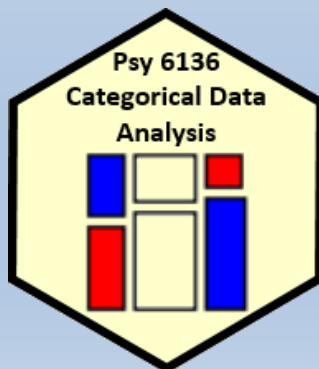


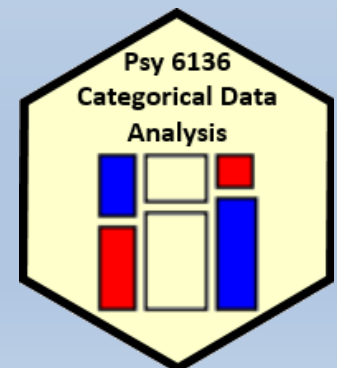
# Discrete distributions



Michael Friendly

Psych 6136

<http://friendly.github.io/psy6136>



# Discrete distributions: Basic ideas

- Quantitative data: often assumed Normal ( $\mu, \sigma^2$ ) – unreasonable for CDA
- Binomial, Poisson, Negative binomial, ... are the building blocks for CDA
- Form the basis for modeling techniques
  - logistic regression, generalized linear models, Poisson regression
- Data:
  - outcome variable ( $k = 0, 1, 2, \dots$ )
  - counts of occurrences ( $n_k$ ): accidents, words in text, males in families of size  $k$

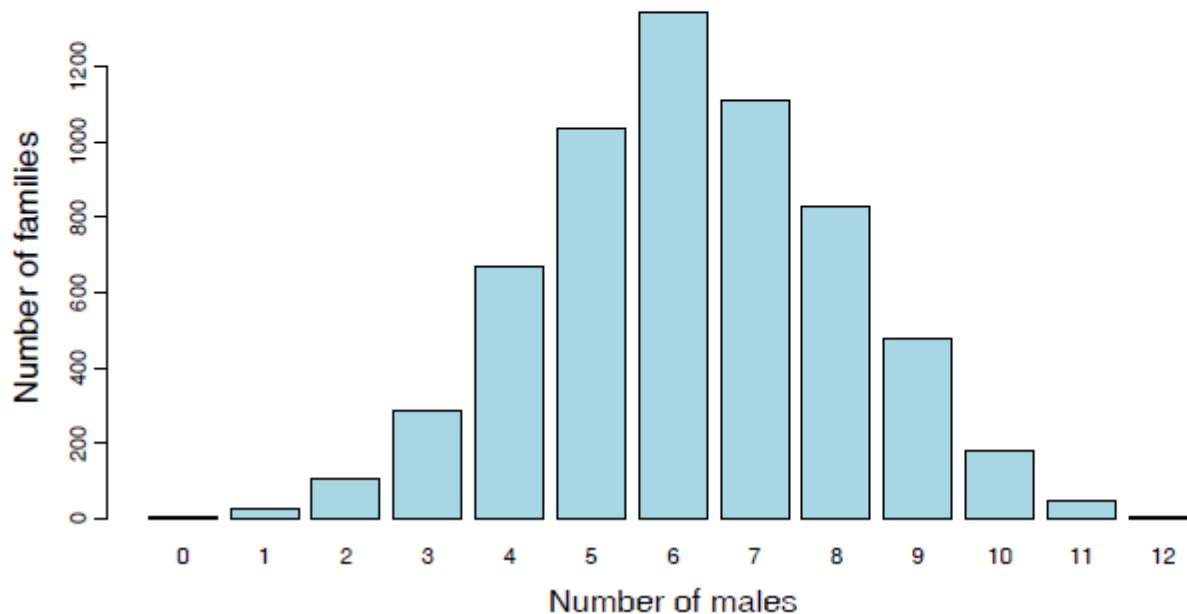
# Examples: binomial

Human sex ratio (Geissler, 1889): Is there evidence that  $\Pr(\text{male}) = 0.5$ ?

## Saxony families

Saxony families with 12 children having  $k = 0, 1, \dots, 12$  sons.

$k$	0	1	2	3	4	5	6	7	8	9	10	11	12
$n_k$	3	24	104	286	670	1033	1343	1112	829	478	181	45	7



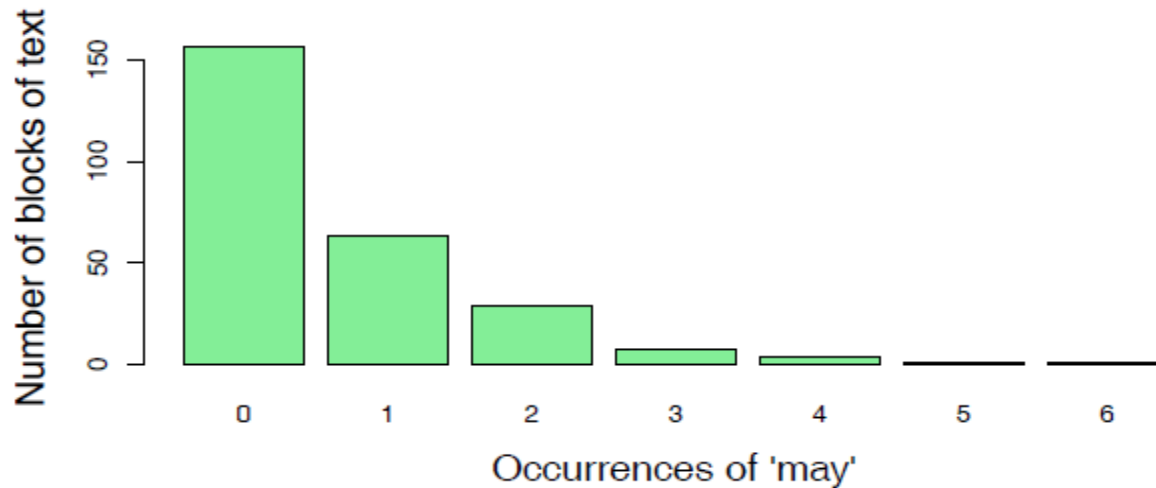
# Examples: count data

## Federalist papers: Disputed authorship

- 77 essays by Alexander Hamilton, John Jay, James Madison to persuade voters to ratify the US constitution, all signed with pseudonym “Publius”
  - Who wrote each?
  - 65 known, 12 disputed (H & M both claimed sole authorship)
- Mosteller & Wallace (1984): analysis of frequency dist<sup>n</sup>s of key “marker” words: from, **may**, whilst, ...
- e.g., blocks of 200 words: occurrences ( $k$ ) of “may” in how many blocks ( $n_k$ )

```
> data(Federalist, package = "vcd")
> Federalist
nMay
  0    1    2    3    4    5    6
156  63  29   8   4   1   1
```

# Count data: models



For each word (“from”, “may”, “whilst”, ...)

- Fit a probability model (Poisson, NegBin)
- Estimate parameters ( $\lambda$ ,  $\theta$ )
- → Calculate log Odds (Hamilton vs. Madison)
- → All 12 disputed papers most likely written by [Madison](#)

# Example: Type-token distributions

- Basic count,  $k$ : number of “types”; frequency,  $n_k$ : number of instances observed
  - Frequencies of distinct words in a book or literary corpus
  - Number of subjects listing words as members of the semantic category “fruit”
  - Distinct species of animals caught in traps
- Differs from other distributions in that the frequency for  $k = 0$  is *unobserved*
- Distribution is often extremely skewed (J-shaped)

Table: Number of butterfly species  $n_k$  for which  $k$  individuals were collected

Individuals ( $k$ )	1	2	3	4	5	6	7	8	9	10	11	12	
Species ( $n_k$ )	118	74	44	24	29	22	20	19	20	15	12	14	
Individuals ( $k$ )	13	14	15	16	17	18	19	20	21	22	23	24	Sum
Species ( $n_k$ )	6	12	6	9	9	6	10	10	11	5	3	3	5

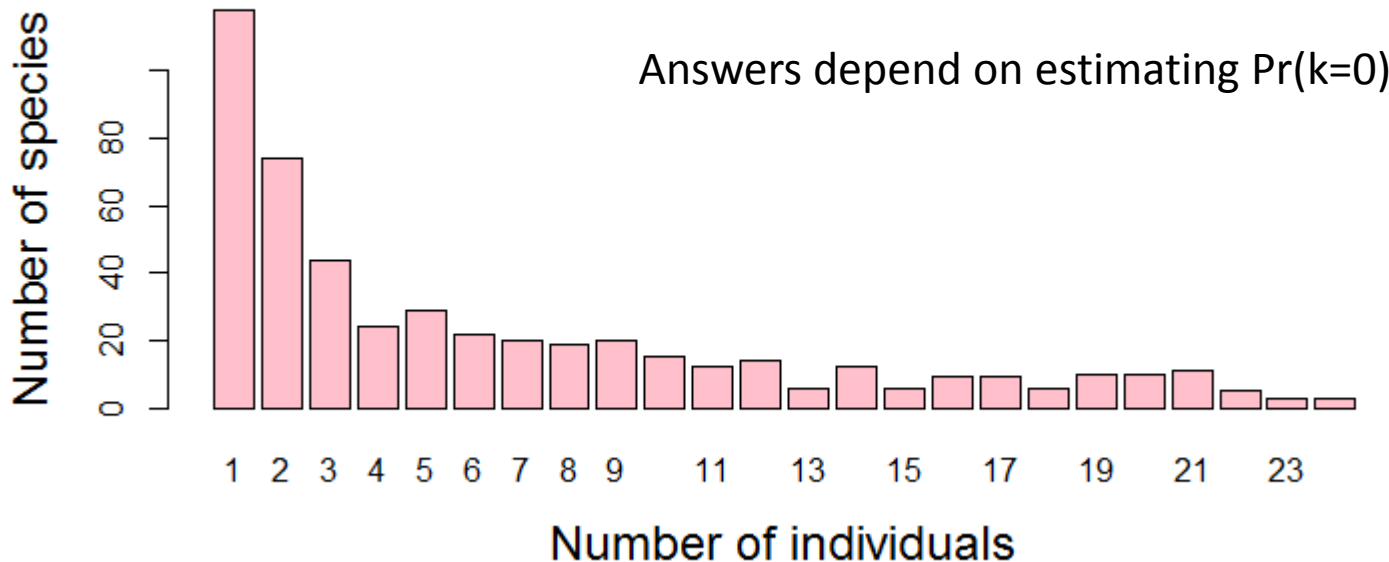
```
data(Butterfly, package="vcd")
barplot(Butterfly,
        xlab = "Number of individuals",
        ylab = "Number of species",
        col = "pink",
        cex.lab = 1.5)
```

### Questions:

What is the total pop. of butterflies in Malaysia?

How many wolves remain in Canada NWT?

How many words did Shakespeare know?



# Discrete distributions: Questions

- General questions
  - What process gave rise to the distribution?
  - What is the form: uniform, binomial, Poisson, negative binomial, ... ?
  - → Fit & estimate parameters
    - Visualize goodness of fit
  - → Use in some larger context to tell a story
- Examples
  - *Families in Saxony*: might expect  $\text{Bin}(n=12, p)$ ;  $p=0.5$ ?
  - *Federalist papers*: Perhaps  $\text{Poisson}(\lambda)$
  - *Butterfly data*: Perhaps a log-series distribution?



# Fitting discrete distributions

## Lack of fit:

- Lack of fit tells us something about the *process* giving rise to the data
- Poisson: assumes constant small probability of the basic event
- Binomial: assumes constant probability and independent trials
- Negative binomial: allows for *overdispersion*, relative to Poisson

## Motivation:

- Models for more complex categorical data use these basic discrete distributions
- Binomial (with predictors) → logistic regression
- Poisson (with predictors) → poisson regression, loglinear models
- ⇒ many of these are special cases of *generalized linear models*

# Common discrete distributions

Discrete distributions are characterized by a probability function,  $\Pr(X = k) \equiv p(k)$ , that the random variable  $X$  has value  $k$ .

- Common discrete distributions have the following forms:

Discrete distribution	Probability function, $p(k)$	Parameters
Binomial	$\binom{n}{k} p^k (1 - p)^{n-k}$	$p = \Pr(\text{success});$ $n = \# \text{ trials}$
Poisson	$e^{-\lambda} \lambda^k / k!$	$\lambda = \text{mean}$
Negative binomial	$\binom{n+k-1}{k} p^n (1 - p)^k$	$p; n = \# \text{ successful trials}$
Geometric	$p(1 - p)^k$	$p$
Logarithmic series	$\theta^k / [-k \log(1 - \theta)]$	$\theta$

# Discrete distributions: R

## R functions: {d, p, q, r}

- d\_\_\_\_\_ density function,  $\Pr(X=k) = p(k)$
- p\_\_\_\_\_ cumulative probability,  $F(k) = \sum_{X \leq k} p(k)$
- q\_\_\_\_\_ quantile function, find  $k = F^{-1}(p)$ , smallest value such that  $F(k) \geq p$
- r\_\_\_\_\_ random number generator

Discrete distribution	Density (pmf) function	Cumulative (CDF)	Quantile $CDF^{-1}$	Random # generator
Binomial	<code>dbinom()</code>	<code>pbinom()</code>	<code>qbinom()</code>	<code>rbinom()</code>
Poisson	<code>dpois()</code>	<code>ppois()</code>	<code>qpois()</code>	<code>rpois()</code>
Negative binomial	<code>dnbinom()</code>	<code>pnbinom()</code>	<code>qnbinom()</code>	<code>rnbinom()</code>
Geometric	<code>dgeom()</code>	<code>pgeom()</code>	<code>qgeom()</code>	<code>rgeom()</code>
Logarithmic series	<code>dlogseries()</code>	<code>plogseries()</code>	<code>qlogseries()</code>	<code>rlogseries()</code>

# Binomial distribution

The binomial distribution,  $\text{Bin}(n, p)$ ,

$$\text{Bin}(n, p) : \Pr\{X = k\} \equiv p(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n, \quad (1)$$

*Diagram annotations:*

- Red arrow from  $\binom{n}{k}$  to "# ways to get k out of n"
- Red arrow from  $p^k$  to "Pr(k events)"
- Red arrow from  $(1-p)^{n-k}$  to "Pr(n-k non-events)"

arises as the distribution of the number of events of interest ("successes") which occur in  $n$  *independent trials* when the probability of the event on any one trial is the *constant* value  $p = \Pr(\text{event})$ .

## Examples

- Toss 10 fair coins— how many heads?  $\text{Bin}(10, \frac{1}{2})$
- Toss 12 fair dice— how many 5s or 6s?  $\text{Bin}(12, 1/3)$

Mean, variance, skewness:

$$\text{Mean}[X] = n p$$

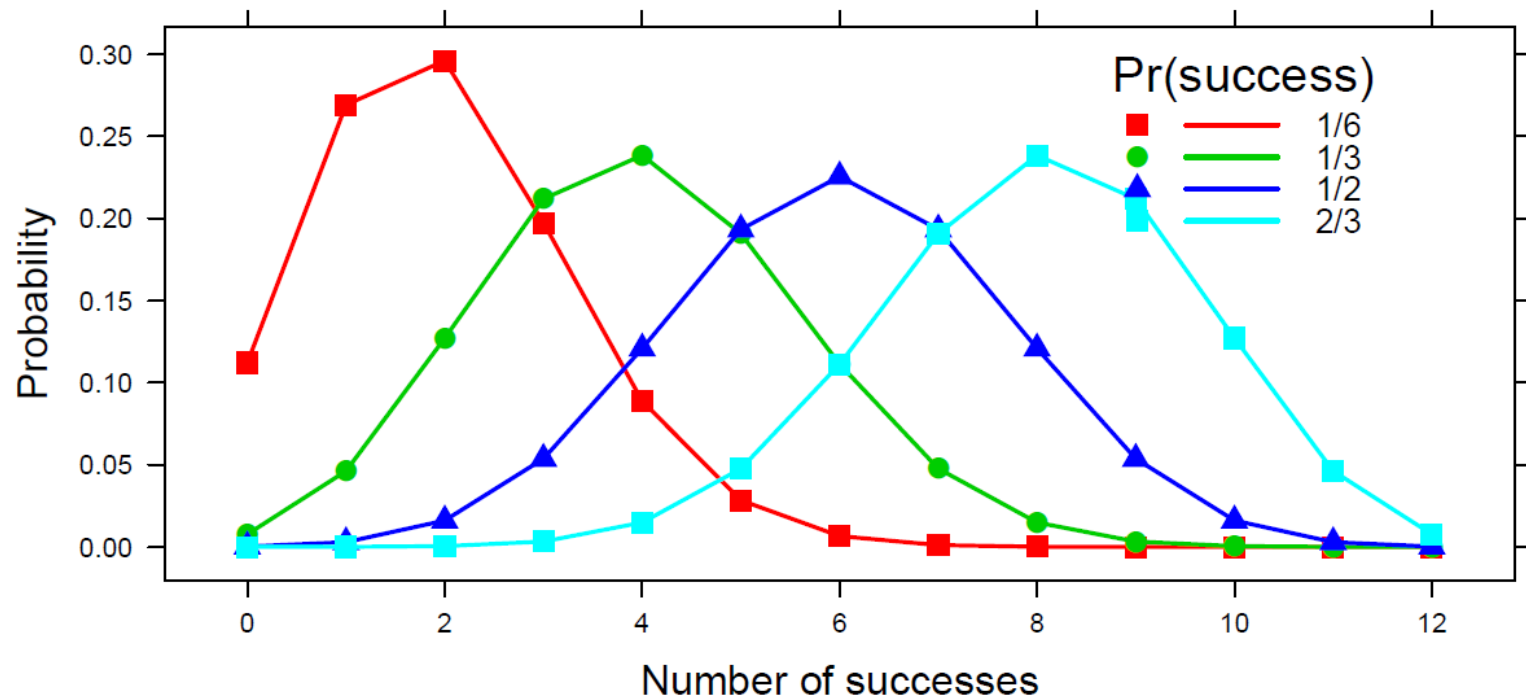
$$\text{Var}[X] = n p (1-p) = n p q$$

$$\text{Skew}[X] = n p q (q-p)$$

$$\text{MLE from data: } \hat{p} = \frac{\bar{x}}{n} = \frac{\sum_k k \times n_k / \sum_k n_k}{n}$$

# Binomial distribution

Binomial distributions for  $k = 0, 1, 2, \dots, 12$  successes in  $n=12$  trials, for 4 values of  $p$



- Mean =  $n p$
- Variance is maximum when  $p = \frac{1}{2}$
- Skewed when  $p \neq \frac{1}{2}$

# Poisson distribution

The Poisson distribution,  $\text{Pois}(\lambda)$ ,

$$\text{Pois}(\lambda) : \Pr\{X = k\} \equiv p(k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad k = 0, 1, \dots \quad (2)$$

gives the probability of an event occurring  $k = 0, 1, 2, \dots$  times over a *large number of independent* trials, when the probability,  $p$ , that the event occurs on any one trial (in time or space) is *small and constant*.

Examples:

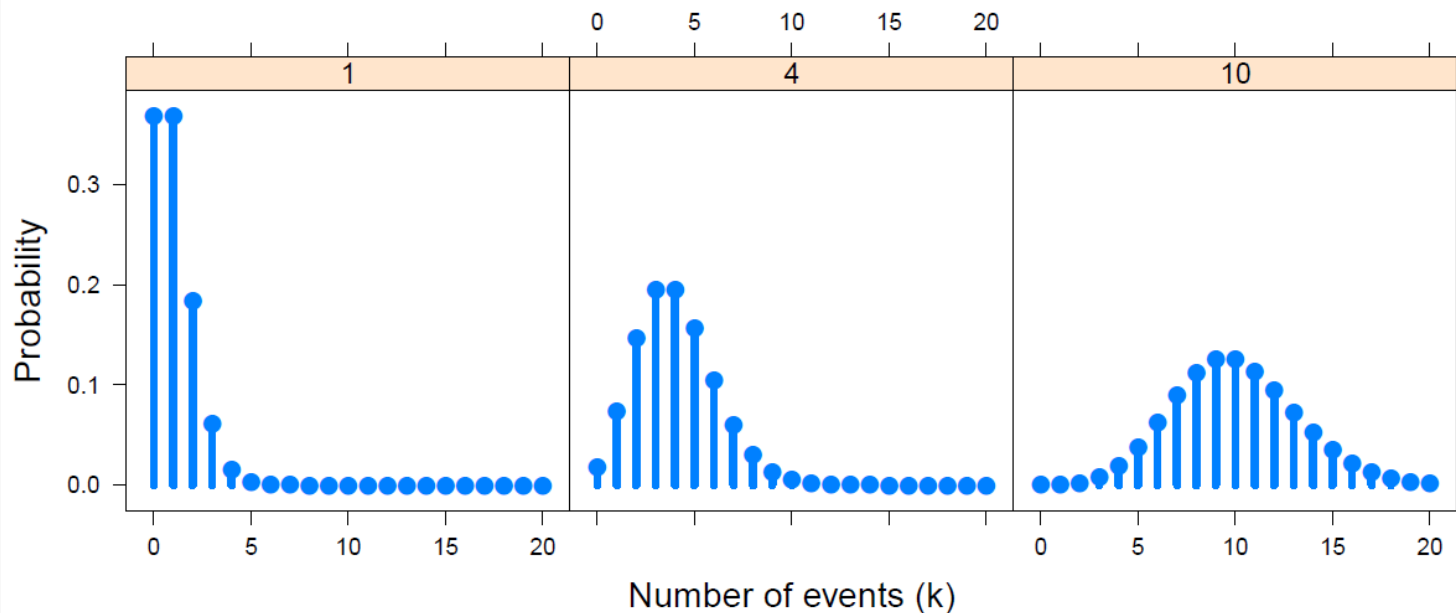
- Number of highway accidents at some given location
- Defects in a manufacturing process
- Number of goals scored in soccer games

**Table:** Total goals scored in 380 games in the Premier Football League, 1995/95 season

Total goals	0	1	2	3	4	5	6	7
Number of games	27	88	91	73	49	31	18	3

# Poisson distribution

Poisson distributions for  $\lambda = 1, 4, 10$



DDAR Fig  
3.10, p 81

Mean, variance, skewness:

$$\text{Mean}[X] = \lambda$$

$$\text{Var}[X] = \lambda$$

$$\text{Skew}[X] = \lambda^{-1/2}$$

$$\text{MLE: } \hat{\lambda} = \bar{x}$$

Properties:

Sum of Pois ( $\lambda_1, \lambda_2, \lambda_3, \dots$ ) = Pois( $\sum \lambda_i$ )

Approaches  $N(\lambda, \lambda)$  as  $n \rightarrow \infty$

# Negative binomial distribution

The Negative binomial distribution,  $\text{NBin}(n, p)$ ,

$$\text{NBin}(n, p) : \Pr\{X = k\} \equiv p(k) = \binom{n+k-1}{k} p^n (1-p)^k \quad k = 0, 1, \dots, \infty$$

is a **waiting time** distribution. It arises when  $n$  trials are observed with constant probability  $p$  of some event, and we ask how many **non-events** (failures),  $k$ , it takes to observe  $n$  **successful events**.

Example: Toss a coin; what is probability of getting  $k = 0, 1, 2, \dots$  tails before  $n = 3$  heads?

This distribution is often used as an alternative to the Poisson when

- constant probability  $p$  or **independence** are violated
- variance is greater than the mean (**overdispersion**:  $\text{Var}[X] > \text{Mean}[X]$  )

$$\text{Mean}(X) = nq/p = \mu$$

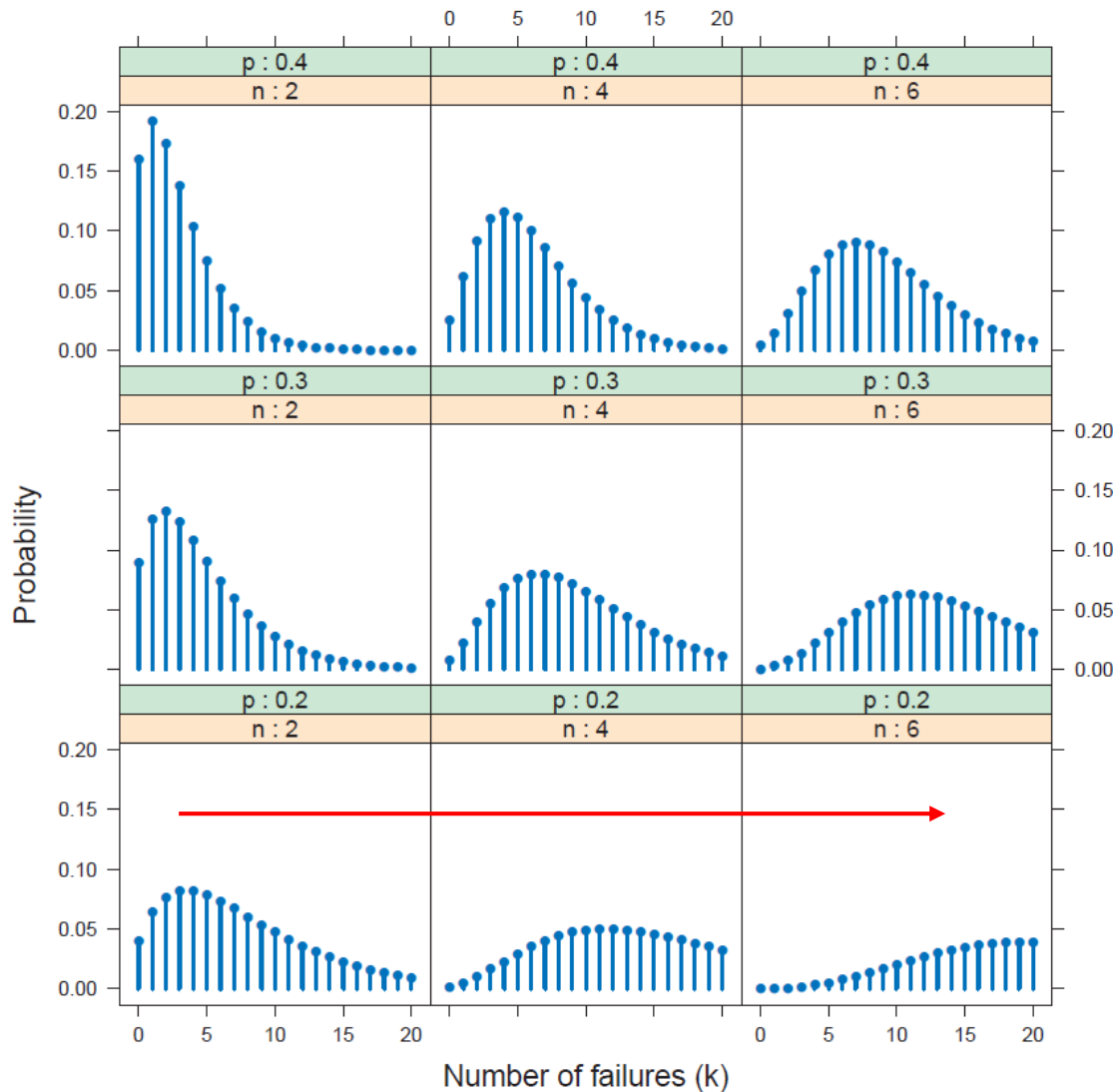
$$\text{Var}(X) = nq/p^2$$

$$\text{Skew}(X) = \frac{2-p}{\sqrt{nq}},$$

$$\text{Mean}(X) = \mu = \frac{n(1-p)}{p} \implies p = \frac{n}{n+\mu},$$

$$\text{Var}(X) = \frac{n(1-p)}{p^2} \implies \text{Var}(X) = \mu + \frac{\mu^2}{n}.$$





Negative binomial  
distributions for  
 $n = 2, 4, 6$   
 $p = 0.2, 0.3, 0.4$

Mean:  
Increases with  $n$   
Decreases with  $p$

DDAR Fig 3.13, p 85

# Fitting discrete distributions

Fitting a discrete distribution involves the following steps:

- 1 Estimate the parameter(s) from the data, e.g.,  $p$  for binomial,  $\lambda$  for Poisson, etc. Typically done using maximum likelihood, but some distributions have simple expressions:
  - Binomial,  $\hat{p} = \sum kn_k / (n \sum n_k) = \text{mean} / n$
  - Poisson,  $\hat{\lambda} = \sum kn_k / \sum n_k = \text{mean}$
- 2 Calculate fitted probabilities,  $\hat{p}(k)$  for the distribution, and then fitted frequencies,  $N\hat{p}(k)$ .
- 3 Assess Goodness of fit: Pearson  $\chi^2$  or likelihood-ratio  $G^2$

$$\chi^2 = \sum_{k=1}^K \frac{(n_k - N\hat{p}_k)^2}{N\hat{p}_k} \quad G^2 = \sum_{k=1}^K n_k \log\left(\frac{n_k}{N\hat{p}_k}\right)$$

Both have asymptotic chisquare distributions,  $\chi^2_{K-s}$  with  $s$  estimated parameters, under the hypothesis that the data follows the chosen distribution.

# Fitting & graphing discrete distributions

In R, the **vcd** and **vcdExtra** packages provide functions to fit, visualize and diagnose discrete distributions

- **Fitting:** `goodfit()` fits uniform, binomial, Poisson, neg bin, geometric, logseries, ...
- **Graphing:** `rootogram()` assess departure between observed, fitted counts
- **Ord plot:** `Ordplot()` diagnose form of a discrete distribution
- **Robust plots:** `distplot()` handle problems with discrepant counts

# Example: Saxony families

```
> data(Saxony, package="vcd")
```

```
> Saxony
```

```
nMales
```

0	1	2	3	4	5	6	7	8	9	10	11	12
3	24	104	286	670	1033	1343	1112	829	478	181	45	7

Use `goodfit()` to fit the binomial; test with `summary()`

```
> Sax.fit <- goodfit(Saxony, type = "binomial", par=list(size=12))
```

```
> summary(Sax.fit)
```

Goodness-of-fit test for binomial distribution

	X^2	df	P(> X^2)
Likelihood Ratio	97	11	6.98e-16

# Example: Saxony families

The `print()` method for **goodfit** objects shows the details

```
> Sax.fit      # print
```

Observed and fitted values for binomial distribution  
with parameters estimated by 'ML'

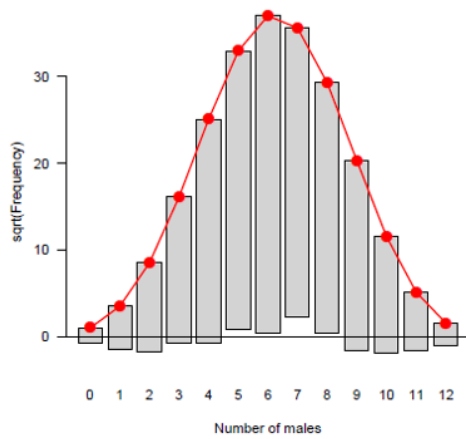
count	observed	fitted	pearson	residual
0	3	0.933		2.140
1	24	12.089		3.426
2	104	71.803		3.800
3	286	258.475		1.712
4	670	628.055		1.674
5	1033	1085.211		-1.585
6	1343	1367.279		-0.657
7	1112	1265.630		-4.318
8	829	854.247		-0.864
9	478	410.013		3.358
10	181	132.836		4.179
11	45	26.082		3.704
12	7	2.347		3.037

Pay attention to the  
**pattern** & magnitudes  
of residuals,  $d_k$

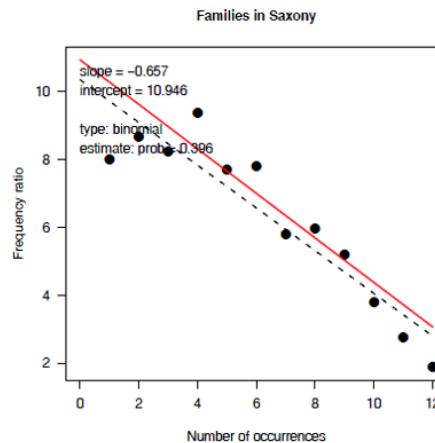
$$\text{Pearson } \chi^2 = \sum d_k^2$$

# Graphing discrete distributions

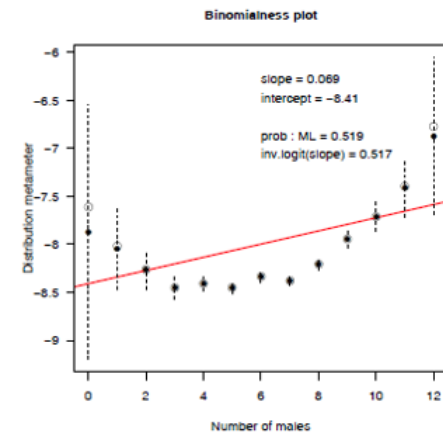
Rootograms



Ord plots



Robust  
distribution  
plots

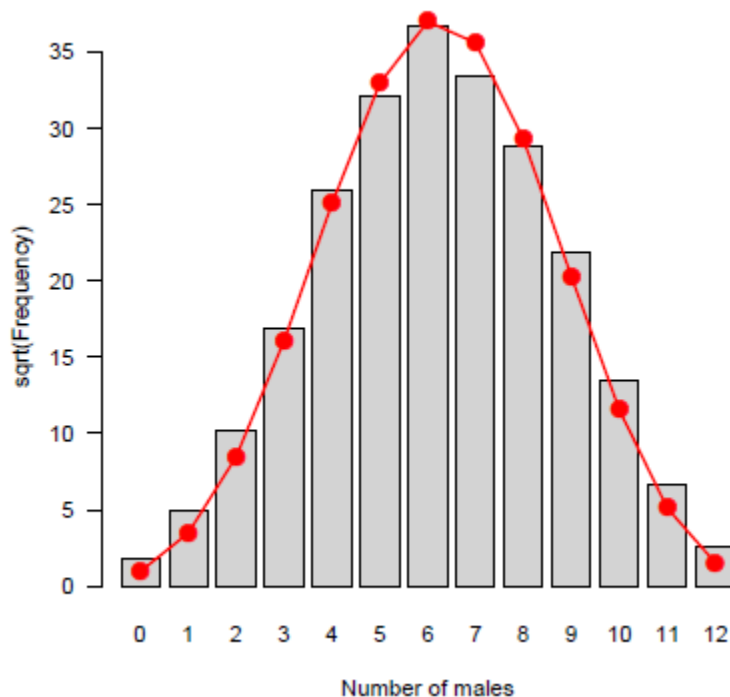


# What's wrong with simple histograms?

Discrete distributions are often graphed as histograms, with a theoretical fitted distribution superimposed

The plot() method for goodfit objects provides some alternatives

```
> plot(Sax.fit, type = "standing", xlab = "Number of males")
```

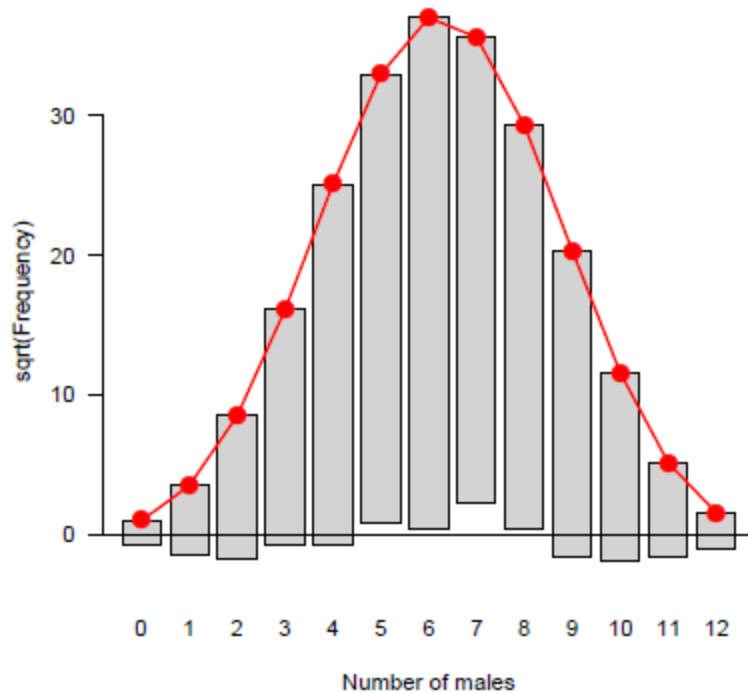


Problems:

- Largest frequencies dominate
- Must assess deviations vs. the fitted curve

# Hanging rootograms

```
> plot(Sax.fit, type = "hanging", xlab = "Number of males") # default
```



Tukey (1972, 1977):

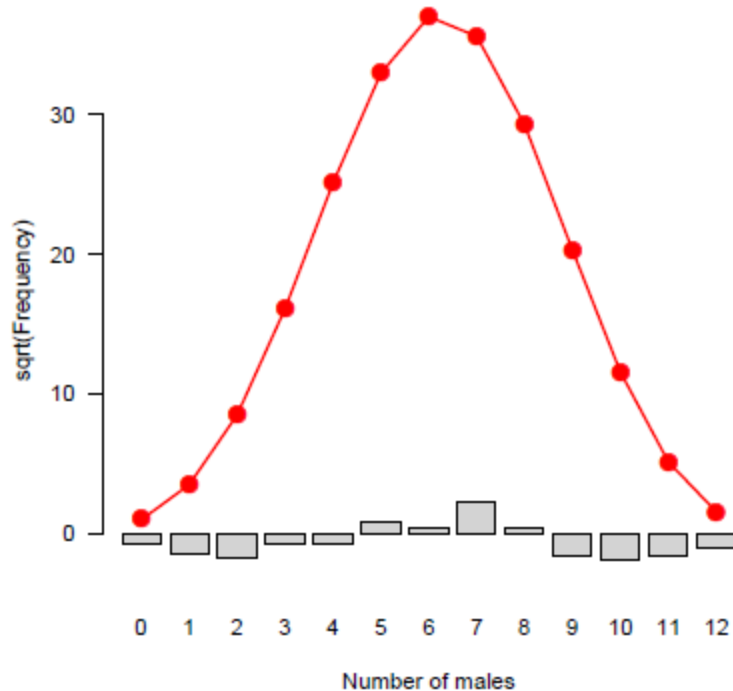
- shift histogram bars to the fitted curve
- → judge deviations vs. horizontal line.
- plot  $\sqrt{\text{freq}}$  → smaller frequencies are emphasized.

We can now see clearly *where* the binomial doesn't fit



# Deviation rootograms

```
> plot(Sax.fit, type = "deviation", xlab = "Number of males")
```



Deviation rootogram:

- emphasize differences between observed and fitted frequencies
- bars now show the residuals (gaps) directly

There are more families with very low or very high number of sons than the binomial predicts.

Q: Why is this so much better than the lack-of-fit test?

# Example: Federalist papers

```
> data(Federalist, package="vcd")
```

```
> Federalist
```

```
nMay
```

0	1	2	3	4	5	6
156	63	29	8	4	1	1

Fit the Poisson distribution

```
> Fed.fit0 <- goodfit(Federalist, type="poisson")
```

```
> summary(Fed.fit0)
```

```
Goodness-of-fit test for poisson  
distribution
```

	X^2	df	P(> X^2)
Likelihood Ratio	25.2	5	0.000125

This fits very poorly!

# Example: Federalist papers

Try the Negative binomial distribution

```
> Fed.fit1<- goodfit(Federalist, type="nbinomial")  
> summary(Fed.fit1)
```

Goodness-of-fit test for nbinomial distribution

	X^2	df	P(> X^2)
Likelihood Ratio	1.96	4	0.742

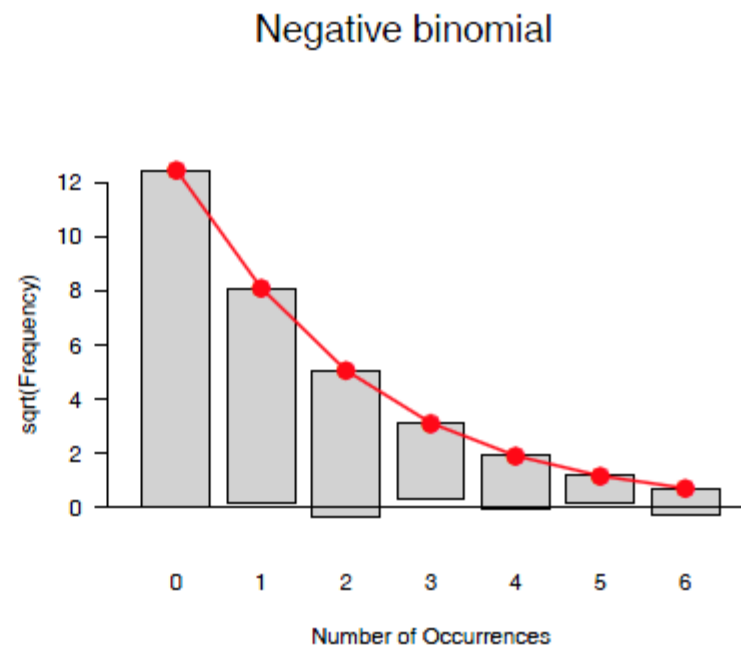
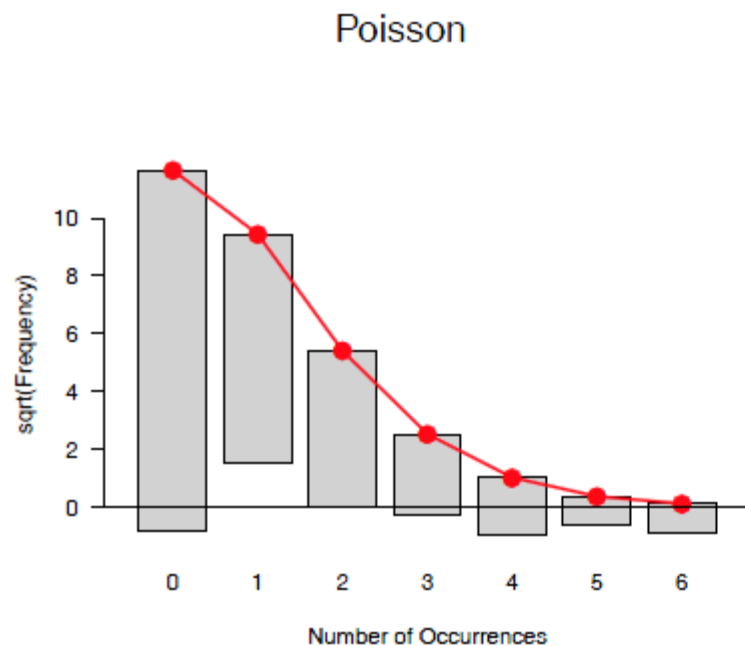
This now fits very well, indeed! Why?

- Poisson assumes that the probability of a given word (“may”) is constant across all blocks of text.
- Negative binomial allows the rate parameter  $\lambda$  to vary over blocks of text

# Federalist papers: Rootograms

Hanging rootograms for the Federalist papers data, comparing Poisson and Negative binomial

```
> plot(Fed.fit0, main = "Poisson")  
> plot(Fed.fit1, main = "Negative binomial")
```

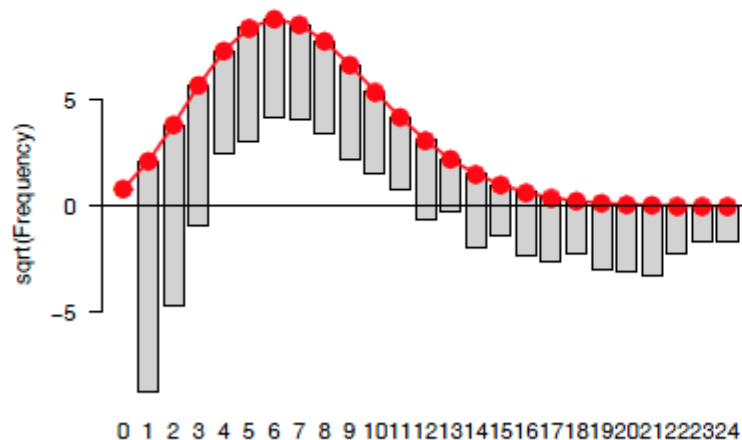


# Butterfly data

Both Poisson and Negative binomial are terrible fits! What to do??

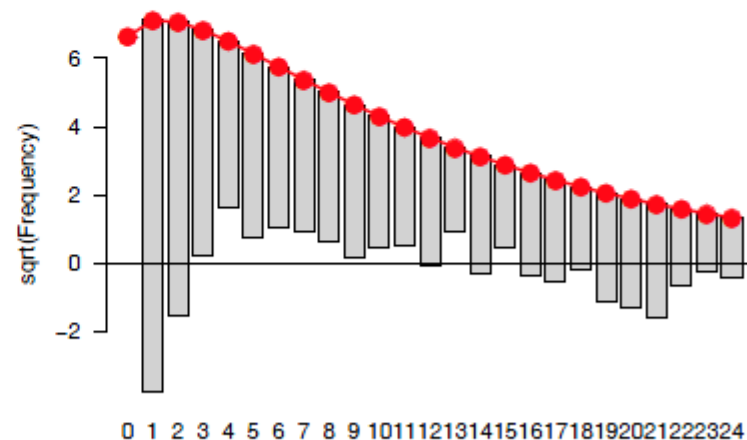
```
But.fit1 <- goodfit(Butterfly, type="poisson")
But.fit2 <- goodfit(Butterfly, type="nbinomial")
plot(But.fit1, main="Poisson")
plot(But.fit2, main="Negative binomial")
```

Poisson



Number of Occurrences

Negative binomial



Number of Occurrences

# Ord plots: Diagnose form of distribution

How to tell which discrete distributions are likely candidates?

- Ord (1967): for each of Poisson, Binomial, Negative binomial, and Logarithmic series distributions,
  - plot of  $kp_k/p_{k-1}$  against  $k$  is linear
  - signs of intercept and slope  $\rightarrow$  determine the form, give rough estimates of parameters

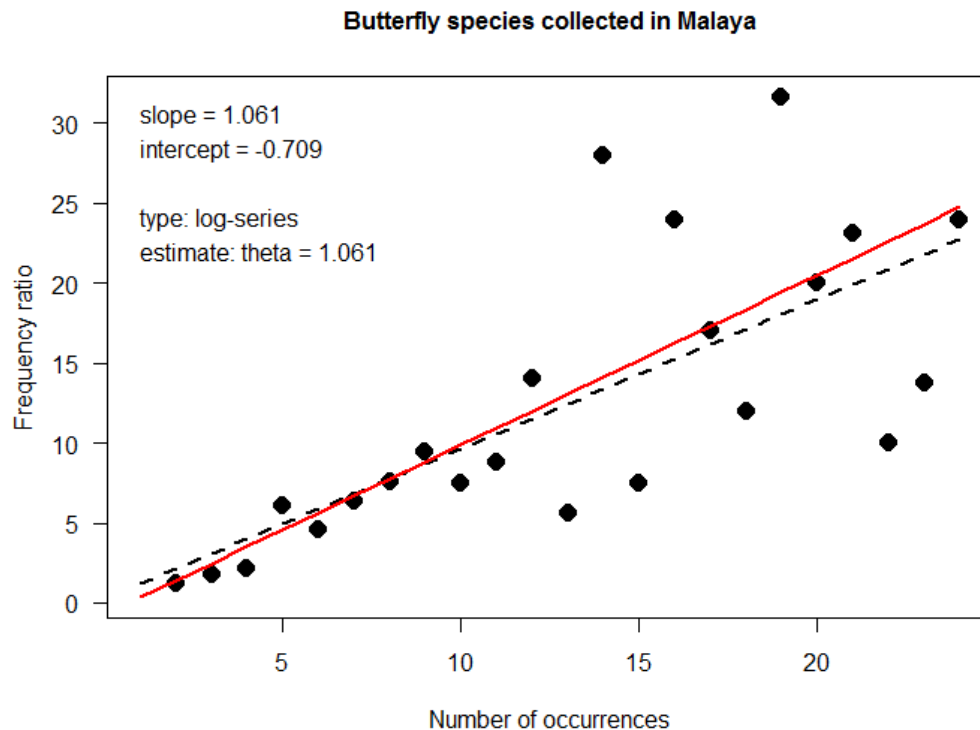
Slope (b)	Intercept (a)	Distribution (parameter)	Parameter estimate
0	+	Poisson ( $\lambda$ )	$\lambda = a$
—	+	Binomial ( $n, p$ )	$p = b/(b - 1)$
+	+	Neg. binomial ( $n, p$ )	$p = 1 - b$
+	—	Log. series ( $\theta$ )	$\theta = b$ $\theta = -a$

- Fit line by WLS, using  $\sqrt{n_k - 1}$  as weights
- A heuristic method: doesn't always work, but often a good start.

# Ord plot: Examples

Butterfly data: The slope and intercept correctly diagnoses the log-series distribution

```
> Ord_plot(Butterfly,  
            main = "Butterfly species collected in Malaya",  
            gp=gpar(cex=1), pch=16)
```

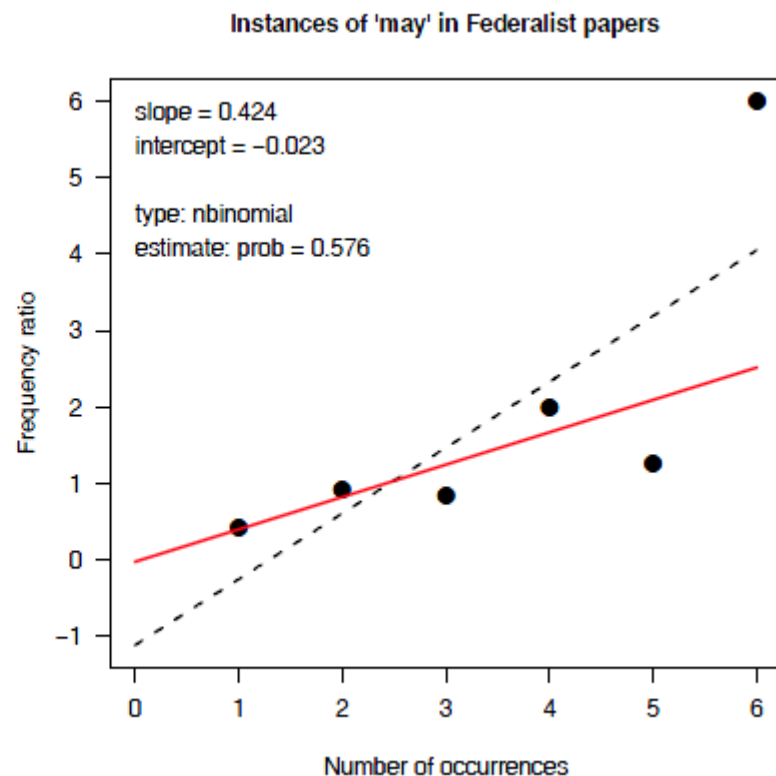
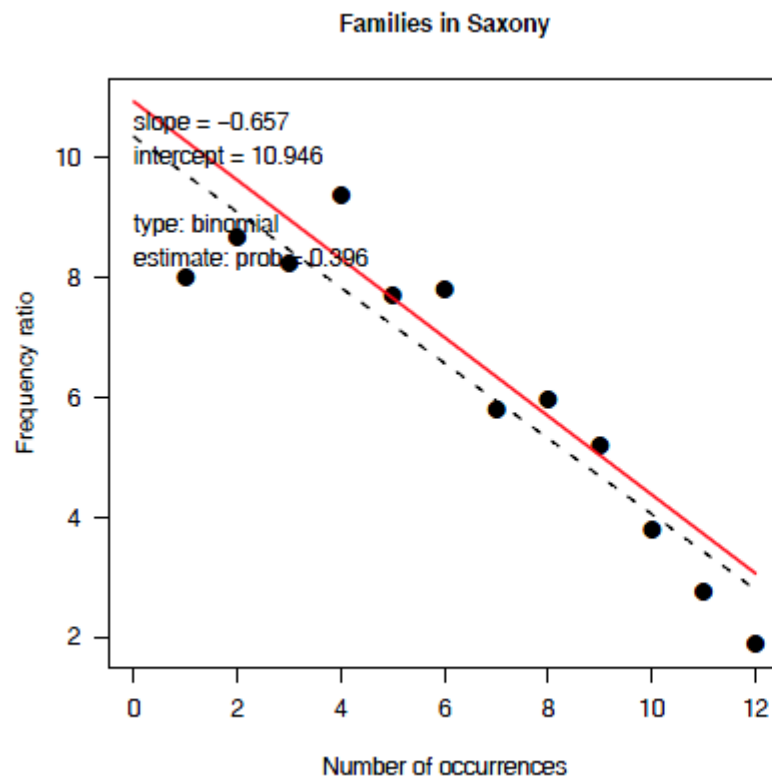


+ slope  
- intercept  
→ log-series

# Ord plots: Examples

## Ord plots for the Saxony and Federalist data

```
> Ord_plot(Saxony, main = "Families in Saxony", gp=gpar(cex=1), pch=16)  
> Ord_plot(Federalist, main = "Instances of 'may' in Federalist papers", gp=gpar(cex=1), pch=16)
```

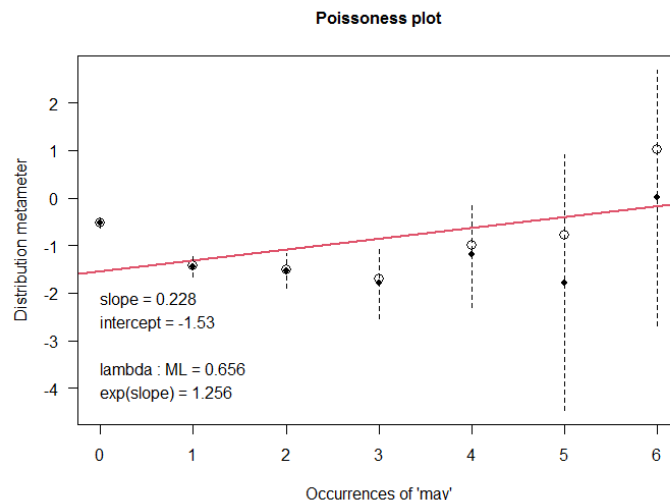




# Robust distribution plots

- Ord plots lack robustness
  - one discrepant frequency,  $n_k$  affects points for both  $k$  and  $k + 1$
  - the use of WLS to fit the line is a small attempt to minimize this
- Robust plots for Poisson distribution (Hoaglin and Tukey, 1985)
  - For Poisson, plot **count metameter**  $= \phi(n_k) = \log_e(k! n_k / N)$  vs.  $k$
  - Linear relation  $\Rightarrow$  Poisson, slope gives  $\hat{\lambda}$
  - CI for points, diagnostic (influence) plot
  - Implemented in `distplot()` in the `vcd` package

For the Poisson distribution, this is called a “poissonness plot”



# Poissonness plot: Details

- If the distribution of  $n_k$  is  $\text{Poisson}(\lambda)$  for some fixed  $\lambda$ , then each observed frequency,  $n_k \approx m_k = Np_k$ .
- Then, setting  $n_k = Np_k = e^{-\lambda} \lambda^k / k!$ , and taking logs of both sides gives

$$\log(n_k) = \log N - \lambda + k \log \lambda - \log k!$$

which can be rearranged to

$$\phi(n_k) \equiv \log \left( \frac{k! n_k}{N} \right) = -\lambda + (\log \lambda) k$$

- $\Rightarrow$  if the distribution is Poisson, plotting  $\phi(n_k)$  vs.  $k$  should give a line with
  - intercept =  $-\lambda$
  - slope =  $\log \lambda$
- Nonlinear relation  $\rightarrow$  distribution is *not* Poisson
- Hoaglin and Tukey (1985) give details on calculation of confidence intervals and influence measures.

# Other distributions

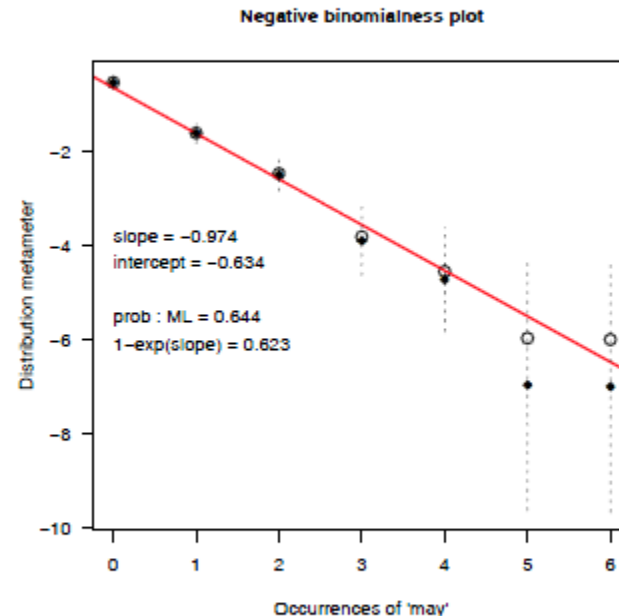
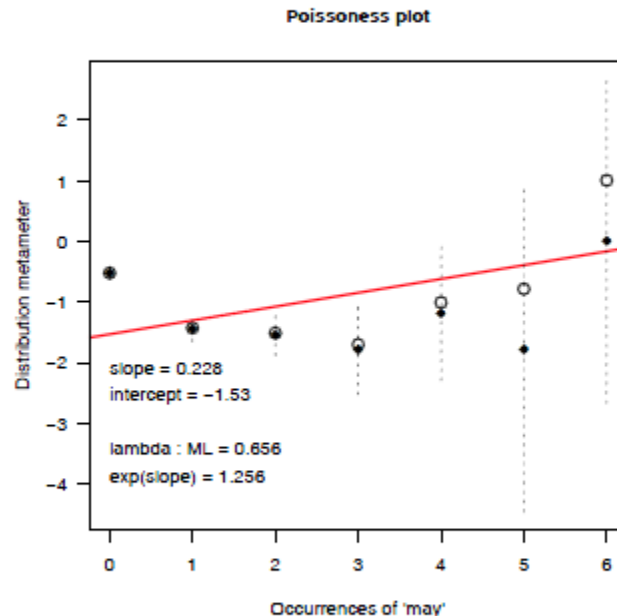
This idea extends readily to other discrete data distributions:

- The binomial, Poisson, negative binomial, geometric and logseries distributions are all members of a general **power series family** of discrete distributions. See: *DDAR*, Table 3.10 for details.
- This allows all of these to be represented in a plot of a suitable count metameter,  $\phi(n_k)$  vs.  $k$ . See: *DDAR*, Table 3.12 for details.
- In these plots, a straight line confirms that the data follow the given distribution.
- Confidence intervals around the points indicate **uncertainty** for the count metameter.
- The slope and intercept of the line give **estimates** of the distribution parameters.

# distplot: Federalist

Try both Poisson & Negative binomial

```
distplot(Federalist, type="poisson", xlab="Occurrences of 'may'")  
distplot(Federalist, type="nbinomial", xlab="Occurrences of 'may'")
```

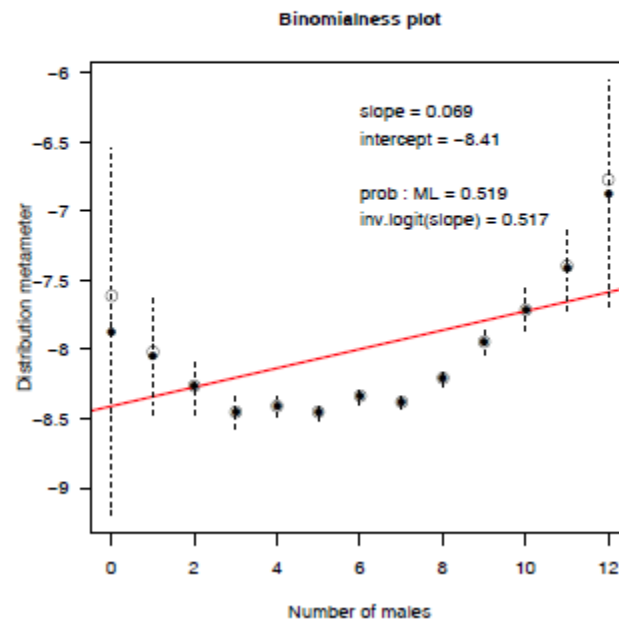
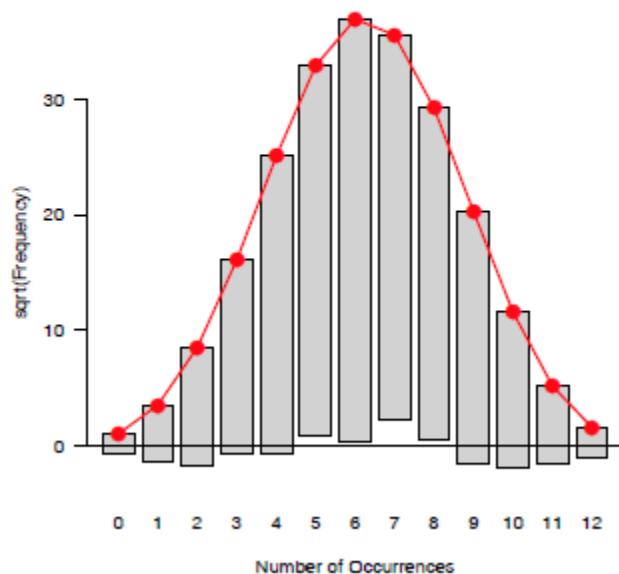


Again, the Poisson distribution is seen not to fit, while the Negative binomial appears reasonable.

# distplot: Saxony

For purported binomial distributions, the result is a “binomialness” plot

```
plot(goodfit(Saxony, type="binomial", par=list(size=12)))  
distplot(Saxony, type="binomial", size=12, xlab="Number of males")
```



Both plots show heavier tails than the binomial distribution. `distplot()` is more sensitive in diagnosing this

# What have we learned?

Main points:

- Discrete distributions involve basic *counts* of occurrences of some event occurring with varying *frequency*.
- The ideas and methods for one-way tables are building blocks for analysis of more complex data.
- Commonly used discrete distributions include the binomial, Poisson, negative binomial, and logarithmic series distributions, all members of a *power series* family.
- Fitting observed data to a distribution  $\rightarrow$  fitted frequencies,  $N\hat{p}_k$ ,  $\rightarrow$  goodness-of-fit tests (Pearson  $X^2$ , LR  $G^2$ )
- R: `goodfit()` provides `print()`, `summary()` and `plot()` methods.
- Plotting with rootograms, Ord plots and generalized distribution plots can reveal *how* or *where* a distribution does not fit.

# What have we learned?

Some explanations:

- The Saxony data were part of a much larger data set from Geissler (1889) (Geissler in [vcdExtra](#)).
  - For the binomial, with families of size  $n = 12$ , our analyses give  $\hat{p} = \Pr(\text{male}) = 0.52$ .
  - Other analyses (using more complex models) conclude that  $p$  varies among families with the same size.
  - One explanation is that family decisions to have another child are influenced by the boy–girl ratio in earlier children.
- As suggested earlier, the lack of fit of the Poisson distribution for words in the Federalist papers can be explained by [context](#) of the writing:
  - Given “marker” words appear more or less often over time and subject than predicted by constant rates ( $\lambda$ ) for a given author (Madison or Hamilton)
  - The negative binomial distribution fit much better.
  - The estimated parameters for these texts allowed assigning all 12 disputed papers to Madison.

# Looking ahead: PhdPubs data

Example 3.24 in DDAR gives data on the number of publications by PhD candidates in the last 3 years of study

```
data("PhdPubs", package = "vcdExtra")  
table(PhdPubs$articles)
```

```
##  
##      0      1      2      3      4      5      6      7      8      9     10     11     12     16     19  
## 275 246 178  84  67  27  17  12   1   2   1   1   2   1   1
```

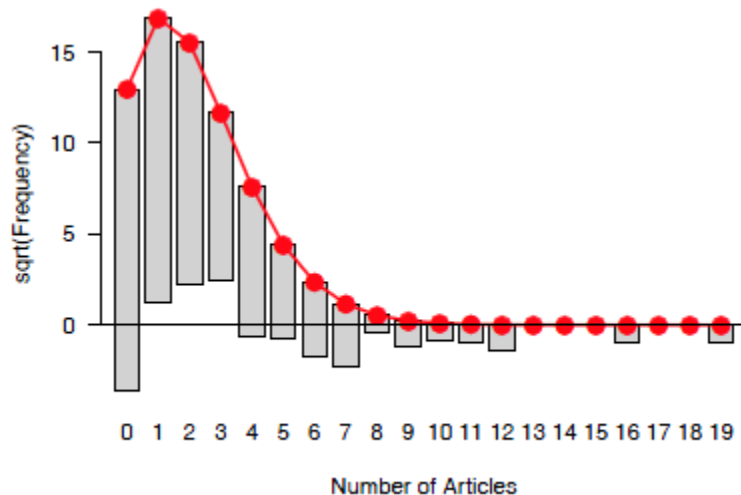
- There are predictors: gender, marital status, number of children, prestige of dept., # pubs by student's mentor
- We fit such models with `glm()`, but need to specify the form of the distribution
- Ignoring the predictors for now, a baseline model could be  
`glm(articles ~ 1, data=PhdPubs, family = "poisson")`



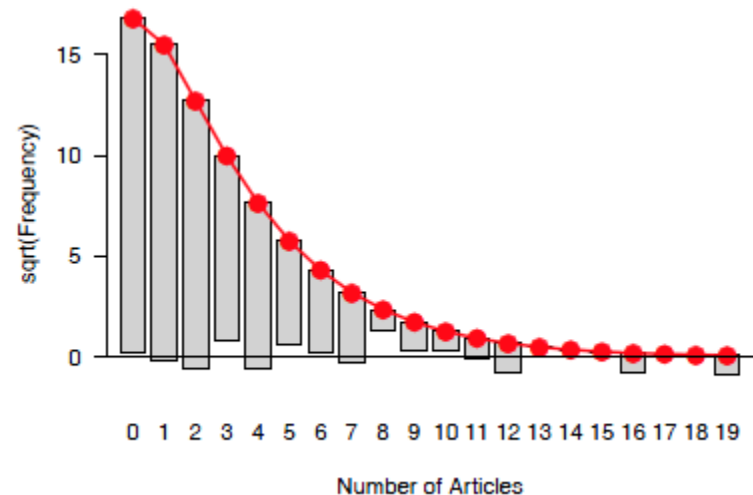
# Looking ahead: PhdPubs

```
plot(goodfit(PhdPubs$articles), xlab = "Number of Articles",  
     main = "Poisson")  
plot(goodfit(PhdPubs$articles, type = "nbinomial"),  
     xlab = "Number of Articles", main = "Negative binomial")
```

Poisson



Negative binomial



Poisson doesn't fit: Need to account for excess 0s (some never published)  
Neg binomial: Sort of OK, but should take predictors into account

# Looking ahead: Count data models

Count data regression models (DDAR Ch 11)

- Include predictors
- Allow different distributions for unexplained variation
- Provide tests of one model vs. another
- Special models handle the problems of excess zeros: `zeroinfl()`, `hurdle()`

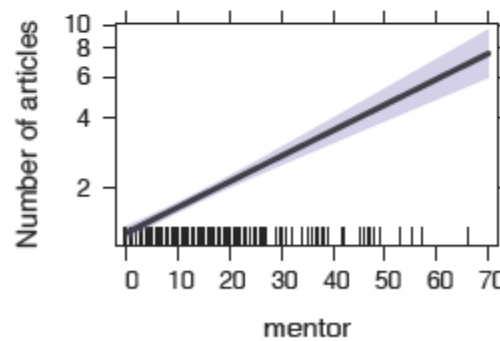
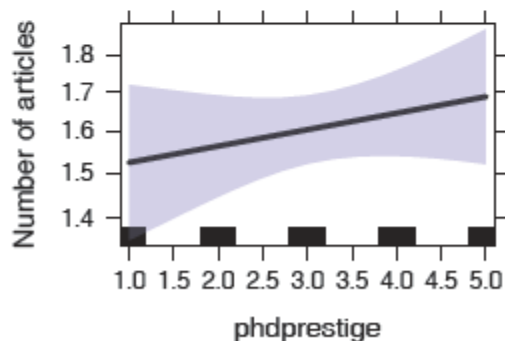
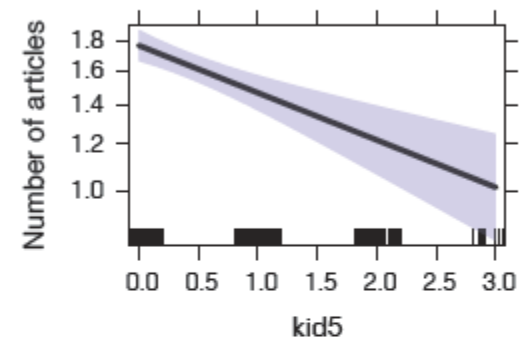
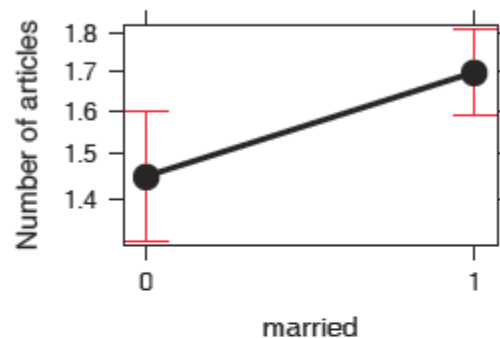
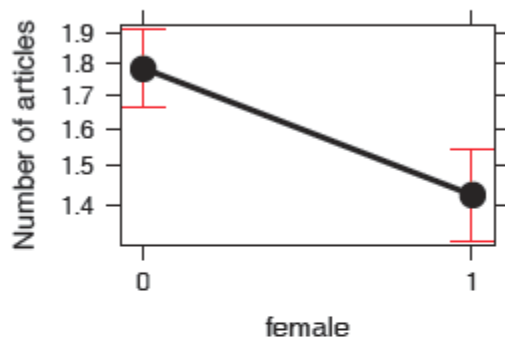
```
# predictors: female, married, kid5, phdprestige, mentor
phd.pois <- glm(articles ~ ., data=PhdPubs, family=poisson)
phd.nbin <- glm.nb(articles ~ ., data=PhdPubs)

LRstats(phd.pois, phd.nbin)

## Likelihood summary table:
##           AIC   BIC LR Chisq  Df Pr(>Chisq)
## phd.pois 3313 3342    1634  909    <2e-16 ***
## phd.nbin 3135 3169    1004  909    0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Looking ahead: Effect plots

Effect plots show the predicted values for each term in a model, averaging over all other factors.



These are better visual summaries for a model than a table of coefficients.

# Summary

- Discrete distributions are the building blocks for categorical data analysis
  - Typically consist of basic counts of occurrences, with varying frequencies
  - Most common: binomial, Poisson, negative binomial
  - Others: geometric, log-series
- Fit with `goodfit()`; plot with `rootogram()`
  - Diagnostic plots: `Ord_plot()`, `distplot()`
- Models with predictors
  - Binomial → logistic regression
  - Poisson → poisson regression; logliner models
  - These are special cases of generalized linear models