STATS 330

Handout 7
A summary of diagnostics and further model extensions

Department of Statistics, University of Auckland

Diagnostics for GLMs

So far we have seen how we can use the deviance statistic and different types of residuals to assess how well our model fits the data.

In this handout we will put these tools to use on our example data sets, showing how they can be used and what can be done when there is evidence for lack-of-fit.

Along the way, we will explore quasi-Poisson and quasi-binomial regression models, and alternative response distributions for count data.

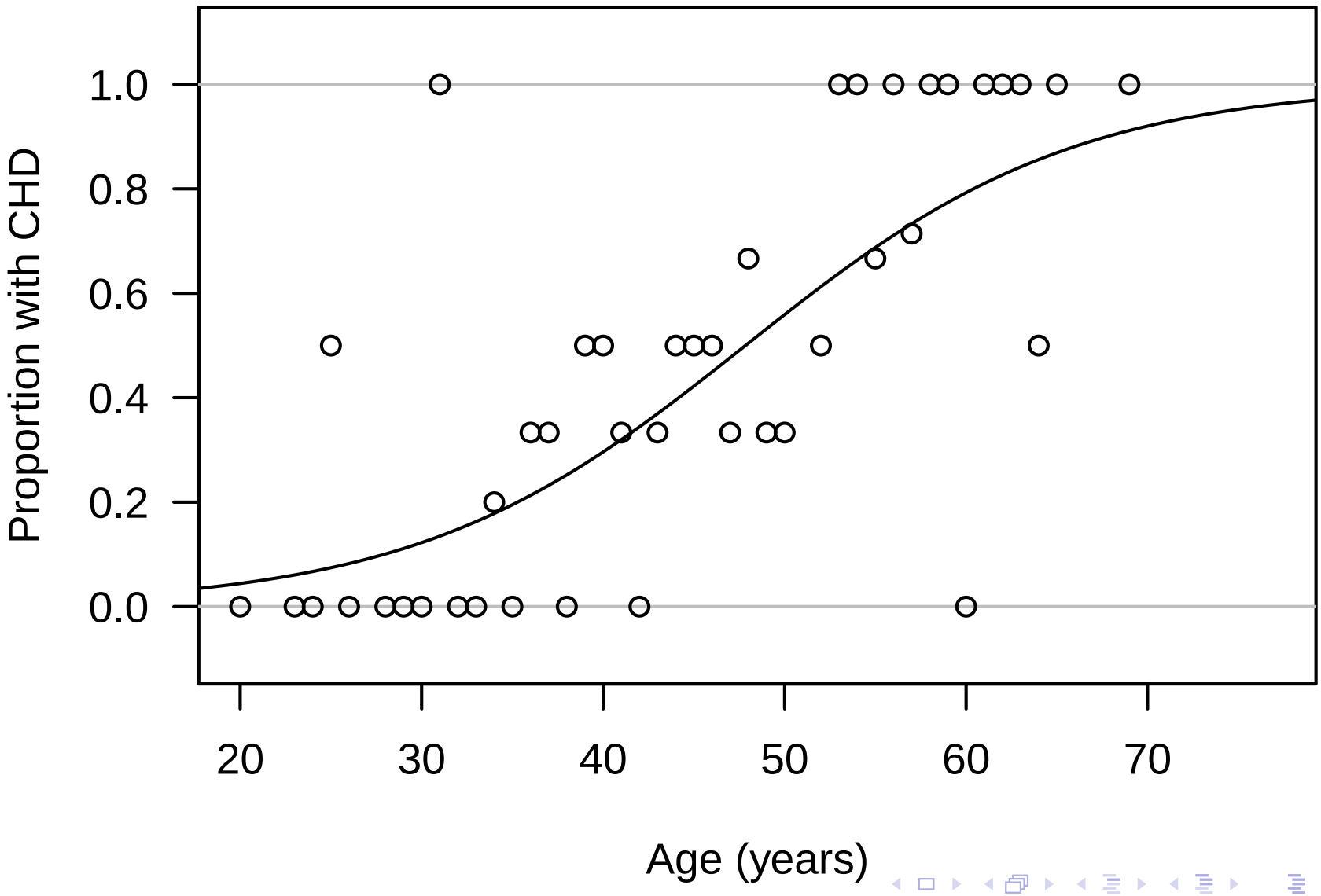
The general strategy

- Test goodness-of-fit using the deviance statistic.
- Inspect the Pearson and/or deviance residual plots. Look for a pattern in the residual plot.
  - If present, attempt to fix it by altering the explanatory terms in the model, for example by including nonlinear terms.
- If the deviance and residual plots look fine, then there is no evidence to suggest your model is inapproriate.
- If the residual plot does not have a pattern but the deviance suggests lack-of-fit, the variance of the response distribution is probably wrong.
- If you expect to see constant variance in your residual plot (see Handout 6) but you do not, then the assumed relationshp between the mean and the variance is probably wrong.
- In either case, the model needs to be changed accordingly.

Diagnostics for the coronary heart disease example

The data

Consider our coronary heart disease analysis, the logistic regression:



Diagnostics for the coronary heart disease example

Deviance

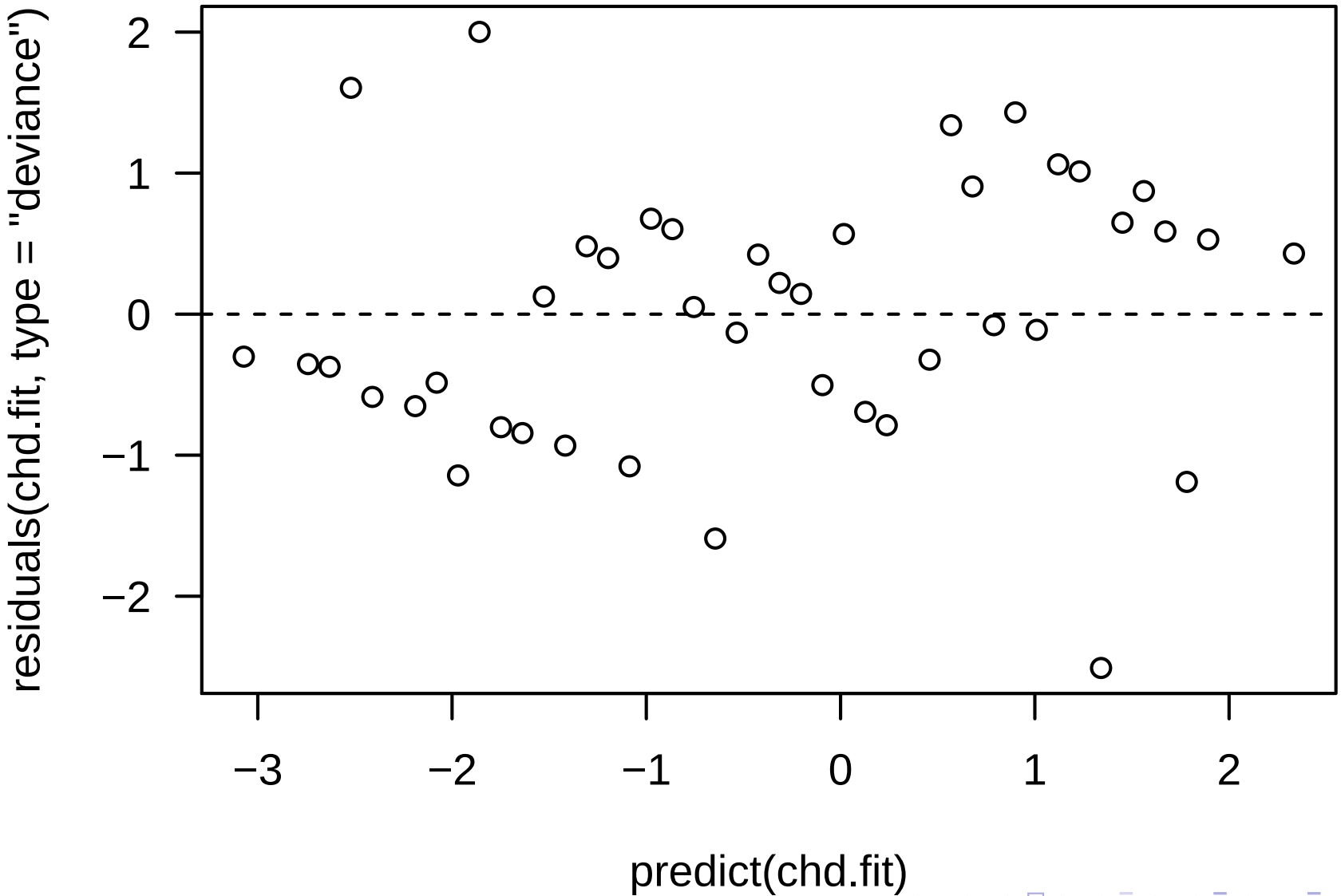
The deviance does not suggest evidence of lack-of-fit:

```
chd.fit <- glm(cbind(y, n - y) ~ age, family = "binomial")
summary(chd.fit)

## Call:
## glm(formula = cbind(y, n - y) ~ age, family = "binomial")
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.27844    1.13053  -4.669 3.03e-06 ***
## age          0.11032    0.02402   4.593 4.36e-06 ***
## ---
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 63.958  on 42  degrees of freedom
## Residual deviance: 34.976  on 41  degrees of freedom

1 - pchisq(chd.fit$deviance, chd.fit$df.residual)

## [1] 0.7344482
```

Diagnostics for the coronary heart disease example

Residuals

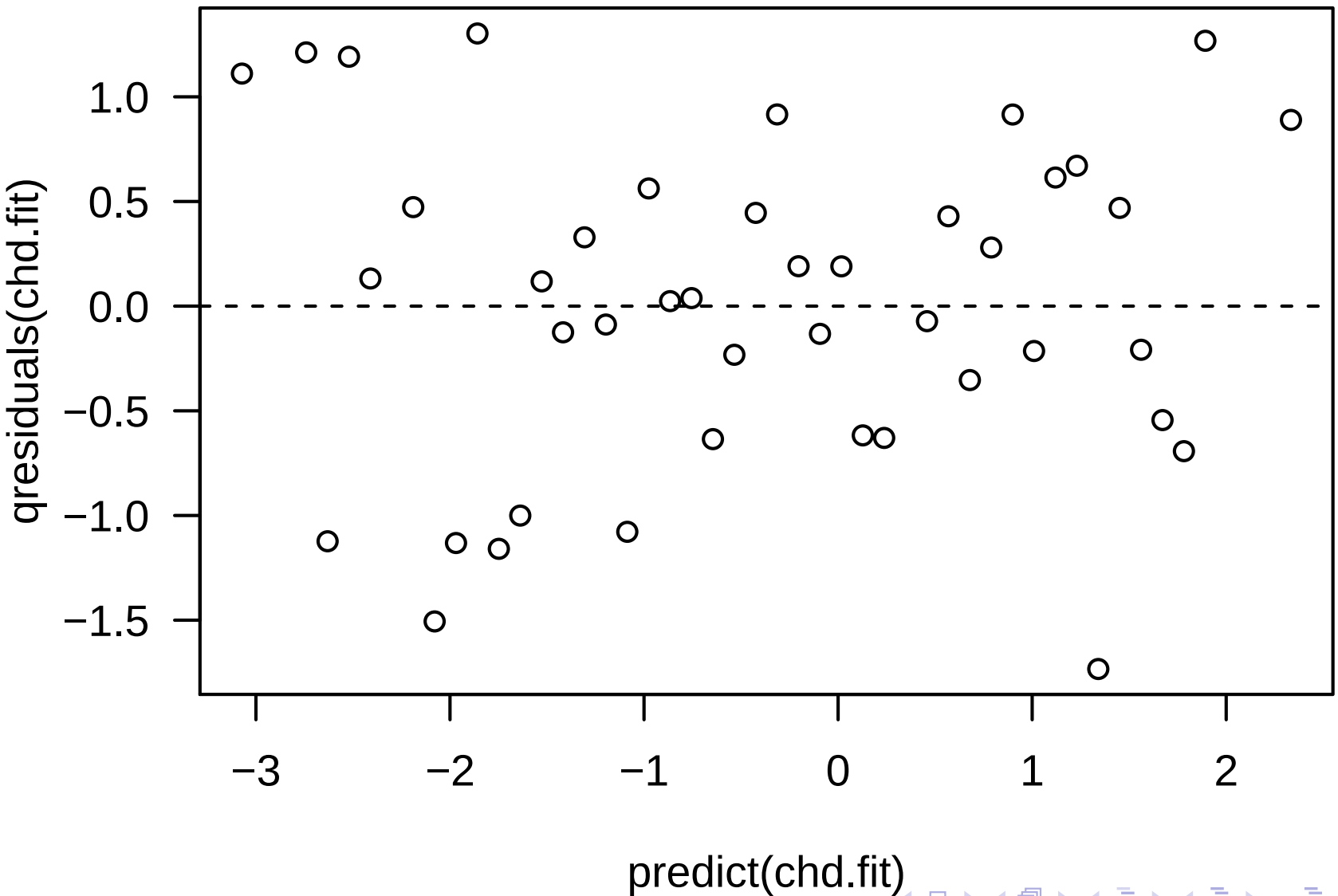
The deviance residuals:



Diagnostics for the coronary heart disease example

Residuals

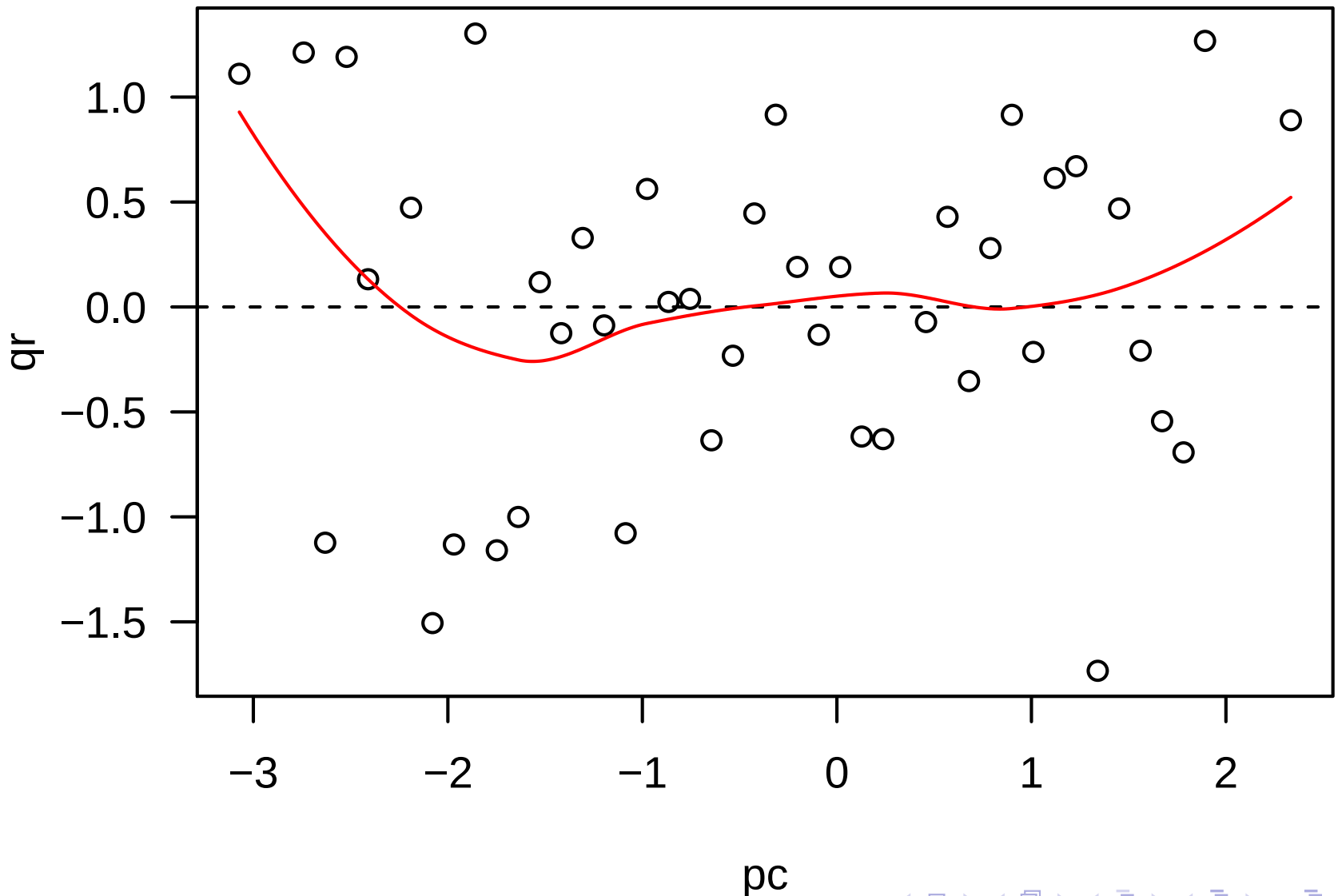
The randomised quantile residuals to account for sparse data:



Diagnostics for the coronary heart disease example

Residuals

After plotting a smoother, perhaps there is a hint of curvature?



Diagnostics for the coronary heart disease example

Checking for curvature

Probably not, but it's easy enough to fit a model with a quadratic term to check:

```
chd.quad.fit <- glm(cbind(y, n - y) ~ age + I(age^2), family = "binomial")
summary(chd.quad.fit)


## Call:
## glm(formula = cbind(y, n - y) ~ age + I(age^2), family = "binomial")
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.9282348  4.3587348  -1.131    0.258
## age          0.0941156  0.1966212   0.479    0.632
## I(age^2)     0.0001786  0.0021542   0.083    0.934
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 63.958  on 42  degrees of freedom
## Residual deviance: 34.969  on 40  degrees of freedom
```

No evidence to suggest we need a quadratic term in the model.

Diagnostics for the coronary heart disease example

Summary

In this case, we are happy with our simple model. Our model assumptions appear to be met, as long as we can assume the individuals were randomly sampled from our population of interest to ensure independence.

(Although recall from Handout 5 that it may not be appropriate to use a chi-squared distribution to get a p-value for the deviance.)

Diagnostics for the *Macrorhabdus ornithogaster* example

The data

In the last few handouts we have only conidered one of the possible explanatory variables. Here's a description of the full data set:

`mo`: The number of *M. ornithogaster* in a sample of ten faecal slides from the chicken

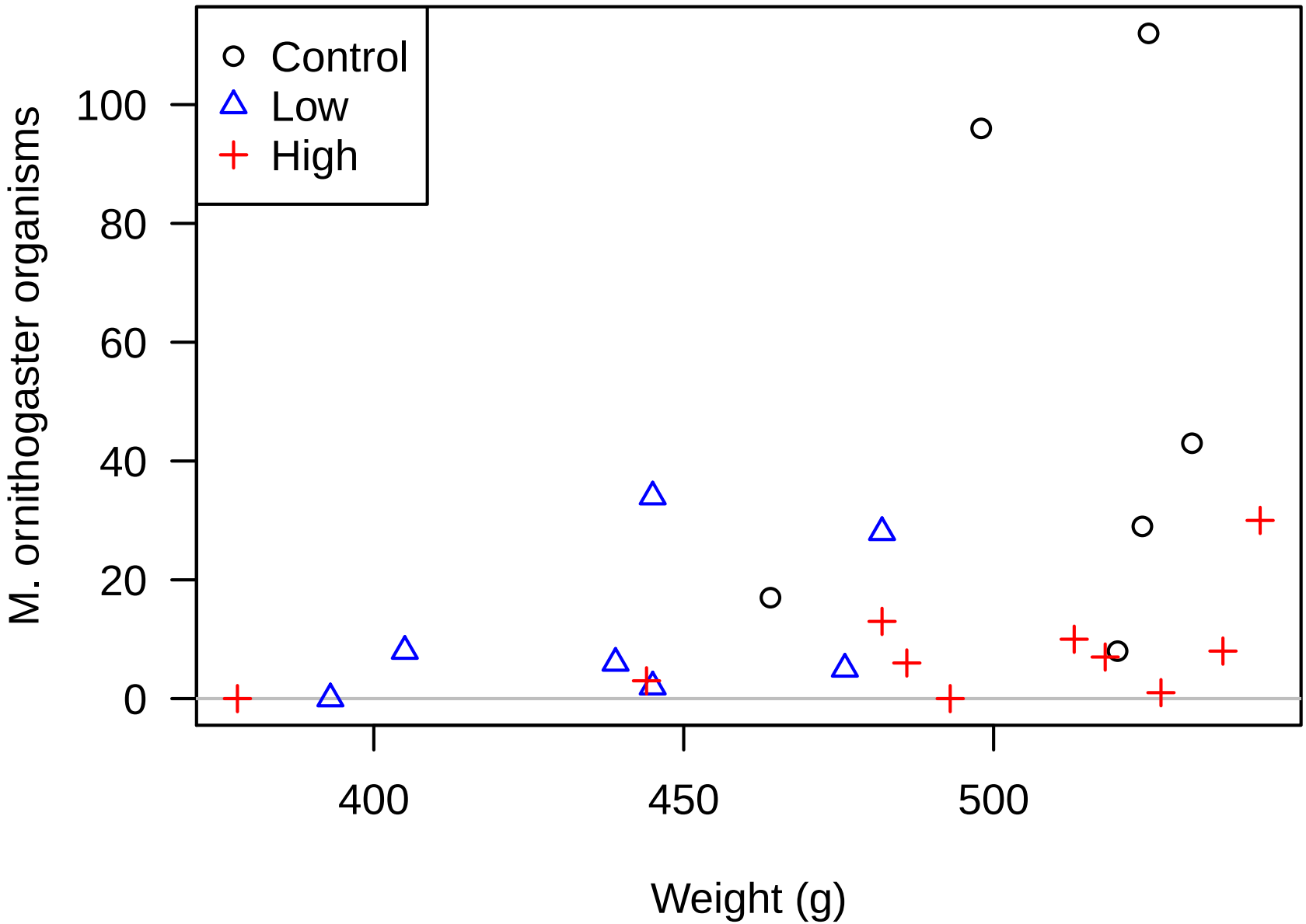
`dose`: The dose of Amphotericin B given to the chicken; either `control`, `low`, or `high`

`weight`: The weight of the chicken in grams

We wish to investigate the relationships between the drug dose and a chicken's weight on the number of *M. ornithogaster* organisms shedded by the chicken.

Diagnostics for the *Macrorhabdus ornithogaster* example

The data



Diagnostics for the *Macrorhabdus ornithogaster* example

Deviance

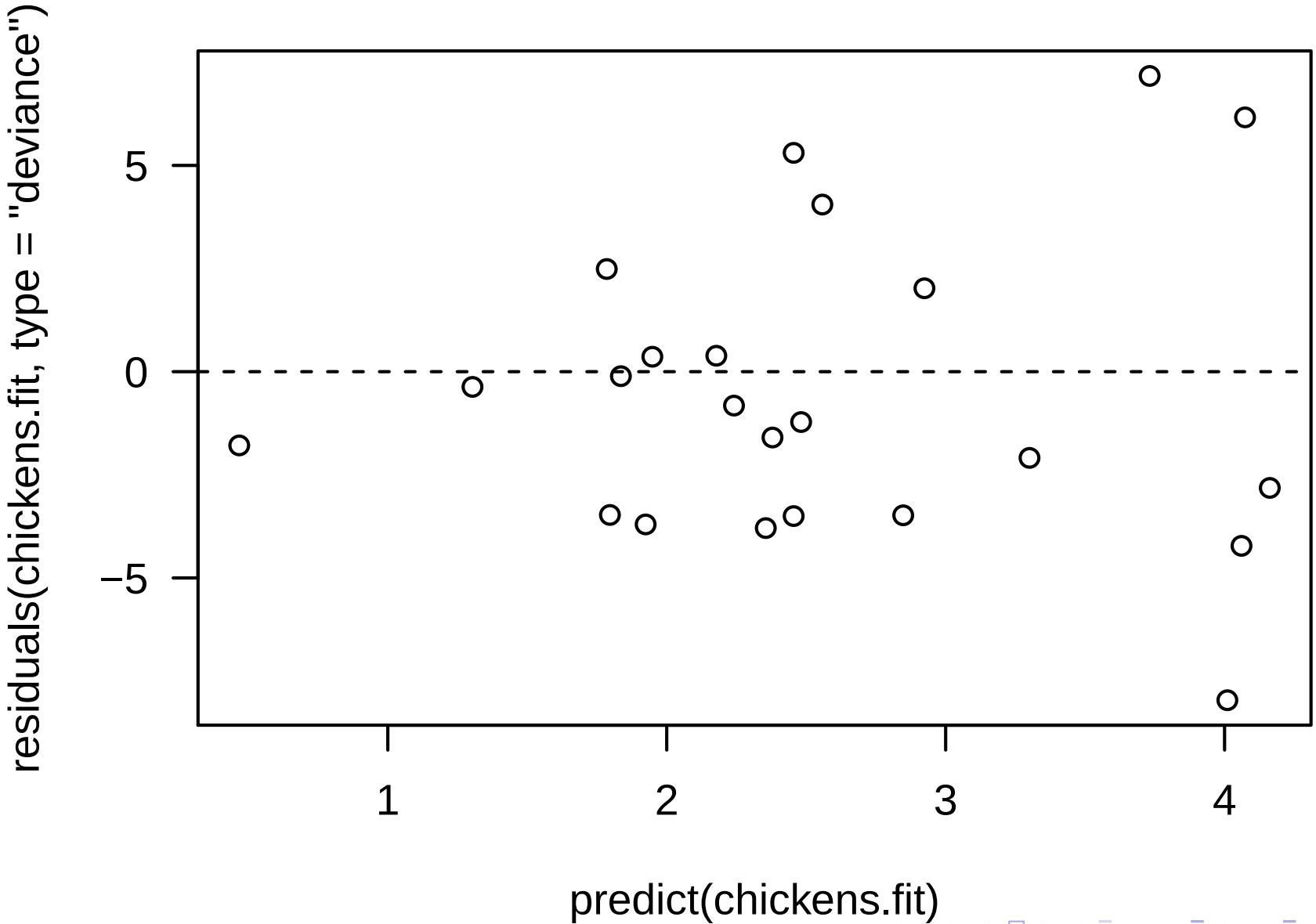
There is evidence of lack-of-fit for the Poisson model:

```
chickens.fit <- glm(mo ~ dose + weight, family = "poisson")
summary(chickens.fit)

## Call:
## glm(formula = mo ~ dose + weight, family = "poisson")
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.578171   1.015763  -2.538 0.011144 *
## doseHigh    -1.743623   0.127181 -13.710  < 2e-16 ***
## doseLow     -0.604736   0.177085  -3.415 0.000638 ***
## weight       0.012670   0.001964   6.450 1.12e-10 ***
## ---
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 683.02  on 22  degrees of freedom
## Residual deviance: 310.61  on 19  degrees of freedom

1 - pchisq(chickens.fit$deviance, chickens.fit$df.residual)

## [1] 0
```

Diagnostics for the *Macrorhabdus ornithogaster* example

Residuals

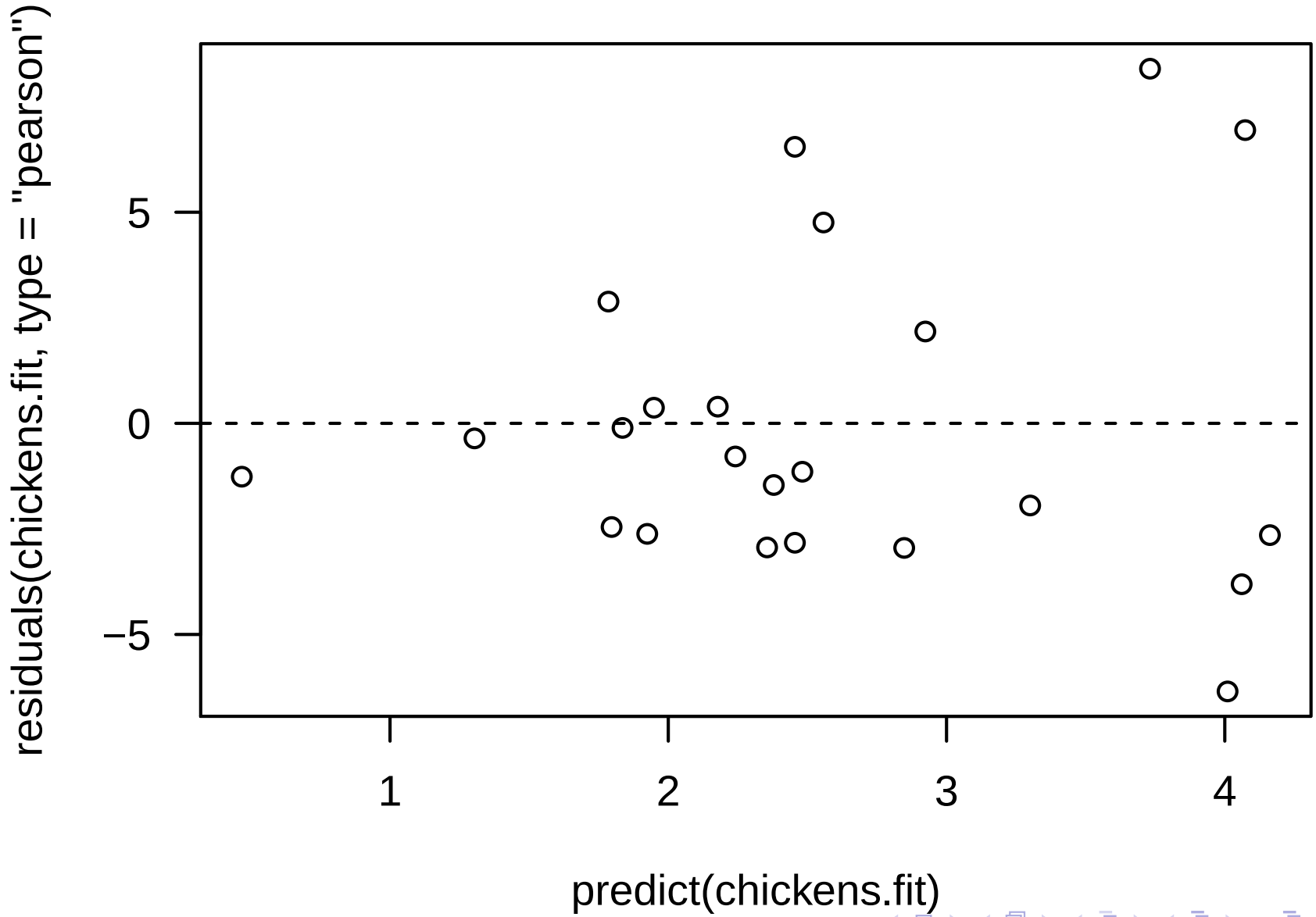
The deviance residuals:



Diagnostics for the *Macrorhabdus ornithogaster* example

Residuals

The Pearson residuals:



Diagnostics for the *Macrorhabdus ornithogaster* example

Some interim conclusions

So far, we have established:

1. The deviance provides very strong evidence to suggest lack-of-fit. We need to change our model in some way.
2. The residual plots to not show a pattern in the *mean* (e.g., curvature). So we won't attempt to fit effects like a quadratic for chicken weight.
3. The deviance and Pearson residuals have much more variance than we would expect—recall from Handout 6 that these should have variance of about 1, so the bulk of the residuals should be between -2 and 2.
4. Moreover, there is an unexpected pattern in the residuals—recall from Handout 6 that these should have constant variance.

Diagnostics for the *Macrorhabdus ornithogaster* example

Some interim conclusions

Our current model assumes

$$\log(\mu_i) = \beta_0 + \beta_1 w_i + \beta_2 h_i + \beta_3 l_i$$
$$Y_i \sim \text{Poisson}(\mu_i),$$

where w_i is the ith chicken's weight and Y_i is the number of organisms shedded, with h_i and l_i being dummy variables signifying whether or not it is in the high and low dose groups, respectively.

These assumptions imply that $\text{Var}(Y_i) = \mu_i$, but the residuals suggest that this is not the case (see Point 3 on the previous slide).

Quasi-Poisson models

When the variance of our response variable is incorrectly specified by a Poisson or logistic regression model, we can fit quasi-Poisson and quasi-binomial models, respectively. We can fit a quasi-Poisson model simply by changing the `family` argument for the `glm()` function:

```
chickens.quasi.fit <- glm(mo ~ dose + weight, family = "quasipoisson")
summary(chickens.quasi.fit)


## Call:
## glm(formula = mo ~ dose + weight, family = "quasipoisson")
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.57817    4.08488  -0.631  0.53546
## doseHigh    -1.74362    0.51146  -3.409  0.00294 **
## doseLow     -0.60474    0.71215  -0.849  0.40636
## weight       0.01267    0.00790   1.604  0.12526
## ---
## (Dispersion parameter for quasipoisson family taken to be 16.17239)
##
##     Null deviance: 683.02  on 22  degrees of freedom
## Residual deviance: 310.61  on 19  degrees of freedom
```

Quasi-Poisson models

Quasi-Poisson and quasi-binomial models make the same assumption as the corresponding Poisson and logistic regression models about the relationship between the expected value and the explanatory variables. So, in our case, we are still assuming

$$\log(\mu_i) = \beta_0 + \beta_1 w_i + \beta_2 h_i + \beta_3 l_i$$

However, quasi-Poisson and quasi-binomials do not make a specific assumption about the distribution of a response. Instead, they simply make an assumption only about its variance. This assumpion comprises multiplying the variance of the corresponding Poisson or logistic regression model by a 'dispersion parameter', k.

Quasi-Poisson models

So, for a quasi-Poisson model, we assume a linear relationship between the mean and the variance:

$$\text{Var}(Y_i) = k\mu_i.$$

In other words, we assume the variance is proportional to the mean.

In our example, we have estimated that the variance is equal to $\hat{k} = 16.17$ times the mean.

A quasi-Poisson model will have the same coefficient estimates in comparison to the corresponding Poisson model. However, anything related to estimate uncertainty (e.g., standard errors and p-values) will be different.

Quasi-Poisson models

Recall the equation for Pearson residuals from Handout 6:

$$r_i = \frac{y_i - \hat{\mathrm{E}}(Y_i)}{\sqrt{\widehat{\mathrm{Var}}(Y_i)}} = \frac{y_i - \hat{\mathrm{E}}(Y_i)}{\widehat{\mathrm{SD}}(Y_i)},$$

So, you might think that R would calculate the Pearson residuals for quasi-Poisson models using

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{k\hat{\mu}_i}},$$

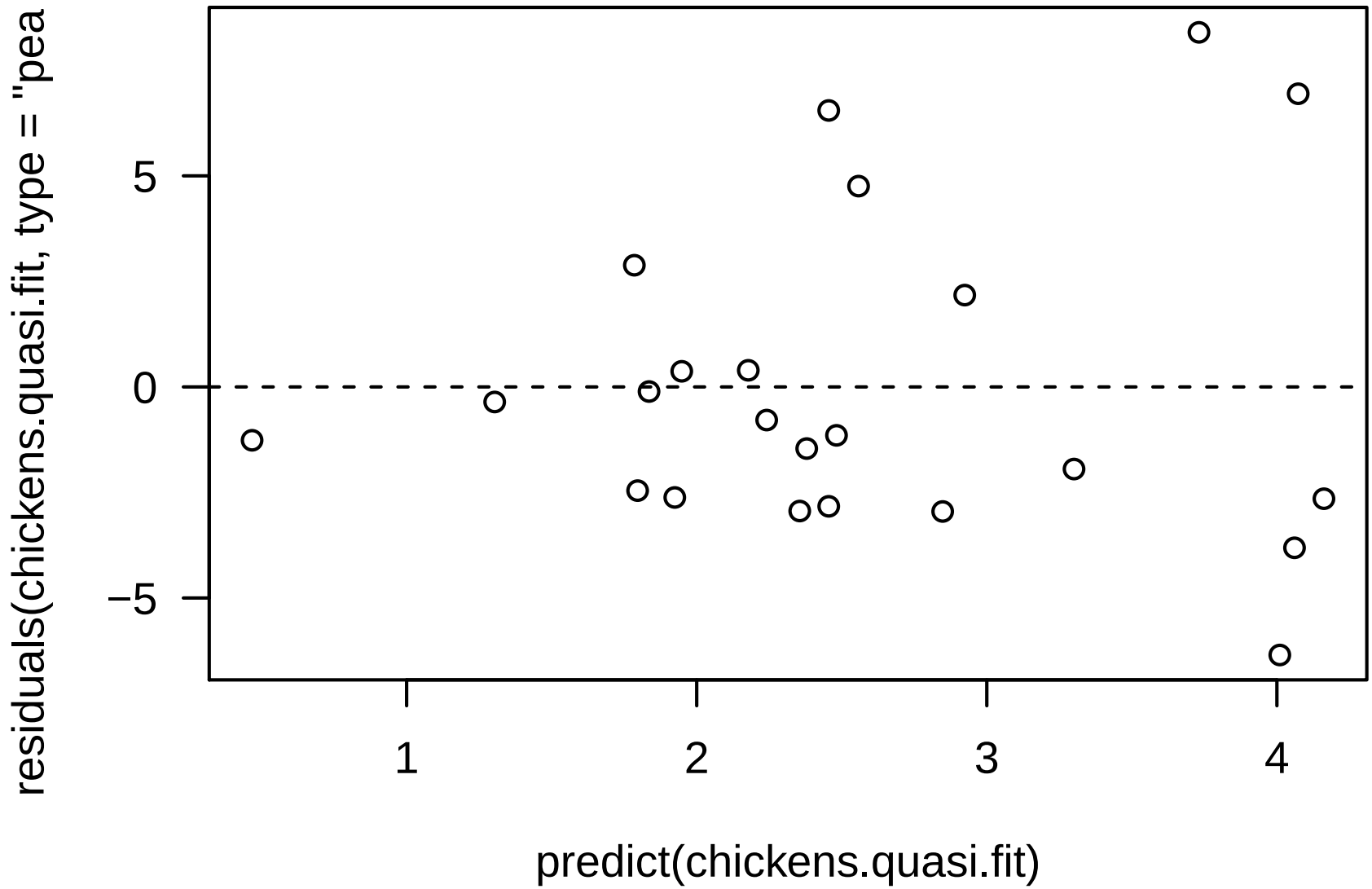
but it doesn't. R ignores the k in the denominator, calculating Pearson residuals using the same equation as it does for a Poisson regression model.

Therefore, Pearson (and deviance) residuals for quasi-Poisson and quasi-binomial models may have variance not equal to 1, even if the model is correct. However, we'd still expect the residuals to have approximately constant variance if the data are not sparse.

Diagnostics for the *Macrorhabdus ornithogaster* example

A quasi-Poisson model?

This means residual plots look exactly the same when we move to a quasi-Poisson model:



Diagnostics for the *Macrorhabdus ornithogaster* example

A quasi-Poisson model?

So we still have nonconstant variance in our Pearson and deviance residual plots, suggesting that the quasi-Poisson model isn't suitable, either.

We have adjusted our model to allow the variance of the response to be proportional to (rather than equal to) the mean. However, Pearson and deviance residuals for observations with large means have much larger variance than those with smaller means: it appears we have not correctly modelled the mean-to-variance relationship.

Another model we can fit assumes the response comes from a negative binomial distribution, rather than a Poisson distribution.

The negative binomial distribution

The negative binomial distribution has two parameters, μ and θ, where

$$\mathrm{E}(Y) = \mu,$$

like the Poisson distribution, and

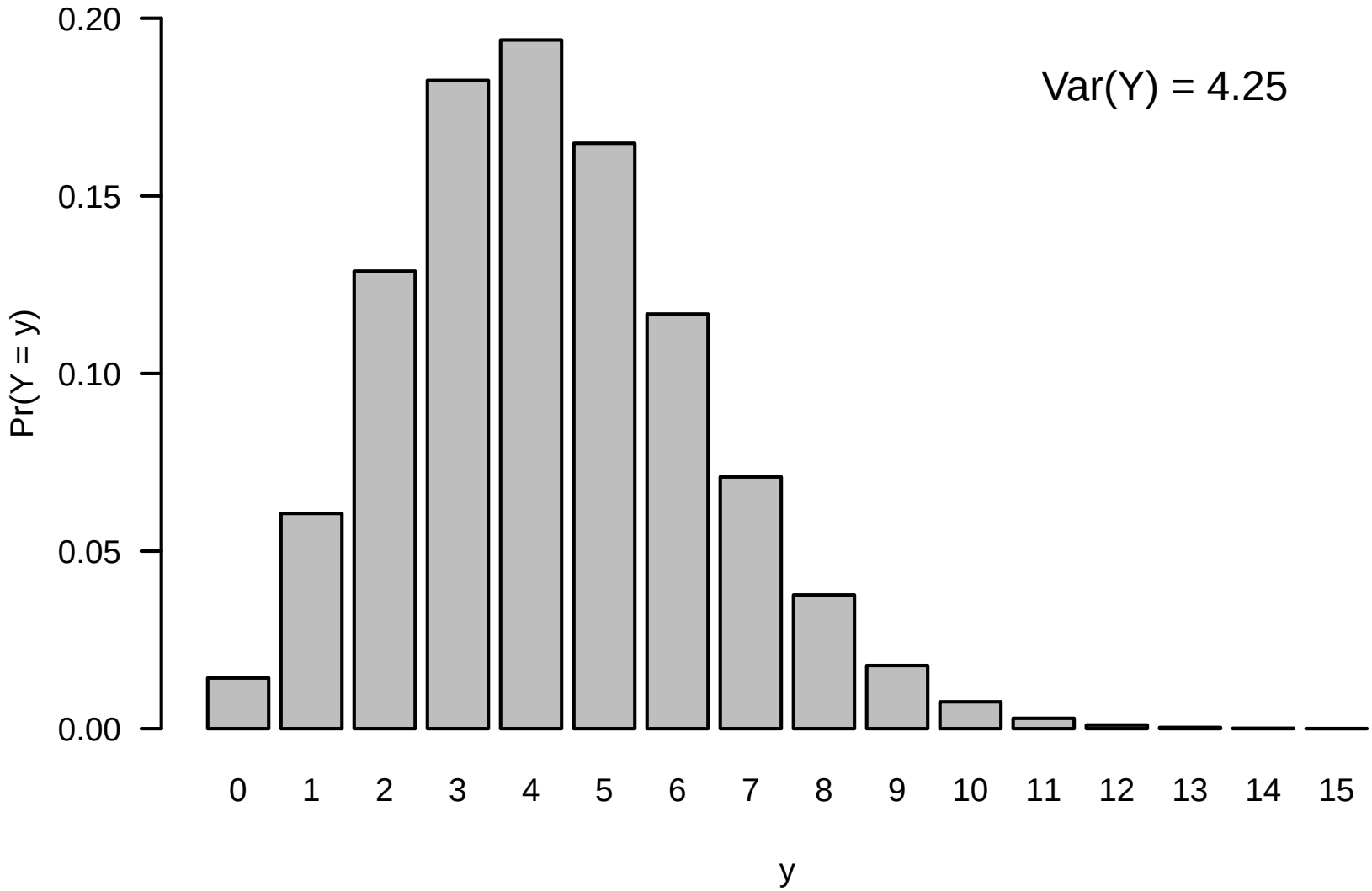
$$\mathrm{Var}(Y) = \mu + \frac{\mu^2}{\theta}.$$

So it assumes that there is a *quadratic* relationship between the mean and the variance.

When $\theta \to \infty$, $\mathrm{Var}(Y) \to \mu$, like the Poisson distribution, but, in general $\mathrm{Var}(Y) > \mu$.

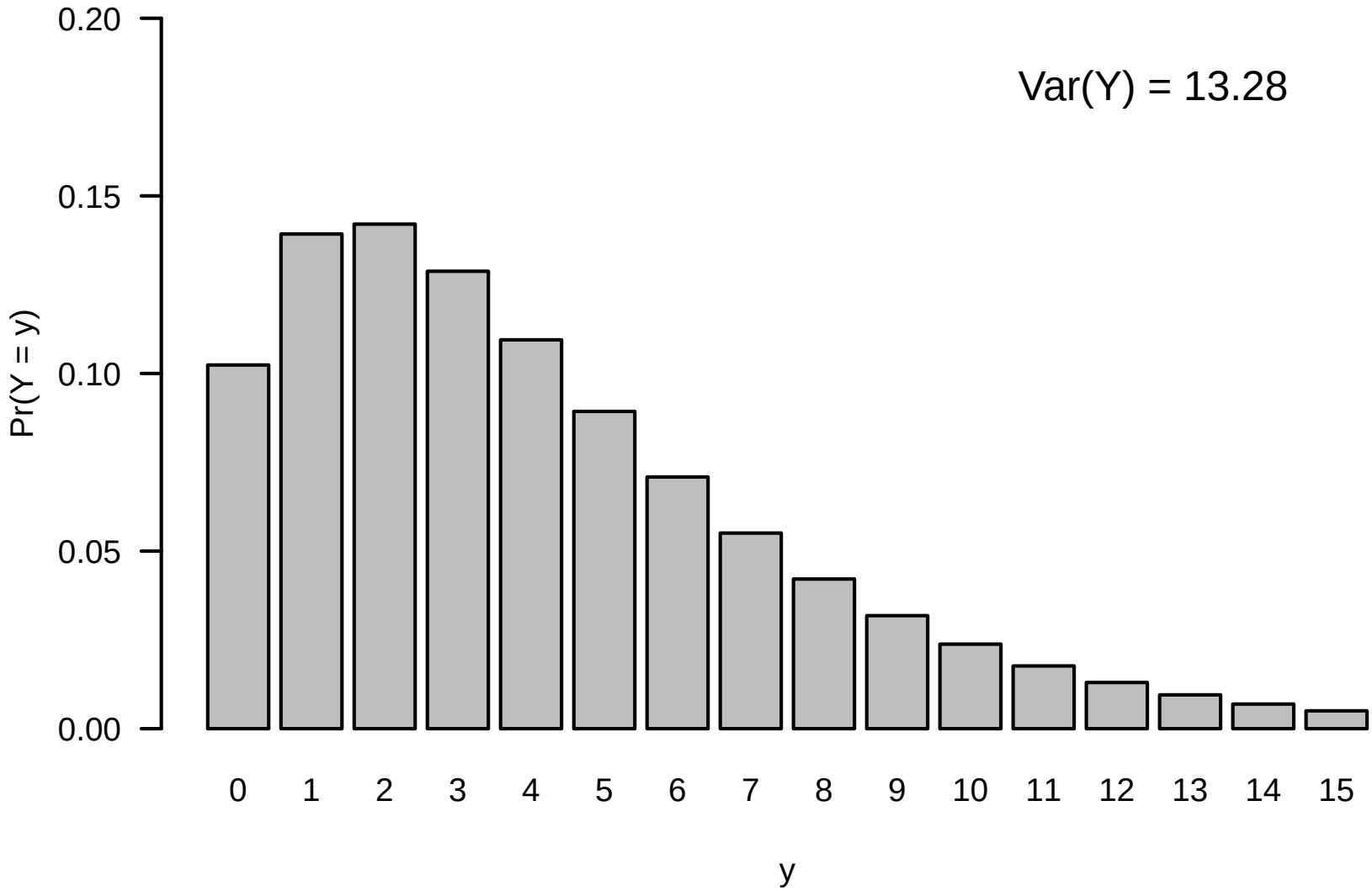
The Poisson distribution

The Poisson distribution with $\mu = 4.25$:



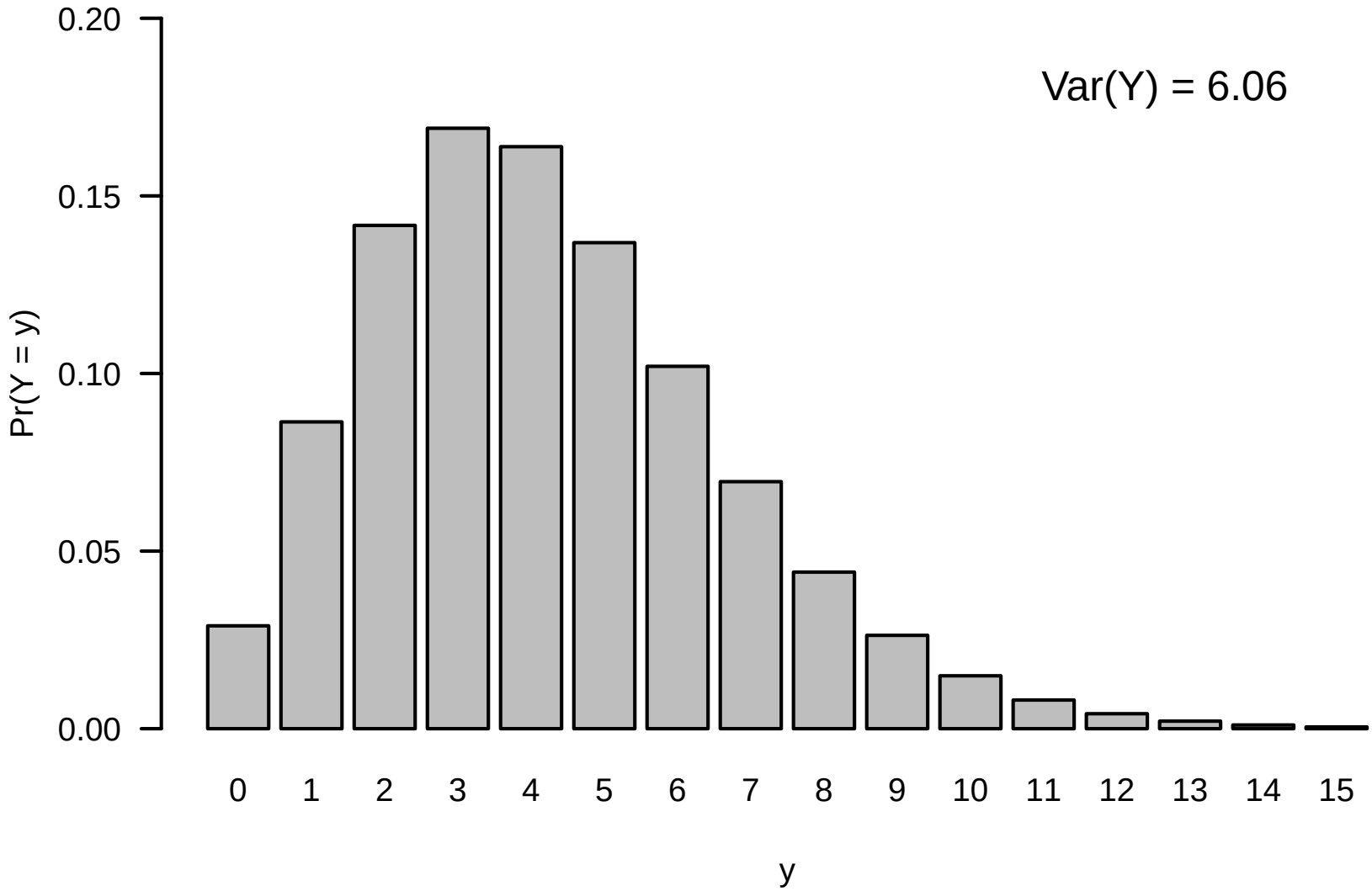
The negative binomial distribution

The negative binomial distribution with $\mu = 4.25$ and $\theta = 2$:



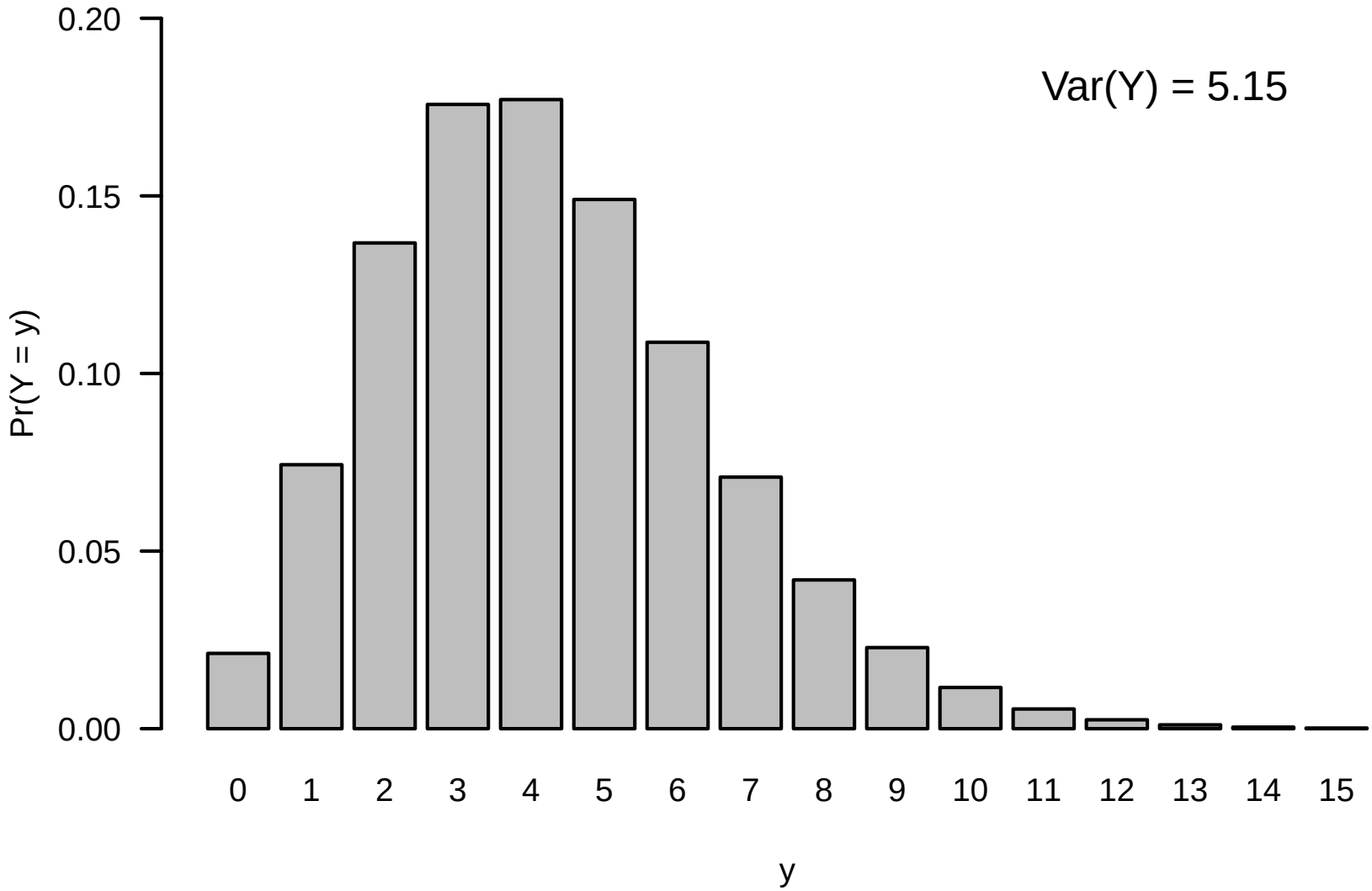
The negative binomial distribution

The negative binomial distribution with $\mu = 4.25$ and $\theta = 10$:



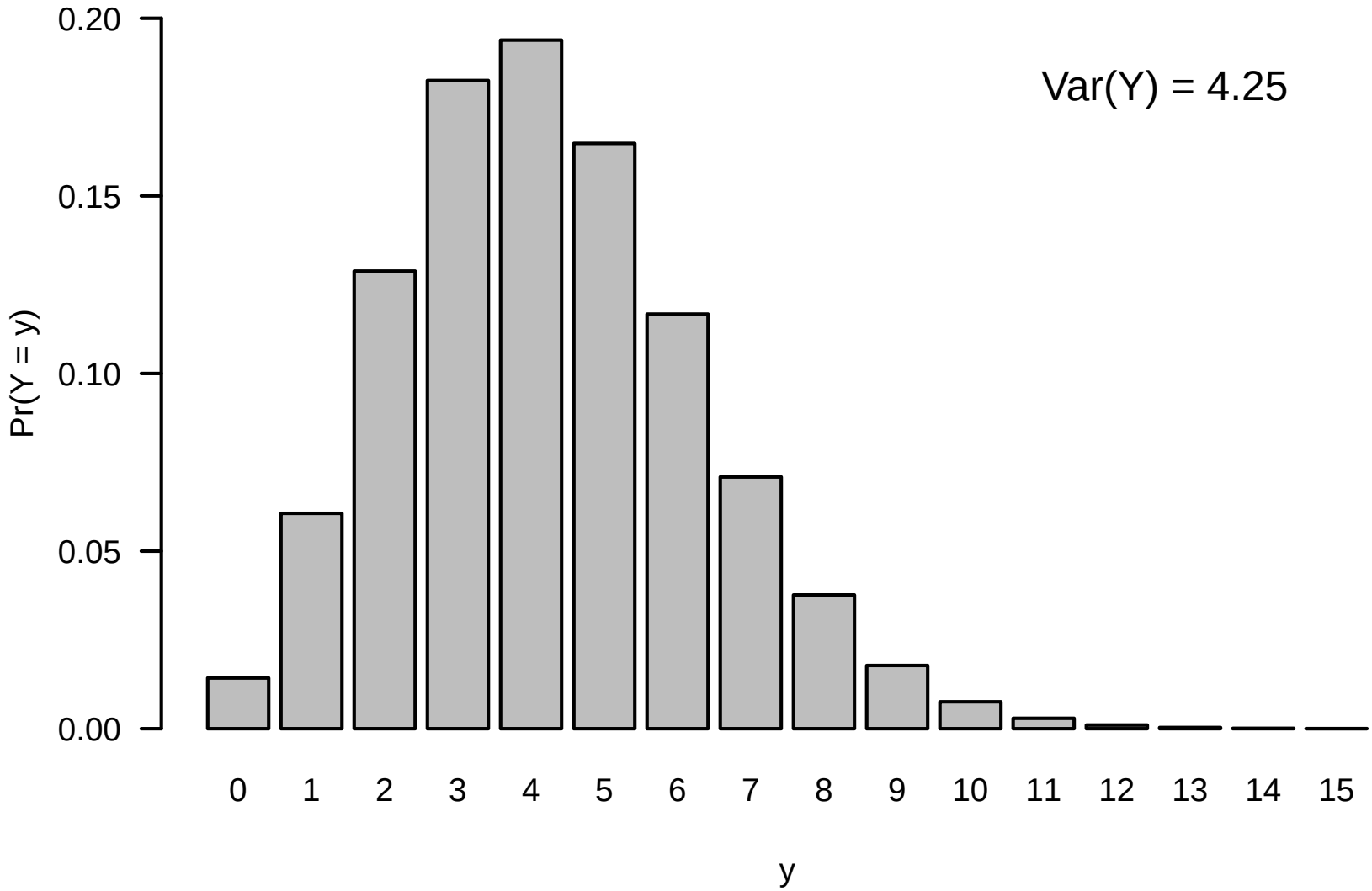
The negative binomial distribution

The negative binomial distribution with $\mu = 4.25$ and $\theta = 20$:



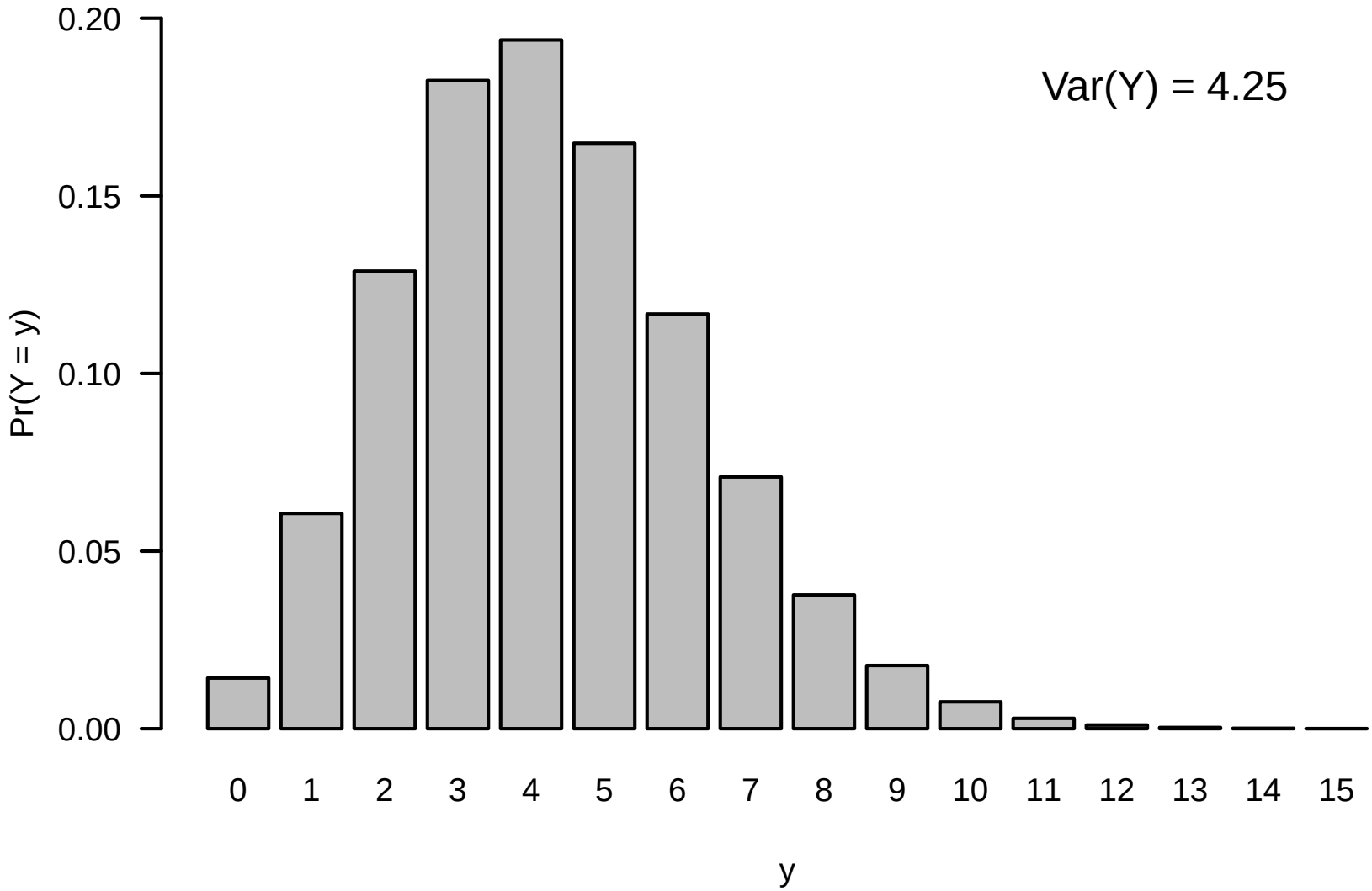
The negative binomial distribution

The negative binomial distribution with $\mu = 4.25$ and $\theta = 10\,000$:



The Poisson distribution

The Poisson distribution with $\mu = 4.25$:



Negative binomial regression

A negative binomial regression model is very similar to a Poisson regression model. We have the same specification for the mean, using a log-link function:

$$\log(\mu_i) = \beta_0 + \beta_1 w_i + \beta_2 h_i + \beta_3 l_i$$

But we make a different assumption for the distribution of the response:

$$Y_i \sim \text{Negative Binomial}(\mu_i, \theta).$$

Just like a quasi-Poisson model, we have a parameter to account for overdispersion (in this case, θ), but here we assume a quadratic relationship between the mean and the variance.

Diagnostics for the *Macrorhabdus ornithogaster* example

A negative binomial model

We can use the `glm.nb()` function from the `MASS` package to fit a negative binomial model:

```
chickens.nb.fit <- glm.nb(mo ~ dose + weight)
summary(chickens.nb.fit)

## Call:
## glm.nb(formula = mo ~ dose + weight, init.theta = 1.2767689,
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.720329   3.189414  -1.794 0.072887 .
## doseHigh    -1.808267   0.486063  -3.720 0.000199 ***
## doseLow     -0.289214   0.663036  -0.436 0.662694
## weight       0.018901   0.006202   3.047 0.002308 **
## ---
## (Dispersion parameter for Negative Binomial(1.2768) family taken to be 1)
##
##     Null deviance: 53.183  on 22  degrees of freedom
## Residual deviance: 26.408  on 19  degrees of freedom

1 - pchisq(chickens.nb.fit$deviance, chickens.nb.fit$df.residual)

## [1] 0.1192234
```

Diagnostics for the *Macrorhabdus ornithogaster* example

A negative binomial model

The deviance does not provide evidence to suggest lack-of-fit. We estimate $\hat{\theta} = 1.28$.

We should also look at some residuals.

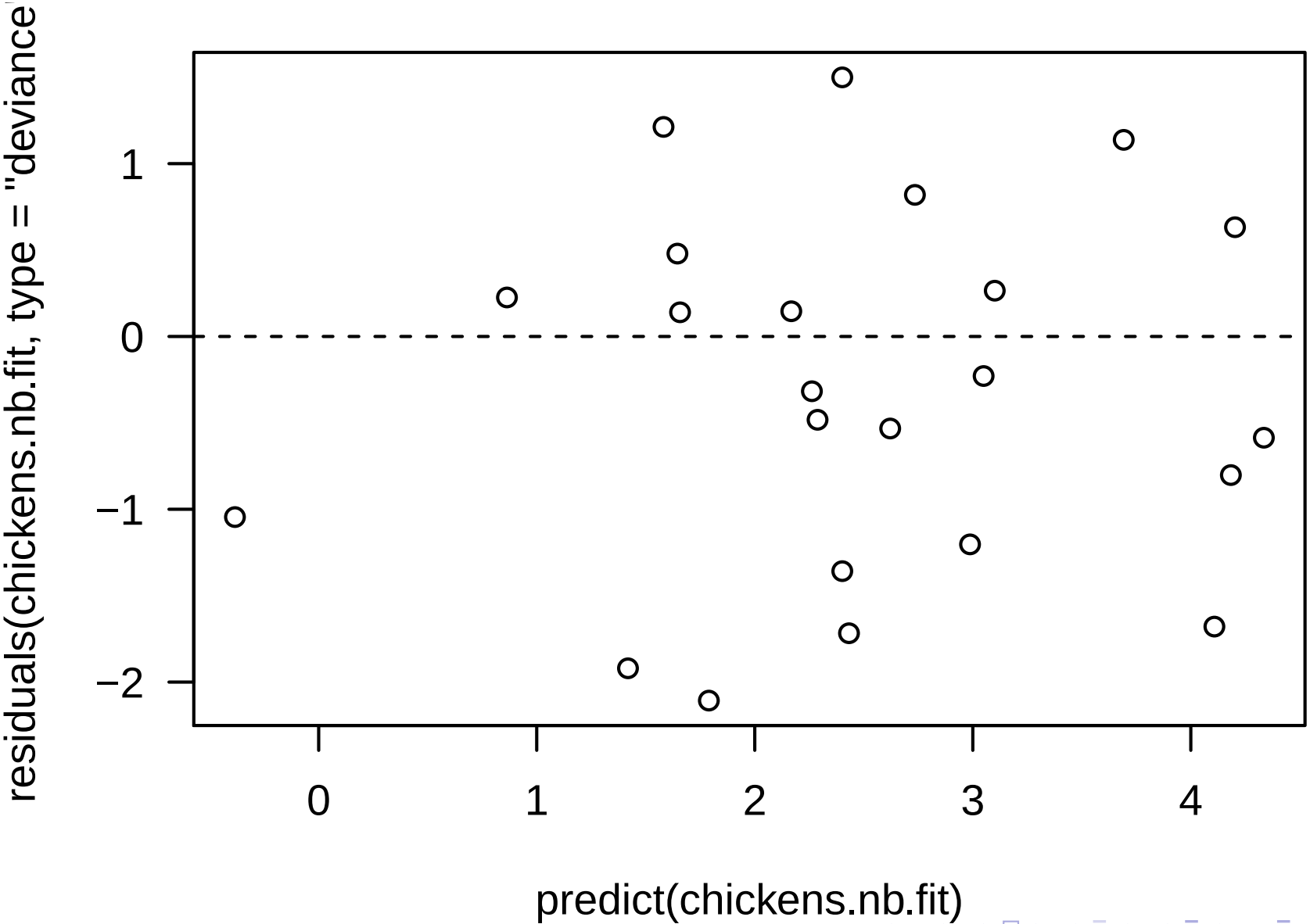
We can calculate deviance residuals in the same general way as we have seen before, via the likelihood function. Likewise, Pearson residuals are calculated as

$$r_i = \frac{y_i - \hat{\mathrm{E}}(Y_i)}{\sqrt{\widehat{\mathrm{Var}}(Y_i)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + \hat{\mu}_i^2/\hat{\theta}}},$$

Diagnostics for the *Macrorhabdus ornithogaster* example

A negative binomial model

The deviance residuals:



Diagnostics for the *Macrorhabdus ornithogaster* example

A negative binomial model

Everything looks pretty good:

- The deviance does not provide evidence for lack-of-fit.
- The deviance residuals do not show a pattern in the mean...
- ... OR the variance.

A quadratic relationship between the mean and the variance appears to be suitable.

Because the negative binomial model has the same specification of the mean, interpretation is the same as the Poisson and quasi-Poisson models: we exponentiate coefficients and interpret multiplicatively.

Diagnostics for the *Macrorhabdus ornithogaster* example

A negative binomial model

```
## CI for 10 g increase in chicken weight.
exp(10*confint(chickens.nb.fit)[4, ])

## Waiting for profiling to be done...

##    2.5 %   97.5 %
## 1.062354 1.381160

## CIs for other parameters.
exp(confint(chickens.nb.fit)[1:3, ])

## Waiting for profiling to be done...

##                     2.5 %    97.5 %
## (Intercept) 3.706294e-06 2.3785881
## doseHigh    6.087552e-02 0.4171181
## doseLow     2.048195e-01 2.7803293
```

Diagnostics for the *Macrorhabdus ornithogaster* example

A negative binomial model

We have strong evidence to suggest that high doses of Amphotericin B is associated with a decrease in the mean number of *M. ornithogaster* organisms shedded, but no such evidence for low doses.

- Chickens on the high dose shed between 58.29 and 93.91% fewer organisms, on average, in comparison to chickens on the control treatment.

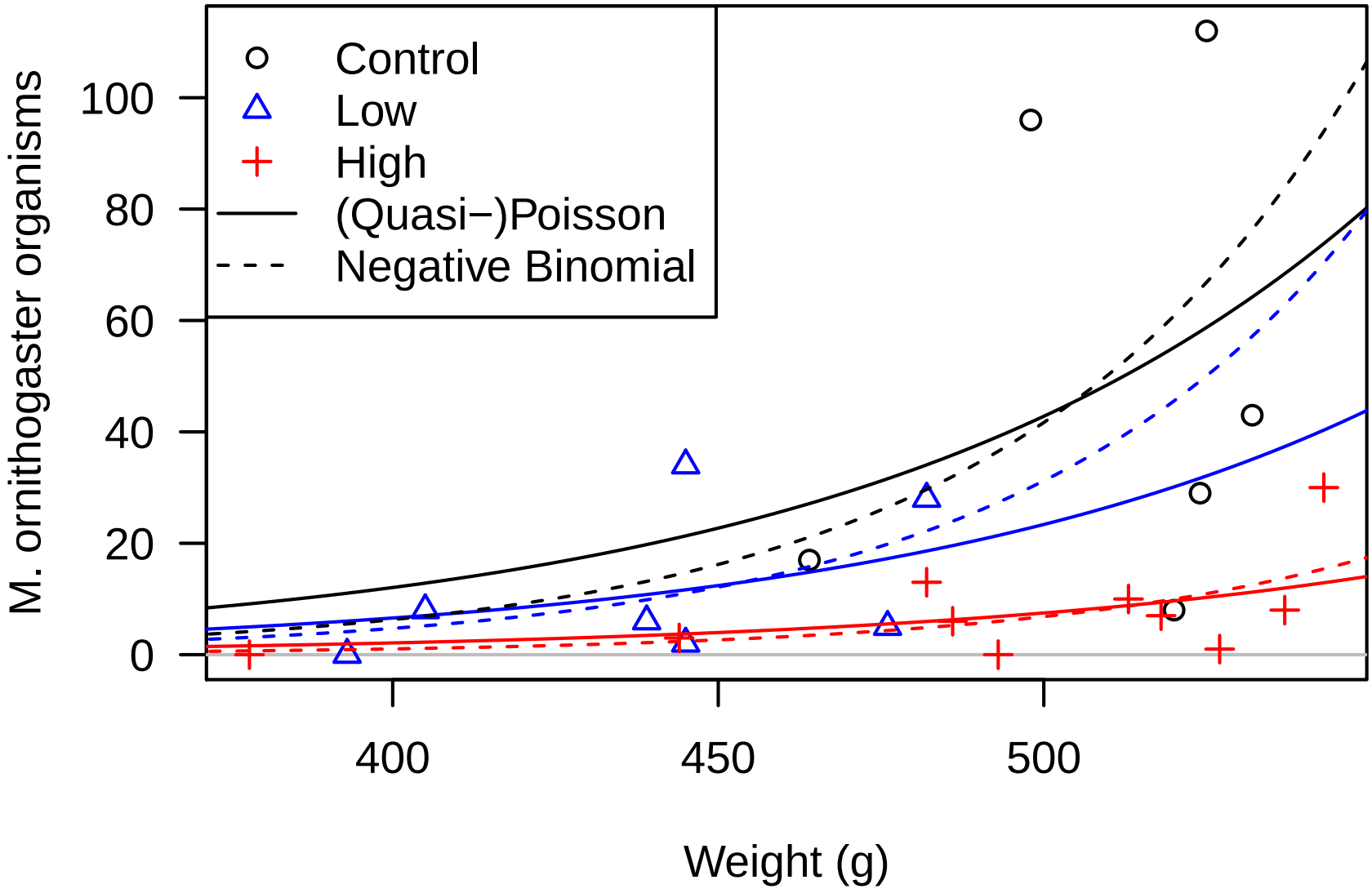
We have strong evidence to suggest that heavier birds shed more organisms, on average.

- For every 10-g increase in chicken weight, the mean number of organisms shedded increases by between 6.24 and 38.12%.

Diagnostics for the *Macrorhabdus ornithogaster* example

Model comparison: Effects of explanatory variables

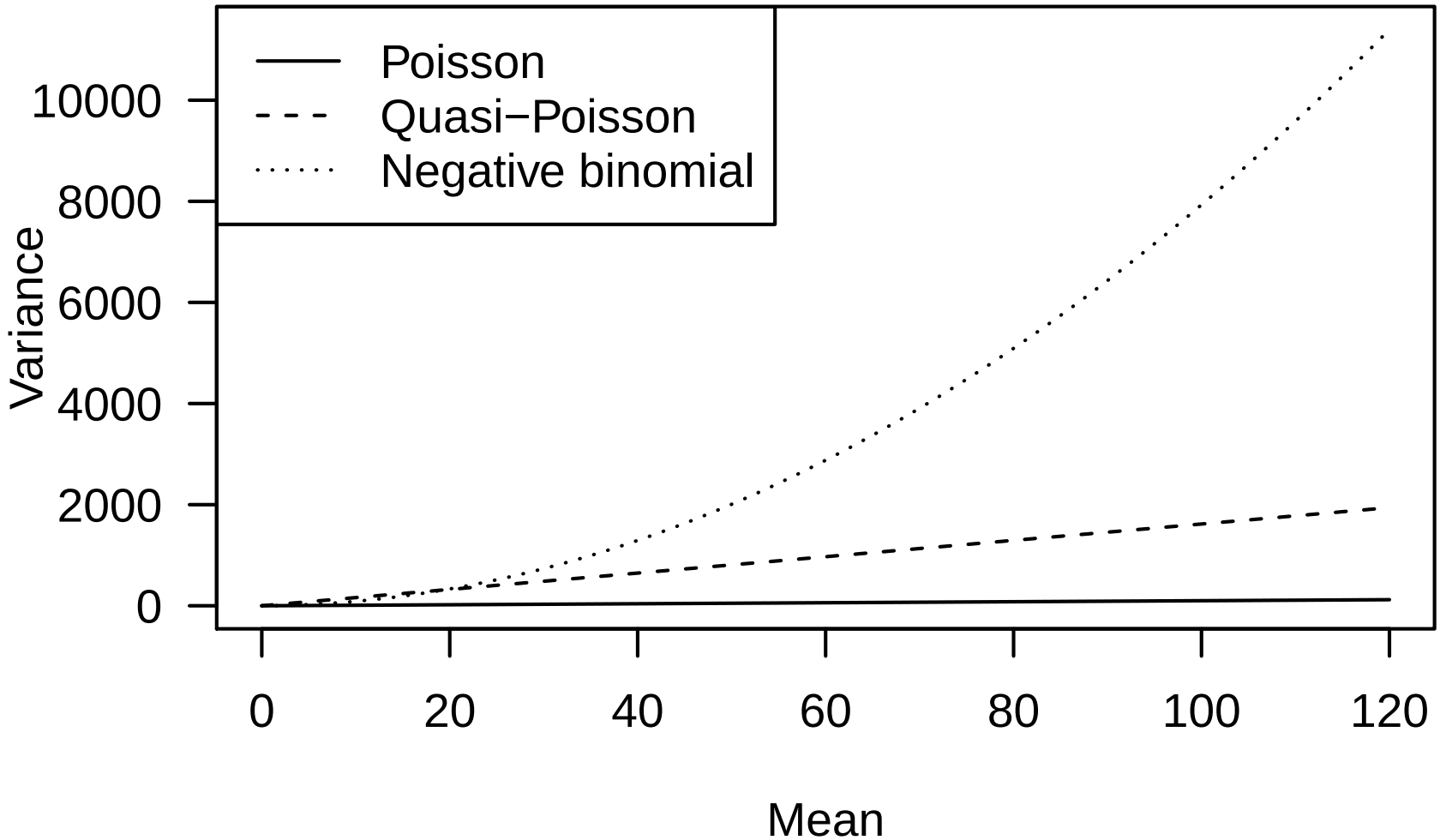
Slight differences in the fitted lines.



Diagnostics for the *Macrorhabdus ornithogaster* example

Model comparison: Mean-to-variance relationships

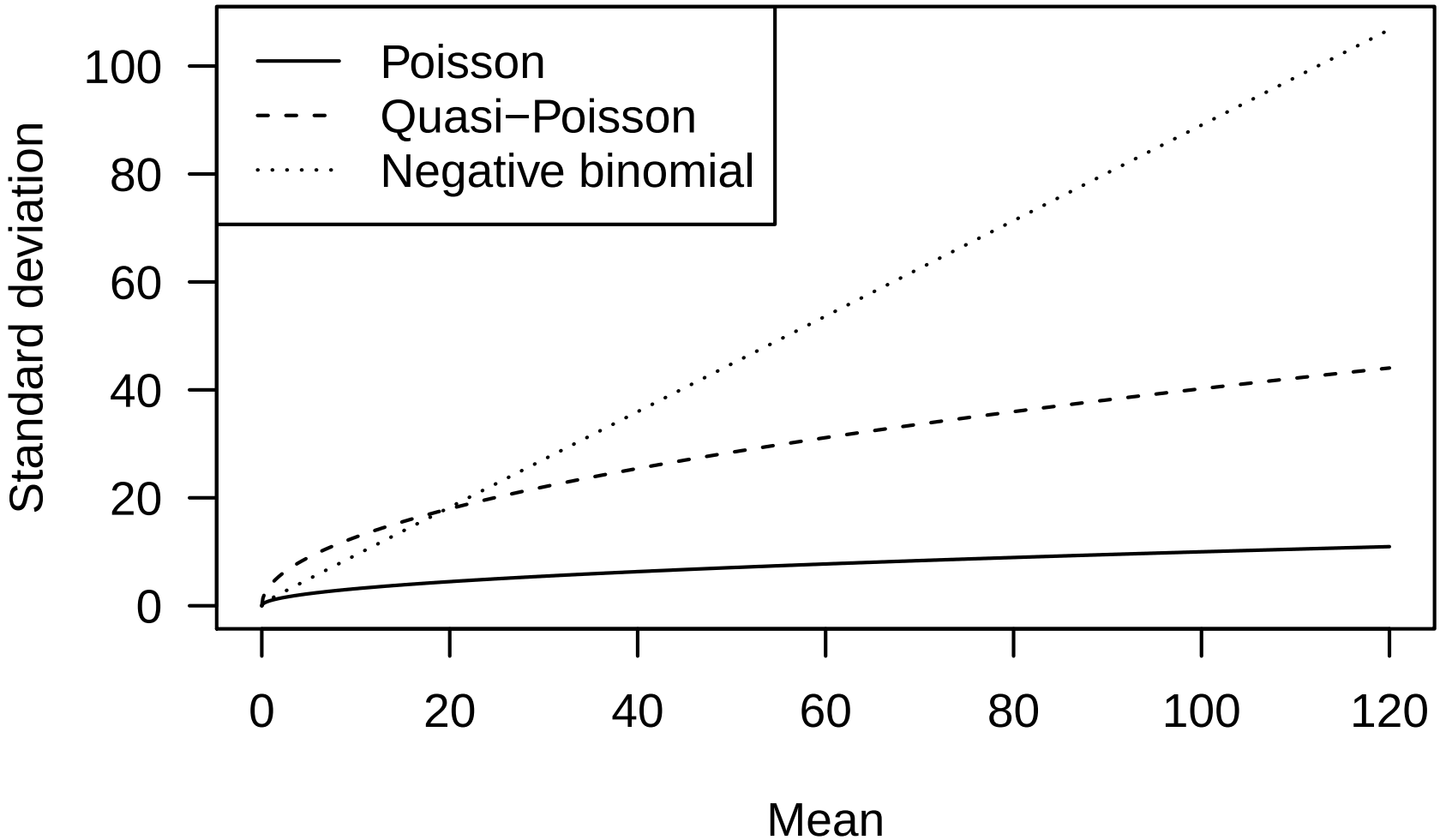
The three models give quite different mean-to-variance relationships. Poisson and quasi-Poisson models appear to severely underestimate variance for observations with large means.



Diagnostics for the *Macrorhabdus ornithogaster* example

Model comparison: Mean-to-variance relationships

The three models give quite different mean-to-variance relationships. Poisson and quasi-Poisson models appear to severely underestimate variance for observations with large means.



Logistic regression extensions

We can use similar extensions to logistic regression models to model different kinds of mean-to-variance relationships.

Logistic regression (see Handout 1):

$$\text{Var}(Y_i) = n_i p_i (1 - p_i)$$

Quasi-binomial regression to deal with overdispersion:

$$\text{Var}(Y_i) = k n_i p_i (1 - p_i)$$

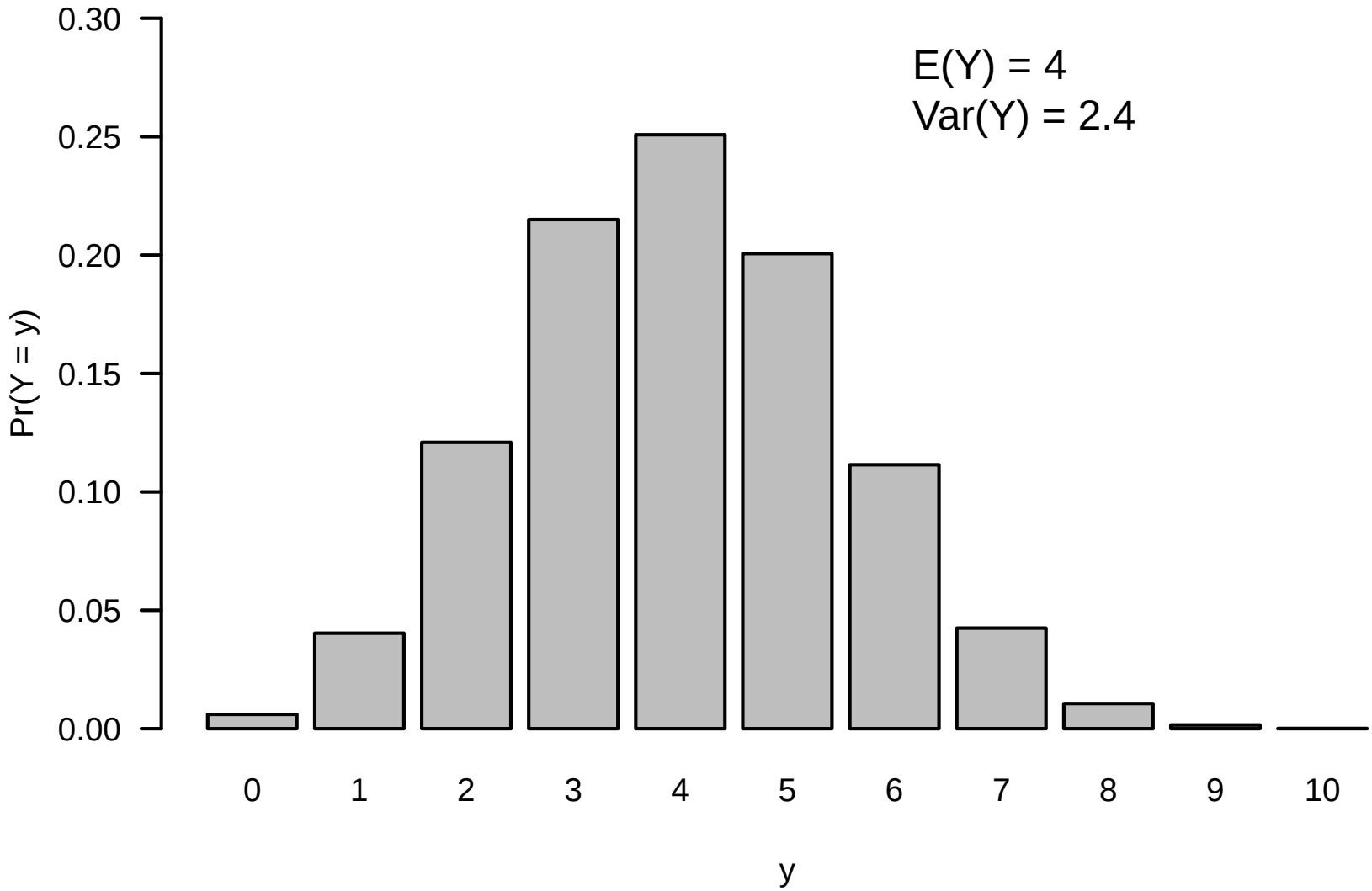
Alternatively, beta-binomial regression to deal with overdispersion:

$$Y_i \sim \text{Beta-binomial}(n_i, p_i, \rho), \text{and}$$
$$\text{Var}(Y_i) = n_i p_i (1 - p_i)(1 + n_i \rho - \rho).$$

The parameter ρ controls the variance such that $\rho = 0$ implies no overdispersion (equivalent to logistic regression) and $0 < \rho \leq 1$ accounts for overdispersion. Implemented in the VGAM package.

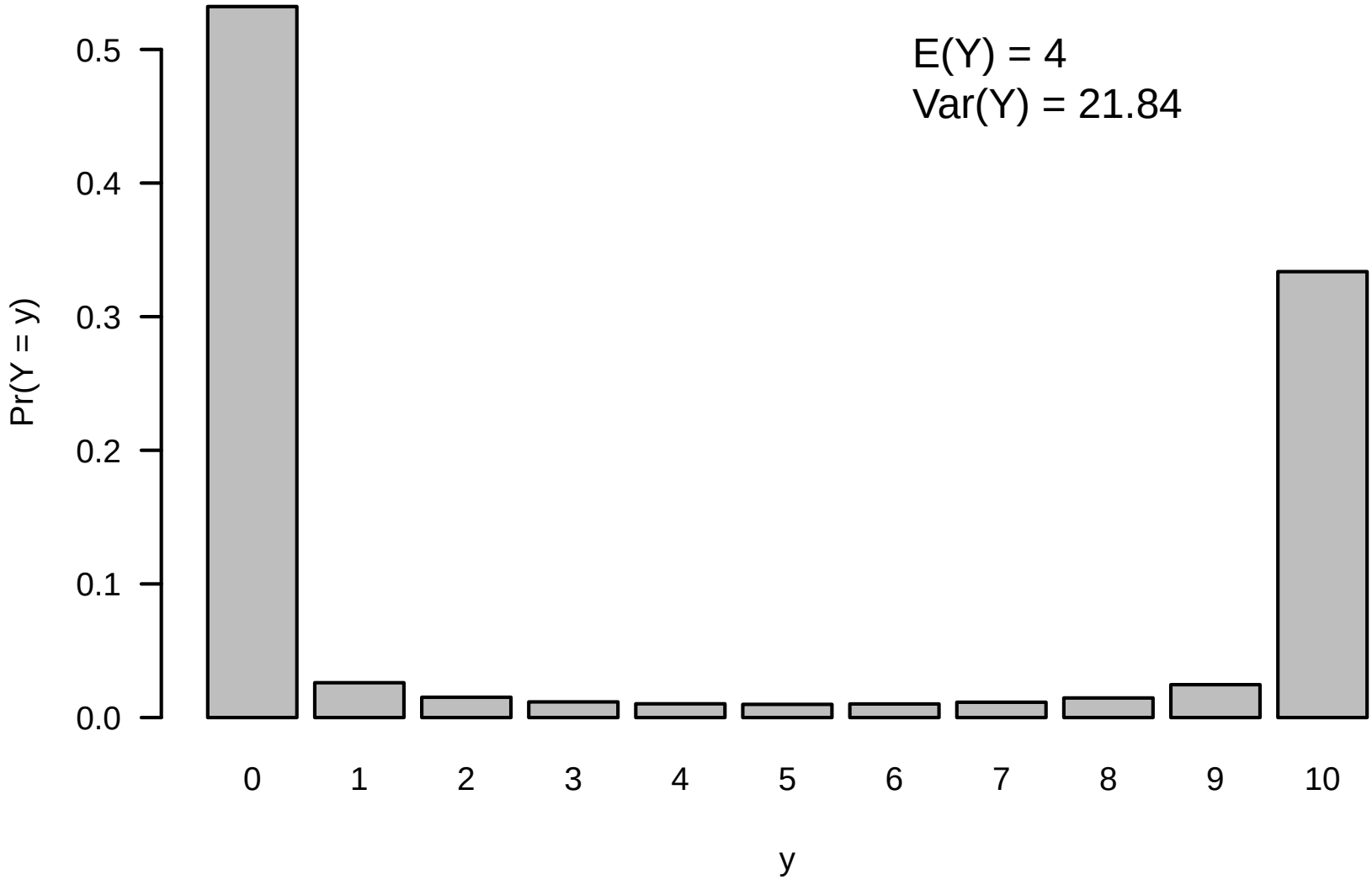
The binomial distribution

The binomial distribution with $n = 10$ and $p = 0.4$.



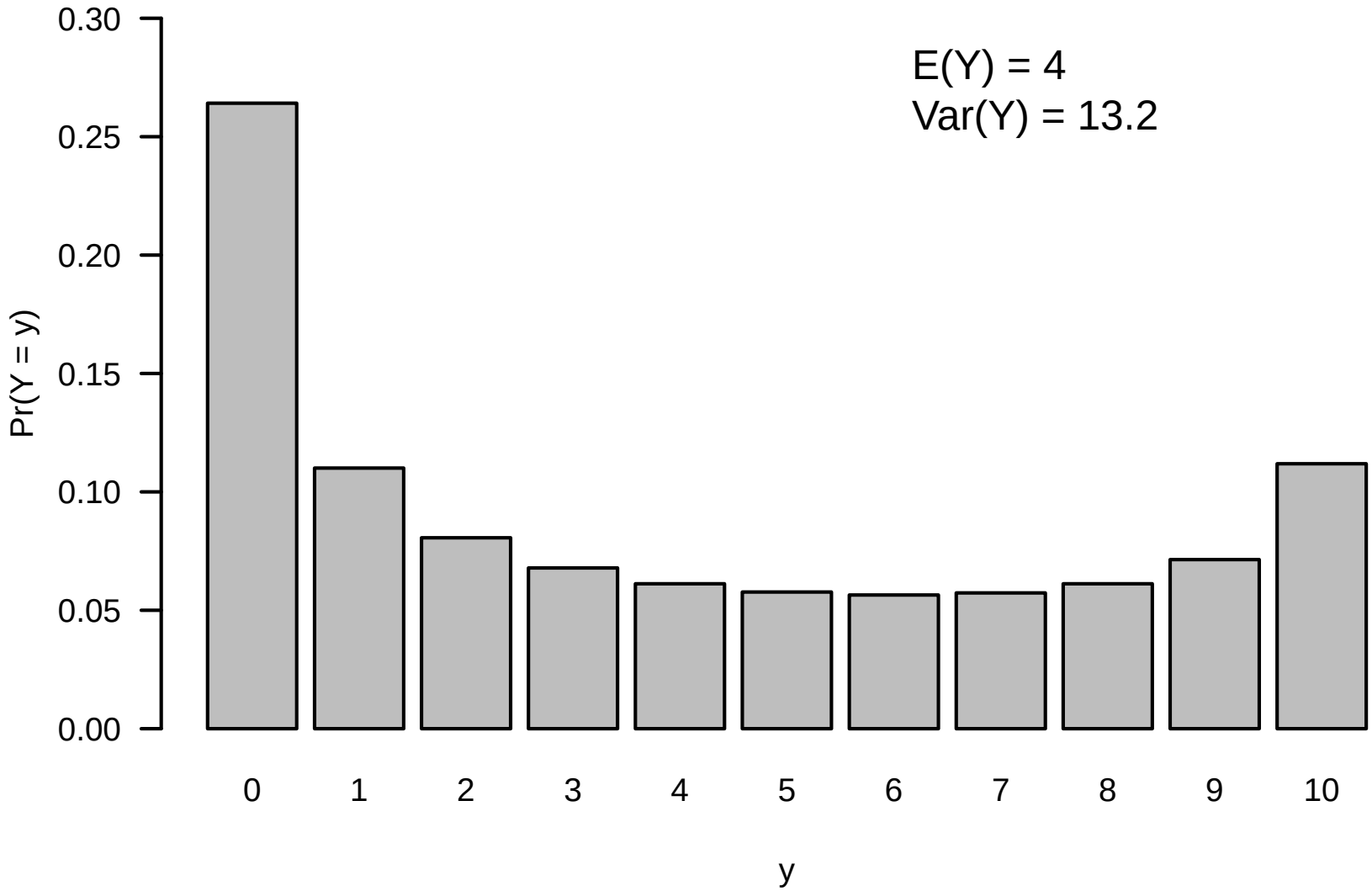
The beta-binomial distribution

The beta-binomial distribution with $n = 10$, $p = 0.4$, and $\rho = 0.9$.



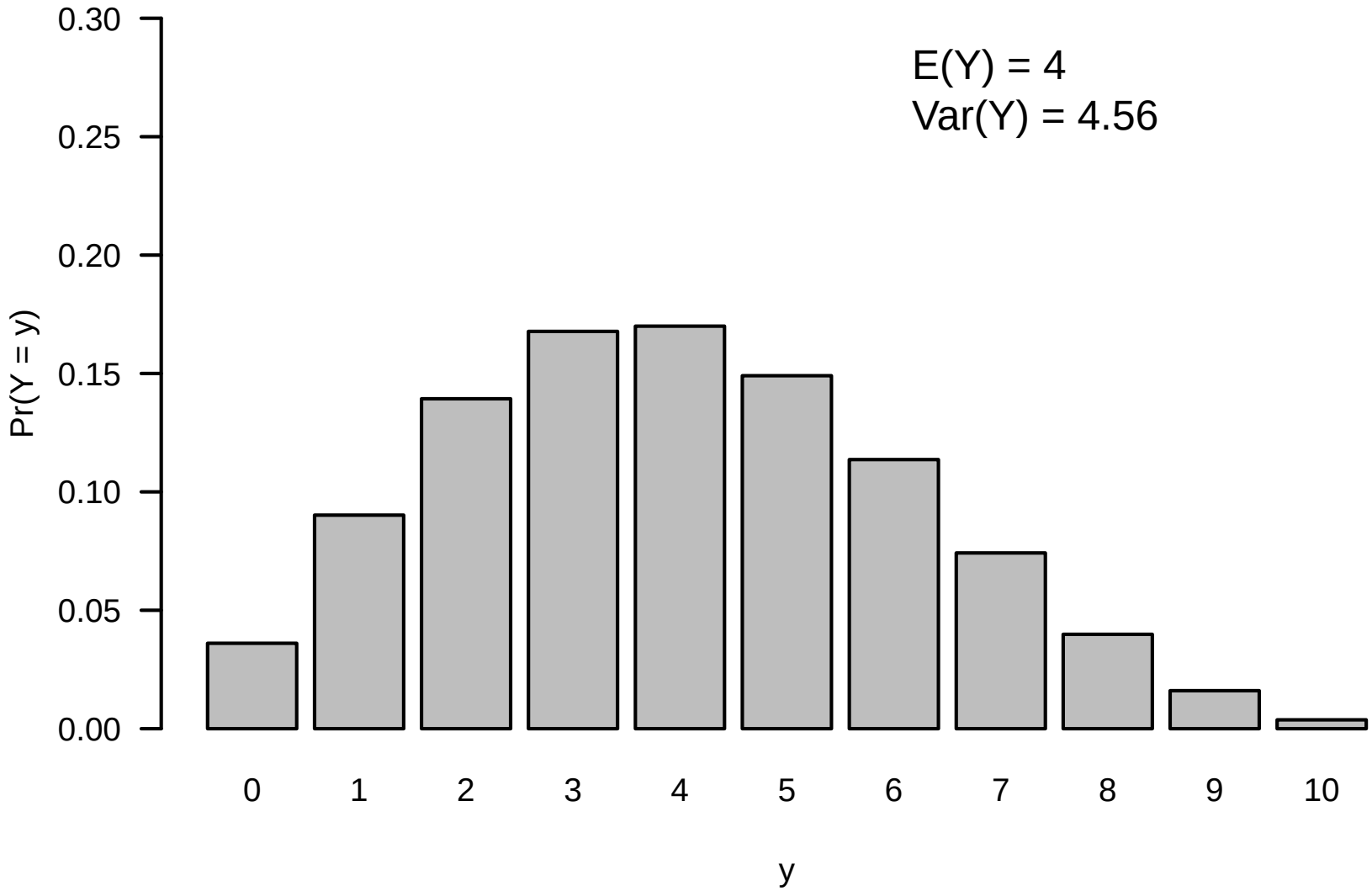
The beta-binomial distribution

The beta-binomial distribution with $n = 10$, $p = 0.4$, and $\rho = 0.5$.



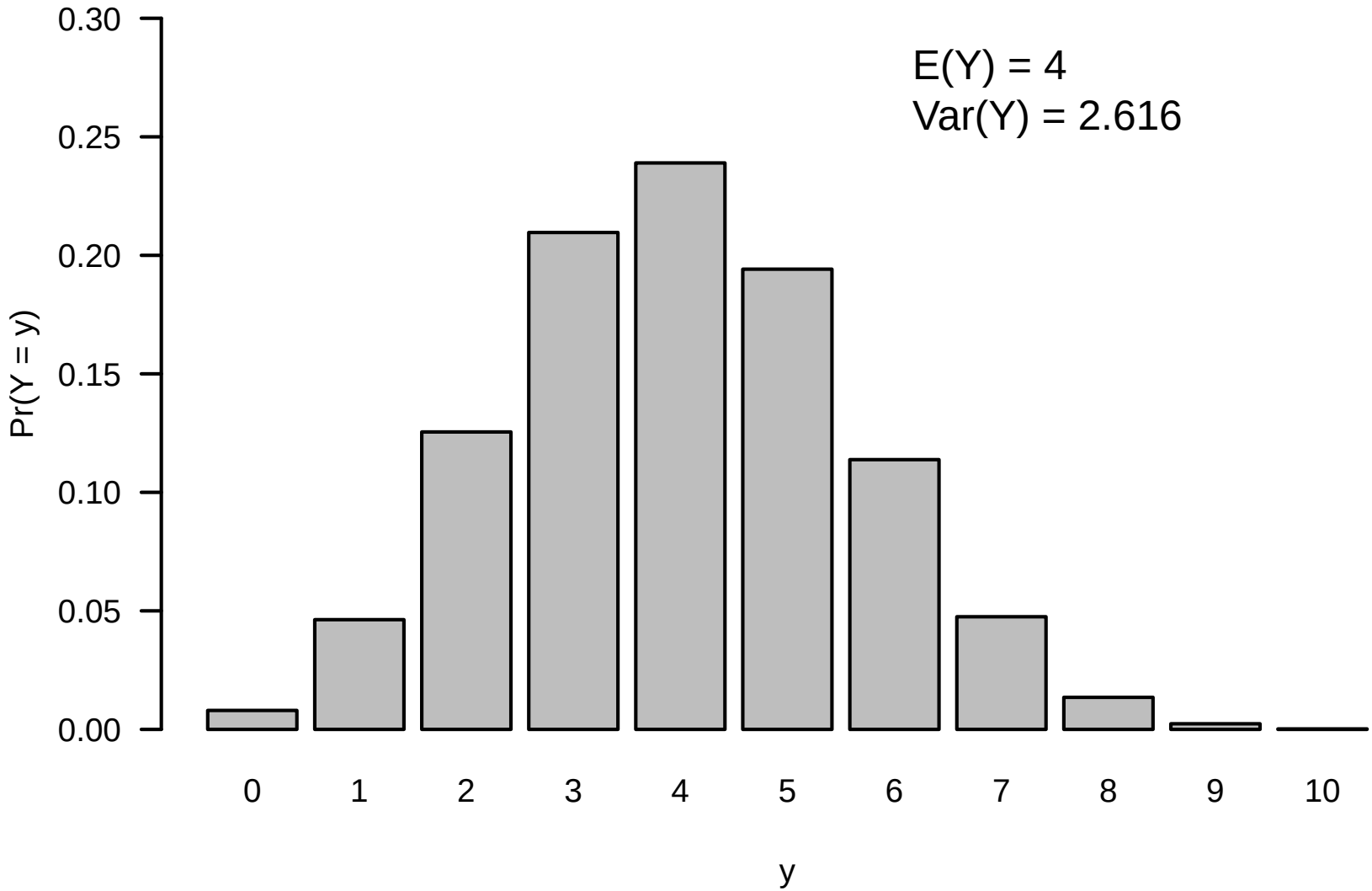
The beta-binomial distribution

The beta-binomial distribution with $n = 10$, $p = 0.4$, and $\rho = 0.1$.



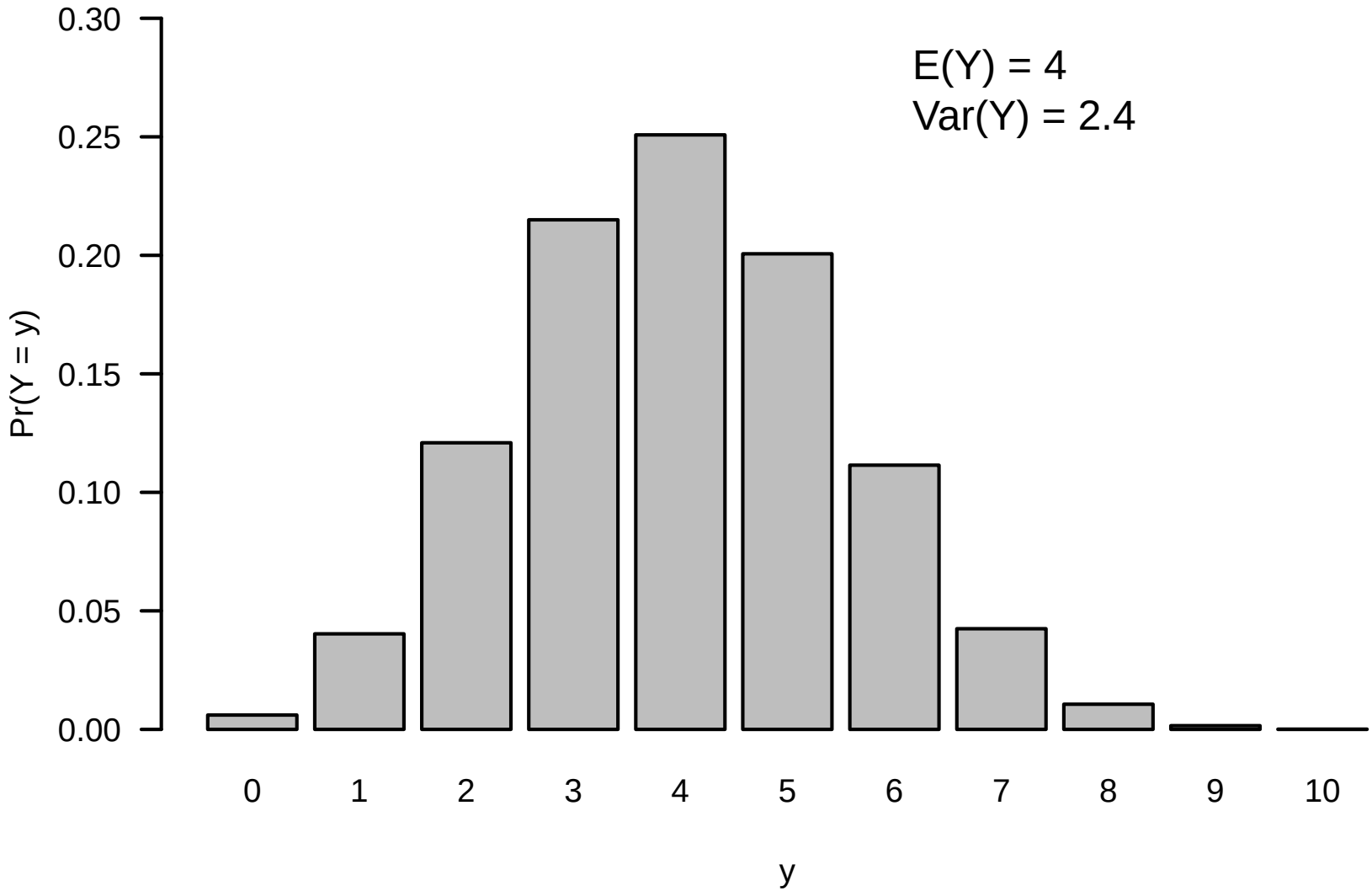
The beta-binomial distribution

The beta-binomial distribution with $n = 10$, $p = 0.4$, and $\rho = 0.01$.



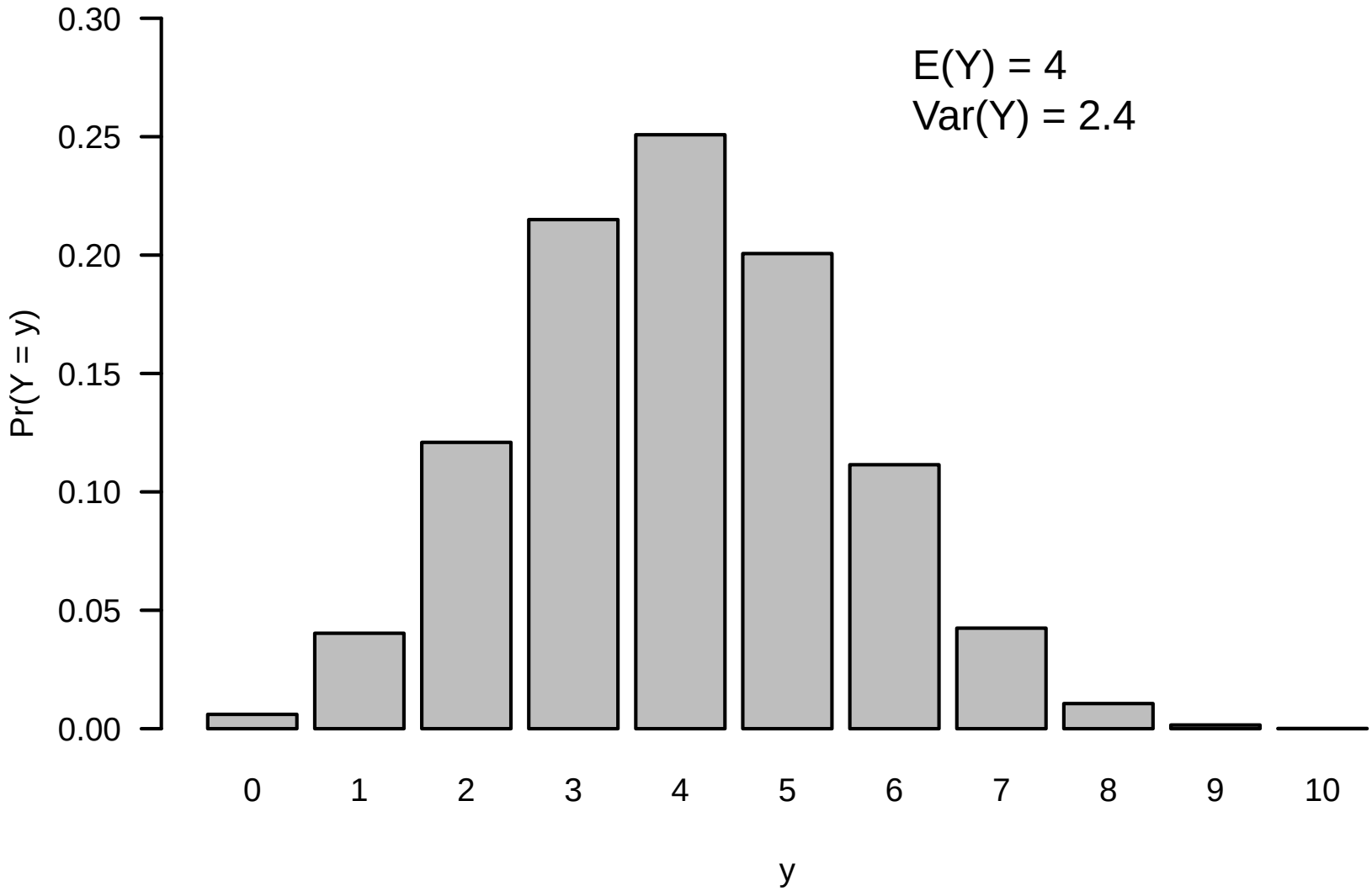
The beta-binomial distribution

The beta-binomial distribution with $n = 10$, $p = 0.4$, and $\rho = 0$.



The beta-binomial distribution

The binomial distribution with $n = 10$ and $p = 0.4$.



Quasibinomial and beta-binomial regression models

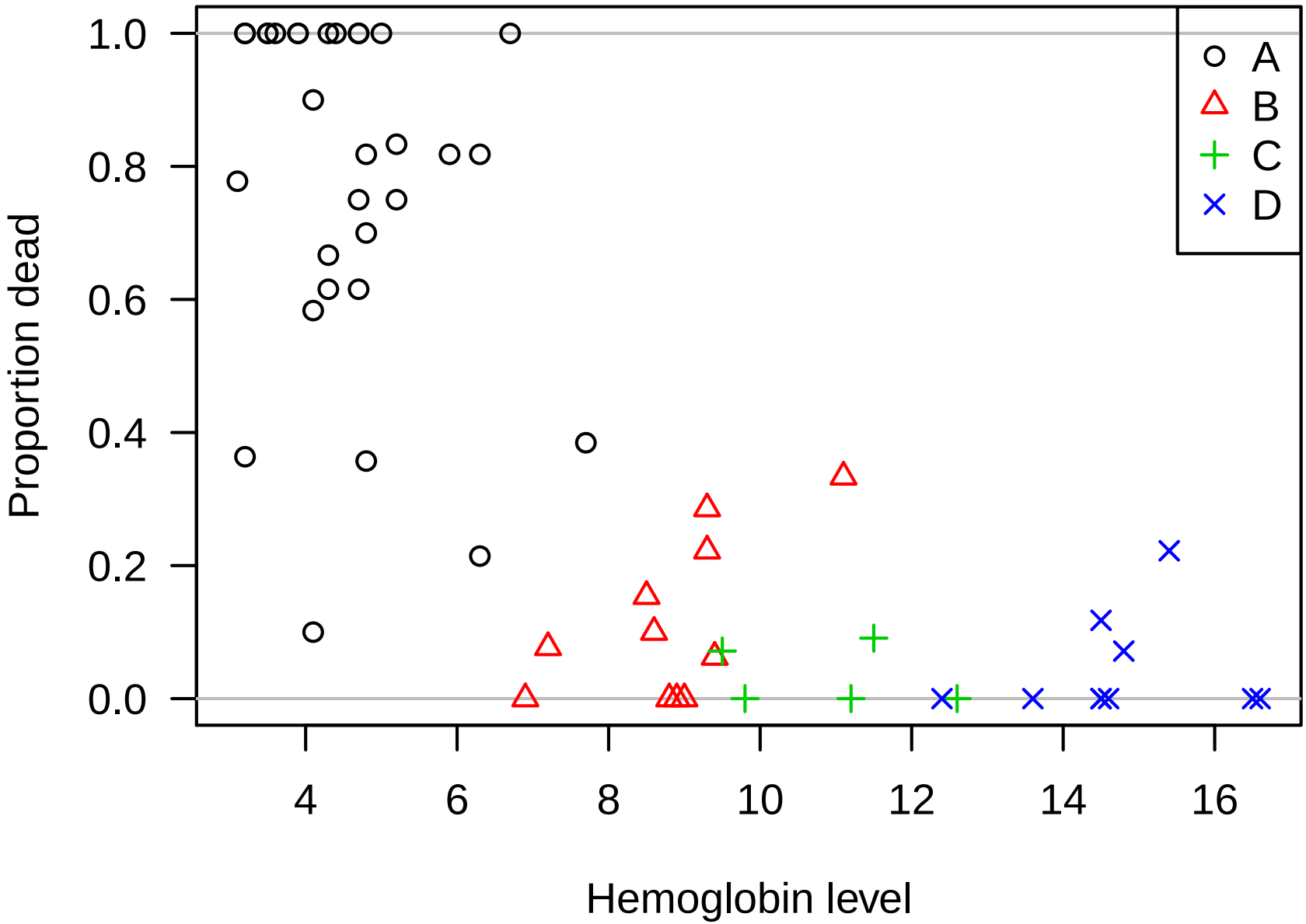
Our analysis of the CHD data set with a logistic regression model did not find any evidence of overdispersion, so we'll briefly look at a new data set to provide examples of quasibinomial and beta-binomial regression models.

The `lirat.df` data set has an observation for each of 58 litters of rats birthed by mothers with low iron levels. The variables are as follows:

- N: The size of the litter
- R: The number of dead fetuses
- hb: The mother's hemoglobin level
- grp: A group corresponding to the treatment given to the mother. Group `A` was given no supplementation, while groups `B`–`D` were given increasing dosages of iron injections.

Quasibinomial and beta-binomial regression models

The data



Logistic regression

A logistic regression model reveals strong evidence to suggest overdispersion:

```
binom.fit <- glm(cbind(R, N - R) ~ grp + hb, family = "binomial")
summary(binom.fit)

## Call:
## glm(formula = cbind(R, N - R) ~ grp + hb, family = "binomial")
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.1715     0.5064   4.288 1.80e-05 ***
## grpB         -2.4753     0.5042  -4.910 9.13e-07 ***
## grpC         -3.1624     0.9442  -3.349 0.000811 ***
## grpD         -2.0675     1.0647  -1.942 0.052150 .
## hb           -0.2174     0.1022  -2.127 0.033380 *
## ---
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 508.27  on 56  degrees of freedom
## Residual deviance: 168.78  on 52  degrees of freedom
```

Logistic regression

We could fit a quasibinomial model instead:

```
qbinom.fit <- glm(cbind(R, N - R) ~ grp + hb, family = "quasibinomial")
summary(qbinom.fit)

## Call:
## glm(formula = cbind(R, N - R) ~ grp + hb, family = "quasibinomial")
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.1715     0.8759   2.479  0.01645 *
## grpB         -2.4753     0.8721  -2.838  0.00645 **
## grpC         -3.1624     1.6333  -1.936  0.05828 .
## grpD         -2.0675     1.8416  -1.123  0.26674
## hb           -0.2174     0.1767  -1.230  0.22425
## ---
## (Dispersion parameter for quasibinomial family taken to be 2.991942)
##
##     Null deviance: 508.27  on 56  degrees of freedom
## Residual deviance: 168.78  on 52  degrees of freedom
```

Beta-binomial regression

Or a beta-binomial model:

```
library(VGAM)
bbinom.fit <- vglm(cbind(R, N - R) ~ grp + hb, family = "betabinomial", data = lirat.df)
summary(bbinom.fit)
```

```
##
## Call:
## vglm(formula = cbind(R, N - R) ~ grp + hb, family = "betabinomial",
##     data = lirat.df)
##
## Pearson residuals:
##                     Min      1Q  Median     3Q   Max
## logitlink(mu)  -2.772 -0.6879 -0.1431 0.9481 1.579
## logitlink(rho) -1.611 -0.9007  0.1699 0.6020 4.250
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1   2.0896     0.8534   2.449 0.014345 *
## (Intercept):2  -1.1718     0.3252  -3.603 0.000314 ***
## grpB           -2.4530     0.8574  -2.861 0.004225 **
## grpC           -2.8840     1.3623  -2.117 0.034255 *
## grpD           -2.3638     1.8113  -1.305 0.191882
## hb             -0.1610     0.1741  -0.925 0.355048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(mu), logitlink(rho)
##
## Log-likelihood: -92.9079 on 108 degrees of freedom
##
## Number of Fisher scoring iterations: 8
##
## No Hauck-Donner effect found in any of the estimates
```

A comparison

Note that the logistic regression model misleadingly finds significance in the effect of hemoglobin:

```
anova(binom.fit, test = "Chisq")

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: cbind(R, N - R)
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                    56     508.27
## grp   3   335.04         53     173.24  < 2e-16 ***
## hb    1     4.46         52     168.78  0.03476 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A comparison

But the models that account for overdispersion do not:

```
anova(qbinom.fit, test = "F")

## Analysis of Deviance Table
##
## Model: quasibinomial, link: logit
##
## Response: cbind(R, N - R)
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev       F    Pr(>F)
## NULL                    56     508.27
## grp   3   335.04        53     173.24 37.3264 5.231e-13 ***
## hb    1     4.46        52     168.78  1.4896    0.2278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A comparison

But the models that account for overdispersion do not:

```
anova(bbinom.fit)

## Analysis of Deviance Table (Type II tests)
##
## Model: 'betabinomial'
##
## Links: 'logitlink', 'logitlink'
##
## Response: cbind(R, N - R)
##
##
##     Df 2 * LogLik Diff. Pr(>Chi)
## grp  3          12.2158 0.006679 **
## hb   1           0.7789 0.377489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Quasi-models or a distributional assumption?

When we wish to model overdispersion we can either fit a quasi-model or assume a different distribution for the response:

- We can change a Poisson regression model with evidence of overdispersion to a quasi-Poisson or negative binomial regression model.
- We can change a logistic regression model with evidence of overdispersion to a quasi-binomial or beta-binomial regression model.

When choosing which to use, one consideration is the assumption we're making about the response variable's variance.

- In the chicken example, the deviance residuals indicated that a negative binomial model was more appropriate than the quasi-Poisson.

Quasi-models or a distributional assumption?

Another consideration is that the theoretical properties of the likelihood function no longer hold when we fit a quasi-model, which is a disadvantage. We have access to a range of useful likelihood-based tools, but they cannot be used for quasi-models.

One such tool is the deviance, which is calculated using the maximised likelihood function. This is a handy tool for checking goodness-of-fit.

- In our chicken example we used the deviance to test goodness-of-fit for the negative binomial model, but we couldn't use it for the quasi-Poisson model.

Later in the course we will meet other likelihood-based tools (e.g., AIC and BIC for model selection), which are available for negative binomial and beta-binomial models, but not quasi-models.

Quasi-models or a distributional assumption?

Negative binomial or beta-binomial regression models do not allow for underdispersion, although this phenomenon is much rarer. We can still use quasi-models, resulting $\hat{k} < 1$.

If we want to avoid quasi-models, then there are further distributions we could consider. For example, the Conway-Maxwell-Poisson distribution generalises the Poisson distribution to allow for either over- or underdispersion, but its probability mass function is impossible to compute exactly. Approximations can be computationally expensive or introduce error (see R package `COMPoissonReg`).