

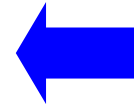
# Introducción a los métodos estadísticos

bayesianos en Ecología

Pablo Inchausti

## Programa del curso:

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



### 3. Análisis bayesiano I:

### Práctico 02

a) GLM con variables categóricas

b) Tests de hipótesis y tests a posteriori

## a) GLM con variables categóricas

En el Teór02 se ajustó un modelo de regresión (GLM) cuyas vars explicativas eran todas numéricas.

En R (y en cualquier otro software), **los modelos con vars explicativas categóricas** (i.e. sexo, sitio) también se ajustan como modelos de regresión con vars explicativas binarias.

← “dummy regression”

**Con estas con vars explicativas binarias**, no se estiman pendientes, **sino diferencias de medias entre grupos.**

**Es CRUCIAL entenderlo para poder interpretar el output**

**El Análisis de varianza de un factor** evalúa las diferencias de  $k$  medias de una var. respuesta  $Y$  entre grupos definidos por una var. explicativa categórica con  $j:1..k > 2$  grupos (o niveles).

**El modelo estadístico es :**  $Y \sim \text{Normal}(\mu_{Y,j}, \sigma_Y)$ , con  $\mu_{Y,j} = \mu_Y + \alpha_j \leftarrow$  diferencial o efecto asociado al grupo  $j$ .

Con  $j$  grupos, **habría** que estimar  $\mu_Y + j\alpha_j + \sigma_Y = j+2$  **parámetros...**

**Not Really** 
$$\mu_Y = \frac{\sum \mu_{j,Y}}{k} = \frac{\sum \mu_{j,Y} + \alpha_j}{k} = \frac{k\mu_Y + \sum \alpha_j}{k} = \mu_Y + \frac{\sum \alpha_j}{k} \Rightarrow \sum \alpha_j = 0$$

Como  $\sum \alpha_j = 0$  hay  $j-1$   $\alpha_j$  a ser estimados, y por ende hay  **$j+1$  parámetros**, y además  $\sigma_Y$  es independiente de  $\mu_Y$ .

Debido a  $\sum \alpha_j = 0$  hay que **imponer restricciones para poder estimar los parámetros**. La más frecuente de éstas es definir  $\alpha_1 = 0$  y denotar  $\alpha_j^* = \mu_{Y,REF} - \mu_{Y,j}$  como el diferencial o efecto del grupo  $j$  con respecto al grupo “de referencia”.

Este procedimiento se aplica de forma genérica a toda var. explicativa categórica de modelos lineales.

**SI**  $Y \sim \text{Normal}(\mu_{REF} + \alpha_j, \sigma_Y^2)$ , y

se emplearía  $g()$ : identidad

$$\rightarrow E(Y) = \mu_{REF} + \alpha_j$$

**SI**  $Y \sim \text{Gama}(\mu_{REF} + \alpha_j, \theta)$ ,

y se emplearía  $g()$ : log,

$$\rightarrow \log(E(Y)) = \mu_{REF} + \alpha_j$$

**En ambos casos**, el modelo bayesiano empleará otra función de enlace para “el otro parámetro”:  $\sigma_Y$  y  $\theta$ .

# Savage & West (2007) sueño de mamíferos según dieta.

```
DF2=read.csv(file="Pr 02 sleep mammals.csv", header=T)
```

## Análisis exploratorio de datos:

```
> summary(DF2[,c("vore", "sleep_total")])
```

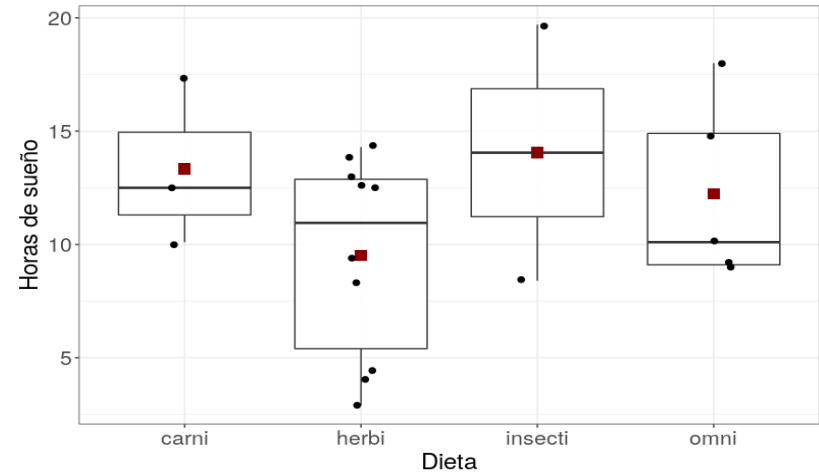
	vore	sleep_total
carni	:19	Min. : 1.90
herbi	:32	1st Qu.: 7.85
insecti	:5	Median :10.10
omni	:20	Mean :10.43
NA's	:7	3rd Qu.:13.75
		Max. :19.90

Conclusions

```
> desc.stats=function(x){c(mean=mean(x), median=median(x),sd=sd(x), n=length(x))}
```

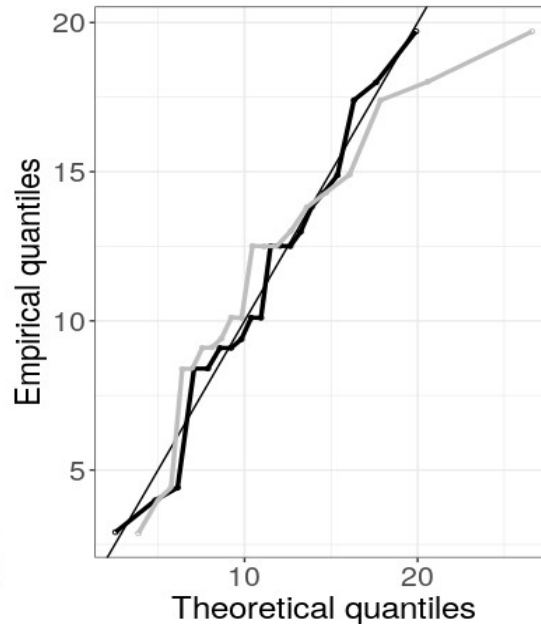
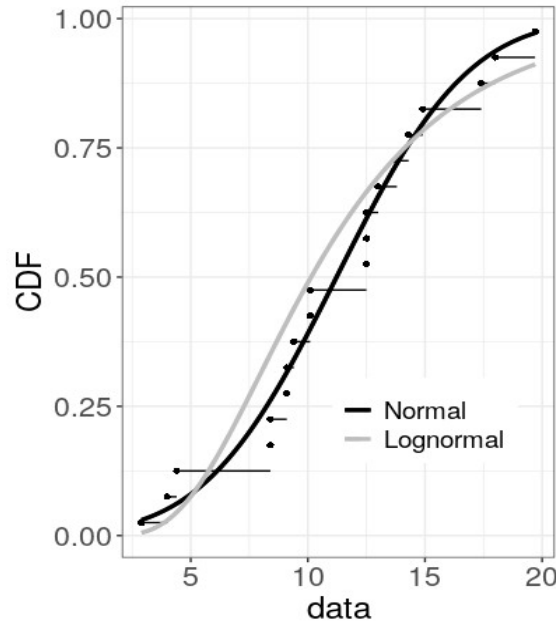
```
> summaryBy(sleep_total~vore, data=DF2, FUN=desc.stats)
```

	vore	sleep_total.mean	sleep_total.median	sleep_total.sd	sleep_total.n
1	carni	13.33	12.5	3.72	3
2	herbi	9.52	10.9	4.38	10
3	insecti	14.05	14.1	7.99	2
4	omni	12.24	10.1	4.02	5



```
ggplot(data=DF2, aes(x=vore, y=sleep_total))+  
  geom_boxplot(size=0.5, alpha=0.9)+  
  geom_jitter(size=2, width=0.1)+  
  stat_summary(fun = "mean", colour = "dark red",  
    size = 4, pch=15, geom = "point") + etc
```

## Distr prob. de la var respuesta:



```
norm=fitdist(DF2$sleep_total, "norm")  
lognorm=fitdist(DF2$sleep_total,"lnorm")
```

```
cdf.sleep=cdfcomp(list(norm,lognorm),  
  plotstyle ="ggplot")+ etc.
```

```
qq.sleep=qqcomp(list(norm,lognorm),  
  plotstyle ="ggplot")+ etc.
```

```
plot_grid(cdf.sleep, qq.sleep, ncol=2)
```

Conclusions

**Likelihood:**  $\text{sleep\_total} \sim \text{Normal}(\mu_{Y,j}, \sigma_Y)$

Hay ahora que definir las distr previas....

```
> get_prior(formula=sleep_total~vore, data=DF2,family=gaussian)
      prior      class      coef group resp dpar nlpar lb ub      source
  (flat)      b              (flat)      b              (vectorized)
  (flat)      b    voreherbi      b              (vectorized)
  (flat)      b  voreinsecti      b              (vectorized)
  (flat)      b    voreomni      b              (vectorized)
student_t(3, 11.3, 4) Intercept
student_t(3, 0, 4)      sigma              0      default
default
```

```
> levels(DF2$vore)
[1] "carni" "herbi" "insecti" "omni"
```

¿Nivel de referencia?

**Supongamos** que un mamífero podría dormir en promedio 9h/día y que es esperable un rango de  $\pm 40\%$  de variación.

i.e.  $9 \times 0.6 = 3.6\text{h}$  y  $9 \times 1.4 = 12.6\text{h}$ . Con este rango, se puede estimar  $\sigma_Y \sim (\text{max-min})/4 = 1.8\text{h}$  (Wan et al 2014).

Entonces, **intercept**  $\sim \text{Lognormal}(\text{mean}=\log(9), \text{sd}=??)$

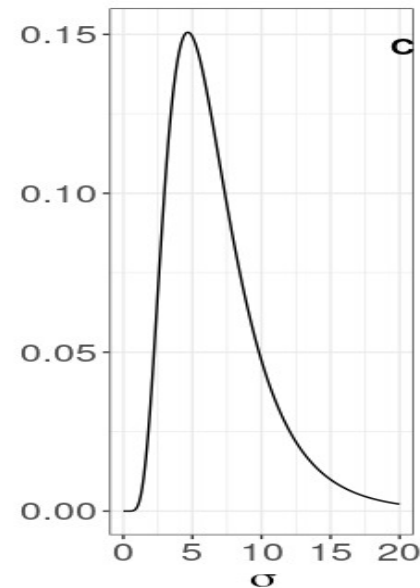
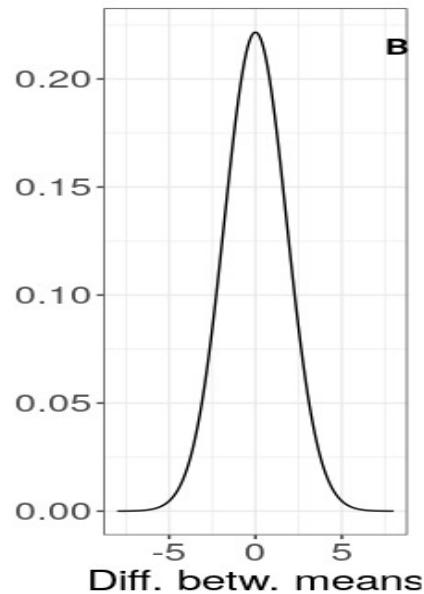
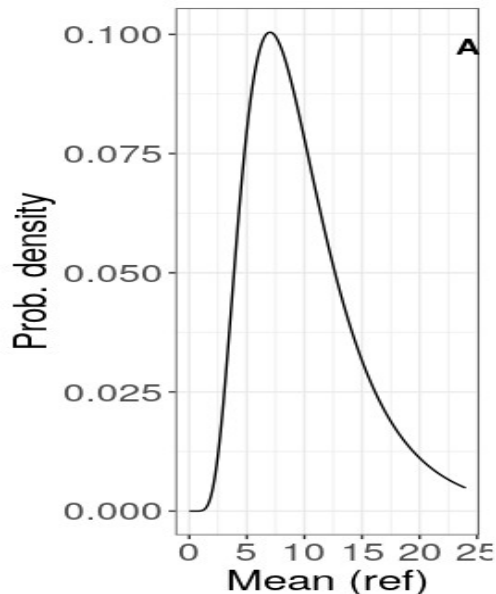
Lo determinaremos un poco por tanteo....



Los  $\alpha_j$  son **diferenciales de medias** con respecto al grupo de referencia (carni)  $\sim$  Normal (mean=0, sd=1.8)

La distr previa de  $\sigma_Y$  es más difícil **→ tanteo con gráficos??**

```
> prior.m3 = c(set_prior("lognormal(log(9), 0.5)", class = "Intercept"),  
+              set_prior("normal(0, 1.8)", class = "b"),  
+              set_prior("lognormal(log(6), 0.5)", class = "sigma"))
```



Es **CLAVE** entender el significado de los parámetros en el modelo!

**54 a 76**

## El modelo a ajustar:

```
m3.brms=brm(formula=sleep_total~vore, data=DF2, family=gaussian, prior=prior.m3,  
·····warmup=1000, future=TRUE, chains=3, iter=2000, thin=3)~
```

```
> summary(m3.brms)
```

Family: gaussian

Links: mu = identity; sigma = identity

Formula: sleep\_total ~ vore

Data: DF2 (Number of observations: 20)

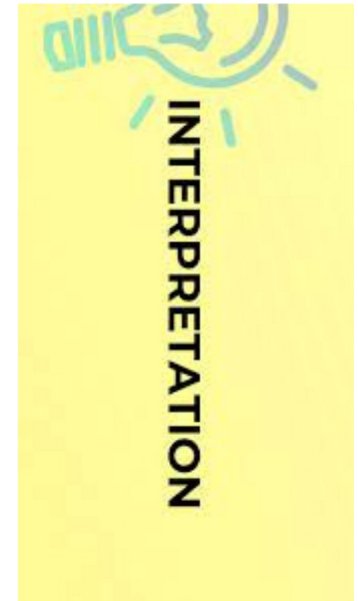
Draws: 3 chains, each with iter = 2000; warmup = 1000; thin = 3;  
total post-warmup draws = 1000

Population-Level Effects:

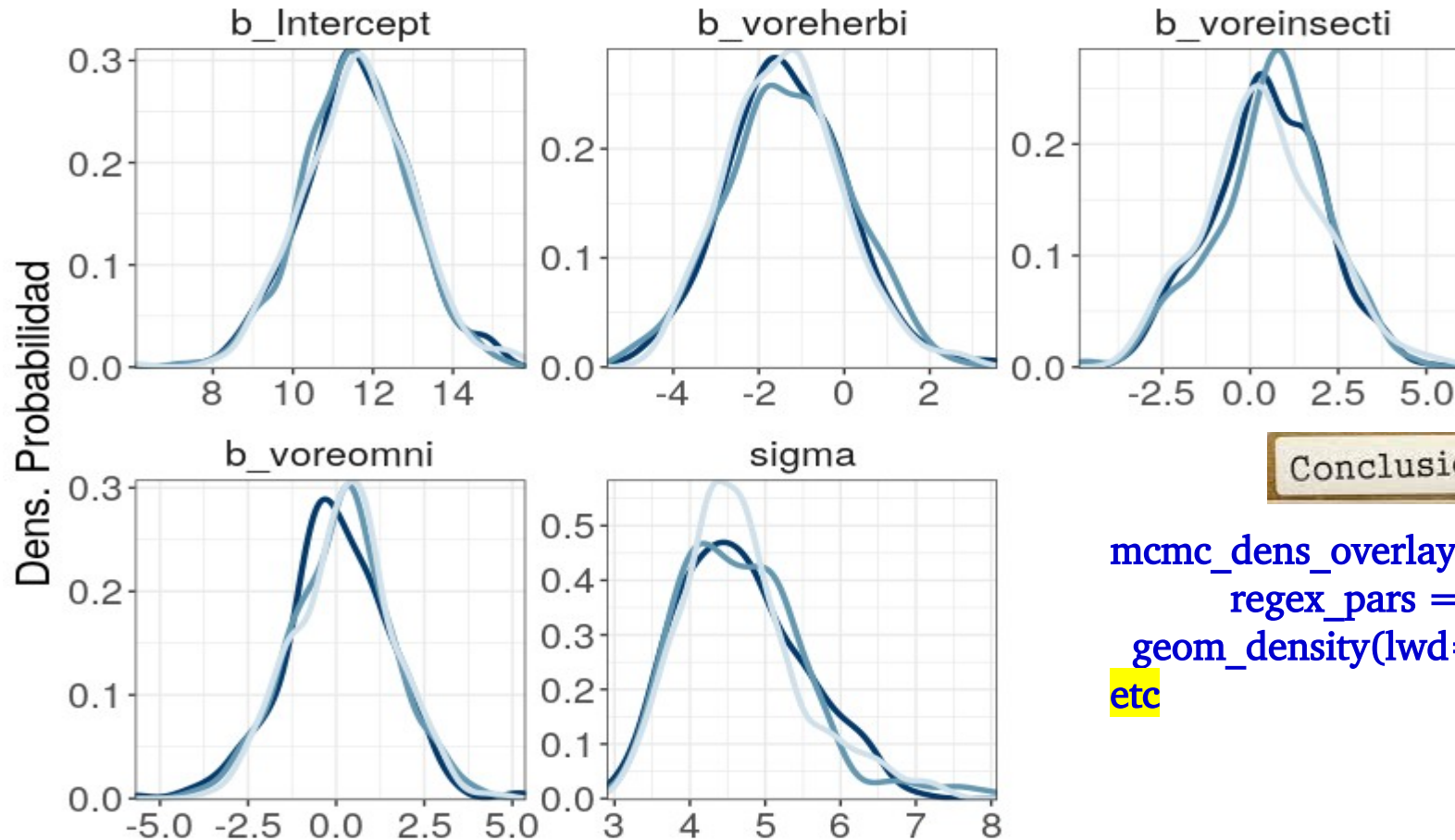
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	11.60	1.34	9.03	14.29	1.00	1045	947
voreherbi	-1.35	1.41	-3.99	1.49	1.00	952	899
voreinsecti	0.48	1.63	-2.65	3.69	1.00	950	955
voreomni	0.13	1.45	-2.76	2.91	1.00	958	899

Family Specific Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	4.69	0.82	3.40	6.58	1.00	1005	877



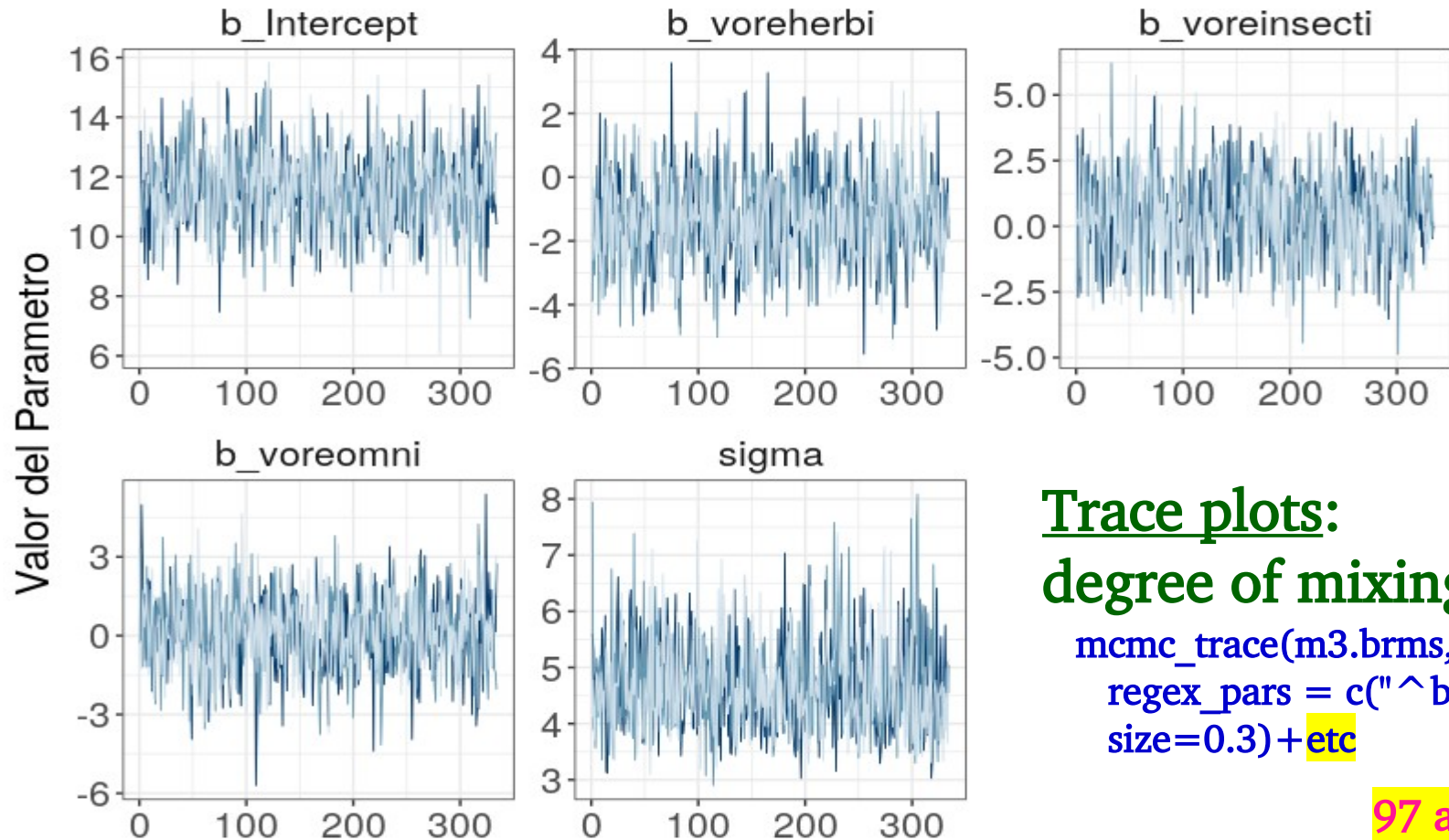
## Visualización de convergencia de las 3 cadenas:



Conclusions

```
mcmc_dens_overlay(m3.brms,  
  regex_pars = c("^b", "sigma"))+  
  geom_density(lwd=1.2, alpha=0.1)+  
  etc
```

85 a 94

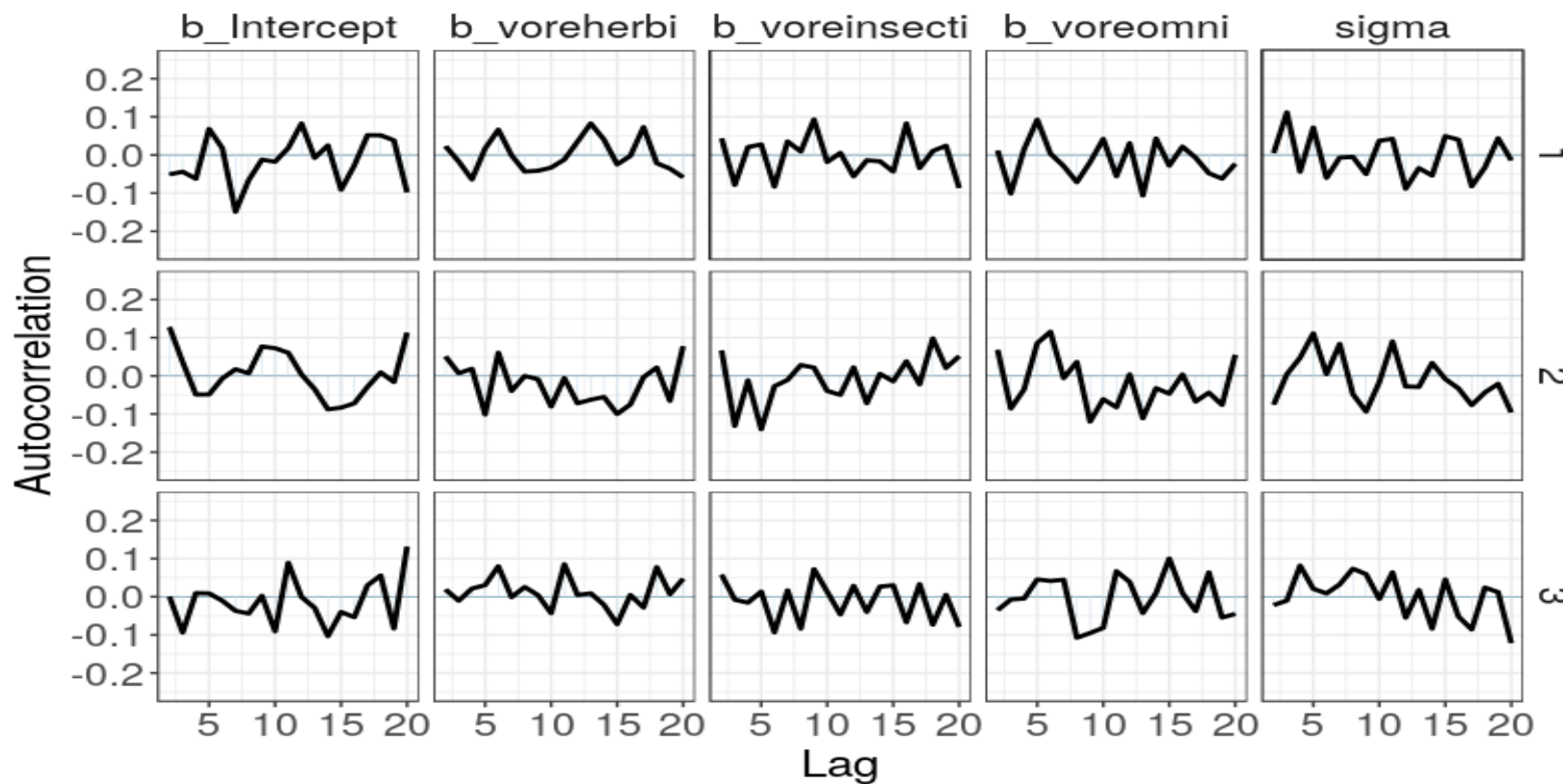


Trace plots:  
degree of mixing of chains

```
mcmc_trace(m3.brms,  
  regex_pars = c("^b", "sigma"),  
  size=0.3)+etc
```

97 a 104

## Grado de autocorrelación de los estimados en cada cadena:

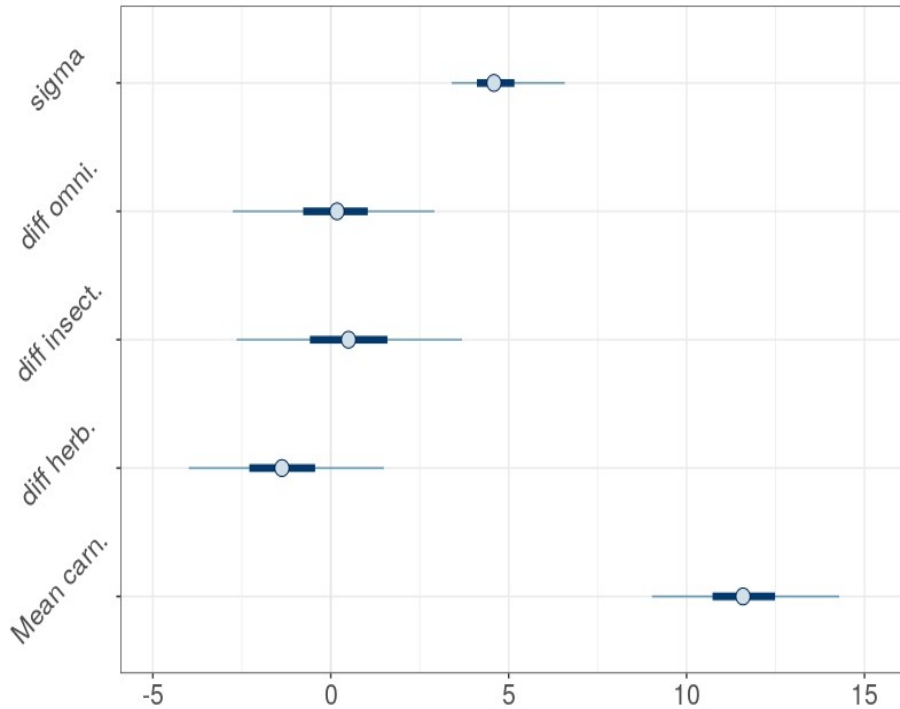


`mcmc_acf(m3.brms, regex_pars = c("^b", "sigma"))+etc`

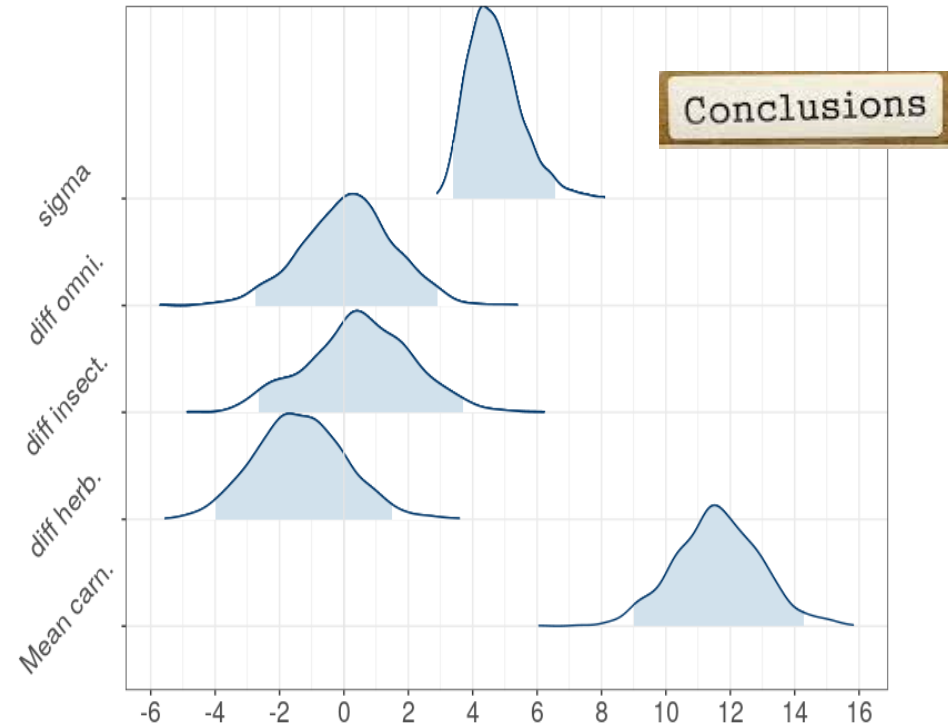
Conclusions

107 a 115

## Una mejor visualización del summary del modelo:



```
mcmc_plot(m3.brms, regex_pars = c("^b", "sigma"),  
  type="intervals", prob_outer = 0.95)+  
  scale_y_discrete(labels=c("Mean carn.", "diff herb.",  
    "diff insect.", "diff omni.", "sigma"))+ etc
```



```
mcmc_areas_ridges(m3.brms, regex_pars = c("^b",  
  "sigma"), prob_outer = 0.95)+ etc
```

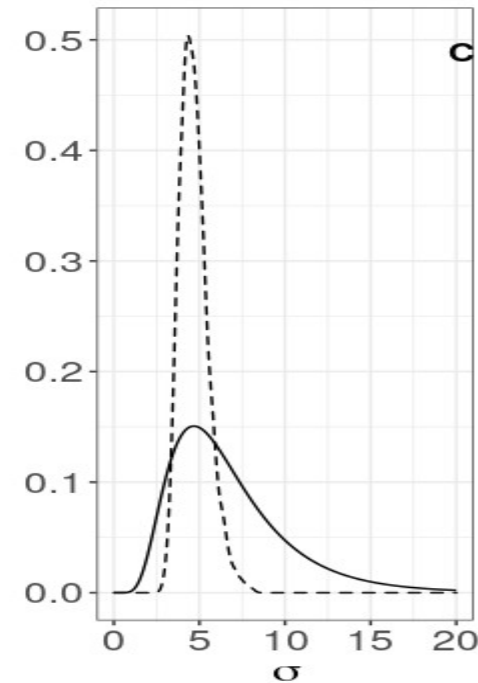
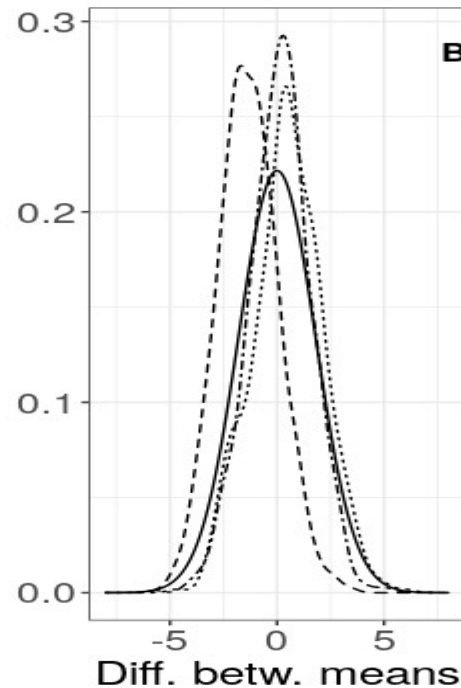
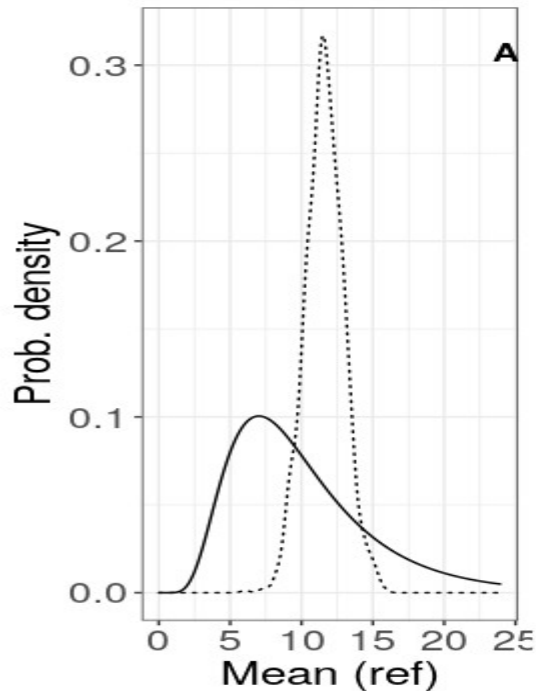
118 a 132



# Una visualización de cuánta información aportan los datos:

```
> head(post.m3brms,2)      post.m3brms=as_draws_df(m3.brms)
# A draws_df: 2 iterations, 1 chains, and 7 variables
  b_Intercept b_voreherbi b_voreinsecti b_voreomni sigma lprior lp__
1          13.6         -1.8           3.5        -0.31   5.6    -12   -69
2           9.8         -1.7          -2.7         4.98   5.3    -14  -74
```

CONCLUSION



142 a 151

A partir de la distr posteriores marginales podemos:  $\mu_{Y,j} = \mu_{REF} + \alpha_j^*$

```
> head(post.m3brms,2)
```

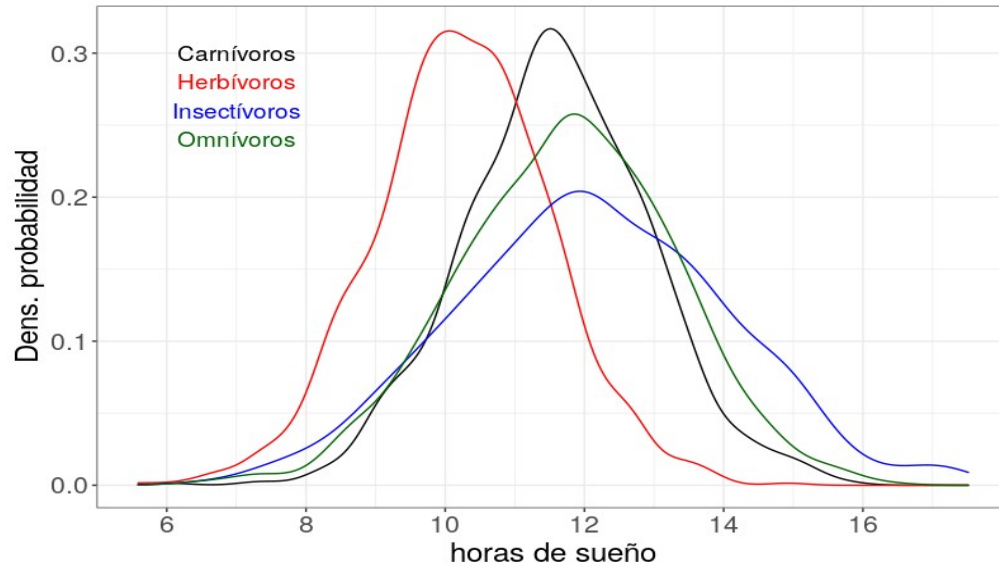
```
# A draws_df: 2 iterations, 1 chains, and 7 variab
```

	b_Intercept	b_voreherbi	b_voreinsecti	b_voreomni
1	13.6	-1.8	3.5	-0.31
2	9.8	-1.7	-2.7	4.98

```
> bayes_R2(m3.brms)
```

	Estimate	Est.Error	Q2.5	Q97.5
R2	0.076	0.0627	0.00388	0.23

```
mediasm3=data.frame(carni=post.m3brms$b_Intercept,~  
.....herbi=post.m3brms$b_Intercept+post.m3brms$b_voreherbi,~  
.....insecti=post.m3brms$b_Intercept+post.m3brms$b_voreinsecti,  
.....omni=post.m3brms$b_Intercept+post.m3brms$b_voreomni)~
```



A partir de la misma distr posterior, se puede así hacer cualquier tipo de cálculo y obtener la distr. de prob. asociada al resultado!

155 a 168



# Análisis de residuos: validación del ajuste a los datos

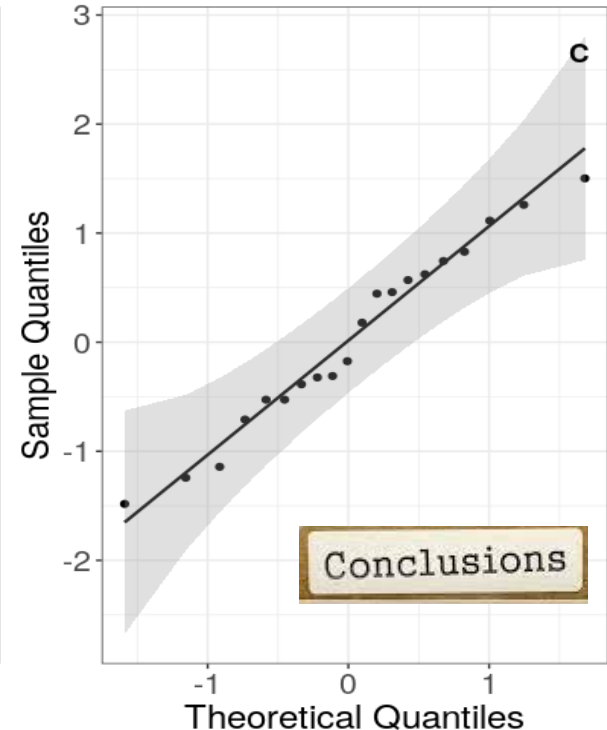
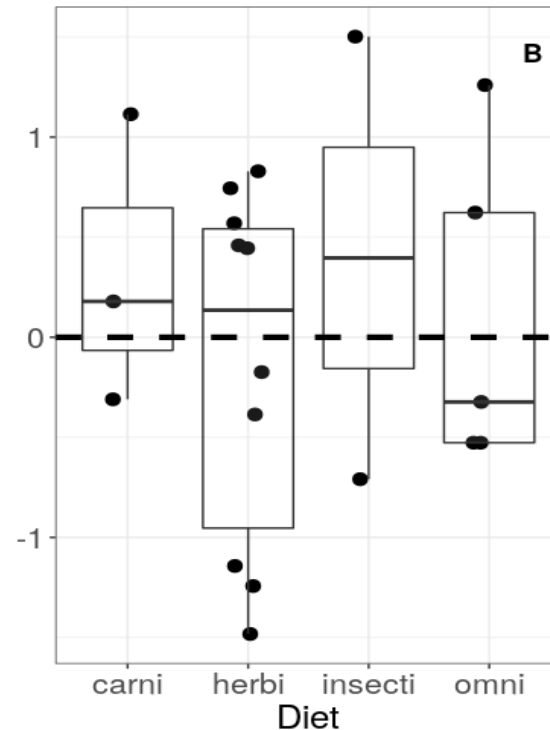
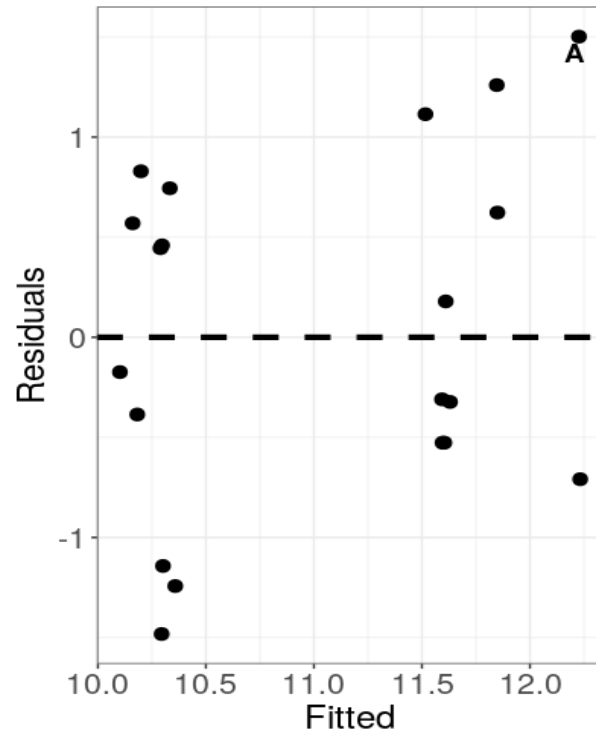
```
res.m3.brms=residuals(m3.brms, type="pearson", ndraws=1000, summary=T)
```

```
fit.m3.brms=fitted(m3.brms, scale="linear", summary=T, ndraws=1000)
```

```
DF2$fit=fit.m3.brms$Estimate
```

```
DF2$resid=res.m3.brms$Estimate
```

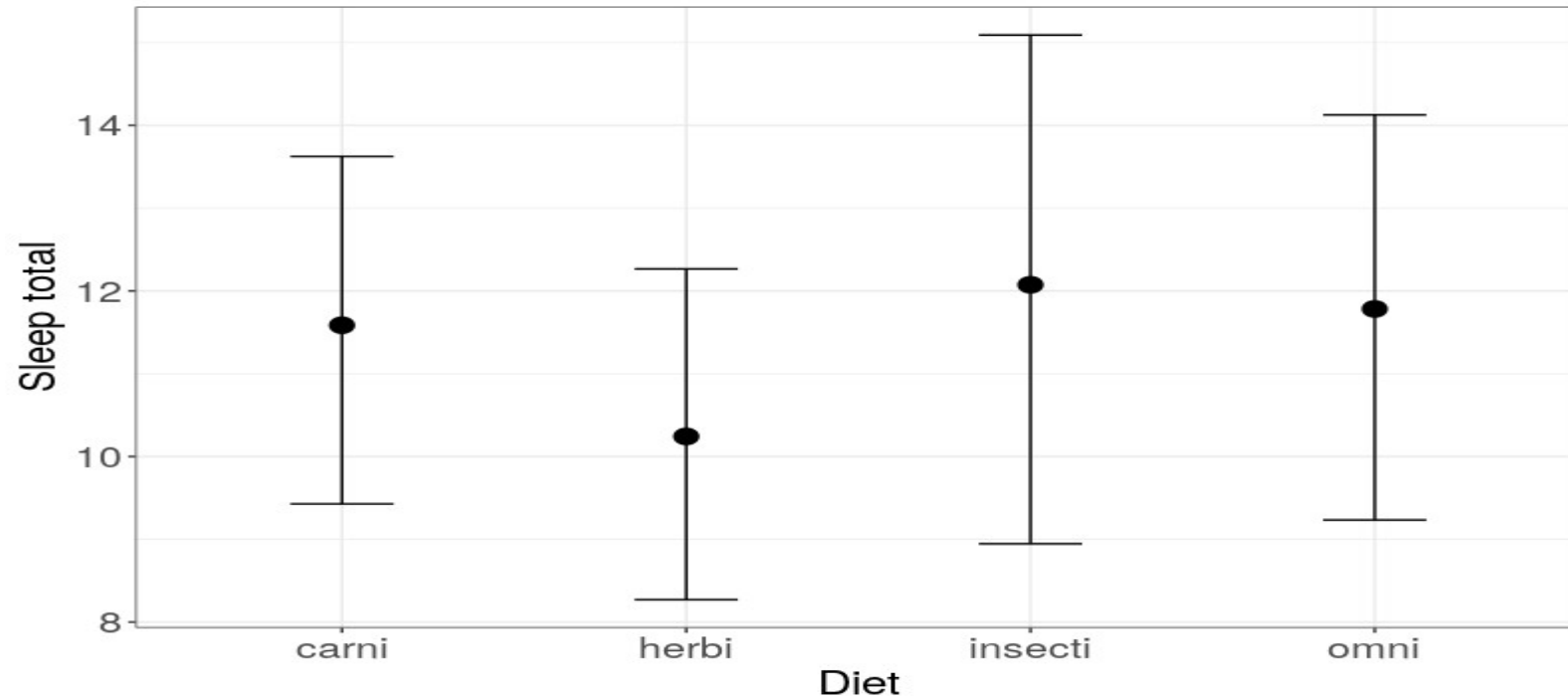
186 a 213



## Curva condicional:

```
plot(conditional_effects(m3.brms, effects="vore", prob=0.89))[[1]] +  
labs(y="Sleep total", x="Diet") + etc
```

215 a 217



## b) Tests de hipótesis y tests a posteriori

Habría un **consenso implícito** de que **no sería necesario** hacer correcciones tipo Bonferroni al hacer tests a posteriori comparando niveles de una var explicativa categórica.

**¿Por qué?**

1. Los marcos teóricos frecuentista y bayesiano tienen flujos de información en dirección contraria.

Frecuentista: el  $p$ -valor =  $\Pr(\text{datos extremos posibles} | H_0 \text{ cierta})$  se obtiene de distr muestrales suponiendo  $H_0$  fija. Los ajustes del  $p$ -valor corrigen por falsos positivos ( $R H_0$ ) cuando los mismos datos potenciales son usados en muchos contrastes.

Bayesiano: no hay p-valor ni distr muestrales. Son los datos ya obtenidos los que están fijos y se usan una sola vez. Se obtiene una distr posterior que es invariante y refleja totalmente nuestro conocimiento actual del problema.

2. En estadística bayesiana se pueden crear modelos jerárquicos con dependencias entre parámetros a varios niveles (shrinkage) que minimiza el riesgo de detectar diferencias entre niveles de la var categórica.

El grado de información de las distr. previas permitiría “regularizar” y así prevenir la sobre-estimación de efectos.

→ Gelman et al 2012, 2014, Kruschke 2015, Banner et al 2019, Lemoine 2019, Mc Elreath 2019 y también Ogle et al 2019...

Y si se desea, se pueden realizar “tests a posteriori”:

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	11.60	1.34	9.03	14.29
voreherbi	-1.35	1.41	-3.99	1.49
voreinsecti	0.48	1.63	-2.65	3.69
voreomni	0.13	1.45	-2.76	2.91

$$\text{herbi} \equiv \mu_{Y, \text{REF}} - \mu_{Y, \text{herb}}$$

$$\text{insecti} \equiv \mu_{Y, \text{REF}} - \mu_{Y, \text{insecti}}$$

$$\text{herbi-insecti} \equiv \mu_{Y, \text{insecti}} - \mu_{Y, \text{herbi}}$$

Que es mejor ver como **contrastes respecto a cualquier valor** (ej.  $\text{herbi} > 1.4$ ), no solamente respecto a cero.

```
> hypothesis(m3.brms, "voreherbi>1.4", class="b")
```

Hypothesis Tests for class b:

	Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio	Post.Prob	Star
1	(voreherbi) - (1.4) > 0	-2.75	1.41	-5.01	-0.36	0.03	0.03	

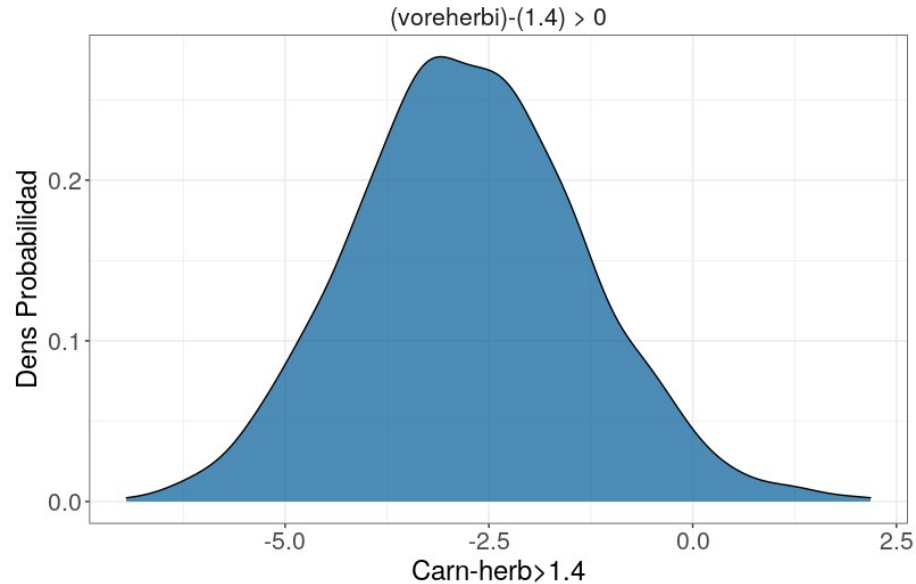
$$\text{Evidence ratio} = \frac{\text{Prob}(\text{hipótesis})}{\text{Pr}(\text{su complemento})} = \frac{\text{Pr}(\mu_{Y, \text{Carn}} - \mu_{Y, \text{Herb}} > 1.4)}{\text{Pr}(\mu_{Y, \text{Carn}} - \mu_{Y, \text{Herb}} < 1.4)}$$

← Bayes Factor

Todo se calcula a partir de la distr posterior antes obtenida

```
plot(hypothesis(m3.brms, "voreherbi > 1.4", class="b"))[[1]] +  
labs(x="Carn-herb > 1.4", y="Dens Probabilidad") + etc
```

224 a 231



Se pueden hacer varios  
contrastes al mismo tiempo  
y visualizar las 3 distr  
posteriores resultantes...

```
> h = c("carn>herb"="Intercept> Intercept+voreherbi",  
+       "carn>insect"="Intercept> Intercept+voreinsecti",  
+       "carn>prom.herb.insect"="Intercept> (Intercept+voreherbi+Intercept+voreinsecti)/2")  
> hypothesis(m3.brms, h, class="b")
```

Hypothesis Tests for class b:

	Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio	Post.Prob	Star
1	carn>herb	1.35	1.41	-1.04	3.61	5.22	0.84	
2	carn>insect	-0.48	1.63	-3.12	2.33	0.55	0.35	
3	carn>prom.herb.insect	0.43	1.12	-1.34	2.32	1.80	0.64	

240 a 247

El test de hipótesis bayesiano se basa en el cálculo de **Bayes factor (BF)** para las 2 hipótesis que se cotejan.

En el primer ejemplo, consideremos:

$$H_0: \text{dif} > 1.4 \text{ vs } H_1: \text{dif} < 1.4$$

Recordemos:

Dist posterior

Verosimilitud

Dist previa

$$Pr(\text{Param}|\text{datos}) \approx Pr(\text{datos}|\text{Param}) * Pr(\text{Param})$$

Evidencia  
actual

Información  
en los datos

Evidencia  
previa

Suponiendo  $H_i$  cierta, para cada hipótesis:

$$Pr(\text{Param si } H_i|\text{datos}) \approx Pr(\text{datos}|\text{Param si } H_i) * Pr(\text{Param si } H_i)$$

## Calculando el cociente para ambas hipótesis:

$$\frac{Pr(\text{Param si } H_0 | \text{datos})}{Pr(\text{Param si } H_1 | \text{datos})} \approx \frac{Pr(\text{datos} | \text{Param si } H_0) * Pr(\text{Param si } H_0)}{Pr(\text{datos} | \text{Param si } H_1) * Pr(\text{Param si } H_1)}$$

**Cociente dists posteriores**  $BF =$  **Cociente de verosimilitudes** **Cociente dists previas**

**BF: cociente de verosimilitudes** → cuan más probable es cada hipótesis después vs. antes de examinar los datos (Jeffreys 1961). **BF > 1:**  $H_0$  más probable que  $H_1$

```
> hypothesis(m3.brms, "voreherbi>1.4", class="b")
```

```
Hypothesis Tests for class b:
```

	Hypothesis	Estimate	Est.Error	CI.Lower	CI.Upper	Evid.Ratio	Post.Prob	Star
1	(voreherbi)-(1.4) > 0	-2.75	1.41	-5.01	-0.36	0.03	0.03	



El BF se puede calcular con el cociente Savage-Dickey (dens posterior/dens previa para valores específicos de los parámetros; Wagenmakers et al 2010, Heck 2019, Ly et al 2020).

Aquí no hay  
significancia estadística,  
ni p-valores, ni \*\*\*

No muy frecuentemente  
utilizado (mi opinión).

Interpretation	
100	Extreme evidence for $M_1$
100	Very strong evidence for $M_1$
30	Strong evidence for $M_1$
10	Moderate evidence for $M_1$
3	Anecdotal evidence for $M_1$
	No evidence
1	Anecdotal evidence for $M_2$
1/3	Moderate evidence for $M_2$
1/10	Strong evidence for $M_2$
1/30	Very strong evidence for $M_2$
1/100	Extreme evidence for $M_2$

### 3. Análisis bayesiano I:

### Práctico 02

a) GLM con variables categóricas

b) Tests de hipótesis y tests a posteriori

## Referencias:

Banner et al (2020) Meth Ecol Evol 11:882-889

Heck (2019) Br J Math and Stat Psyc 72: 316-333

Jeffreys (1961) Theory of probability (2<sup>nd</sup> ed) Cambridge Univ Press

Ly et al (2020) Comput Brain & Behavior 3: 153-161

Gelman et al (2012) J Res Educ Effect 5:189-211

Gelman et al (2014) Bayesian data analysis (2nd ed). CRC Chapman & Hall

Kruschke (2015) Doing Bayesian Data Analysis a tutorial with R, JAGS and Stan. (2<sup>nd</sup> ed). Academic Press

Lemoine (2019) Oikos 128: 912–928.

Mc Elreath (2019) Statistical rethinking (2<sup>nd</sup> ed) CRC Chapman & Hall

Ogle et al (2019) Meth Ecol Evol 10:553–564

Savage & West (2007) PNAS 104:1051-1056.

Wagenmakers et al (2010) Cognitive Psych. 60:158-189.

Wan et al (2004) BMC Medical Res Meth 14:135-147