


Introducción a los métodos estadísticos bayesianos en Ecología

Programa del curso:

Pablo Inchausti

1. Introducción general
2. Elementos básicos del análisis bayesiano
3. Análisis bayesiano I
4. Análisis bayesiano II
5. Modelos bayesianos jerárquicos 

Modelos Jerárquicos

Práctico 04

a) Pichones de Lechuzas en bloques aleatorizados

a) Pichones de Lechuzas en bloques aleatorizados

Obj: cómo la tasa de aprovisionamiento **SiblingNegotiation** de pichones varía según el **sexo** del padre y la cantidad de comida **FoodTreatment**.



(Roulin & Bersier 2007)

```
> DF=read.csv("lechuzasPr04.csv", header = T)
> str(DF)
'data.frame': 599 obs. of 6 variables:
 $ Nest          : Factor w/ 27 levels "AutavauxTV"
 $ FoodTreatment : Factor w/ 2 levels "Deprived","S
 $ SexParent     : Factor w/ 2 levels "Female","Mal
 $ ArrivalTime   : num 22.2 22.4 22.5 22.6 22.6 ..
 $ SiblingNegotiation: int 4 0 2 2 2 2 18 4 18 0 ...
 $ BroodSize     : int 5 5 5 5 5 5 5 5 5 5 ...
```



```
> summary(DF)
```

	Nest	FoodTreatment	SexParent	ArrivalTime	SiblingNegotiation	BroodSize
Oleyes	: 52	Deprived:320	Female:245	Min. :21.71	Min. : 0.00	Min. :1.000
Montet	: 41	Satiated:279	Male :354	1st Qu.:23.11	1st Qu.: 0.00	1st Qu.:4.000
Etrabloz	: 34			Median :24.38	Median : 5.00	Median :4.000
Yvonnand	: 34			Mean :24.76	Mean : 6.72	Mean :4.392
Champmartin	: 30			3rd Qu.:26.25	3rd Qu.:11.00	3rd Qu.:5.000
Lucens	: 29			Max. :29.25	Max. :32.00	Max. :7.000
(Other)	:379					

Estructura de los datos:

Hay datos en diferentes localidades: `table(DF$Nest)`

```
> with (DF, table(SexParent, FoodTreatment))
```

	FoodTreatment	
SexParent	Deprived	Satiated
Female	123	122
Male	197	157

```
with(DF, ftable(Nest, FoodTreatment)))
```

```
with(DF, ftable(Nest, SexParent, FoodTreatment)))
```

¿qué se concluye?

A priori, las observaciones realizadas en la misma localidad son “similares entre sí” en éstas y otras variables que afecten Y: ← Nest como efecto de grupo

Además, el valor de Y (SiblingNegotiation) es afectado por el tamaño de la nidada **BroodSize** y se debe controlar por ello. ← offset.

Los GLM de conteos típicamente modelan frecuencias, pero en ocasiones es más relevante modelar tasas (frec./tiempo o esfuerzo) en lugar de frecuencias.

Un Proceso Poisson ocurre a una tasa λ constante y se espera observar $\mu = \lambda t$ eventos en un lapso de tiempo (o espacio o esfuerzo) t .

La tasa λ es simplemente conteo/unidad de tiempo, área, esfuerzo, o por inds. expuestos....

En este caso, la tasa sería el valor observado de $Y/\text{tamaño de la nidada ese día}$ ← offset



GLM conteos: función de enlace log $\rightarrow \log(E(Y)) = X\beta$.

Si ahora $Y = \text{tasa (SiblingNegotiation/BroodSize)}$,
 $\log(E(\text{SiblingNegotiation/BroodSize})) = X\beta$, y además:

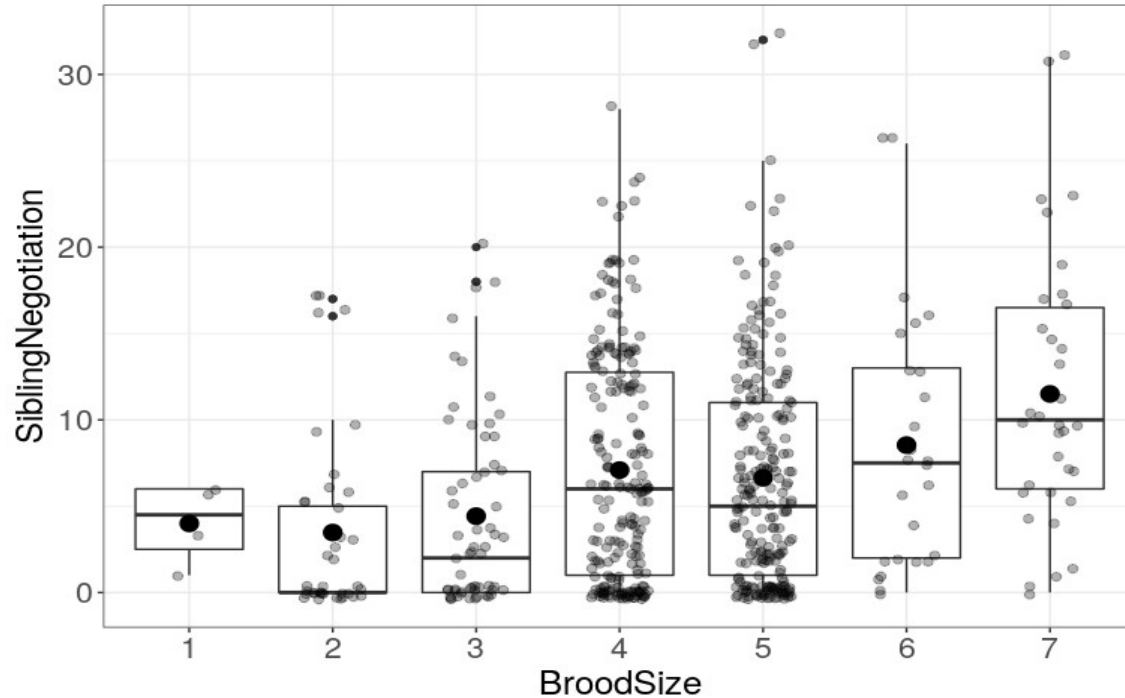
$$\log(E(\text{SiblingNegotiation})) = X\beta + \log(\text{BroodSize})$$

El último término es el offset que se supone con
coeficiente=1. (i.e. no se estima)

¿Cuándo usar offset en un GLM(M)?

- * Si hay una covariable que afecta la tasa λ .
- * Heterogeneidad en el esfuerzo de observación o susceptibilidad entre unidades muestrales.

Análisis exploratorio:



```
ggplot(data=DF, aes(x=BroodSize, y=SiblingNegotiation)) +  
  geom_boxplot(aes(x=as.factor(BroodSize)))+  
  geom_jitter(alpha=0.3, size=2, position = position_jitter(width = .2))+etc
```

**To offset or
not to offset?**

```
summaryBy(SiblingNegotiation~BroodSize,  
BroodSize SiblingNegotiation.mean  
1 4.00  
2 3.47  
3 4.42  
4 7.09  
5 6.64  
6 8.54  
7 11.50
```

**Si, pero ahora se
modela una tasa.**

Conclusions

Análisis exploratorio de datos:

```
> summaryBy(SiblingNegotiation~SexParent+FoodTreatment, data=DF, FUN=c(mean, sd))
```

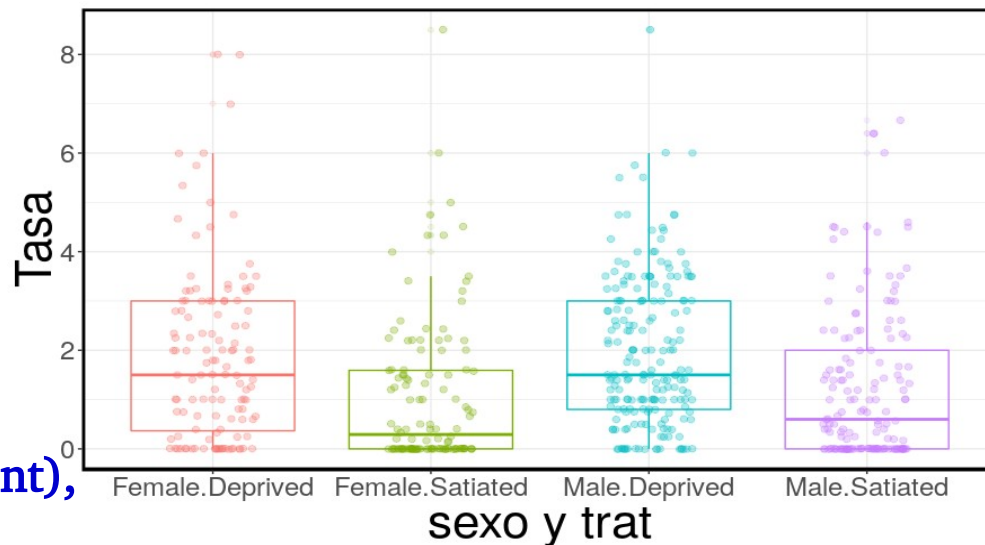
	SexParent	FoodTreatment	SiblingNegotiation.mean	SiblingNegotiation.sd
1	Female	Deprived	7.26	6.33
2	Female	Satiated	4.75	6.53
3	Male	Deprived	8.73	6.49
4	Male	Satiated	5.31	6.56

```
> DF$Tasa=DF$SiblingNegotiation/DF$BroodSize
```

```
> summaryBy(Tasa~SexParent+FoodTreatment, data=DF, FUN=c(mean, sd))
```

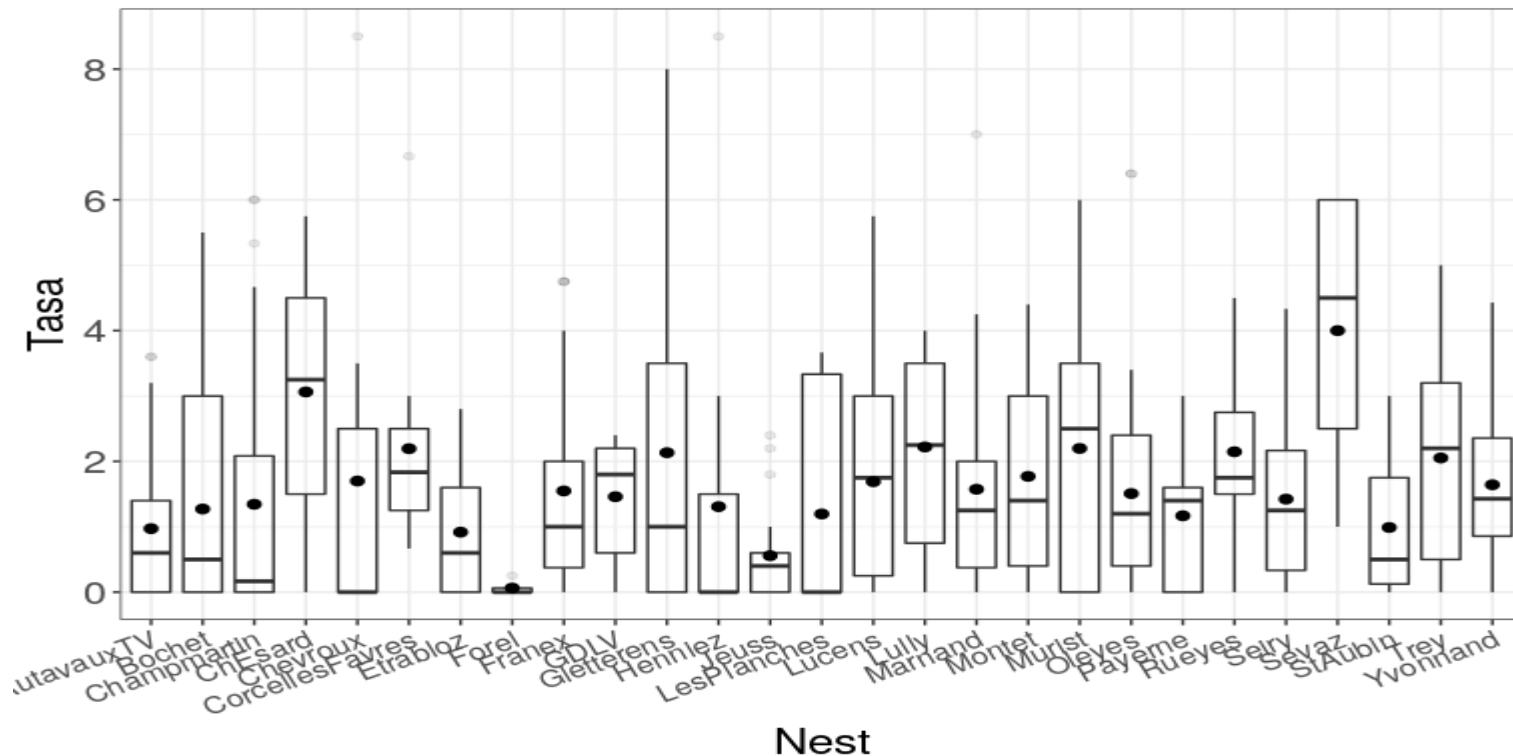
	SexParent	FoodTreatment	Tasa.mean	Tasa.sd
1	Female	Deprived	1.86	1.75
2	Female	Satiated	1.03	1.49
3	Male	Deprived	1.97	1.52
4	Male	Satiated	1.24	1.54

¿Interacción?



```
ggplot(data=DF,  
aes(x=interaction(SexParent, FoodTreatment),  
y=Tasa)) + etc
```

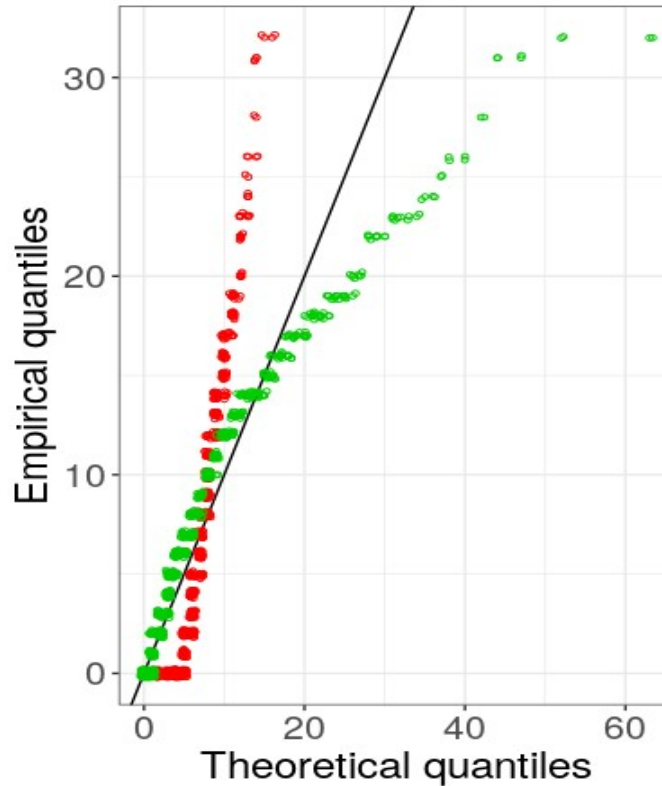
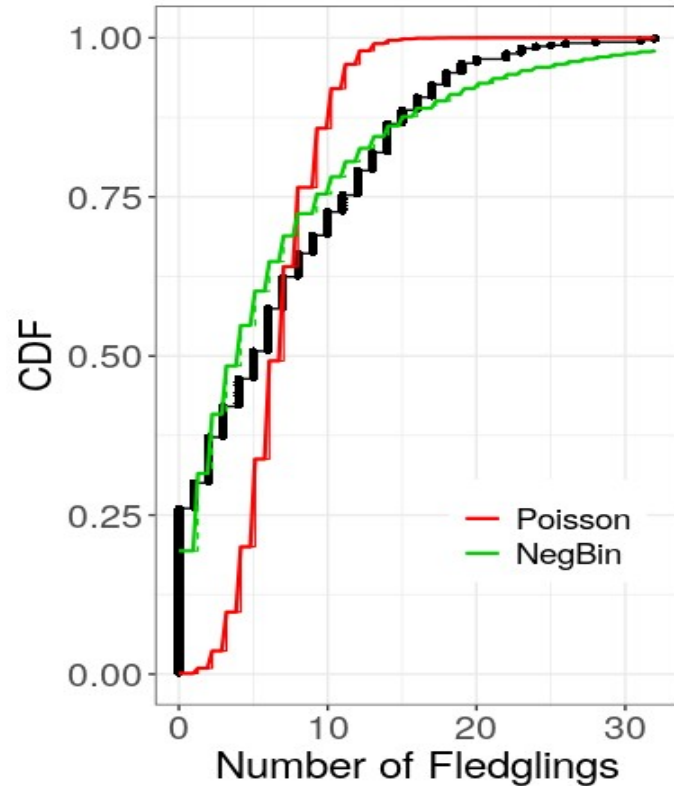

Variación entre zonas:



42 a 49

```
ggplot(DF, aes(x=Nest, y=Tasa)) + geom_boxplot() + etc.
```

Distr de prob. de Y: para la verosimilitud del modelo



```
poisson=fitdist(DF$SiblingNegotiation,"pois")  
negbinom=fitdist(DF$SiblingNegotiation,"nbinom")  
CDF=cdfcomp0  
QQ=qqcomp0  
etc
```

54 a 70

Nada muy claro: ¿NegBinom?

Modelo estadístico: $Y \sim \text{BiNeg}(\mu_Y, \phi)$ con $\log(E(Y)) = \beta X$.

$\log(E(Y)) \sim \text{SexParent} + \text{FoodTreatment} + \text{offset}(\log(\text{BroodSize}))$

Falta la variación entre Nests ← efecto de grupo.

Cuando las vars explicativas son TODAS categóricas, no tiene sentido ajustar pendientes variables entre grupos...

¿Cuántos niveles tienen las vars explicativas categóricas?

2

$\log(E(Y)) = \beta_0 + \beta_1 \text{SexParent} + \beta_2 \text{FoodTreatment} + b_0 + \text{offset}()$

β_0 : $\log(E(Y))$ para niveles de referencia de ambas variables.

β_i : diferencial de $\log(E(Y))$ entre nivel de ref. y el otro nivel.

b_j : diferencial de β_0 para cada sitio $\sim N(0, \sigma^2_{\text{sitio}})$

```

> require(brms)
> get_prior(formula=SiblingNegotiation~SexParent+FoodTreatment+offset(log(BroodSize))+(1|Nest),
+           data=DF,family=negbinomial())

```

	prior	class	coef	group	resp	dpar	nlpar	bound
1		b						
2		b	FoodTreatmentSatiated					
3		b	SexParentMale					
4	student_t(3, 2, 10)	Intercept						
5	student_t(3, 0, 10)	sd						
6		sd			Nest			
7		sd	Intercept	Nest				
8	gamma(0.01, 0.01)	shape						

Ajustando el modelo:

También es posible: `brm(y~x1* x2+(1 | gr(id,dist= "student"))`

Con `prior.m = c(set_prior("",class = "b", coef= ""),
set_prior("", class = "Intercept"),
set_prior("", class = "shape")) definan 2 tipos de previas para los ef. poblacionales:`

a) `prior.m11`: no informativos, con sd “grande” (sd=5)

b) `prior.m12`: muy informativos, con sd “pequeño” (sd=0.5).

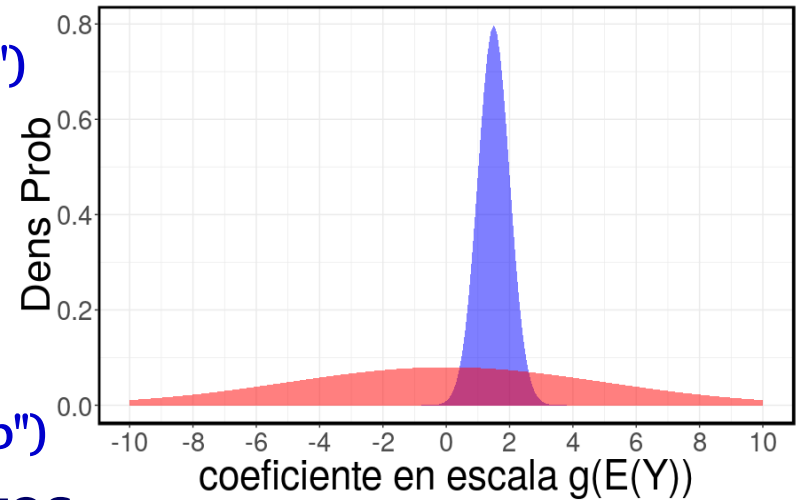
```
prior.m11 = c(set_prior("normal(0,5)",class = "b")
```

prior.m11 tiene poca información previa. Los datos tendrán mucha importancia en la distr posterior.

```
prior.m12=c(set_prior("normal(1.5,0.5)",class="b")
```

Al contrario, con **prior.m12** los datos **podrían tener POCA** importancia en la distr posterior.

```
ggplot(data=data.frame(x=seq(from=-10, to=10, by=.1)),  
aes(x = x)) + geom_ribbon(aes(ymin = 0, ymax=  
dnorm(x,1.5, 0.5)),fill = "blue", alpha=0.5)+ geom_ribbon  
(aes(ymin = 0, ymax = dnorm(x, 0, 5)),fill = "red", alpha=0.5) +etc
```



El modelo evalúa posibles valores de los parámetros tomando en cuenta la evidencia previa.

```

m11=brm(SiblingNegotiation~SexParent+FoodTreatment+offset(log(BroodSize))+(1|Nest),~
.....data=DF,family=negbinomial(),cores=mc.cores,prior=prior.m11,warmup=1000,-
.....chains=3,iter=2000,thin=2,control=list(adapt_delta=0.95))~
m12=brm(SiblingNegotiation~SexParent+FoodTreatment+offset(log(BroodSize))+(1|Nest),~
.....data=DF,family=negbinomial(),cores=mc.cores,prior=prior.m12,warmup=1000,-
.....chains=3,iter=2000,thin=2,control=list(adapt_delta=0.95))~

```

Group-Level Effects:

~Nest (Number of levels: 27)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.40	0.10	0.22	0.61	1.00	807	1032

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.67	0.13	0.44	0.95	1.00	1056	952
SexParentMale	0.04	0.11	-0.16	0.24	1.00	1315	1373
FoodTreatmentSatiated	-0.68	0.11	-0.90	-0.47	1.00	983	1088

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
shape	0.84	0.07	0.71	0.98	1.00	1108	1056

Previas vagas, no informativas

Group-Level Effects:

~Nest (Number of levels: 27)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.36	0.11	0.18	0.59	1.00	675	842

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.58	0.12	0.34	0.83	1.00	1119	1124
SexParentMale	0.12	0.11	-0.09	0.32	1.00	1451	1421
FoodTreatmentSatiated	-0.58	0.11	-0.80	-0.37	1.00	895	1266

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
shape	0.83	0.07	0.70	0.98	1.00	1148	1036

Previas precisas, muy informativas

Que los previas puedan afectar la dist posterior **no es un defecto ni una aberración** de la estadística bayesiana.

$$\text{Post}(\text{param} | Y) \propto \text{Vero}(Y | \text{param}) \text{Previa}(\text{param})$$

Si los previas son “demasiado informativas”, el modelo no puede incorporar adecuadamente (hacer un update) la información contenida en los datos.

Si los previas son “demasiado vagas o muy poco informativas”, cada análisis se realiza como si no existieran antecedentes o conocimiento acumulado.



Esta es una vieja, inconclusa e irresoluble discusión...

Últimos comentarios generales sobre los previas :

- 1) Siempre explicitar detalladamente los previas usadas.
- 2) Siempre hacer un análisis de sensibilidad mínimo a los previas empleados → robustez de las conclusiones.
- 3) Los “previas informativos” tienen mayor influencia con set de datos pequeños. Permiten incorporar información de diversos orígenes (expertos), ayudan a la convergencia e imponen “limitaciones útiles”.
- 4) Los previas “vagas o no informativas” son provisionarias.
- 5) No hay un método general para compilar y elicitar previas (ver Hobbs & Hooten 2015, OTROS).

Efectos poblacionales:

```
> summary(m11)
Family: negbinomial
Links: mu = log; shape = identity
Formula: SiblingNegotiation ~ SexParent + FoodTreatment + offset(log(BroodSi
Data: DF (Number of observations: 599)
Samples: 2 chains, each with iter = 2000; warmup = 1000; thin = 2;
total post-warmup samples = 1000
```

Group-Level Effects:

~Nest (Number of levels: 27)

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
sd(Intercept)	0.40	0.10	0.24	0.61	510	1.00

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
Intercept	0.67	0.12	0.43	0.92	593	1.00
SexParentMale	0.04	0.11	-0.18	0.24	882	1.00
FoodTreatmentSatiated	-0.68	0.11	-0.90	-0.47	681	1.00

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Eff.Sample	Rhat
shape	0.83	0.07	0.71	0.97	1014	1.00

Female: $\log(E(Y))=0.67$

Male: $\log(E(Y))=0.67+0.04$

Dist posterior de m11:

```
> post.m11=as_draws_df(m11)
```

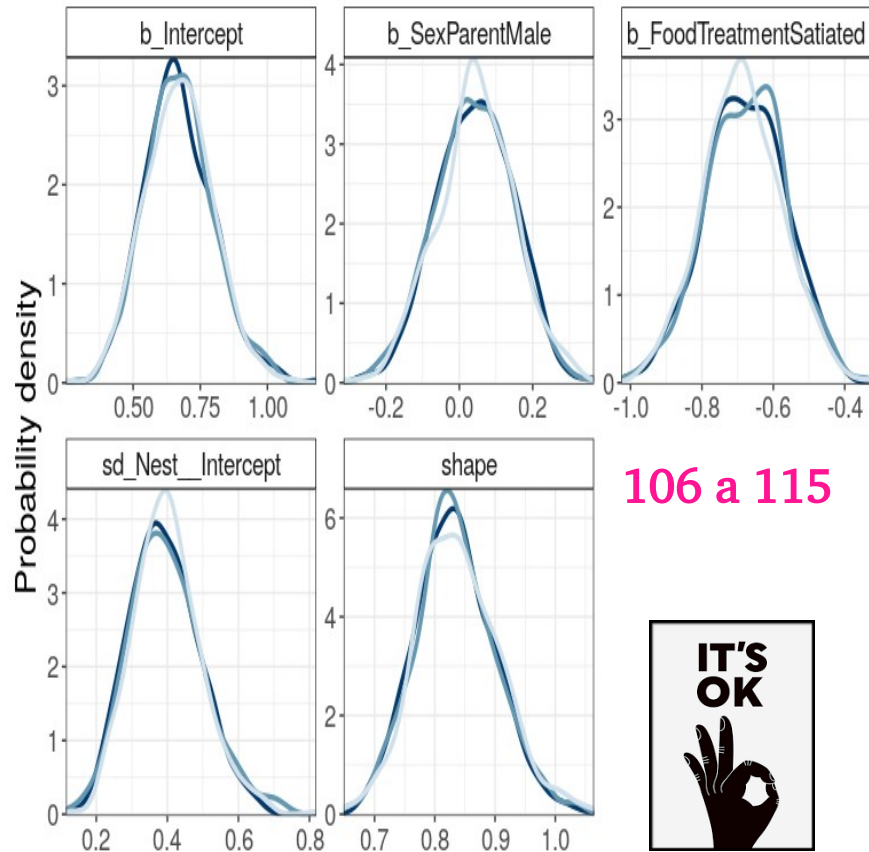
```
> names(post.m11)
```

[1] "b_Intercept"	"b_SexParentMale"
[3] "b_FoodTreatmentSatiated"	"sd_Nest_Intercept"
[5] "shape"	"r_Nest[AutavauxTV,Intercept]"
[7] "r_Nest[Bochet,Intercept]"	"r_Nest[Champmartin,Intercept]"
[9] "r_Nest[ChEsard,Intercept]"	"r_Nest[Chevroux,Intercept]"
[11] "r_Nest[CorcellesFavres,Intercept]"	"r_Nest[Etrabloz,Intercept]"
[13] "r_Nest[Forel,Intercept]"	"r_Nest[Franex,Intercept]"
[15] "r_Nest[GDLV,Intercept]"	"r_Nest[Gletterens,Intercept]"
[17] "r_Nest[Henniez,Intercept]"	"r_Nest[Jeuss,Intercept]"
[19] "r_Nest[LesPlanches,Intercept]"	"r_Nest[Lucens,Intercept]"
[21] "r_Nest[Lully,Intercept]"	"r_Nest[Marnand,Intercept]"
[23] "r_Nest[Montet,Intercept]"	"r_Nest[Murist,Intercept]"
[25] "r_Nest[Oleyes,Intercept]"	"r_Nest[Payerne,Intercept]"
[27] "r_Nest[Rueyes,Intercept]"	"r_Nest[Seiry,Intercept]"
[29] "r_Nest[Sevaz,Intercept]"	"r_Nest[StAubin,Intercept]"
[31] "r_Nest[Trey,Intercept]"	"r_Nest[Yvonnand,Intercept]"

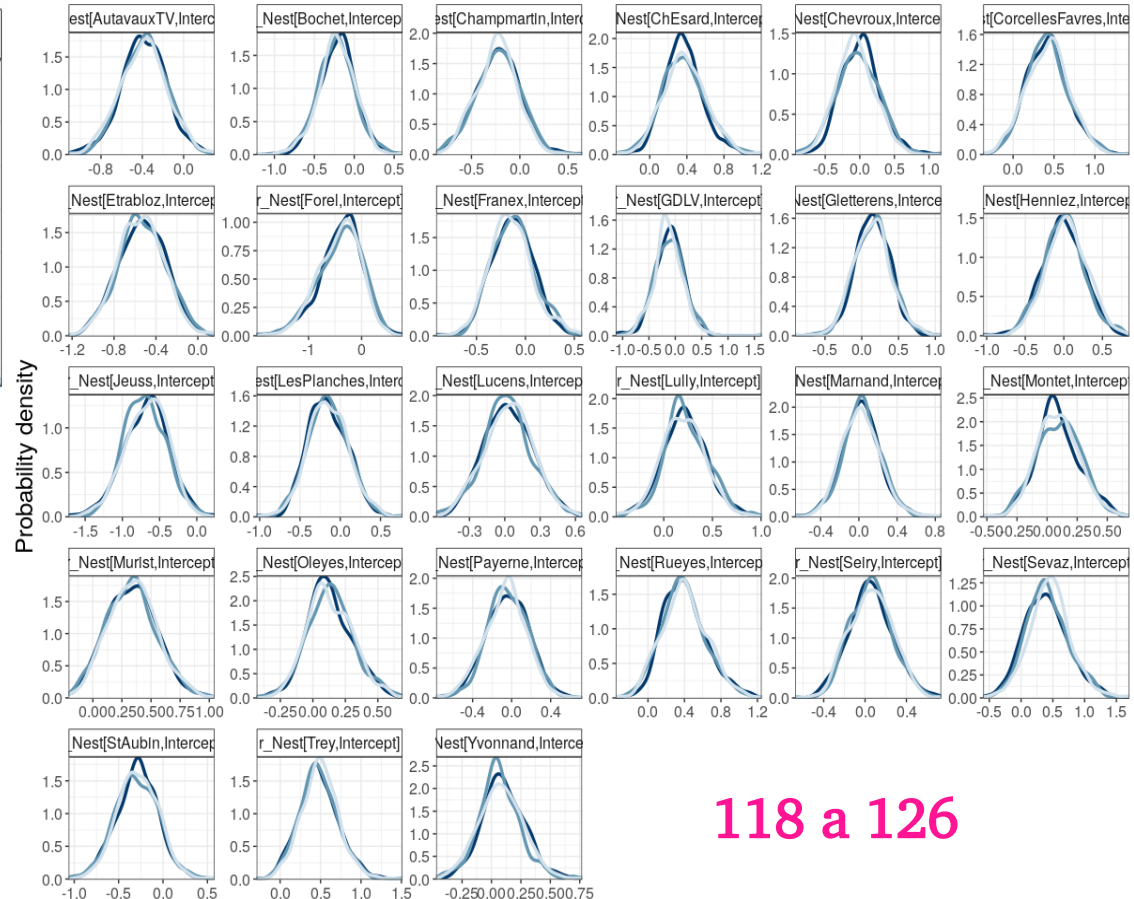
Deprived: $\log(E(Y))=0.67$

Satiated: $\log(E(Y))=0.67-0.68$

Convergencia a una distr posterior estacionaria:

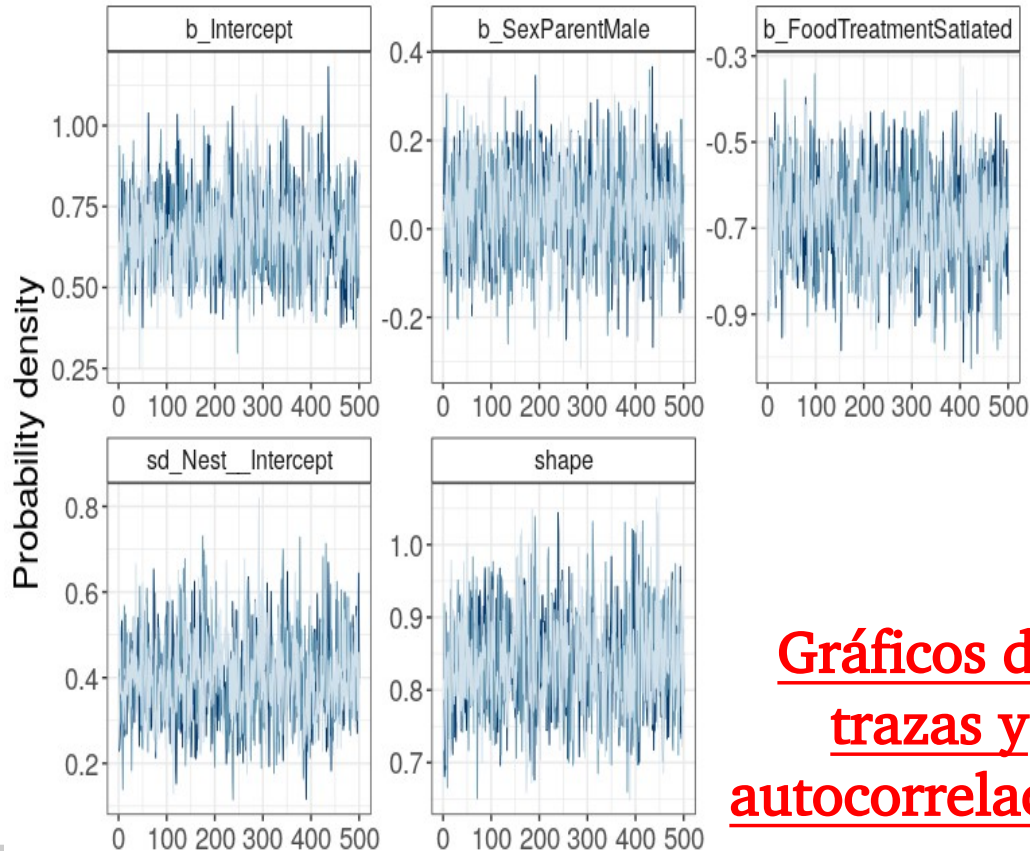


106 a 115

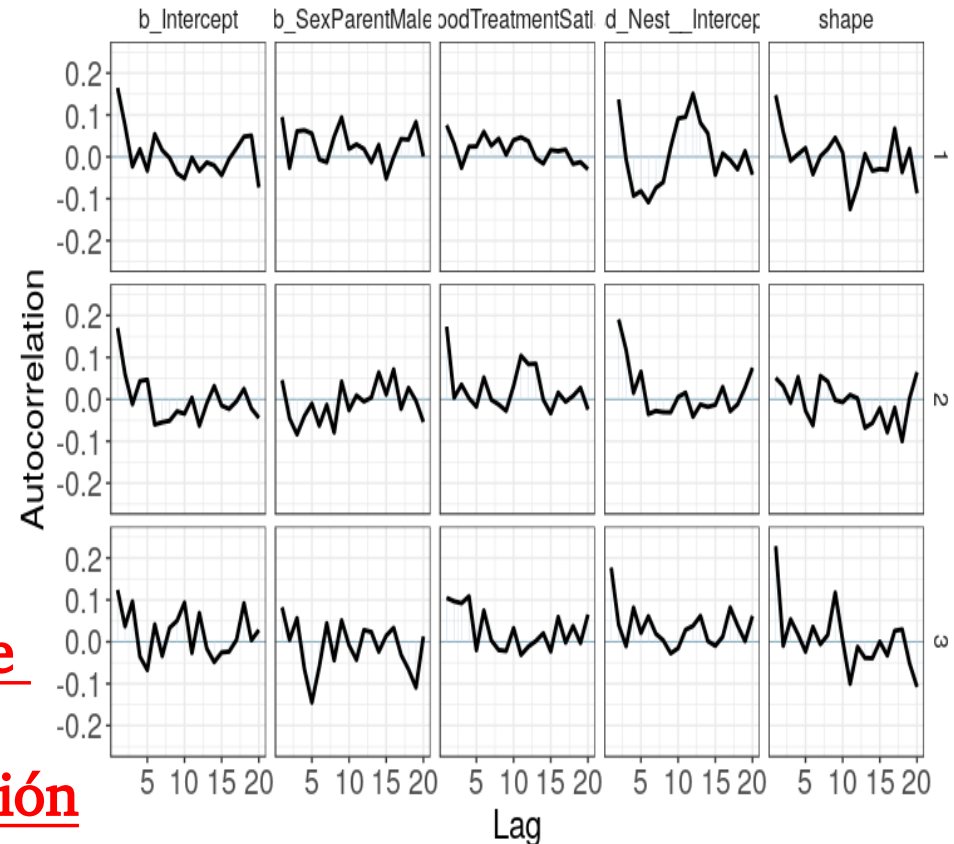


118 a 126

`mcmc_trace(m11, regex_pars = c("^b", "sd", "shape"))`
 +etc 129 a 136



`mcmc_acf(m11, regex_pars = c("^b", "sd", "shape"))`
 +etc 139 a 147



Gráficos de
trazas y
autocorrelación

	E(Y)
♀	$\exp(0.67)=1.95$
♂	$\exp(0.71)=2.03$

Treatment	E(Y)
Deprived	$\exp(0.67)=1.95$
Satiated	$\exp(-0.01)=0.99$

$E(Y)$ es $2.03/1.95=1.04$ 4.0% de diferencia entre sexos
 $\rightarrow \log(0.99/1.95)*100 \rightarrow 67.8\%$ de diferencia entre Treatments

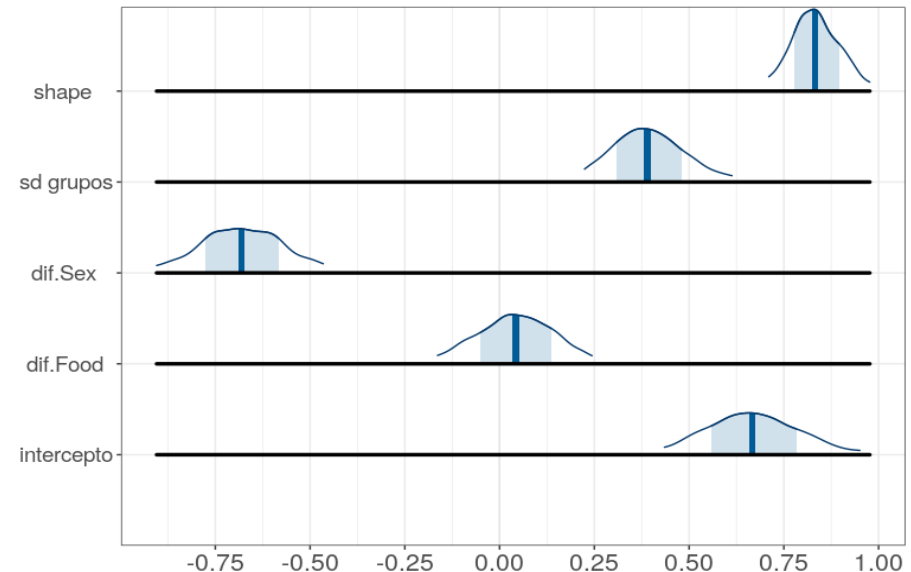
Efectos poblacionales:

```
mcmc_areas(m11, prob=0.62, prob_outer = 0.95,
  regex_pars = c("^b", "sd", "shape"))+
  scale_y_discrete(labels=c("intercepto", "dif.Food",
    "dif.Sex", "sd grupos", "shape"))+
```

etc.

146 a 152

```
> bayes_R2(m11)
  Estimate Est.Error Q2.5 Q97.5
R2      0.251      0.0475 0.16 0.352
```



```
interc=variables(m11)[grep("Intercept", variables(m11))]
```

Efectos de grupo:

```
mcmc_areas(m11, prob=0.62,  
  prob_outer = 0.95, pars = interc)+  
scale_y_discrete(labels=  
  rev(unique(DF$Nest)))+
```

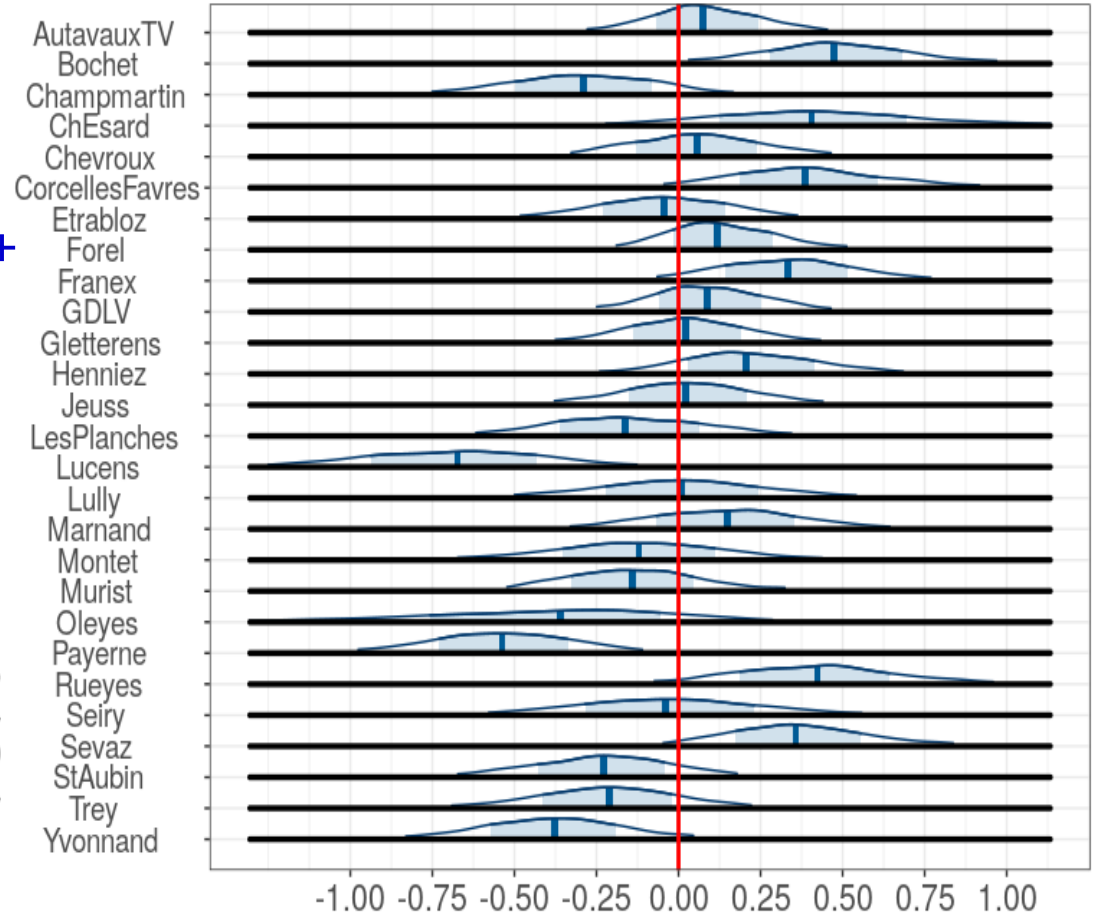
etc.

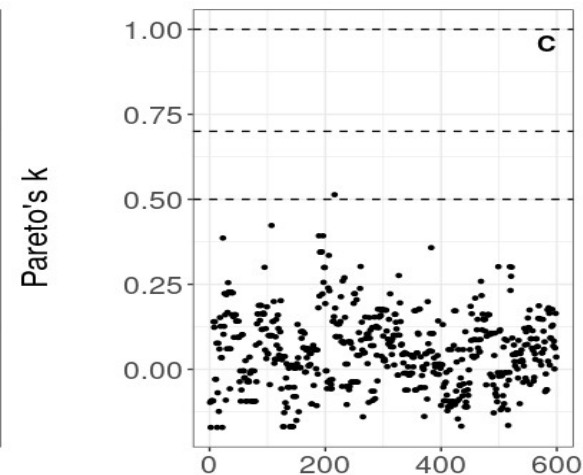
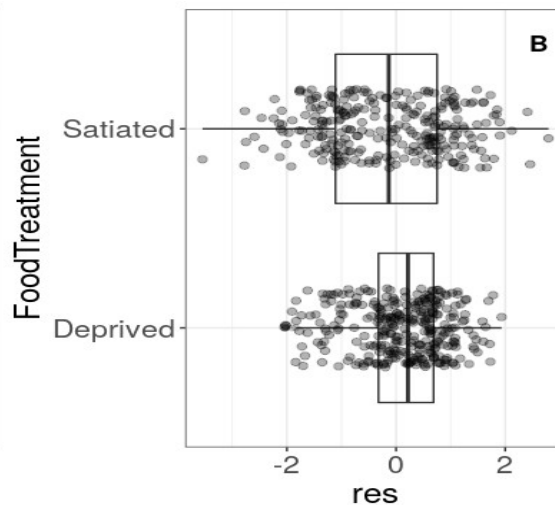
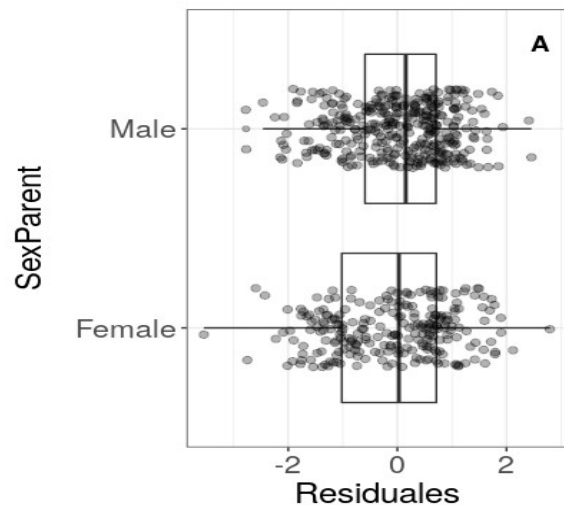
154 a 162

```
> ranef(m11)  
$Nest  
, , Intercept
```

Conclusions

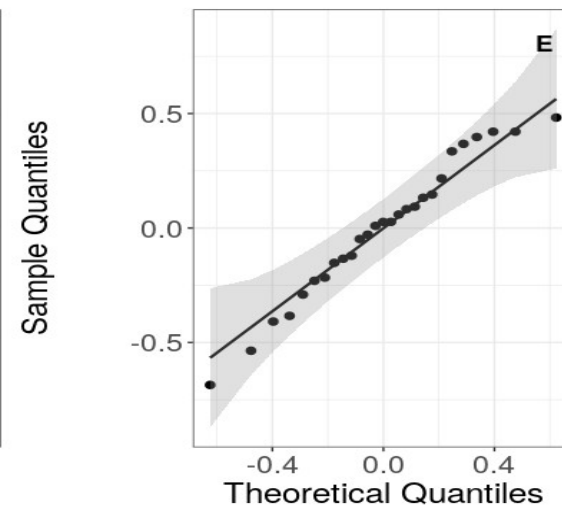
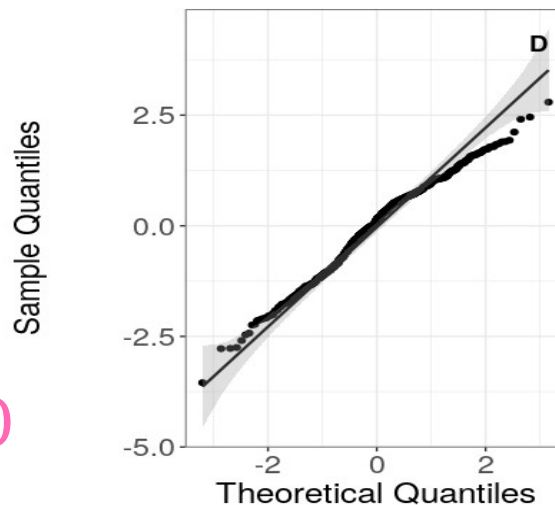
	Estimate	Est.Error	Q2.5	Q97.5
AutavauxTV	-0.38363	0.220	-0.8309	0.0454
Bochet	-0.21664	0.230	-0.6910	0.2219
Champmartin	-0.23043	0.222	-0.6724	0.1784





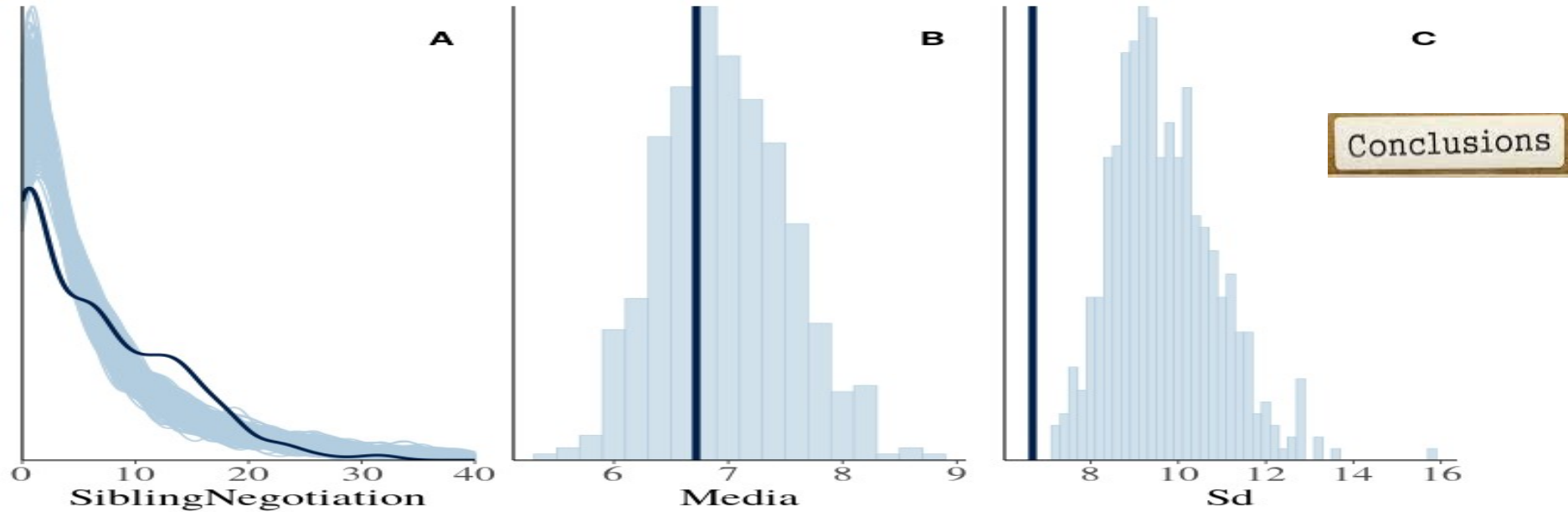
Análisis de residuos:

- 1) Obtener residuales y loo
- 2) Graficar resid vs vars explicativas y loo
- 3) QQplot datos y ef. grupo
(líneas 185 a 253 aprox.)



Distribución predictiva posterior: `m11.pred.post=posterior_predict(m11, ndraws = 500)`

pp-checks:



```
ppc.density=ppc_dens_overlay(y=DF$SiblingNegotiation,  
  yrep=m11.pred.post) + etc
```

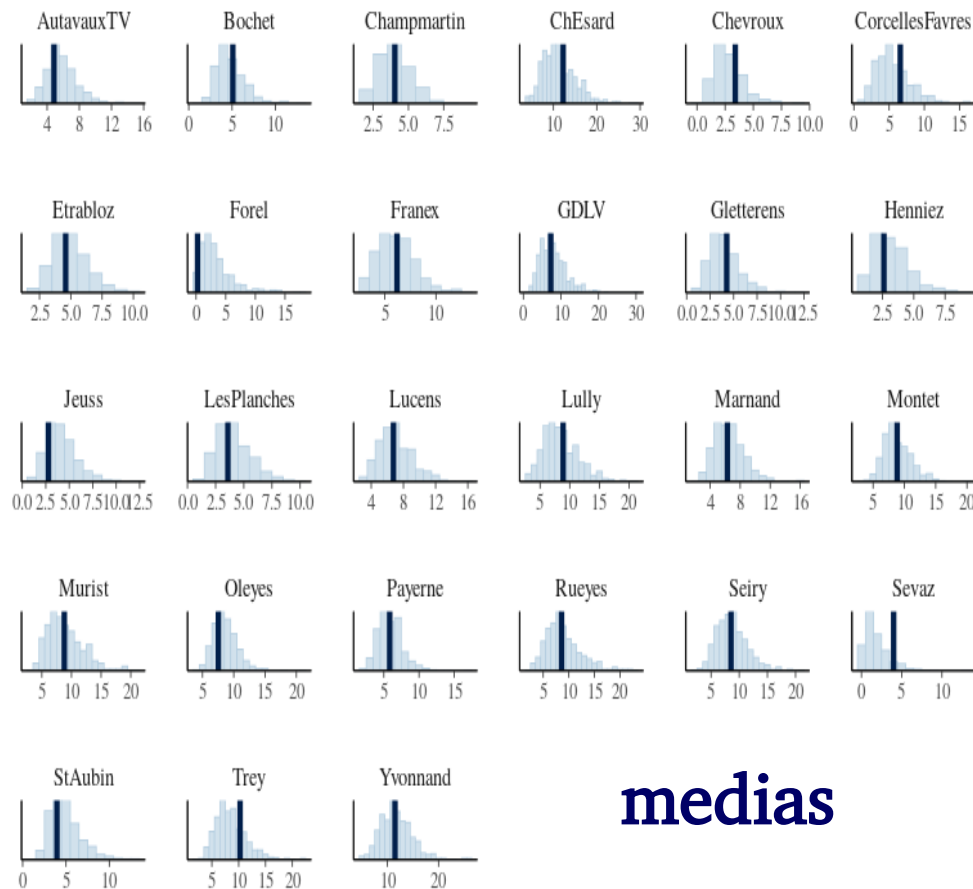
```
ppc.sd=ppc_stat(y=DF$SiblingNegotiation,  
  yrep=m11.pred.post, stat = sd,binwidth = 0.2) + etc
```

```
ppc.media=ppc_stat(y=DF$SiblingNegotiation,  
  yrep=m11.pred.post, stat = mean,binwidth = 0.2) + etc
```

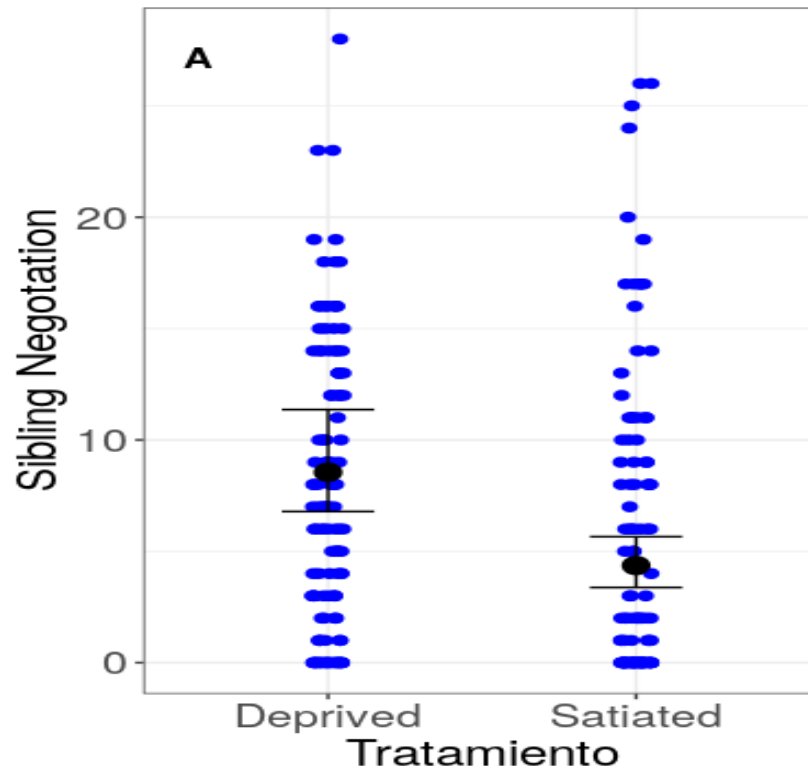
260 a 277

pp-checks por grupo:

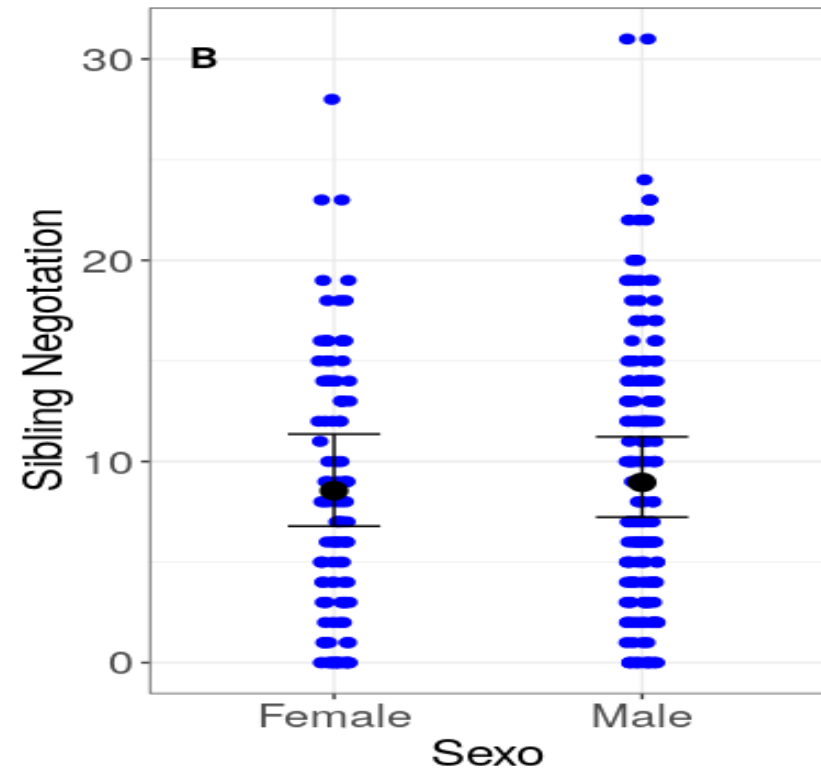
280 a 285



Curvas condicionales: `m11.cond.eff=conditional_effects(m11)` 290 a 303



```
> names(m11.cond.eff)  
[1] "SexParent" "FoodTreatment"
```



```
cond.m11.sexo=plot(m11.cond.eff, plot = F, points=T,  
point_args = list(width =0.05, col="blue"))[[1]]+etc
```

Modelos Jerárquicos

Práctico 04

a) Pichones de Lechuzas en bloques aleatorizados