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

2. Elementos básicos del análisis bayesiano

Teórico 02

a. Modelo Lineal General: regresión múltiple.
b. Distribuciones previas I:
c. GLM en 6 dispositivas.

a. Modelo Lineal General: regresión múltiple.

El Modelo Lineal General supone que $Y \sim Normal(\mu_Y, \sigma_Y)$ y que μ_Y tiene relación lineal con vars. explicativas categóricas y/o numéricas. μ_Y y σ_Y son independientes.

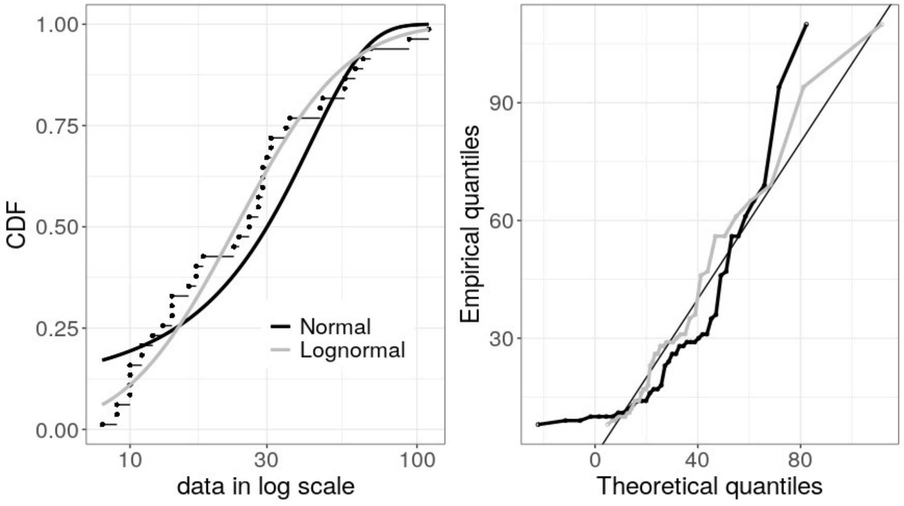
Forma general de proceder: 1) distr. de var de respuesta, 2) relaciones entre var respuesta y explicativa(s), 3) formular modelo y las dist previas, 4) ajustar el modelo, 5) evaluar la calidad del ajuste, 6) curvas condicionales.

Hagamos un 1er análisis bayesiano de regresión múltiple.

Los datos: $S0_2$ en 42 ciudades de USA en función de vars. explicativas numéricas (clima e industria).

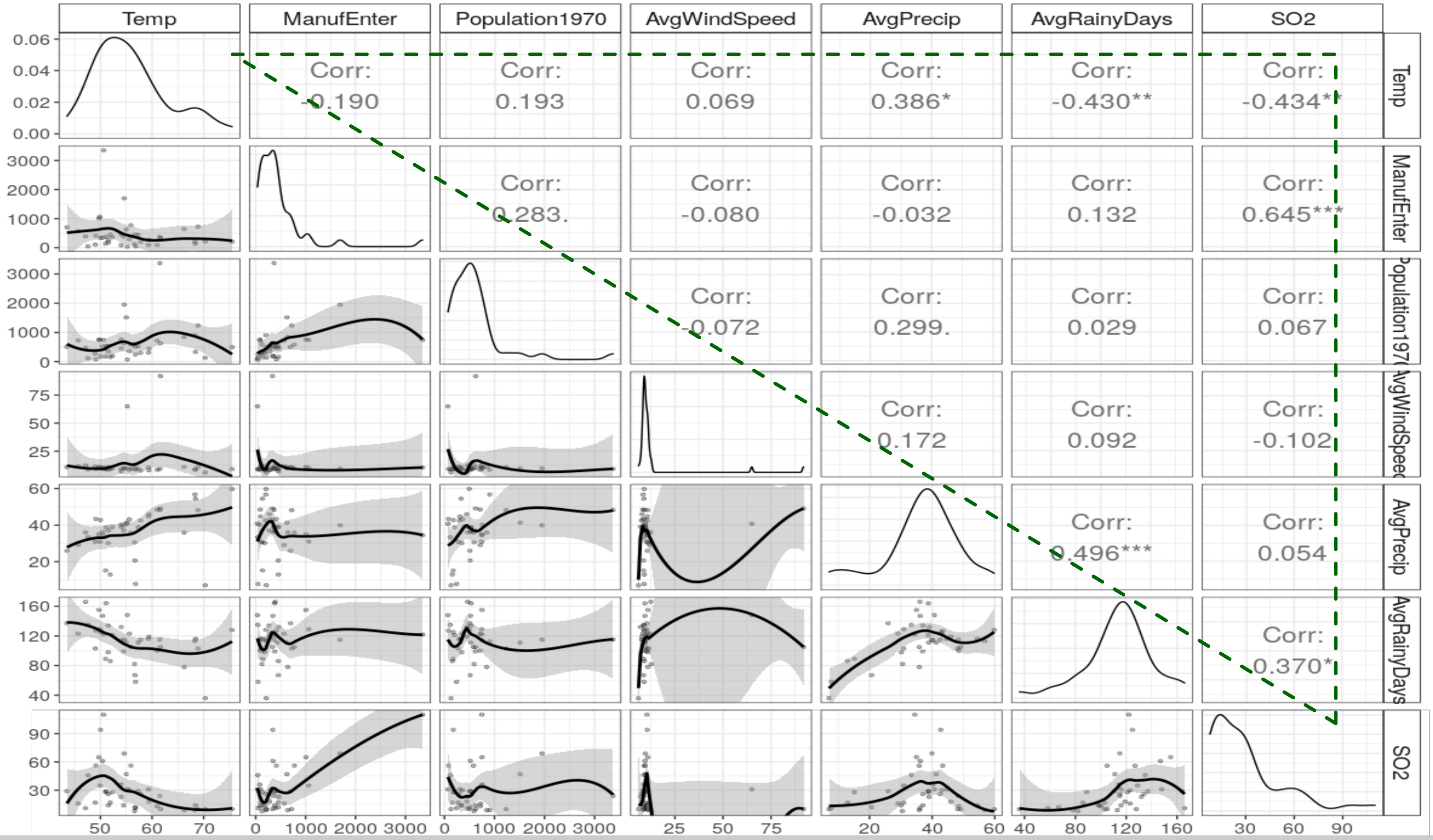
```
> DF1= read.csv("Teo 02 regr multiple.csv", header=T)
> summary(DF1)
         City         SO2              Temp         ManufEnter      Population1970  AvgWindSpeed     AvgPrecip       AvgRainyDays
 Albany     : 1   Min.   :  8   Min.   :43.5   Min.   :  35   Min.   :  71   Min.   : 6.0   Min.   : 7.0   Min.   : 36
 Albuquerque: 1   1st Qu.: 13   1st Qu.:50.6   1st Qu.: 181   1st Qu.: 299   1st Qu.: 8.8   1st Qu.:31.0   1st Qu.:103
 Atlanta    : 1   Median : 26   Median :54.6   Median : 347   Median : 515   Median : 9.4   Median :38.7   Median :115
 Baltimore  : 1   Mean   : 30   Mean   :55.8   Mean   : 463   Mean   : 609   Mean   :12.9   Mean   :36.8   Mean   :114
 Buffalo    : 1   3rd Qu.: 35   3rd Qu.:59.3   3rd Qu.: 462   3rd Qu.: 717   3rd Qu.:10.6   3rd Qu.:43.1   3rd Qu.:128
 Charleston : 1   Max.   :110   Max.   :75.5   Max.   :3344   Max.   :3369   Max.   :92.0   Max.   :59.8   Max.   :166
```

1) Distr. de var de respuesta: → Likelihood del modelo




```
require(fitdistrplus)
norm.m2=fitdist(DF1$SO2,"norm")
lognorm.m2=fitdist(DF1$SO2,"lnorm")
cdf.m2=cdfcomp(list(norm.m2,lognorm.m2),
   xlogscale=T,  plotstyle ="ggplot")+etc
qq.m2=qqcomp(list(norm.m2,lognorm.m2),
     plotstyle ="ggplot"))+etc
grid.arrange(cdf.m2, qq.m2, ncol=2)
```

2) Relaciones entre var respuesta y explicativa(s)




```
ggpairs(DF1[,c("Temp","ManufEnter","Population1970",
"AvgWindSpeed","AvgPrecip","AvgRainyDays", "SO2")],
lower=list(continuous = wrap("smooth_loess", alpha =
0.3,size=1)), upper = list(continuous = wrap("cor",
size=6)))+etc
```

3) Modelo estadístico y distr previas:

Y~Lognormal, pero hagamos logY~Normal.

8 distr previas a especificar

DF1$logSO2=log(DF1$SO2)   logSO2~Normal(μ_Y,σ_Y)

$\mu_Y=\beta_0+\beta_1X_1+\ldots+\beta_6X_6$: 7 parámetros + σ_Y

Estandardizar las vars explicativas a media=0, sd=1.

```
> DF1s=scale(DF1[,c("Temp","ManufEnter","Population1970","AvgWindSpeed","AvgPrecip","AvgRainyDays")], center=T, scale=T)
> summary(DF1s)
      Temp            ManufEnter      Population1970     AvgWindSpeed        AvgPrecip          AvgRainyDays
 Min.   :-1.697   Min.   :-0.76   Min.   :-0.93   Min.   :-0.45   Min.   :-2.525   Min.   :-2.939
 1st Qu.:-0.714   1st Qu.:-0.50   1st Qu.:-0.53   1st Qu.:-0.27   1st Qu.:-0.493   1st Qu.:-0.411
 Median :-0.161   Median :-0.21   Median :-0.16   Median :-0.23   Median : 0.167   Median : 0.041
 Mean   : 0.000   Mean   : 0.00   Mean   : 0.00   Mean   : 0.00   Mean   : 0.000   Mean   : 0.000
 3rd Qu.: 0.489   3rd Qu.: 0.00   3rd Qu.: 0.19   3rd Qu.:-0.15   3rd Qu.: 0.539   3rd Qu.: 0.532
 Max.   : 2.731   Max.   : 5.11   Max.   : 4.77   Max.   : 5.14   Max.   : 1.956   Max.   : 1.965
> DF1s=as.data.frame(cbind(logSO2=DF1$logSO2, DF1s))
```

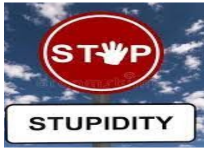
Permite comparar los efectos relativos de cada var explicativa, y el intercepto es ahora interpretable.

```
> get_prior(formula=logSO2~Temp+ManufEnter+Population1970+AvgWindSpeed+AvgPrecip+AvgRainyDays,
+           data=DF1s,family=gaussian)
                prior     class           coef group resp dpar nlpar lb ub       source
               (flat)         b                                                 default
               (flat)         b      AvgPrecip                             (vectorized)
               (flat)         b   AvgRainyDays                             (vectorized)
               (flat)         b   AvgWindSpeed                             (vectorized)
               (flat)         b     ManufEnter                             (vectorized)
               (flat)         b Population1970                             (vectorized)
               (flat)         b           Temp                             (vectorized)
 student_t(3, 3.3, 2.5) Intercept                                               default
   student_t(3, 0, 2.5)     sigma                                   0           default
```

Útil para ver las previas a especificar

brms: por defecto "previas débilmente informativas"

Ello implica que en brms se podría ajustar un modelo sin especificar las distr previas.



No hay razón para renunciar al aspecto clave (pero siempre debatido y discutido) de la inferencia bayesiana.

El comando set_prior permite especificar dist previas para todas set_prior("normal(0, 2)", class = "b") las pendientes o para cada una set_prior("normal(0,1)", class = "b", coef = "Temp") por separado.

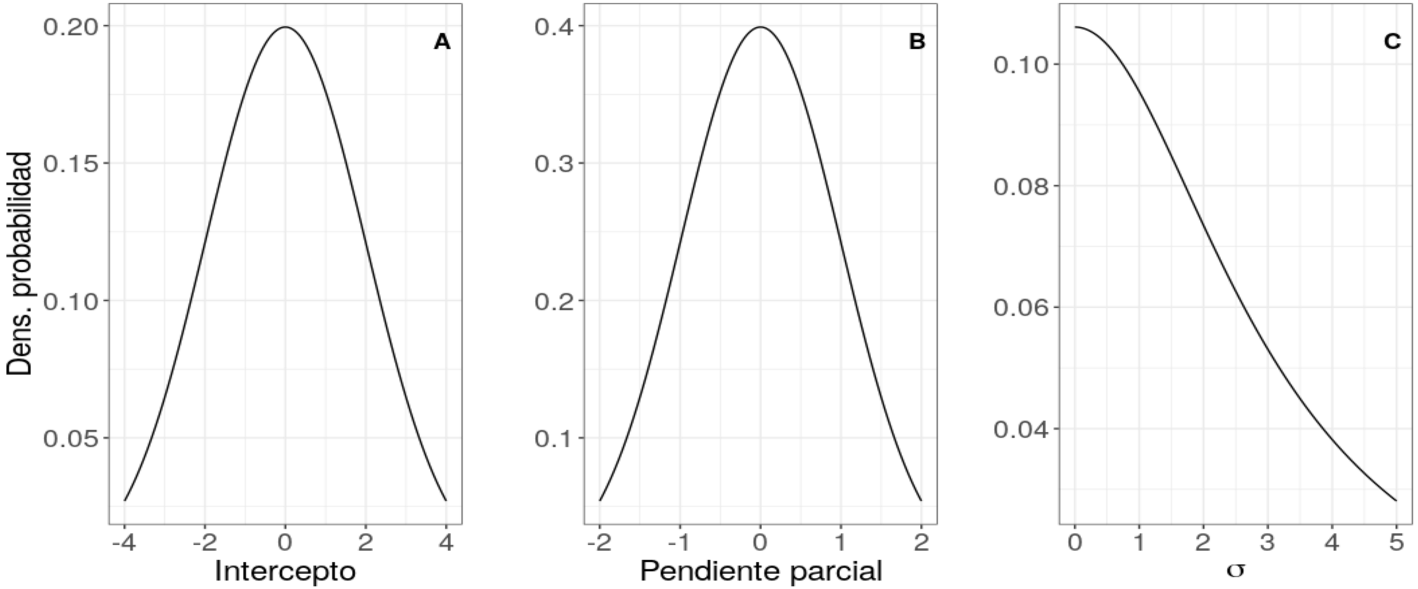
Las dist previas se definen ANTES Y SIN ver los datos.

Reflejan el conocimiento previo acumulado del problema bajo estudio reflejado en el conjunto de valores plausibles de los parámetros de un modelo estadístico a ajustar.

Definí set_prior("normal(0, 2)", class = "intercept") → 95% de los valores de logSO2 estarán en [-4; 4]→ [-exp(4)= 0.183; exp(4)=54.6], ¿un amplio intervalo?

Definí `set_prior("normal(0, 2)", class = "b")` ➞ 95% de los valores de las pendientes (estandarizadas) estarán en [-2; 2]. efectos pueden ser positivos o negativos...

Definí `set_prior("cauchy(0, 3)", class = "sigma")` ➞ como $\sigma_Y > 0$, se usa "half Cauchy". IMPORTANTE GRAFICAR PREVIAS



THE CODE.

```
int.m2=ggplot(data.frame(x = c(-4, 4)), aes(x)) +
  stat_function(fun = dnorm, n = 1000, args = list(mean=0, sd=2)) + etc
slope.m2=ggplot(data.frame(x = c(-2, 2)), aes(x)) +
  stat_function(fun = dnorm, n = 1000, args = list(mean=0, sd=1)) + etc
sd.m2=ggplot(data.frame(x = c(0, 5)), aes(x)) +
  stat_function(fun = dcauchy, n = 1000, args = list(location=0, scale=3)) + etc
plot_grid(int.m2, slope.m2, sd.m2, ncol=3,labels = LETTERS[1:3],
        align="hv",label_x=0.90, label_y=0.95)
```

$$\frac{(2000-1000)*3 \text{ cad}}{\text{thin}=3} = \frac{1000}{\text{valores}}$$

```
prior.m2 = c(set_prior("normal(0, 2)", class = "Intercept"),
             set_prior("normal(0, 1)", class = "b"),
             set_prior("cauchy(0,3)", class = "sigma"))
```

← Las distr previas:

```
m2.brms=brm(formula=logSO2~Temp+ManufEnter+Population1970+
AvgWindSpeed+AvgPrecip+AvgRainyDays,  data=DF1s, family=gaussian, prior =
prior.m2, warmup = 1000, chains=3, iter=2000, thin=3, future=T)
```

Se generarán 3 cadenas (de Markov) de 2000 iteraciones, descartándose las 1eras 1000 de cada una, y reteniendo 1 de cada 3 valores en cada cadena.

4) Ajustar el modelo.

```
> summary(m2.brms)
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: logSO2 ~ Temp + ManufEnter + Population1970 + AvgWindSpeed + AvgPr
   Data: DF1s (Number of observations: 41)
  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          3.15      0.08     2.98     3.30 1.00      927     1039
Temp              -0.34      0.14    -0.62    -0.07 1.00      998      789
ManufEnter         0.27      0.10     0.06     0.45 1.00     1015      990
Population1970    -0.03      0.10    -0.22     0.16 1.00      960      955
AvgWindSpeed      -0.08      0.09    -0.26     0.09 1.00      960      945
AvgPrecip          0.15      0.15    -0.14     0.43 1.00      945      920
AvgRainyDays       0.09      0.15    -0.19     0.37 1.00      981      856

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.54      0.07     0.43     0.71 1.00      980      938

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Una distr posterior de 8 dimensiones (parámetros)

↓

Estadísticos de las distr. marginales de cada parámetro.

↓

media, sd, cuantiles 2.5 y 97.5% (Int Credibilidad 95%).

```
> summary(m2.brms)
Population-Level Effects:
               Rhat Bulk_ESS Tail_ESS
Intercept      1.00      927     1039
Temp           1.00      998      789
ManufEnter     1.00     1015      990
Population1970 1.00      960      955
AvgWindSpeed   1.00      960      945
AvgPrecip      1.00      945      920
AvgRainyDays   1.00      981      856

Family Specific Parameters:
      Rhat Bulk_ESS Tail_ESS
sigma 1.00      980      938
```

Convergencia de las cadenas:

$$R=\sqrt{\frac{Var_{Within}+\frac{1}{iter}\left(Var_{Betw}-Var_{Within}\right)}{Var_{Within}}}$$

$1<R<1.1 \rightarrow$ OK

$$ESS=\frac{iter}{1+2\sum\rho_i}$$

ESS~1000→ OK

Opciones: aumentar iter y/o thin

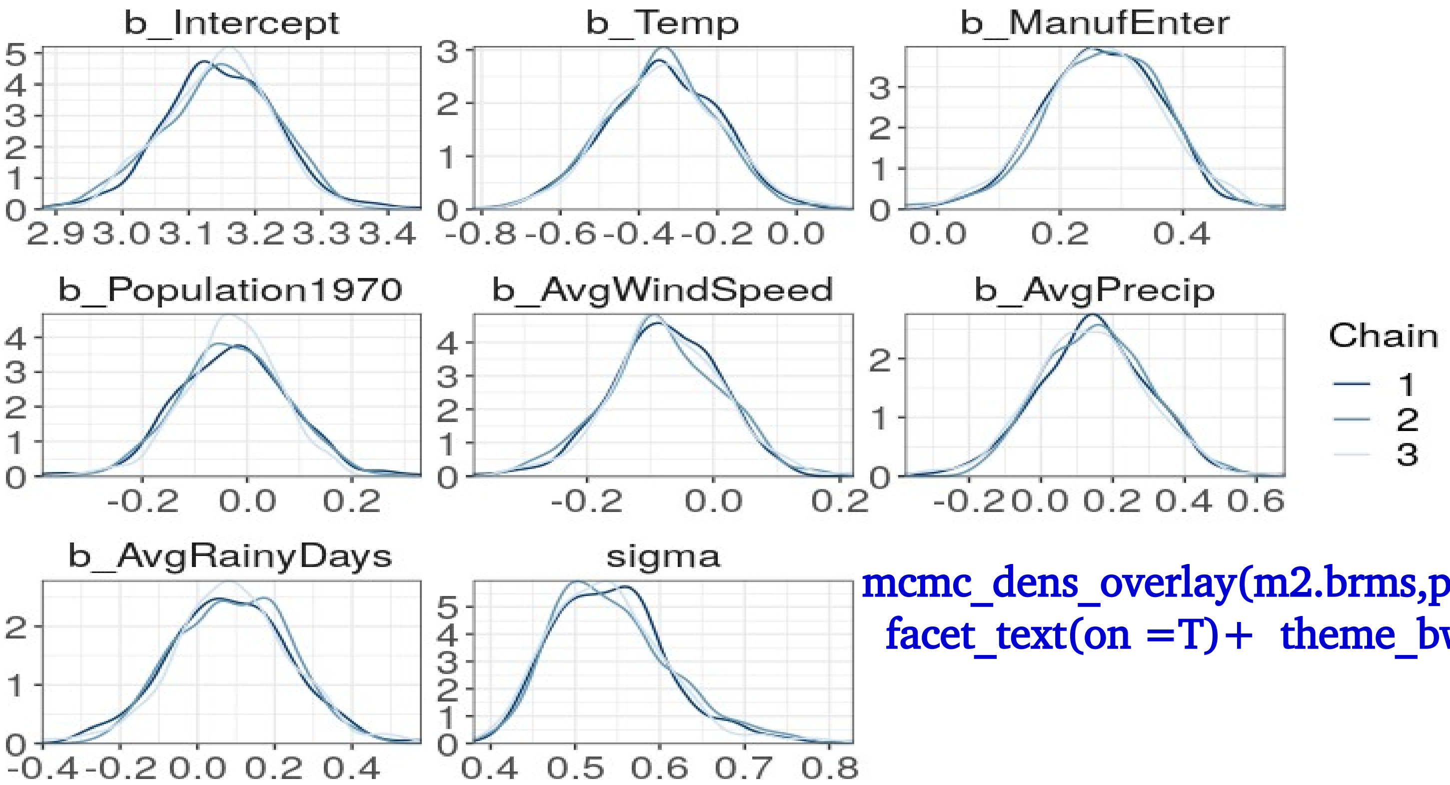
Las 3 cadenas deben converger a la misma distr. posterior estacionaria y tener muchos estimados de los parámetros con baja (ej. $<|0.143|$) autocorrelación.

→ verificación de convergencia del algoritmo es necesaria para fusionar las 3 cadenas empleadas.

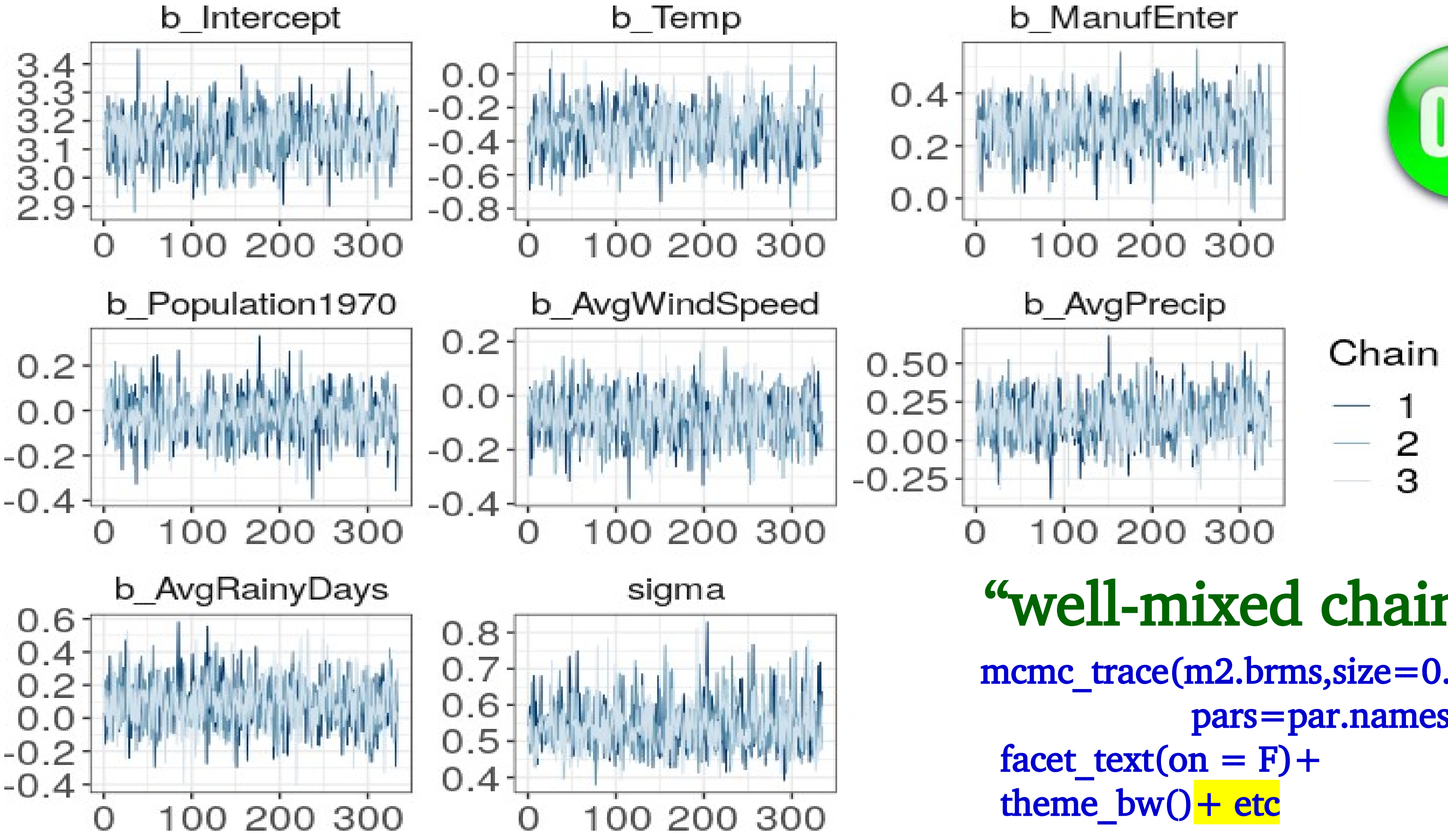
GRÁFICOS

```
> variables(m2.brms)
 [1] "b_Intercept"      "b_Temp"           "b_ManufEnter"      "b_Population1970" "b_AvgWindSpeed"
 [6] "b_AvgPrecip"      "b_AvgRainyDays"   "sigma"
```

par.names.m2=variables(m2.brms)[1:length(variables(m2.brms))-2]



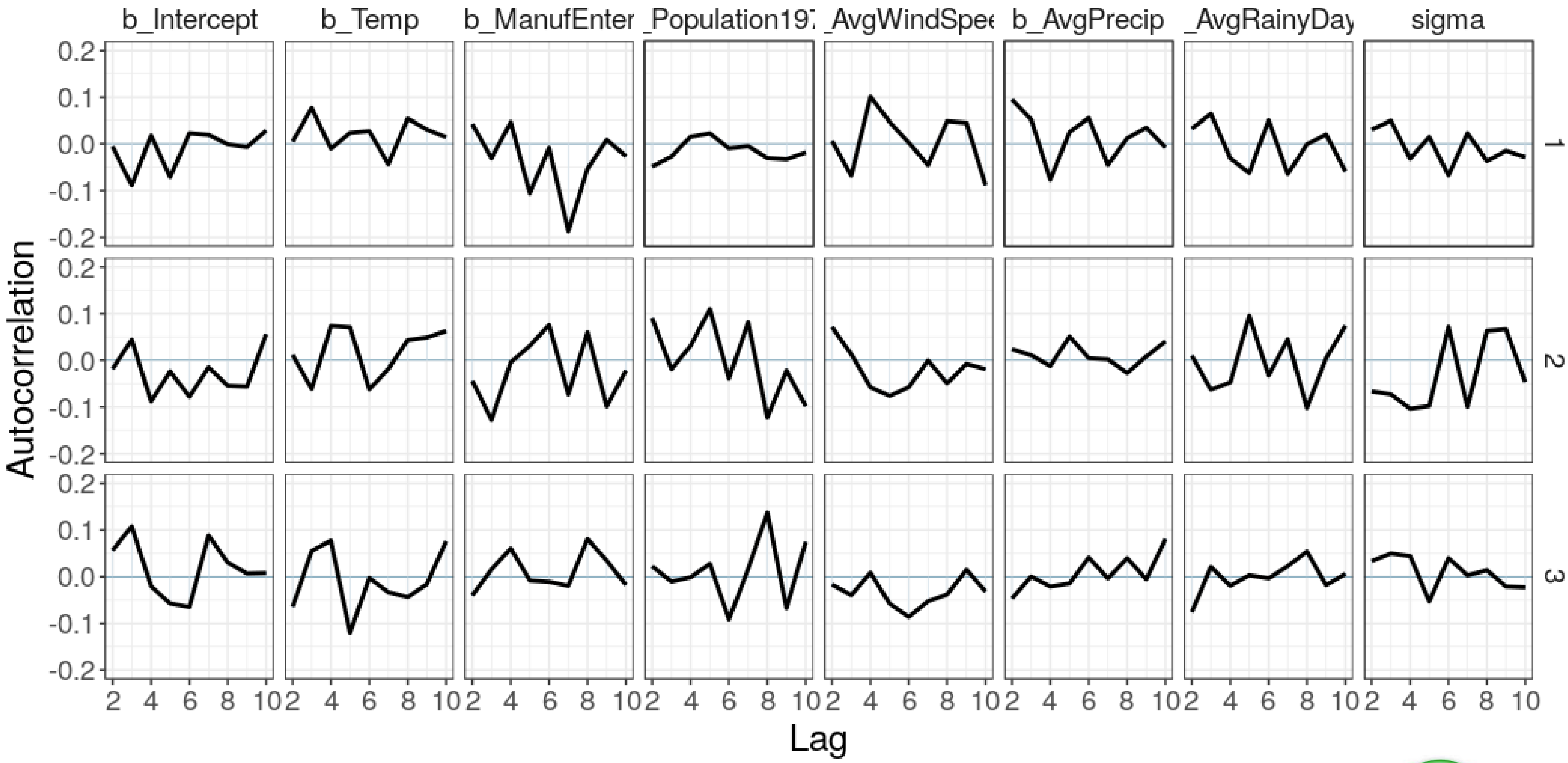
mcmc_dens_overlay(m2.brms,pars=par.names.m2)+
facet_text(on =T)+ theme_bw()+ etc



OK

“well-mixed chains”

```
mcmc_trace(m2.brms,size=0.3,
           pars=par.names.m2)+
  facet_text(on = F)+
  theme_bw()+ etc
```



mcmc_acf(m2.brms, pars=par.names.m2)+ scale_x_continuous(limits=c(2,10))+ scale_y_continuous(limits=c(-0.2,0.2))+ etc

OK

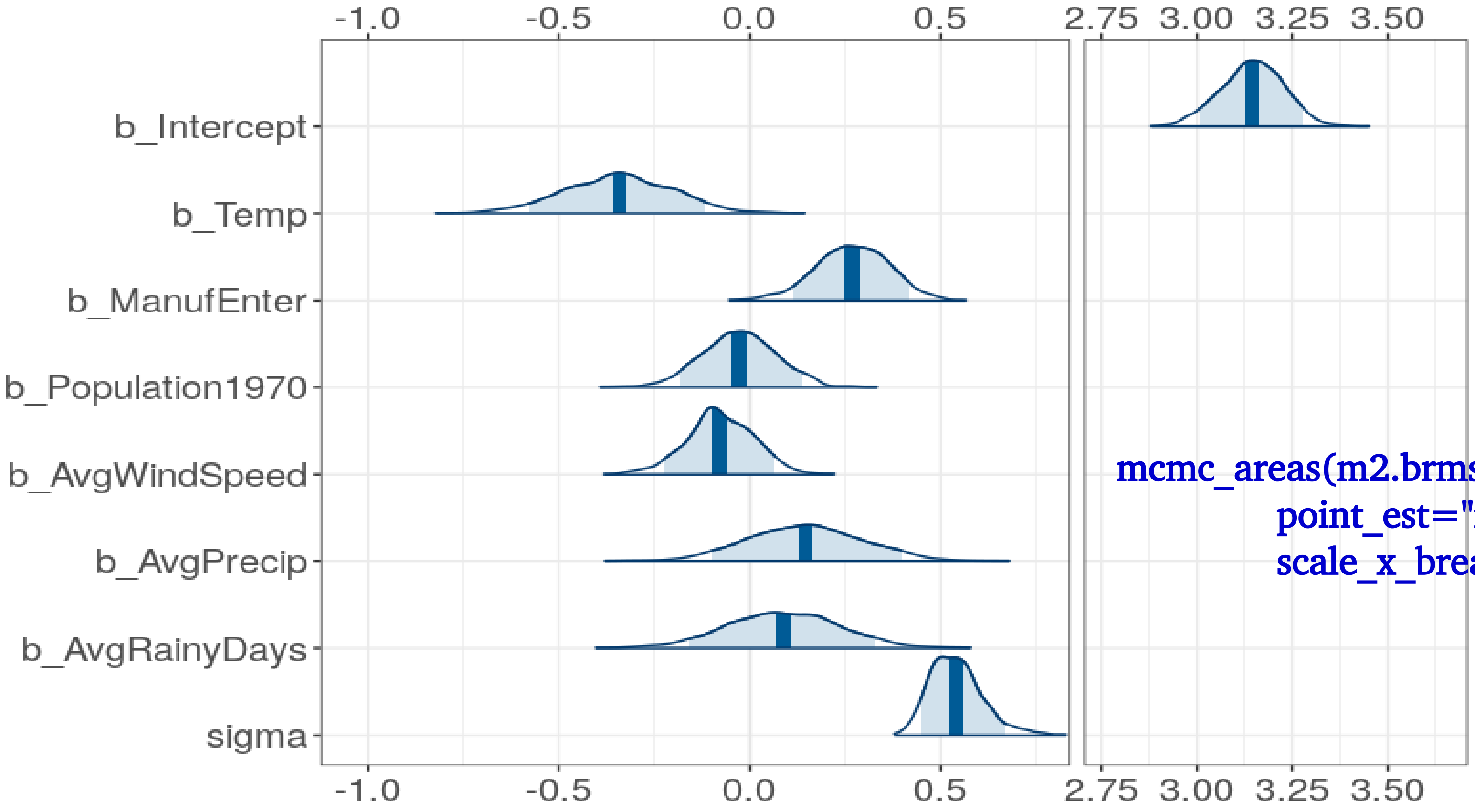
Exploremos un poco más la convergencia de las cadenas:

```
> dist.post.m2=as_draws_df(m2.brms,pars=pars.m2.brms, add_chain = F)
> names(dist.post.m2)
 [1] "b_Intercept"      "b_Temp"           "b_ManufEnter"     "b_Population1970" "b_AvgWindSpeed"
 [6] "b_AvgPrecip"      "b_AvgRainyDays"   "sigma"            "lprior"           "lp__"
[11] ".chain"           ".iteration"       ".draw"
```

```
> summaryBy(b_Intercept+b_Temp+b_ManufEnter+b_Population1970+b_AvgWindSpeed+b_AvgPrecip+
+           b_AvgRainyDays+sigma~.chain, FUN=c(mean), data=dist.post.m2)
# A tibble: 3 × 9
  .chain b_Intercept.mean b_Temp.mean b_ManufEnter.mean b_Population1970.mean b_AvgWindSpeed.mean b_AvgPrecip.mean b_AvgRainyDays.mean sigma.mean
   <int>            <dbl>       <dbl>             <dbl>                 <dbl>               <dbl>            <dbl>               <dbl>      <dbl>
1      1             3.15      -0.339             0.268               -0.0265             -0.0777            0.146              0.0828      0.543
2      2             3.15      -0.349             0.274               -0.0259             -0.0792            0.156              0.0922      0.544
3      3             3.14      -0.337             0.266               -0.0286             -0.0726            0.135              0.0903      0.538
> summaryBy(b_Intercept+b_Temp+b_ManufEnter+b_Population1970+b_AvgWindSpeed+b_AvgPrecip+
+           b_AvgRainyDays+sigma~.chain, FUN=sd, data=dist.post.m2)
# A tibble: 3 × 9
  .chain b_Intercept.sd b_Temp.sd b_ManufEnter.sd b_Population1970.sd b_AvgWindSpeed.sd b_AvgPrecip.sd b_AvgRainyDays.sd sigma.sd
   <int>          <dbl>     <dbl>           <dbl>               <dbl>             <dbl>          <dbl>             <dbl>    <dbl>
1      1         0.0842     0.143          0.0921               0.106            0.0839          0.155             0.154   0.0674
2      2         0.0855     0.140          0.0973              0.0988            0.0909          0.144             0.141   0.0706
3      3         0.0784     0.148          0.0968              0.0876            0.0885          0.156             0.152   0.0687
```

No es necesario hacer esto cada vez...

Distribuciones posteriores de los parámetros:



mcmc_areas(m2.brms,pars=par.names.m2,
point_est="mean", prob = 0.9)+
scale_x_break(c(0.75,2.75))+ etc

Gelman et al (2018) propusieron un métrico:

$$R^2 = \frac{Variance\left(Y_{fitted}\right)}{Variance\left(Y_{fitted}\right) + Variance\left(Y_{residuals}\right)}$$

Obtenida de la distr predictiva posterior (que explicaré luego)

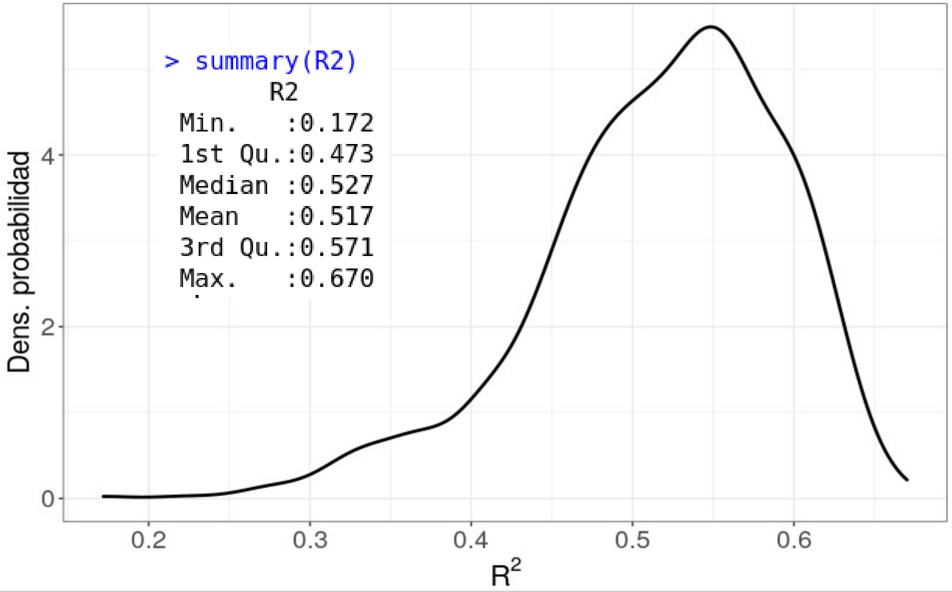
```
> str(posterior_predict(m2.brms))
 num [1:1002, 1:41] 2.258 0.745 2.519 1.479 1.996
 - attr(*, "dimnames")=List of 2
```

```
> str(DF1)
'data.frame':   41 obs. of  8 variables:
```

```
R2=data.frame(R2=bayes_R2(m2.brms,
                          summary=F))

ggplot(data=R2, aes(x=R2))+
  geom_density(size=1)+
  labs(x=expression(paste("R"^2)),
       y="Dens. probabilidad")+etc
```

```
> bayes_R2(m2.brms)
   Estimate Est.Error  Q2.5 Q97.5
R2    0.517     0.076 0.335 0.632
```



5) Evaluar la calidad del ajuste

Análisis de residuos y posterior predictive checks (luego)

res.m2.brms=as.data.frame(residuals(m2.brms, type="pearson", ndraws=1000, summary=T))

```
> str(res.m2.brms)
'data.frame':	41 obs. of  4 variables:
 $ Estimate : num  0.825 -0.463 -0.482 -0.247 0.723
 $ Est.Error: num  0.58 0.309 0.329 0.349 0.349 ...
 $ Q2.5     : num  -0.3249 -1.0494 -1.1397 -0.9378 (
 $ Q97.5    : num  1.984 0.144 0.161 0.469 1.42 ...
```

```
> head(fit.m2.brms,2)
  Estimate Est.Error Q2.5 Q97.5
1     1.77     0.377 1.01  2.51
2     2.84     0.182 2.48  3.18
```

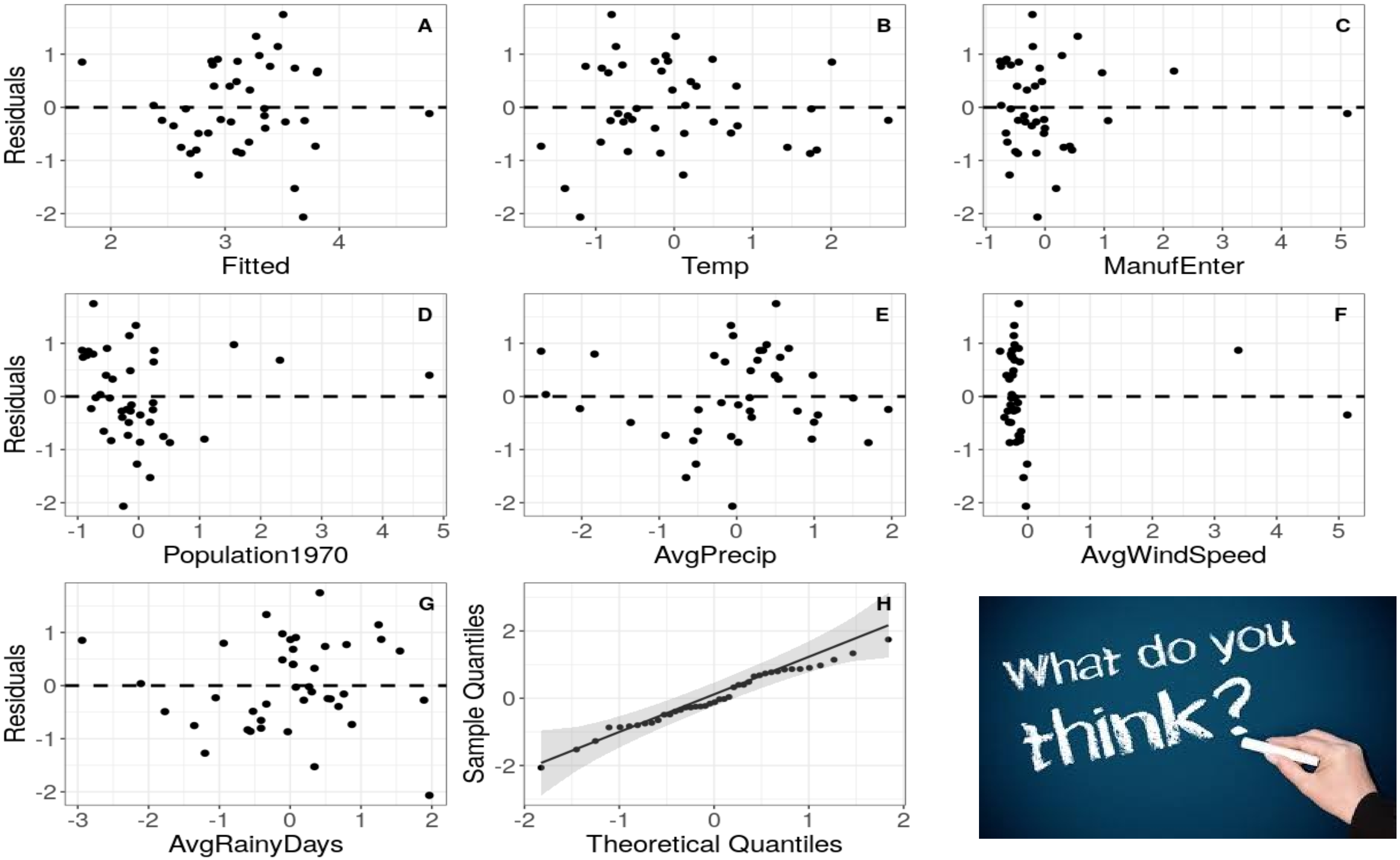
fit.m2.brms=as.data.frame(fitted(m2.brms, scale="linear", summary=T, ndraws=1000))

Poniendo todo en DF1s para hacer los gráficos:

DF1s$resid=res.m2.brms$Estimate
DF1s$resid.Q2.5=res.m2.brms$Q2.5
DF1s$resid.Q97.5=res.m2.brms$Q97.5
DF1s$fit=fit.m2.brms$Estimate
DF1s$fit.Q2.5=fit.m2.brms$Q2.5
DF1s$fit.Q97.5=fit.m2.brms$Q97.5

Los gráficos:

Resid vs fitted
Resid vs vars explicativas
QQplot Resid


A
B
C
D
E
F
G
H
Residuals
Fitted
Temp
ManufEnter
Population1970
AvgPrecip
AvgWindSpeed
AvgRainyDays
Sample Quantiles
Theoretical Quantiles
What do you think?

THE CODE.

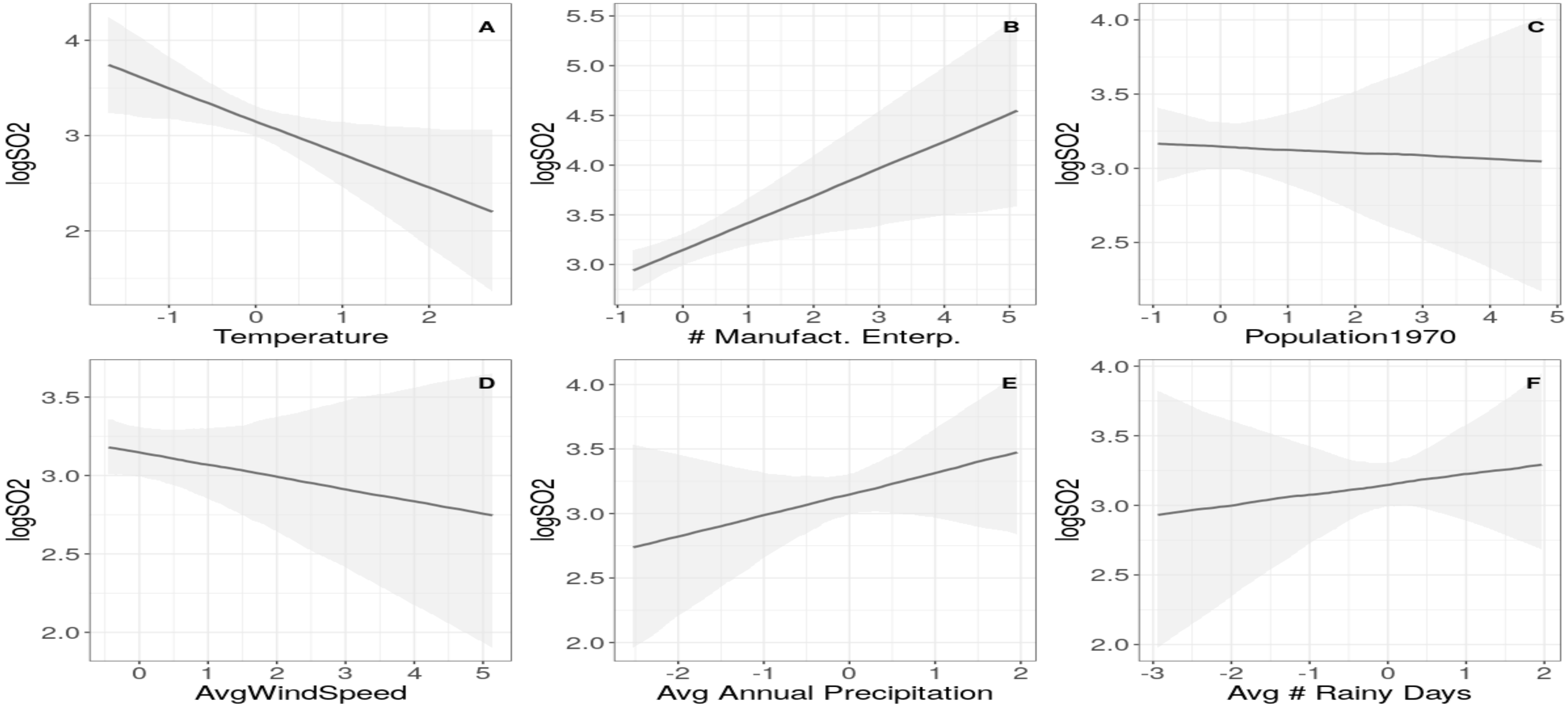
```
res.fit.m2.brms=ggplot(data=DF1s, aes(x=fit,y=resid))+
  geom_point(col="black", size=2)+
  geom_hline(yintercept =0,linetype = 2, size=1.1)+etc
res.Temp.m2.brms=ggplot(data=DF1s, aes(x=Temp,y=resid))+
  geom_point(col="black", size=2)+
  geom_hline(yintercept =0,linetype = 2, size=1.1)+etc
(idem para las otras vars explicativas)
qq.m2.brms=ggplot(data=DF1s, mapping=aes(sample = resid)) +
  stat_qq_point()+ stat_qq_line()+ stat_qq_band(alpha=0.3)+          qqplotr
  labs(x = "Theoretical Quantiles", y = "Sample Quantiles")+etc
plot_grid(res.fit.m2.brms, res.Temp.m2.brms,res.ManufEnter.m2.brms,
        res.Population1970.m2.brms,res.AvgPrecip.m2.brms,res.AvgWindSpeed.m2.brms,
        res.AvgRainyDays.m2.brms, qq.m2.brms,ncol=3,labels = LETTERS[1:8],
        align="hv",label_x=0.90, label_y=0.95)
```

6) Curvas condicionales

```
m2.brms.cond.eff=conditional_effects(m2.brms)
```

Relaciones predichas para cada var. explicativa.

```
> names(m2.brms.cond.eff)
[1] "Temp"            "ManufEnter"      "Population1970" "AvgWindSpeed"    "AvgPrecip"
[6] "AvgRainyDays"
```



ggplot(data=m2.brms.cond.eff$Temp, aes(x=Temp, y=estimate__))+ geom_line(size=1)+
geom_ribbon(aes(ymin=lower__,ymax=upper__),fill = "grey90", alpha=0.5) + etc

b. Distribuciones previas I:

La gran discusión es cómo moderar la influencia relativa del conocimiento previo y de los aspectos particulares de los datos en los resultados, i.e. en las distr. posteriores.

El mundo NO COMIENZA DE NUEVO con cada análisis como se pretende de "forma objetiva" la estadística frecuentista → dist previas uniformes. (todo es posible)

La estadística bayesiana carece de un procedimiento general para definir las dist previas de clases de modelos.

Hay un inevitable grado de discrecionalidad o arbitrariedad remanente al definir las distr. previas.

Hay al menos 3 grandes razones para ello:

a) No es obvio decidir qué es un antecedente válido a la hora de definir las dist. previas de los parámetros.

b) No es trivial cómo armonizar/ponderar diferentes fuentes de evidencia previa de "expertos" en un tema.

c) Dista de ser evidente cómo transformar evidencia previa (i.e. media, sd, rangos) en distr. de probabilidad.

No obstante, hay enfoques e ideas recientes al respecto algunas provenientes de la psicología: O'Hagan (2019), Paoli & van de Shoot (2017), Veen et al (2017) y otros.

Las dist. previas podrían ser "demasiado importantes" :
* cuando hay pocos datos para la complejidad del modelo a ajustar ← ¿frecuente en ciencias naturales y sociales?
* cuando los datos tienen gran variabilidad e "detalles idiosincráticos" en modelos estadísticos complejos.

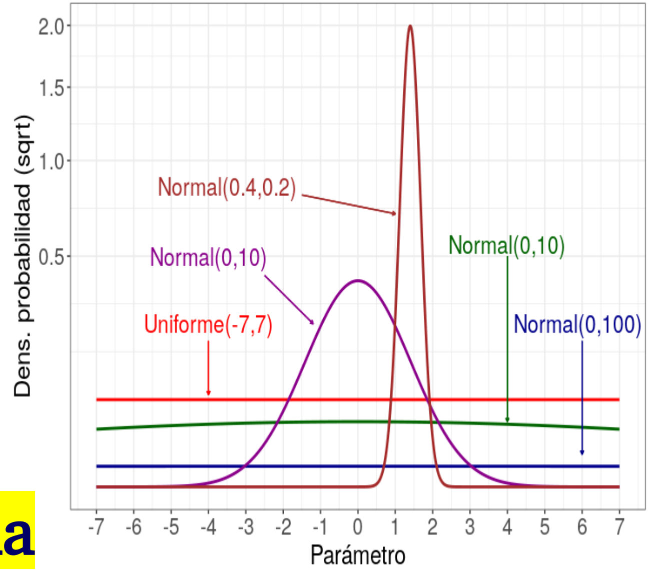
↳ Riesgo: un modelo de buen ajuste, pero no generalizable.

Regularización del modelo: usar las distr. previas para "impedir que el modelo aprenda demasiado de los datos", una noción bastante difusa y/o metafísica en mi opinión!

En esta visión, sería aún más importante tener principios claros y defendibles en la definición de las distr. previas...

De forma MUY general, se puede categorizar las distr. previas de acuerdo al grado de información que proveen sobre los valores plausibles de cada parámetro.

* no informativas: ~ Uniforme(min, max) ← "impropias"
* vagas: ~ Normal(0, 100)
* débilmente informativas: ~ Normal(0, 10)
* más informativas: ~ Normal(0, 1)
* específica, fuertemente informativas: ~ Normal(0.4, 0.2) o escaladas.



Ver: van Zwet & Gelman (2021), DePaoli & van de Shoot (2017), Gelman et al (2008), Banner et al (2020), Seaman et al (2011), Lemoine (2019), etc.

Tomar en cuenta el impacto de la función de enlace para cada parámetro.

Hay un par de enfoques no excluyentes en relación a las distr previas que iremos viendo a lo largo del curso.

1. <u>Análisis de sensibilidad de los resultados</u> a dist. previas con diferente cantidad de información.

Simplemente ajustar el modelo con diferentes dist. previas y evaluar las diferencias entre casos.

2. <u>Prior predictive distribution</u> (Teo 03): permite evaluar los valores de la var de respuesta plausibles según las distr previas (Box 1980, Gabry el al 2019).

Descartar distr previas que generarían valores absurdos o imposibles de la var respuesta.

c. GLM en 6 dispositivas:

Hasta 1972 (y aún en cursos "antediluvianos de estadística"), el análisis de datos casi requería Y~Normal

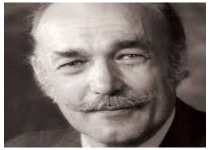
Habían varios métodos de tortura ("transformaciones") para que Y~Normal: $\text{sen}^{-1}(\sqrt{p})$, log, $\sqrt{(y+3/8)}$, Box-Cox

En ocasiones, Y claudicaba y transf(Y)~Normal...

Pero en general, estas transformaciones no funcionan: Warton & Hui 2011, O'Hara & Kotze 2010, Stroup 2014.

Últimas opciones: a) estadística no paramétrica b) tests de aleatorización diseñados para cada problema.

Los Modelos Lineales Generalizados (GLM; Nelder & Wedderburn 1972) permiten analizar distintos tipos de var. respuesta con un único método de estimación de parámetros e inferencia estadística.



	Tipo datos	FuncDesProb
Y	Binaria	Binomial
	Conteos	Poisson, Binomial Neg.
	Proporciones	Beta
	Reales >0	Gama
	Reales	Normal (y otras)

(cada FDensProb tiene su media y otros parám.)

Progresivamente emplearemos éstas (y otras) FDensProb para diferentes tipos de datos.

Table 2.1: *Properties of some distributions in the exponential family*

	Poisson	Binomial	Normal	Gamma	Inverse Gaussian	Negative Binomial
$a(\phi)$	1	1	σ^2	ν^{-1}	σ^2	1
$b(\cdot)$	e^θ	$n\log\left(1+e^\theta\right)$	$\theta^2/2$	$-\log(-\theta)$	$-(-2\theta)^{1/2}$	$-\frac{\log(1-e^\theta)}{k}$
$c(y;\phi)$	$-\log y!$	$\log\binom{n}{y}$	$-\frac{1}{2}\left(\frac{y^2}{\phi}+\log(2\pi\phi)\right)$	$\nu\log(\nu y)$ $-\log y-\log\Gamma(\nu)$	$-\frac{1}{2}\log\left(2\pi\phi y^3+\frac{1}{\phi y}\right)$	
$\mu(\theta)=E(Y;\theta)$	e^θ	$n\cdot e^\theta/\left(1+e^\theta\right)$	θ	$-\frac{1}{\theta}$	$-\frac{1}{\theta^2}$	$-\frac{e^\theta}{(-1+e^\theta)k}$
Canonical link	log	logit	identity	$-\frac{1}{\mu}$	$-\frac{1}{\mu^2}$	$\log\left(\frac{k\mu}{1+k\mu}\right)$
$V(\mu)$	μ	$np(1-p)$	σ^2	μ^2	μ^3	$\mu+k\mu^2$
$Var(y)$	μ	$n\mu(1-\mu)$	σ^2	μ^2/ν	$\sigma^2\mu^3$	$\mu+k\mu^2$

En estadística frecuentista, se formuló un algoritmo (IWLS) para $f(y,\theta,\phi)=\exp\left(\frac{y\,\theta-b(\theta)}{a(\phi)}+c(y,\phi)\right)$ de aplicación general. En estadística bayesiana, se emplea MCMC.

2) El predictor lineal: las var. explicativas X (numéricas y categóricas) predicen μ_Y = E(Y)=Xβ.

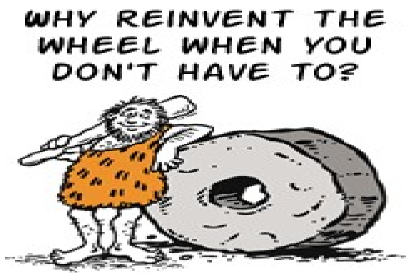
Componentes de los GLM:

1) Función de probabilidad de Y (random component)
2) El predictor lineal de las var.explicativas
3) Función de enlace (link function) g()

1) Función de probabilidad: se usa una FDensPr que pertenece a la familia exponencial ("modelo generador").

$$f(y,\theta,\phi)=\exp\left(\frac{y\,\theta-b(\theta)}{a(\phi)}+c(y,\phi)\right)$$

i.e. combinaciones de las funciones a, b y c permiten modelar las diferentes FDensProb como parte de la familia exponencial.



3) La función de enlace (link function) g():

Modela E(Y)=(Xβ) en la escala de g() adecuada a cada tipo de datos. $g(E(Y))=(X\beta)$

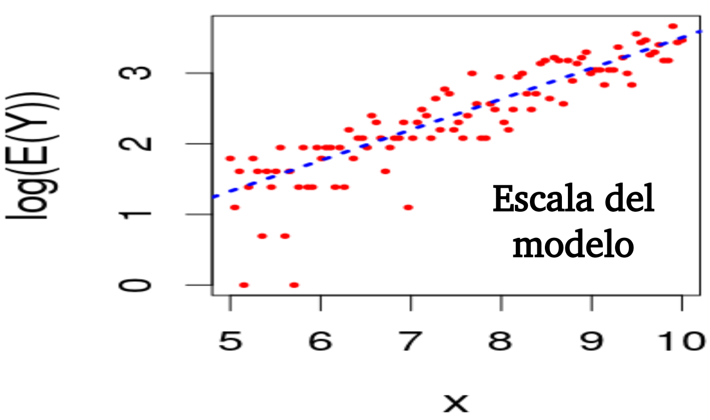
LOS GLM NO TRANSFORMAN Y, sino que modelan la relación E(Y)= Xβ en la escala definida por la función de conexión g() para cada tipo de variable Y.

	Tipo datos	FuncDesProb	Función de enlace g()
Y	Binaria	Binomial	logit, probit
	Conteos	Poisson, Binomial Neg.	log
	Proporciones	Beta	logit
	Reales >0	Gama	inverso, log
	Reales	Normal (y otras)	Identidad, log

Ej: se relacionan conteos (Y) con una var. explicativa numérica X→ regresión Poisson. Se emplea g()=log.

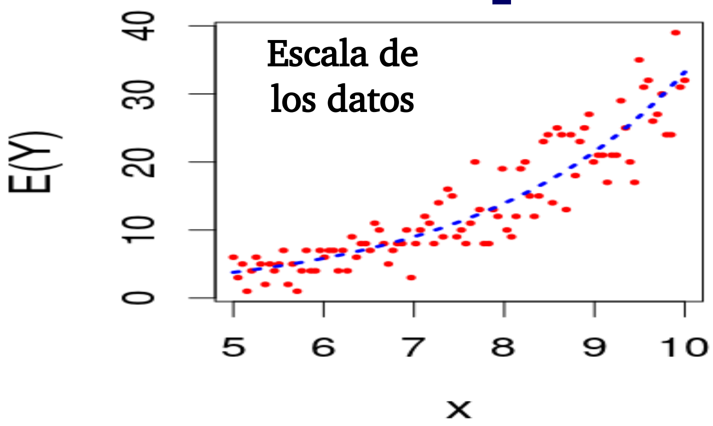
$$\log(E(Y)) = b_0 + b_1 X$$

log(E(Y))=-0.98+0.45X



$$E(Y) = g^{-1}(b_0 + b_1 X)$$

$E(Y) = \exp^{(-0.98+0.45X)}$



E(Y) es una función lineal de la var. explicativas X **únicamente** en la escala de la función de enlace.

En este curso veremos algunas de las distr. de probab. empleadas para modelar diferentes clases de vars. de respuesta, sin llegar a examinarlas todas.

Binaria	Binomial
Conteos	Poisson, Binomial Neg. y modelos de mezcla
Proporciones	Beta
Reales >0	Gama, Lognormal (y modelos de mezcla)
Reales	Normal

stancode(m2.brms)

El objetivo es entender los principios del ajuste y la interpretación de resultados de los modelos estadísticos desde la perspectiva bayesiana.

Cada modelo depende de las vars explicativas y sus eventuales interacciones que se desprenden de las hipótesis o del análisis exploratorio de los datos.

Tres ejemplos:

a. Y: conteos ~ Poisson(μ) → g()= log(). X: numérica y si una relación cuadrática: $\log(E(Y))=\beta_0+\beta_1 X+\beta_2 X^2$

b. Y: binaria ~ Binomial(π)→ g()= logit(). X_1: numérica y X_2: categórica → $\mathrm{logit}(E(Y))=\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_1 * X_2$.

c. Y: Reales >0, ~ Gama(μ,ϕ)→ g()= log(). X_1: categórica y X_2: categórica → $\log(E(Y))=\beta_0+\beta_1 X_1+\beta_2 X_2$.

Hay que definir dist previas para TODOS los parámetros.

Algunos parámetros de la FuncDensProb de Y tienen rangos de valores plausibles que deben cumplirse.

Trivialmente, sd o var >0 en Normal y Lognormal, pero ϕ (shape) en NegBinom, Beta y Gama también deben ser >0, y obviamente $-1 \leq$ corr ≤ 1.

Las funciones de enlaces de cada parámetro permiten/deben garantizar estas consistencias.

Sin embargo, los valores de los hiper-parámetros deben también ser "razonables y consistentes"...

2. Elementos básicos del análisis bayesiano

Teórico 02

a. Modelo Lineal General: regresión múltiple.

b. Distribuciones previas I:

c. GLM en 6 dispositivas.

REFERENCIAS DE ESTA CLASE:

Banner et al (2020) Methods in Ecology and Evolution 11:882-889.
Box (1980) Journal of the Royal Statistical Society A 143: 383-430
DePaoli & van de Shoot (2017) Psychological Methods 22: 240 –261
Gabry el al (2019) Journal of the Royal Statistical Society A 182:389-402
Gelman et al (2008)Annals of Statistics 2: 1360-1383
Gelman et al (2017) Entropy 19: 555-568
Gelman et al (2018) American Statistician 73: 307-309
Lemoine (2019) Oikos 128: 912–928
O'Hara & Kotze (2010) Methods in Ecology and Evolution 1: 118-122
Seaman et al (2011) American Statistician 6t6:77-84
Stroup (2014) Agronomy Journal 106:1-17
Warton & Hui (2011) Ecology 92:3-10.
van Zwet & Gelman (2021) American Statistician (in press)
van de Shoot & DePaoli (2014) European Journal of Health Psychology 16:76-85