

Tidy models

Ben Bolker

October 23, 2018



Licensed under the Creative Commons attribution-noncommercial license (<http://creativecommons.org/licenses/by-nc/3.0/>). Please share & remix noncommercially, mentioning its origin.

Introduction

The broom package is the core of the **tidymodels** ecosystem, a set of tools for handling models in a “tidy” fashion. (`broom.mixed` contains extensions for mixed models, and a few other odds and ends.)

- `tidy` produces coefficient tables (estimates, standard errors, t/Z statistics, p values, confidence intervals, etc.)
 - use `conf.int=TRUE` to get confidence intervals (not just SEs)
 - defaults to Wald CIs; use `conf.method="profile"` for profile CIs
 - use `exponentiate=TRUE` to back-transform log to response scale, logit to odds scale, ...
- `glance` produces model summaries (R^2 , residual df, AIC, BIC, etc.)
- `augment` produces predictions, residuals, etc..

Revisiting the AIDS data

```
library(glmTMB)
library(dotwhisker)
library(ggplot2)
library(broom)
library(broom.mixed)
aids <- read.csv("../data/aids.csv")
aids <- transform(aids,
                  date=year+(quarter-1)/4,
                  index=seq(nrow(aids)))

g1 <- glm(cases ~ poly(date,2), data=aids, family=poisson)
g2 <- update(g1, family=quasipoisson)
g3 <- glmTMB::glmTMB(cases~poly(date,2), data=aids, family=nbinom2)
```

```
## Warning in fitTMB(TMBStruc): Model convergence problem;
non-positive-definite Hessian matrix. See vignette('troubleshooting')
## Warning in fitTMB(TMBStruc): Model convergence problem;
false convergence (8). See vignette('troubleshooting')
```

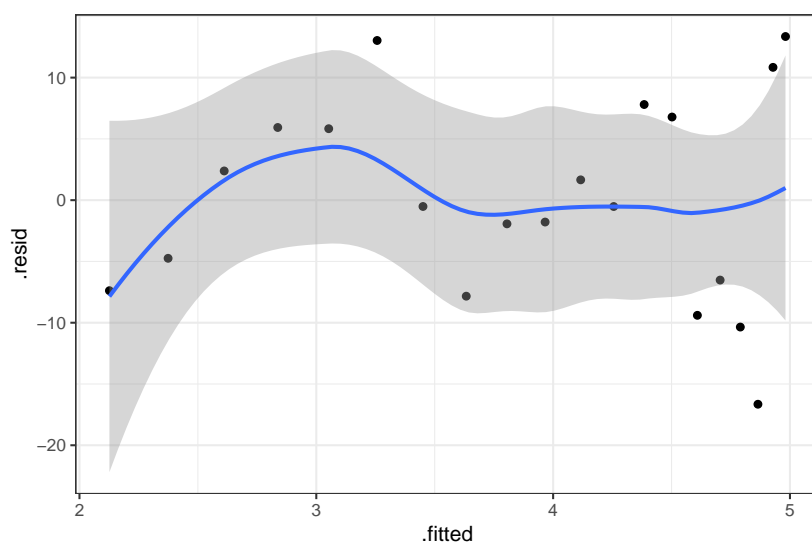
```
g4 <- glmmTMB::glmmTMB(cases~poly(date,2), data=aids, family=nbinom1)
```

```
aa <- augment(g4,data=aids)
```

```
## fitted-residual plot
```

```
ggplot(aa,aes(.fitted,.resid))+geom_point()+geom_smooth()
```

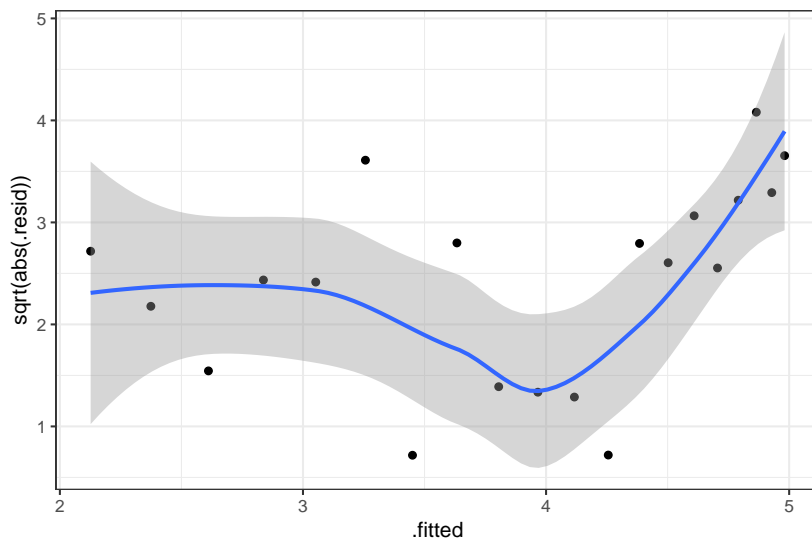
```
## 'geom_smooth()' using method = 'loess' and formula 'y ~
x'
```



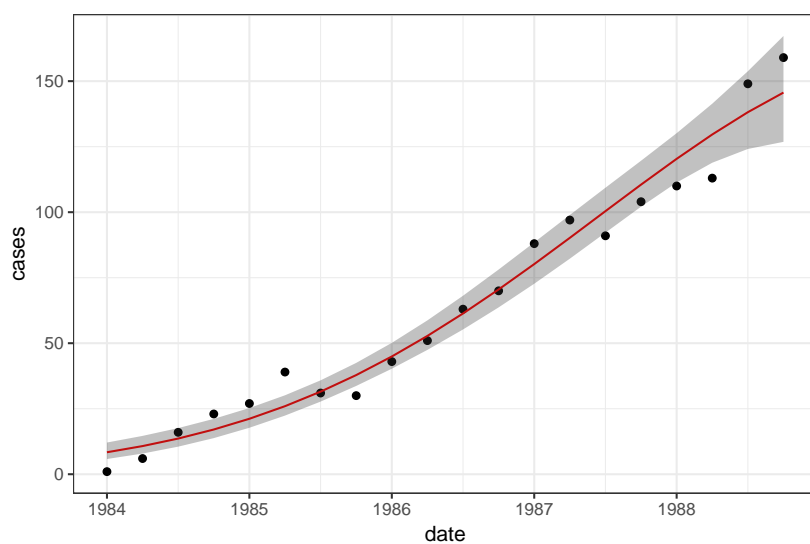
```
## scale-location plot
```

```
ggplot(aa,aes(.fitted,sqrt(abs(.resid))))+geom_point()+geom_smooth()
```

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~
x'
```



```
## manual addition of predictions to the plot
ggplot(aa, aes(date, cases)) + geom_point() +
  geom_line(aes(y = exp(.fitted)), colour = "red") +
  geom_ribbon(aes(ymin = exp(.fitted - 1.96 * se.fit),
                 ymax = exp(.fitted + 1.96 * se.fit)),
            alpha = 0.3)
```



dotwhisker::dwplot

`dotwhisker::dwplot` creates *coefficient plots*

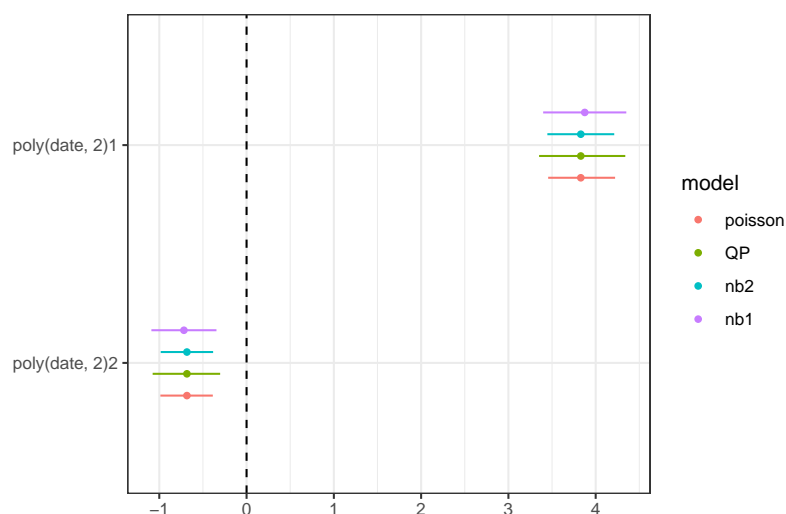
- works under the hood by calling `tidy` for one or more models and combining the results

- can pass arguments to `tidy` (e.g. `conf.method`)
- by default, scales continuous predictors by 2σ ; use `by_2sd=FALSE` to turn this off
- drops intercept by default
- given a (named) list of models, plots all the coefficients side-by-side (use to compare different modeling approaches, or models with different subsets of predictors)
- creates a `ggplot` object that can be post-processed by adding components, faceting, changing colour scales, etc..

Example:

```
dwplot(list(poisson=g1,QP=g2,nb2=g3,nb1=g4),
        by_2sd=FALSE)+
  geom_vline(xintercept=0,lty=2)

## Warning in sqrt(diag(object$cov.fixed)): NaNs produced
```



huxtable::huxreg

If your boss insists on parameter tables:

```
h0 <- huxtable::huxreg(poisson=g1,QP=g2,nbinom2=g3,nbinom1=g4)

## Warning in huxtable::huxreg(poisson = g1, QP = g2, nbinom2
= g3, nbinom1 = g4): Unrecognized statistics: r.squared
## Try setting "statistics" explicitly in the call to
huxreg()
```

```
cat(to_latex(h0, tabular_only=TRUE))  
  
## Error in to_latex(h0, tabular_only = TRUE): could not  
find function "to_latex"
```