087  
## S100 4 -674.7096 77.6881583  
## S000 3 -655.0052 97.3925634

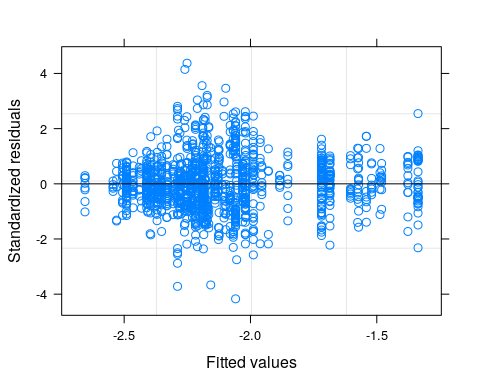
A continuación el resumen del mejor modelo. Los efectos fijos sugieren que hay mayor duración en la isla que en tierra firme. La varianza está repartida en los tres componentes (localidades, grupos y residuales), y la función de varianza residual es mayor en tierra firme que en las islas.

summary(S111)

## Linear mixed-effects model fit by maximum likelihood  
## Data: dts   
## AIC BIC logLik  
## -752.3977 -721.2899 382.1989  
##   
## Random effects:  
## Formula: ~1 | LocCode  
## (Intercept)  
## StdDev: 0.1980626  
##   
## Formula: ~1 | IndivGroup %in% LocCode  
## (Intercept) Residual  
## StdDev: 0.1583451 0.1495826  
##   
## Variance function:  
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Region   
## Parameter estimates:  
## isl main   
## 1.000000 1.221859   
## Fixed effects: log(S1\_dur) ~ Region   
## Value Std.Error DF t-value p-value  
## (Intercept) -2.2595704 0.09608198 1226 -23.517108 0.000  
## Regionisl 0.2127516 0.13400374 8 1.587654 0.151  
## Correlation:   
## (Intr)  
## Regionisl -0.717  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -4.16922331 -0.58164784 -0.05387406 0.51619096 4.37313849   
##   
## Number of Observations: 1319  
## Number of Groups:   
## LocCode IndivGroup %in% LocCode   
## 10 93

Los residuales del modelo tienen una distribución más o menos simétrica pero con valores extremos en ambas colas ( o ), y cierta estructura residual (menor variabilidad en el extremo izquierdo, variabilidad entre $-2.3 $ y ). Esto parece indicar que no se han incluido fuentes de variación importantes en el modelo (multimodalidad en la variable respuesta).

plot(S111)



Los intervalos de confianza aproximados sugieren que la diferencia en duración entre islas y continente NO es significativa. Recordemos que el modelo alternativo S011 sin efecto de la region tiene un soporte similar.

intervals(S111,which="fixed")

## Approximate 95% confidence intervals  
##   
## Fixed effects:  
## lower est. upper  
## (Intercept) -2.4479307 -2.2595704 -2.0712100  
## Regionisl -0.0960272 0.2127516 0.5215304  
## attr(,"label")  
## [1] "Fixed effects:"

Los intervalos de confianza aproximados para los componentes de varianza muestran: \* componente de varianza entre localidades es ligeramente superior a la varianza entre grupos y residual, \* el componente de heterocedasticidad indica diferencias significativas, con una variabilidad 12 – 32 % mayor en tierra firma que en islas.

intervals(S111,which="var-cov")

## Approximate 95% confidence intervals  
##   
## Random Effects:  
## Level: LocCode   
## lower est. upper  
## sd((Intercept)) 0.1223479 0.1980626 0.320633  
## Level: IndivGroup   
## lower est. upper  
## sd((Intercept)) 0.133927 0.1583451 0.1872151  
##   
## Variance function:  
## lower est. upper  
## main 1.128969 1.221859 1.322392  
## attr(,"label")  
## [1] "Variance function:"  
##   
## Within-group standard error:  
## lower est. upper   
## 0.1416158 0.1495826 0.1579977

Aproximadamente 45% de la varianza se puede atribuir a diferencias entre localidades, y el resto se reparte en diferencias entre grupos y residual (dentro de grupos o “no explicada”).

vs <- varcomp(S111)  
round(vs\*100/sum(vs),1)

## LocCode IndivGroup Within   
## 45.3 28.9 25.8   
## attr(,"class")  
## [1] "varcomp"

La partición de varianza del modelos alternativo es similar.

vs <- varcomp(S011)  
round(vs\*100/sum(vs),1)

## LocCode IndivGroup Within   
## 51.7 25.5 22.8   
## attr(,"class")  
## [1] "varcomp"

### Resultados para Tcall

Todos los modelos con heterocedasticidad tienen buen soporte de los datos (), y la evidencia favorece un efecto fijo de la region () pero no hay diferencias que apoyen una partición de varianza en dos o tres niveles de efectos aleatorios (diferencias en AIC equivalentes al número de parámetros añadidos).

aic.tab.T[order(aic.tab.T$AIC),]

## df AIC delta.AIC  
## T101 6 801.5656 0.000000  
## T111 7 803.5656 2.000001  
## T121 8 805.5656 4.000001  
## T011 6 807.6481 6.082521  
## T021 7 808.9029 7.337307  
## T001 5 813.9550 12.389391  
## T100 5 881.0596 79.493965  
## T110 6 883.0596 81.493966  
## T120 7 885.0596 83.493966  
## T010 5 887.3564 85.790825  
## T020 6 888.5811 87.015528  
## T000 4 891.5924 90.026848

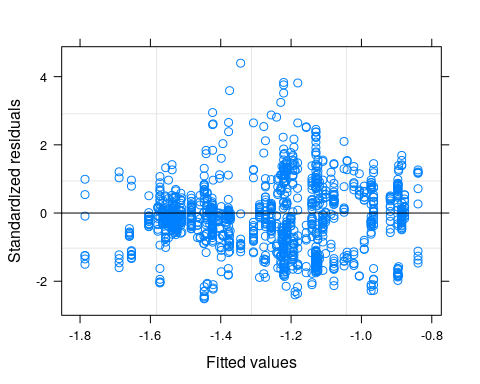
A continuación el resumen del mejor modelo. Los efectos fijos sugieren que hay mayor duración en la isla que en tierra firme. La varianza está repartida en dos componentes (grupos y residuales), y la función de varianza residual es menor en tierra firme que en las islas.

summary(T101)

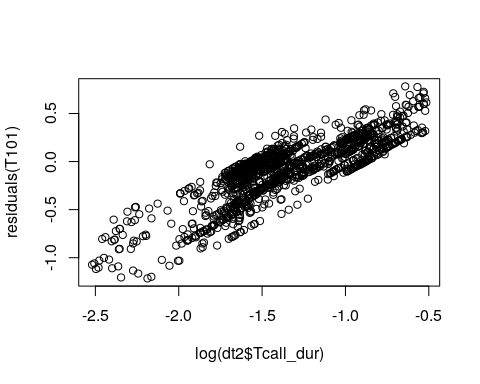
## Linear mixed-effects model fit by maximum likelihood  
## Data: dt2   
## AIC BIC logLik  
## 801.5656 832.5817 -394.7828  
##   
## Random effects:  
## Formula: ~1 | IndivGroup  
## (Intercept) Residual  
## StdDev: 0.2120211 0.1984472  
##   
## Combination of variance functions:   
## Structure: Different standard deviations per stratum  
## Formula: ~1 | Region   
## Parameter estimates:  
## isl main   
## 1.000000 0.687496   
## Structure: Power of variance covariate  
## Formula: ~Tcall\_dur   
## Parameter estimates:  
## power   
## -0.4522134   
## Fixed effects: log(Tcall\_dur) ~ Region   
## Value Std.Error DF t-value p-value  
## (Intercept) -1.401555 0.03292194 1206 -42.57206 0e+00  
## Regionisl 0.195771 0.04866983 91 4.02243 1e-04  
## Correlation:   
## (Intr)  
## Regionisl -0.676  
##   
## Standardized Within-Group Residuals:  
## Min Q1 Med Q3 Max   
## -2.5161618 -0.6696313 -0.1445784 0.3539859 4.3950388   
##   
## Number of Observations: 1299  
## Number of Groups: 93

Los residuales del modelo tienen una distribución ligeramente asimétrica con valores extremos en ambas colas ( o ), y *mucha* estructura residual (muchos grupos densos de puntos). Esto parece indicar que no se han incluido fuentes de variación importantes en el modelo. Hay que considerar seriamente la multimodalidad en la variable respuesta y la correlación con otras variables.

plot(T101)



plot(residuals(T101)~log(dt2$Tcall\_dur))

 Los intervalos de confianza aproximados confirman que la diferencia en duración entre islas y continente es significativa.

intervals(T101,which="fixed")

## Approximate 95% confidence intervals  
##   
## Fixed effects:  
## lower est. upper  
## (Intercept) -1.46609548 -1.401555 -1.3370137  
## Regionisl 0.09916879 0.195771 0.2923732  
## attr(,"label")  
## [1] "Fixed effects:"

Los intervalos de confianza aproximados para los componentes de varianza muestran: \* componente de varianza residual es claramente mayor a la varianza entre grupos, \* el componente de heterocedasticidad indica diferencias significativas, con una variabilidad 23 – 34 % menor en tierra firma que en islas.

intervals(T101,which="var-cov")

## Approximate 95% confidence intervals  
##   
## Random Effects:  
## Level: IndivGroup   
## lower est. upper  
## sd((Intercept)) 0.1769507 0.2120211 0.2540422  
##   
## Variance function:  
## lower est. upper  
## A.main 0.6345572 0.6874960 0.7448512  
## B.power -0.6108212 -0.4522134 -0.2936057  
## attr(,"label")  
## [1] "Variance function:"  
##   
## Within-group standard error:  
## lower est. upper   
## 0.1590438 0.1984472 0.2476129

Aproximadamente una cuarta parte de la varianza se puede atribuir a diferencias entre grupos, y el resto es varianza residual (dentro de grupos o “no explicada”).

vs <- varcomp(T101)  
round(vs\*100/sum(vs),1)

## IndivGroup Within   
## 53.3 46.7   
## attr(,"class")  
## [1] "varcomp"

La partición de varianza es similar aunque se agreguen más niveles aleatorios en el modelo.

vs <- varcomp(T111)  
round(vs\*100/sum(vs),1)

## LocCode IndivGroup Within   
## 0.0 53.3 46.7   
## attr(,"class")  
## [1] "varcomp"

vs <- varcomp(T121)  
round(vs\*100/sum(vs),1)

## Region LocCode IndivGroup Within   
## 0.0 0.0 70.7 29.3   
## attr(,"class")  
## [1] "varcomp"