Advertising case

`sale`: Total number of sold items.

`normalsale`: A proxy of the normal sale in the same week.

`store`: The id-number of the store.

`ad`: Advertising ($0 =$ no advertising, $1 =$ advertising).

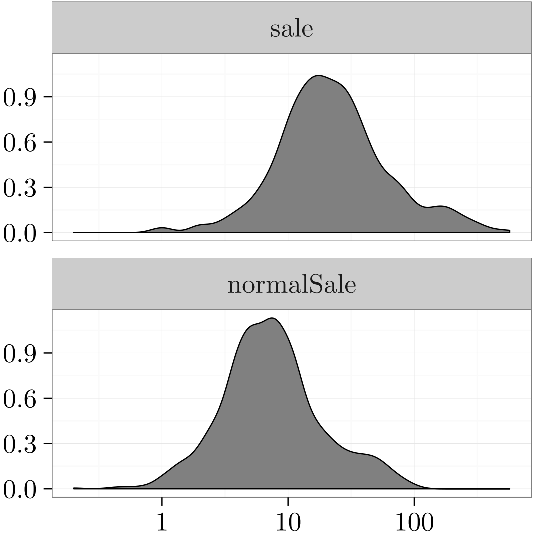
`discount`: Discount in percent.

`discountSEK`: Discount in Swedish kroner.

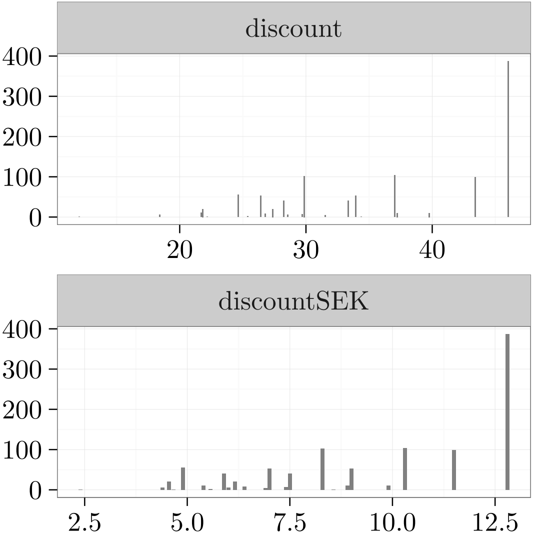
`week`: Week number (2, 4, 5, 7, 8, 9).

Objective: Predict sale for the individual store based on the normal sale and information about the advertising campaign.

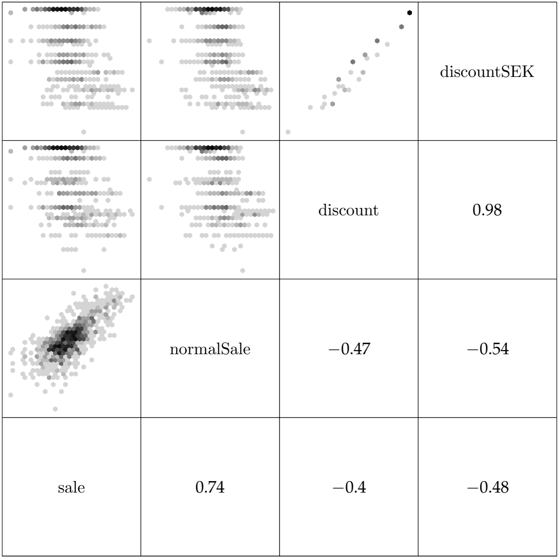
Sale distributions



Discount distributions



Scatter plot matrix



Week and advertising cross tabulation

	0	1
2	25	0
4	1	164
5	0	44
7	344	0
8	317	0
9	0	171

Advertising case

A log-linear model with a Poisson response and

$$E(Y \mid N, X) = Ne^{X^T\beta} = e^{\log(N) + X^T\beta} \tag{1}$$

with N the normal sale and X the other predictors is suggested.

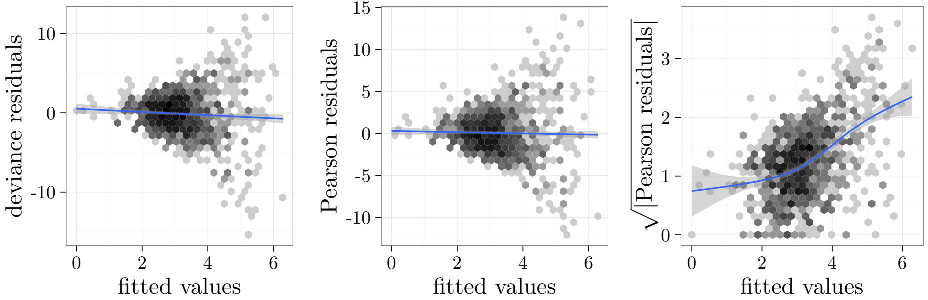
```
form <- sale ~ offset(log(normalSale)) + store + ad +
        discount + discountSEK - 1

vegetablesGlm <- glm(form,
                     family = poisson,
                     data = vegetables)
```

Advertising case

	Estimate	Std. Error	z value	Pr(>|z|)
ad1	0.32	0.03	9.90	4.3e−23
discount	−0.08	0.02	−4.43	9.4e−06
discountSEK	0.42	0.06	7.01	2.4e−12

Question: Why does discount have the "wrong" sign? Why is it significant?

Diagnostic plot



Quasi Poisson

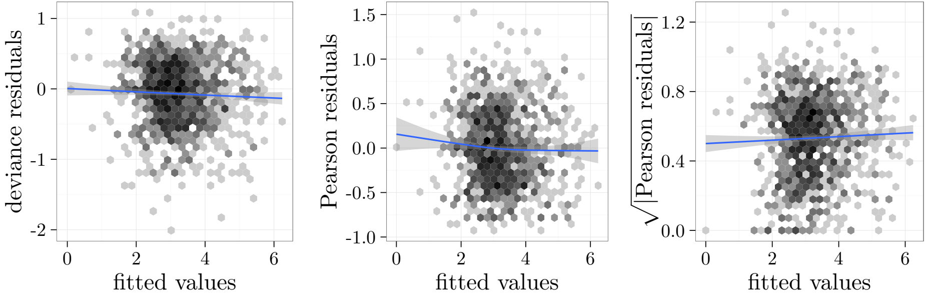
```
vegetablesGlm2 <- glm(form,
                      family = quasipoisson,
                      data = vegetables)
```

	Estimate	Std. Error	t value	Pr(>|t|)
ad1	0.32	0.11	2.89	0.004
discount	−0.08	0.06	−1.29	0.2
discountSEK	0.42	0.21	2.04	0.041

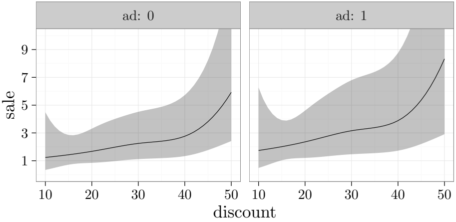
Gamma log-linear model

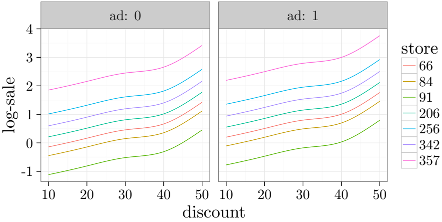
```
form <- sale ~ offset(log(normalSale)) + store  + ad +
  ns(discount, knots = c(20, 30, 40),
     Boundary.knots = c(0, 50)) - 1

vegetablesGlm3 <- glm(form,
                      family = Gamma("log"),
                      data = vegetables)
```

Diagnostic plot – Gamma model



Reporting the model





Standard combinants

For the parameter $\gamma = a^T\beta$ (for fixed a), the standard combinants are

$$Z_a = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}\sqrt{a^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}a}}.$$

or

$$Z_a^2 = \frac{(\hat{\gamma} - \gamma)^2}{\hat{\sigma}^2 a^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}a}$$

which both give confidence intervals of the form

$$\hat{\gamma} \pm z\hat{\sigma}\sqrt{a^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}a}$$

for z a suitable quantile in the (approximate) distribution of Z_a.

The profile log-likelihood

Define $\ell_{\max} = \sup_\beta \ell(\beta)$ and

$$\ell(\gamma) = \sup_{\beta : \gamma(\beta) = \gamma} \ell(\beta)$$

as the profile log-likelihood for the parameter of interest $\gamma(\beta) \in \mathbb{R}$.

The distribution of the combinant

$$R(\mathbf{Y}, \gamma) = 2(\ell_{\max} - \ell(\gamma))$$

can be approximated by the χ^2-distribution with 1 degrees of freedom if the model is true.

The 95%-quantile is 3.841 yielding the profile likelihood interval

$$\{\gamma \mid \ell(\gamma) \geq \ell_{\max} - 1.96\}$$

with approximate coverage 95%.

Quadratic approximations

$$\ell(\beta) \simeq \ell_{\max} - \frac{1}{2}(\hat{\beta} - \beta)^T \mathcal{J}^{\text{obs}}(\hat{\beta} - \beta).$$

With $\mathcal{J}^{\text{obs}} = -D^2\ell(\hat{\beta})$ the positive (semi)definite observed Fisher information.

From this

$$2(\ell_{\max} - \ell(\beta)) \simeq (\hat{\beta} - \beta)^T \mathcal{J}^{\text{obs}}(\hat{\beta} - \beta)$$

and approximate profile log-likelihood intervals can be computed by profiling the quadratic approximation.

This gives the same intervals as using a squared Z-score with

$$\hat{\text{se}} = \sqrt{((\mathcal{J}^{\text{obs}})^{-1})_{pp}}$$

for a standard confidence interval construction.