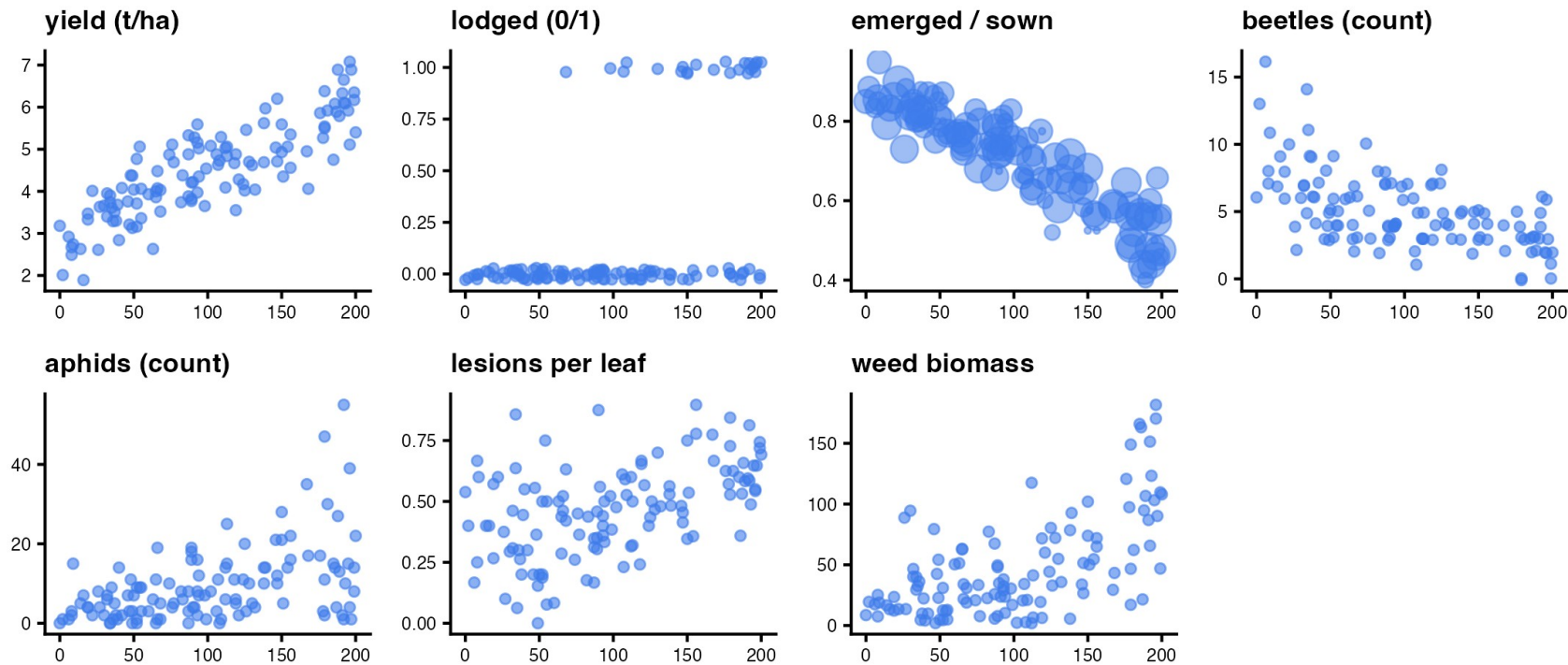


The GLM family

One recipe, seven kinds of response • bonus

Seven responses, one predictor



Same x. **Seven kinds of y.**

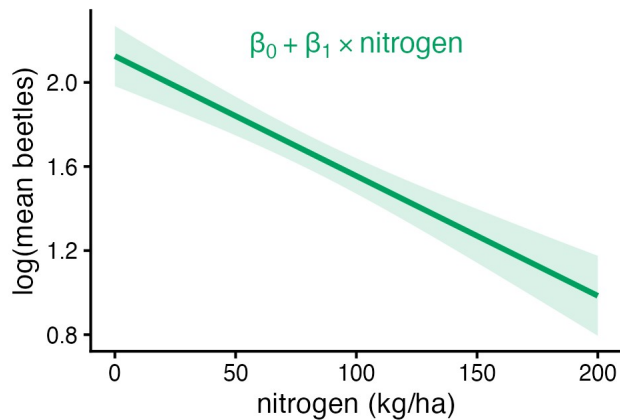
Every GLM is three choices

1. **distribution** $y \sim$ Normal, Binomial, Poisson, Negative binomial, Gamma

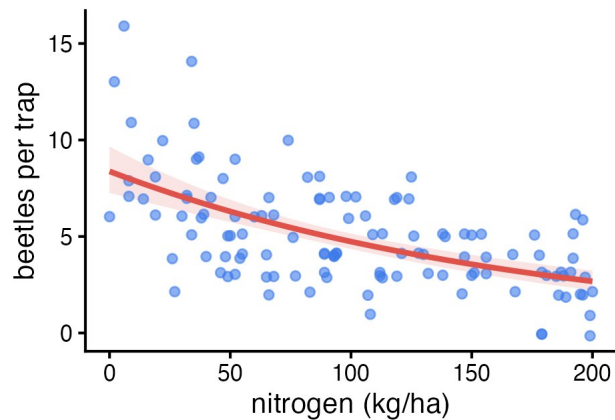
2. **linear predictor** $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$

3. **link** $g(\mu) = \eta \quad \mu = g^{-1}(\eta)$

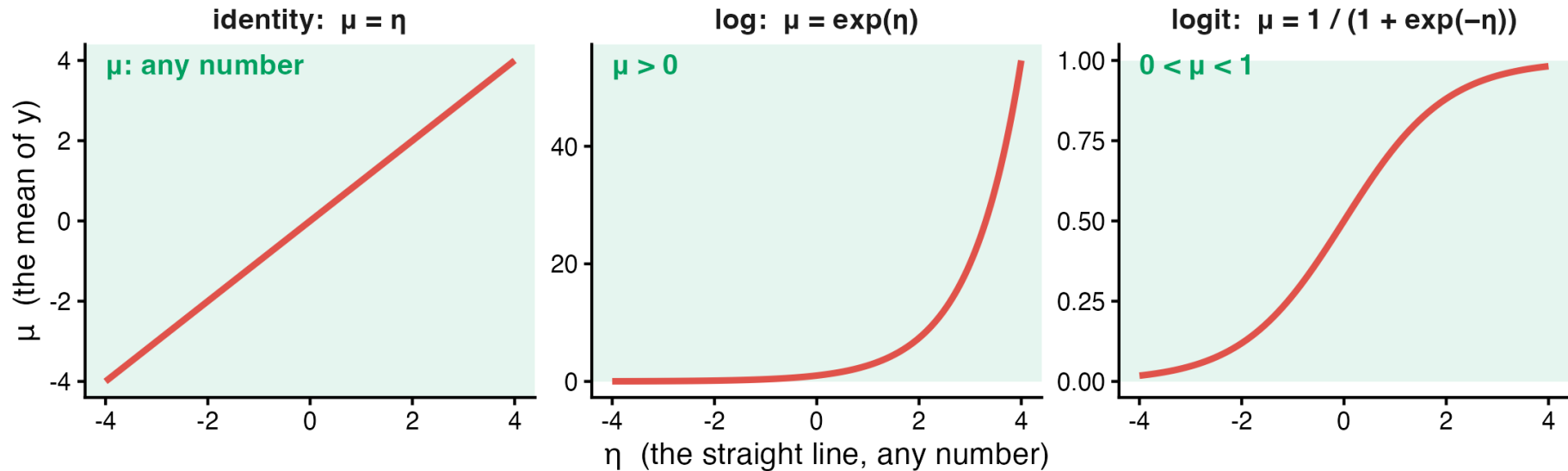
1. a straight line



2. back-transformed: mean = exp(line)

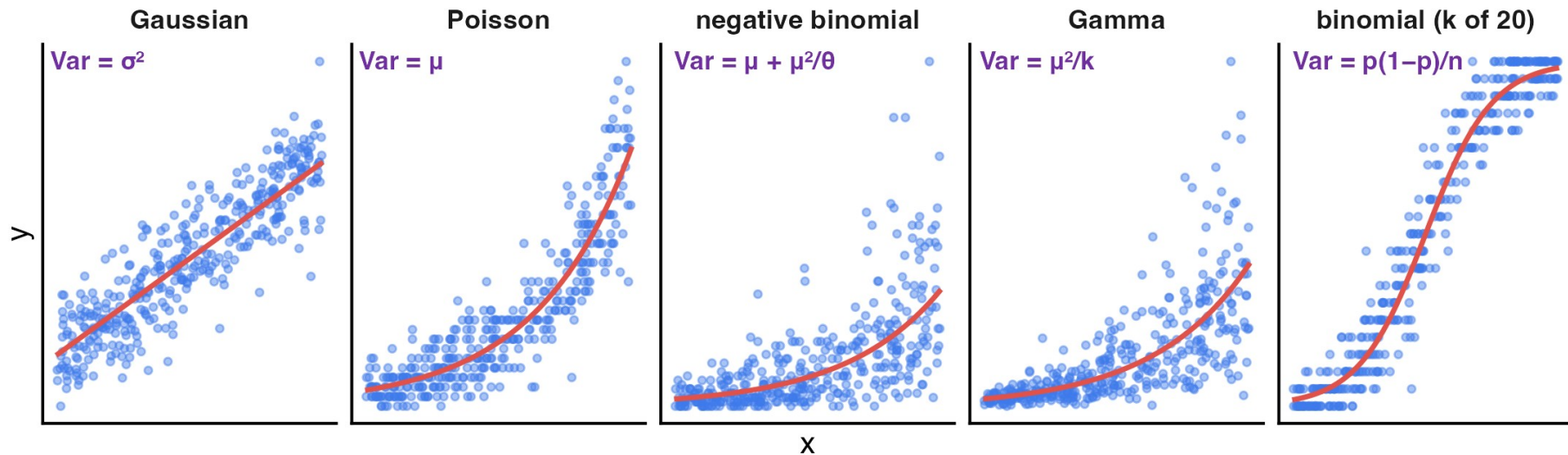


The link: keep the mean where it can live



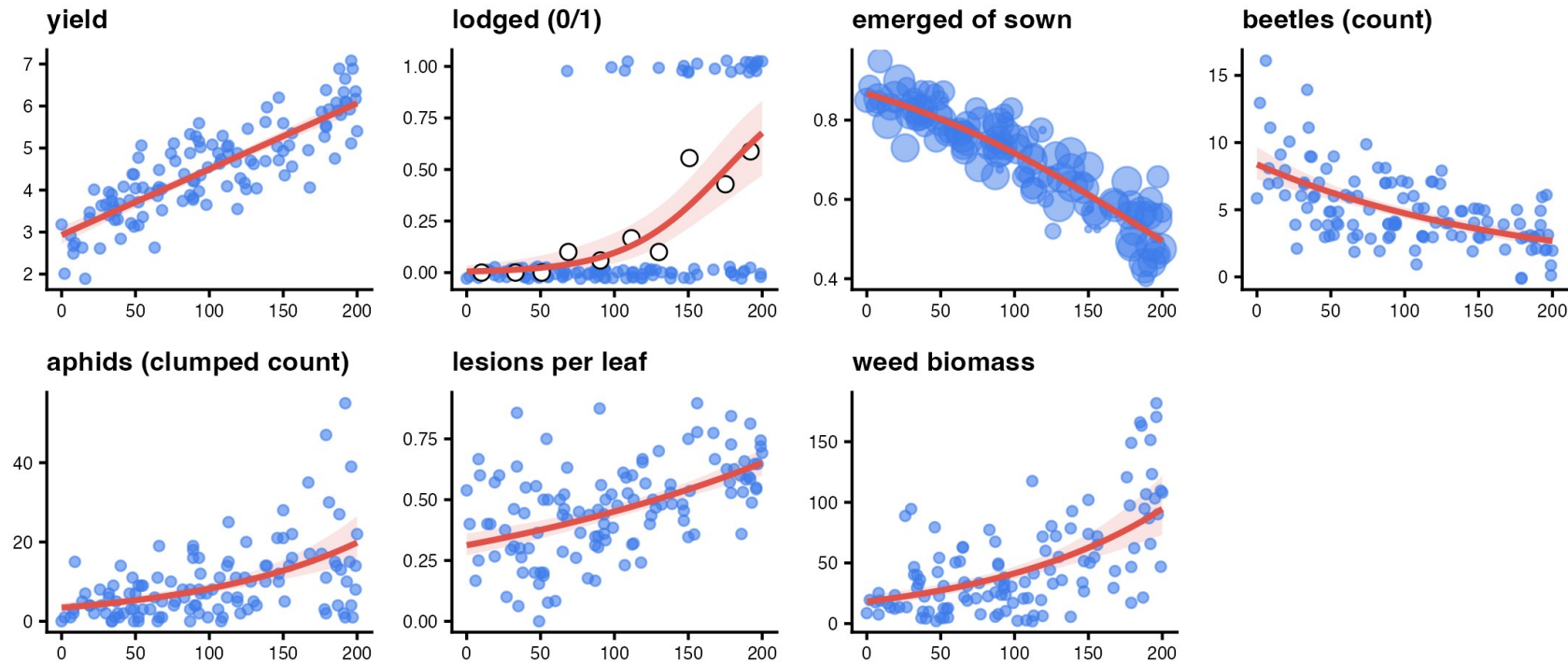
A straight line goes in, **a mean that makes sense** comes out.

The family: how spread grows with the mean

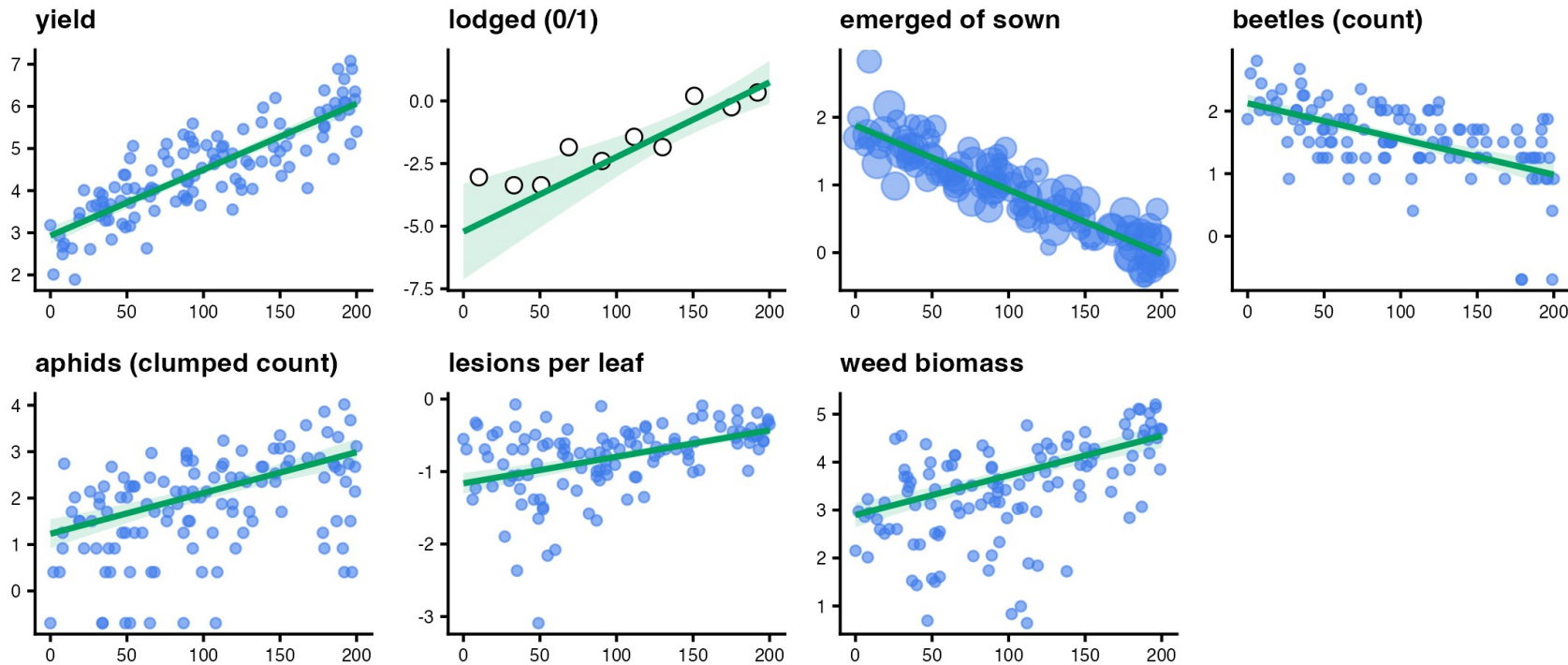


Pick the family whose **spread** looks like your data.

Seven fits on the data: curves



... straight lines on the link scale



Every GLM is a **linear model** on the link scale.

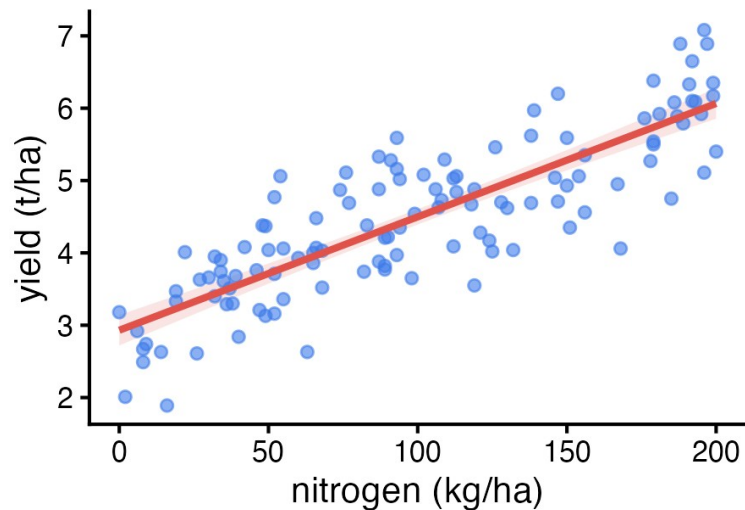
Normal: yield

distribution $\text{yield} \sim \text{Normal}(\mu, \sigma^2)$

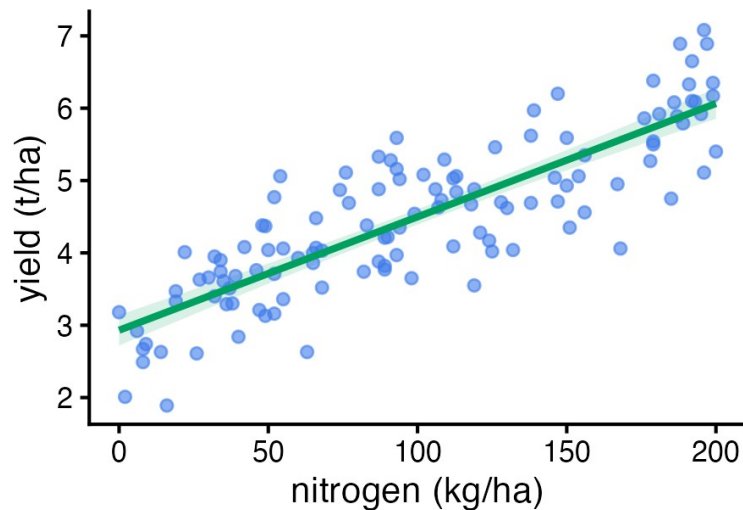
link $\mu = \beta_0 + \beta_1 N$

spread $\text{Var} = \sigma^2$

the data



the link scale



Normal in R

```
> m1 <- glm(yield ~ nitrogen, family = gaussian, data = survey)
> coef(summary(m1))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.9308407	0.1073437827	27.30331	7.922147e-53
nitrogen	0.0156731	0.0009275542	16.89723	2.716035e-33

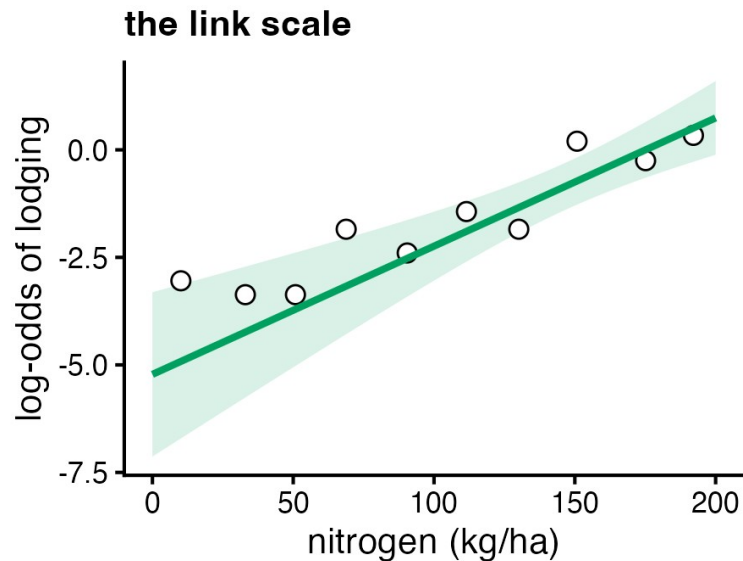
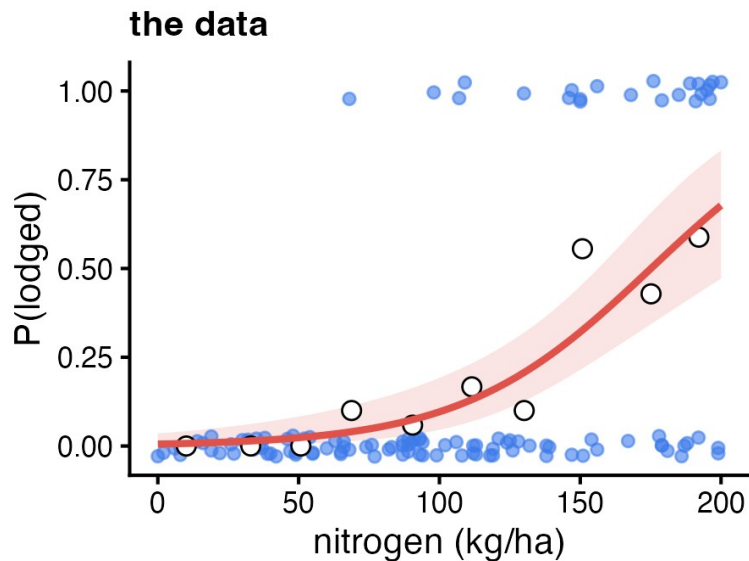
$\beta_1 = 0.0157$: +1.57 t/ha per 100 kg N

Binomial: did it lodge? (0/1)

distribution $\text{lodged} \sim \text{Bernoulli}(p)$

link $\log \frac{p}{1-p} = \beta_0 + \beta_1 N$

spread $\text{Var} = p(1-p)$



Binomial in R

```
> m2 <- glm(lodged ~ nitrogen, family = binomial, data = survey)
> coef(summary(m2))
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.22043380	0.974492412	-5.357080	8.457768e-08
nitrogen	0.02981359	0.006350583	4.694622	2.670995e-06

```
> exp(coef(m2)["nitrogen"] * 100) # odds ratio, per 100 kg N
nitrogen
19.71459
```

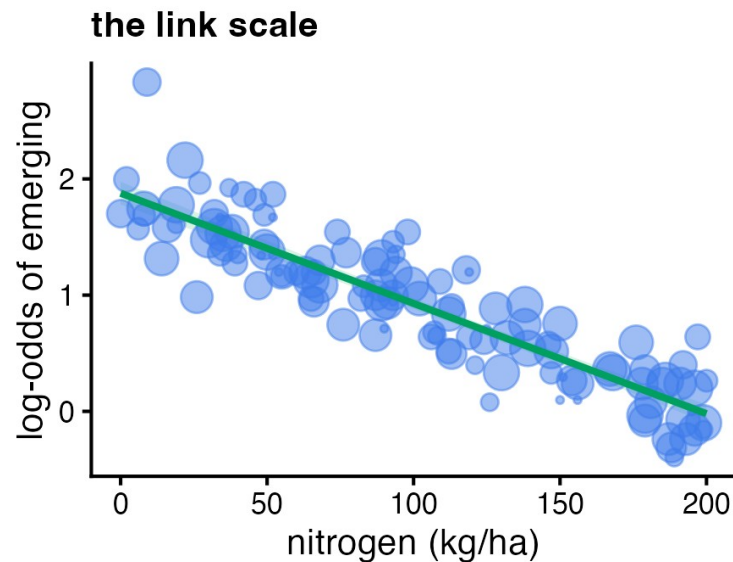
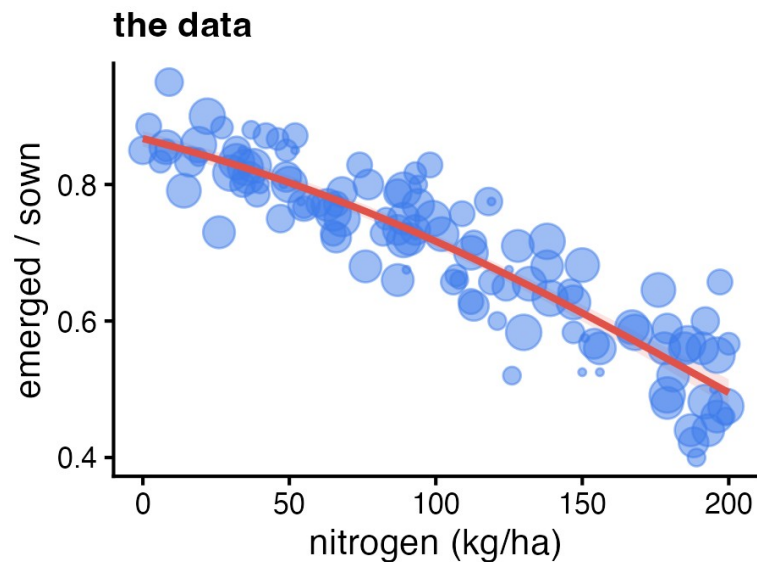
$e^{100\beta_1} = 19.7$: odds of lodging $\times 19.7$ per 100 kg N

Binomial: k emerged of n sown

distribution $\text{emerged} \sim \text{Binomial}(\text{sown}, p)$

link $\log \frac{p}{1-p} = \beta_0 + \beta_1 N$

spread $\text{Var} = \text{sown} \cdot p(1-p)$



Binomial in R

```
> m3 <- glm(cbind(emerged, sown - emerged) ~ nitrogen,  
+           family = binomial, data = survey)  
> coef(summary(m3))
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.876734509	0.0500493886	37.49765	1.005943e-307
nitrogen	-0.009483404	0.0003892029	-24.36622	3.902679e-131

```
> exp(coef(m3)["nitrogen"] * 100) # odds ratio, per 100 kg N  
nitrogen  
0.3873834
```

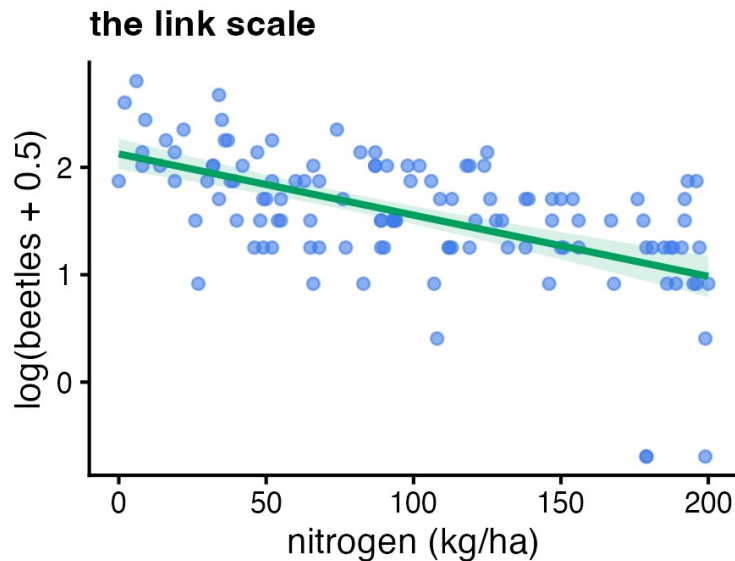
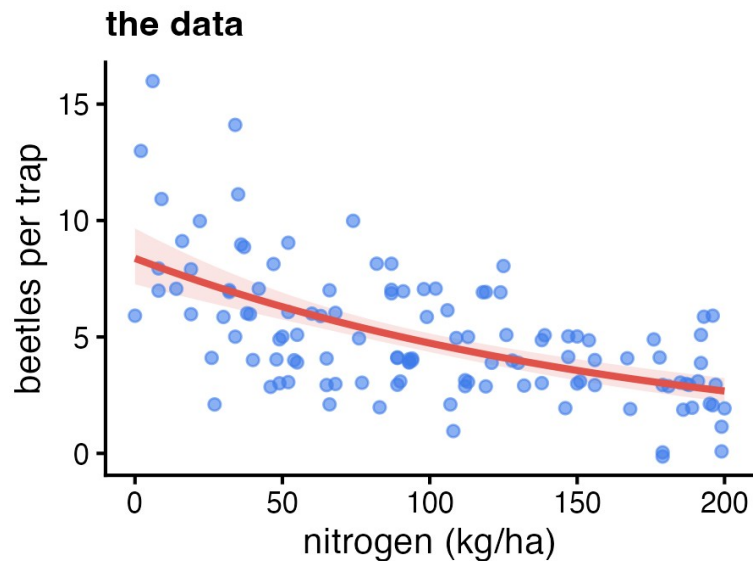
$e^{100\beta_1} = 0.39$: odds of emerging $\times 0.39$ per 100 kg N

Poisson: beetles per trap

distribution $\text{beetles} \sim \text{Poisson}(\mu)$

link $\log \mu = \beta_0 + \beta_1 N$

spread $\text{Var} = \mu$



Poisson in R

```
> m4 <- glm(beetles ~ nitrogen, family = poisson, data = survey)
> coef(summary(m4))
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.125287480	0.0729219240	29.144698	9.751810e-187
nitrogen	-0.005702565	0.0007423705	-7.681562	1.571607e-14

```
> exp(coef(m4)["nitrogen"] * 100)    # rate ratio, per 100 kg N
nitrogen
0.5653804
```

$$e^{100\beta_1} = 0.57: \text{ beetles } \times 0.57 \text{ per } 100 \text{ kg N}$$

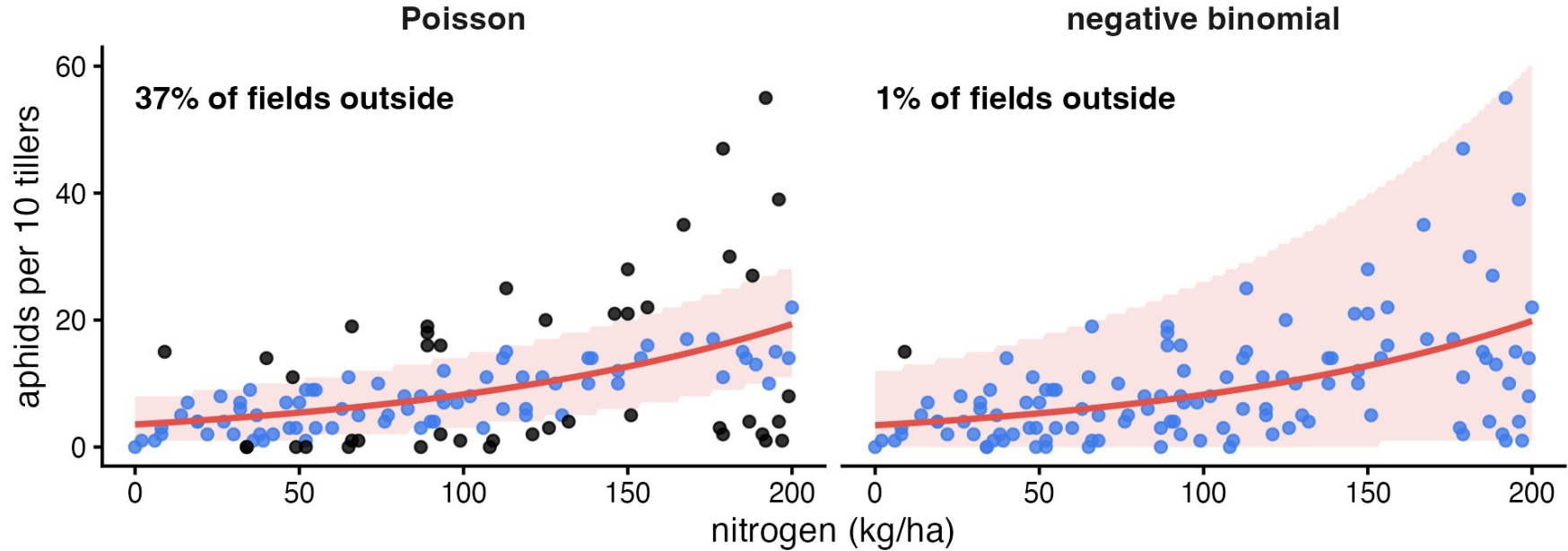
Aphids: is Poisson enough?

```
> m5p <- glm(aphids ~ nitrogen, family = poisson, data = survey)
> deviance(m5p) / df.residual(m5p)    # about 1 if Poisson fits
[1] 5.934917
> deviance(m4) / df.residual(m4)      # the beetles, for comparison
[1] 0.9770481
```

$$\frac{\text{residual deviance}}{df} \approx 1 \quad \text{beetles: } 0.98 \quad \text{aphids: } 5.93$$

Aphids: **~6x more spread** than a Poisson allows.

Poisson vs negative binomial



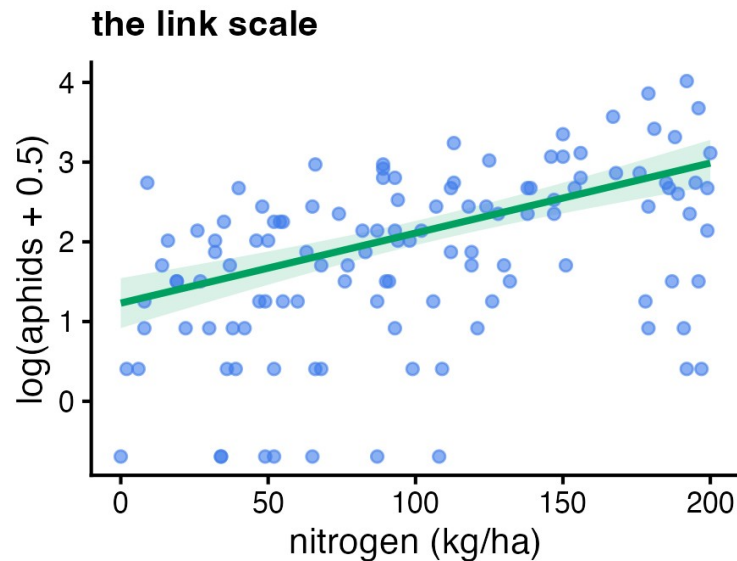
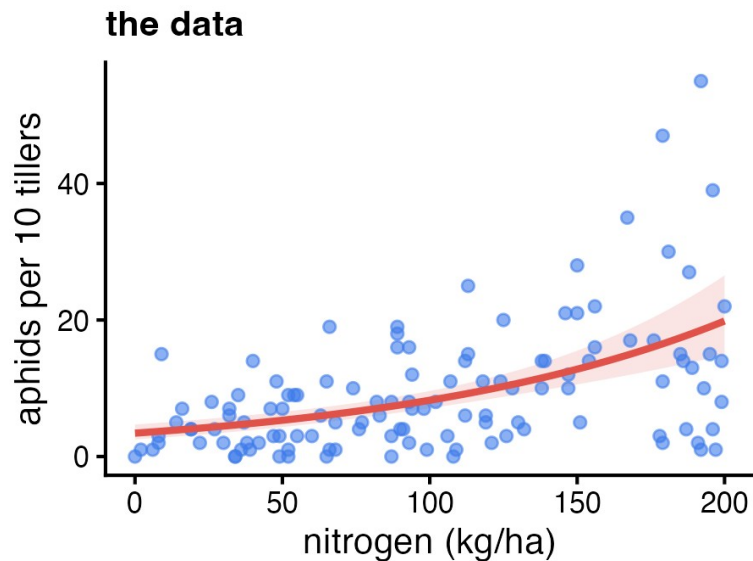
Band = where the model says 95% of fields should fall

Negative binomial: aphids

distribution aphids $\sim \text{NegBin}(\mu, \theta)$

link $\log \mu = \beta_0 + \beta_1 N$

spread $\text{Var} = \mu + \mu^2/\theta$



Negative binomial in R

```
> library(MASS)
> m5 <- glm.nb(aphids ~ nitrogen, data = survey)
> coef(summary(m5))
```

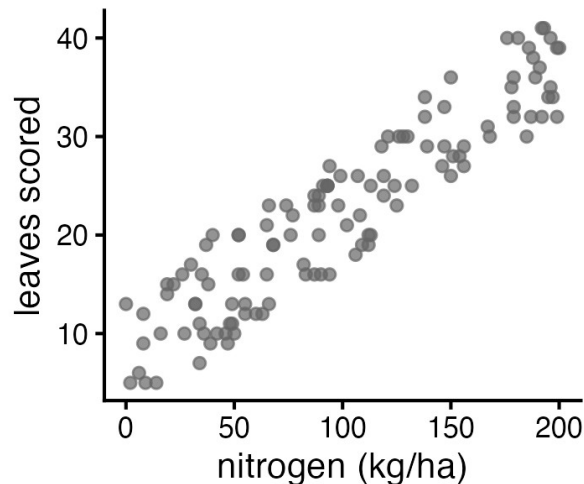
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.231085761	0.159436330	7.721488	1.149791e-14
nitrogen	0.008786187	0.001327348	6.619353	3.607748e-11

```
> m5$theta # the extra spread
[1] 1.702964
> exp(coef(m5)["nitrogen"] * 100) # rate ratio, per 100 kg N
nitrogen
2.407572
```

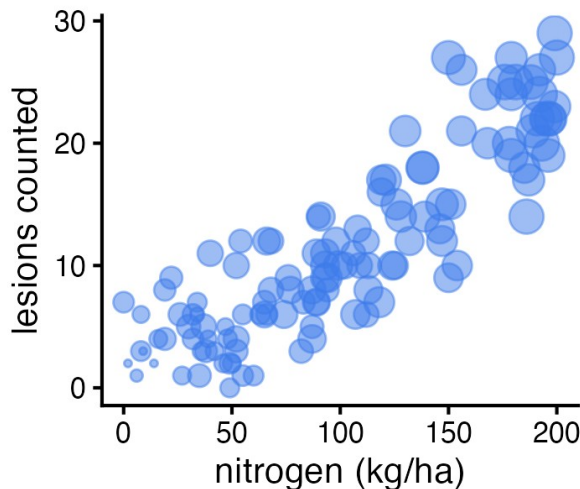
$$e^{100\beta_1} = 2.41: \text{ aphids } \times 2.41 \text{ per 100 kg N}$$

Counts per effort: the offset

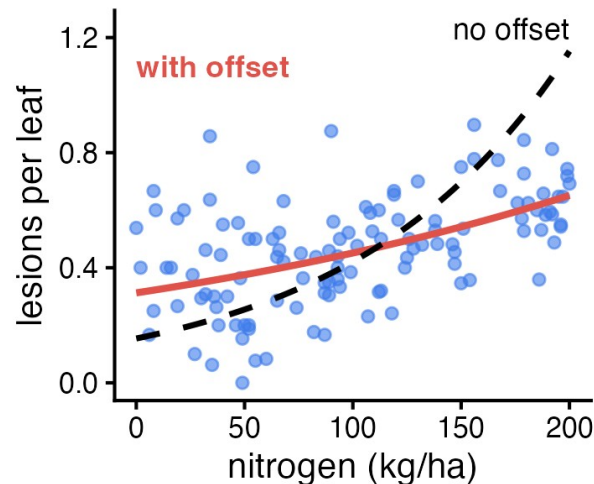
1. effort rises with N



2. raw counts



3. per leaf scored



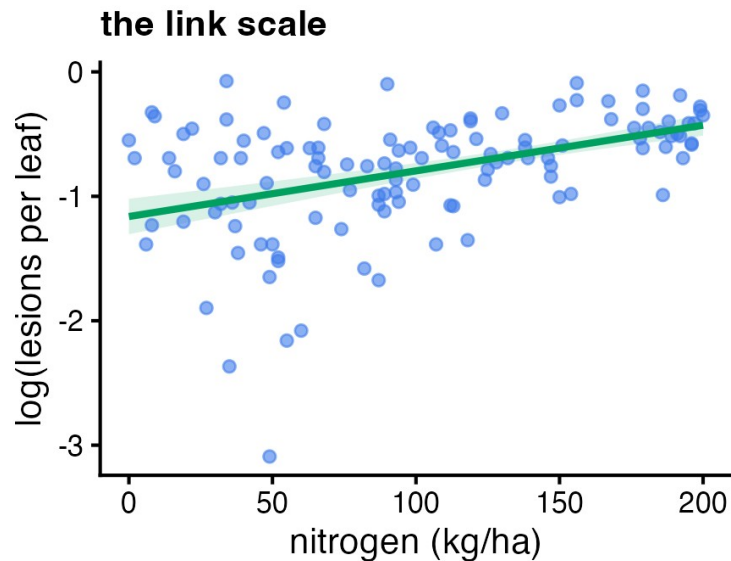
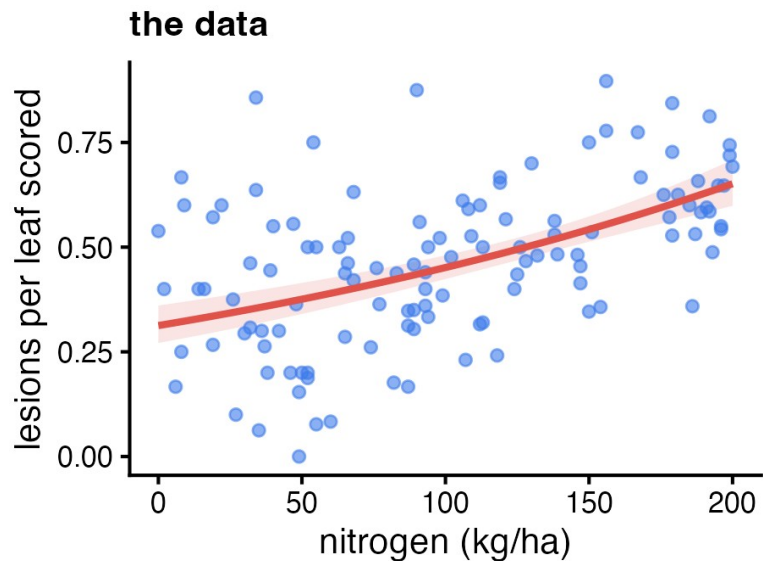
$$\log \mu = \log(\text{leaves}) + \beta_0 + \beta_1 N \quad \Longleftrightarrow \quad \log \frac{\mu}{\text{leaves}} = \beta_0 + \beta_1 N$$

Poisson + offset: lesions per leaf

distribution lesions $\sim$ Poisson(μ)

link $\log \mu = \log(\text{leaves}) + \beta_0 + \beta_1 N$

spread $\text{Var} = \mu$



Poisson + offset in R

```
> m6 <- glm(lesions ~ nitrogen + offset(log(leaves)),  
+           family = poisson, data = survey)  
> coef(summary(m6))
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.162398954	0.0726126759	-16.008210	1.119857e-57
nitrogen	0.003668435	0.0005034928	7.285972	3.193588e-13

```
> exp(coef(m6)["nitrogen"] * 100) # rate ratio, per 100 kg N  
nitrogen  
1.443172
```

$e^{100\beta_1} = 1.44$: lesions per leaf $\times 1.44$ per 100 kg N

Without the offset

```
> m6x <- glm(lesions ~ nitrogen, family = poisson, data = survey)
> exp(coef(m6x)["nitrogen"] * 100)  # no offset
nitrogen
2.735118
> exp(coef(m6)["nitrogen"] * 100)   # with offset
nitrogen
1.443172
```

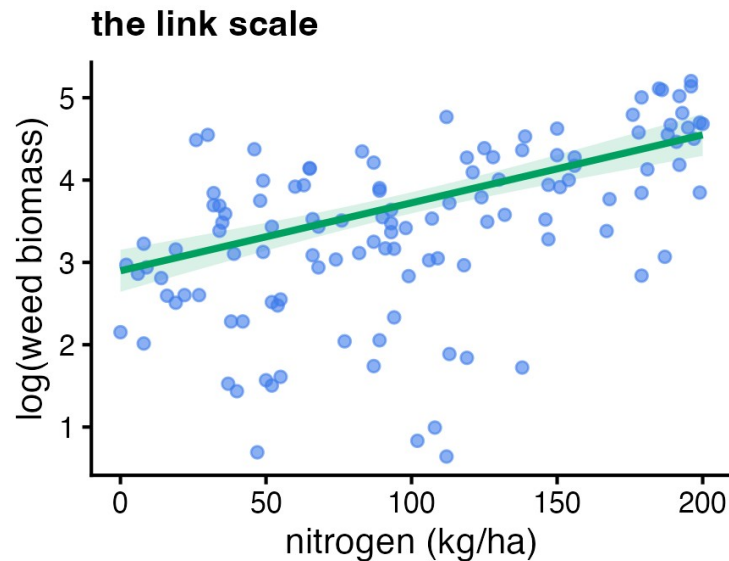
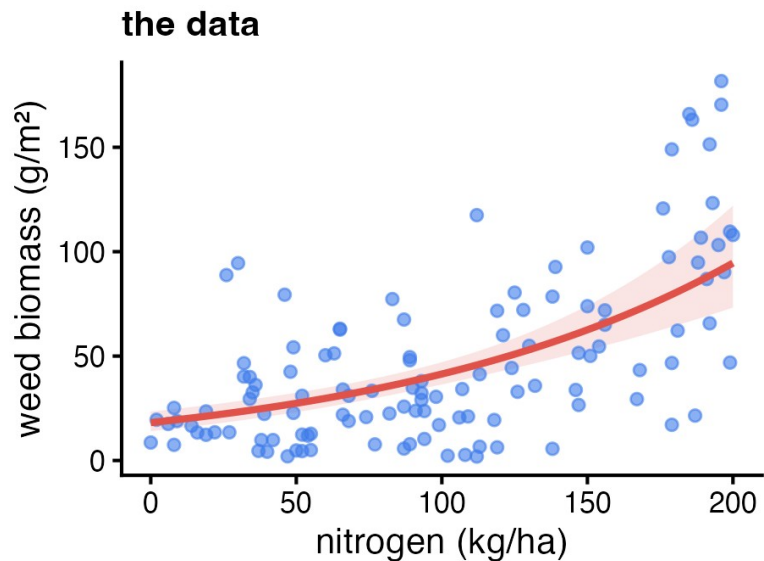
×2.74 vs ×1.44: effort masquerading as disease.

Gamma: weed biomass

distribution weeds $\sim \text{Gamma}(\mu, k)$

link $\log \mu = \beta_0 + \beta_1 N$

spread $\text{Var} = \mu^2/k$



Gamma in R

```
> m7 <- glm(weed_biomass ~ nitrogen, family = Gamma(link = "log"),
+           data = survey)
> coef(summary(m7))
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.898090364	0.130031201	22.287654	4.156874e-44
nitrogen	0.008254916	0.001123595	7.346875	2.867309e-11

```
> exp(coef(m7)["nitrogen"] * 100) # ratio of means, per 100 kg N
nitrogen
2.283003
```

$e^{100\beta_1} = 2.28$: mean weed biomass $\times 2.28$ per 100 kg N

Reading β_1 in any GLM

identity $\mu = \beta_0 + \beta_1 x$ $x + 1$: mean $+ \beta_1$

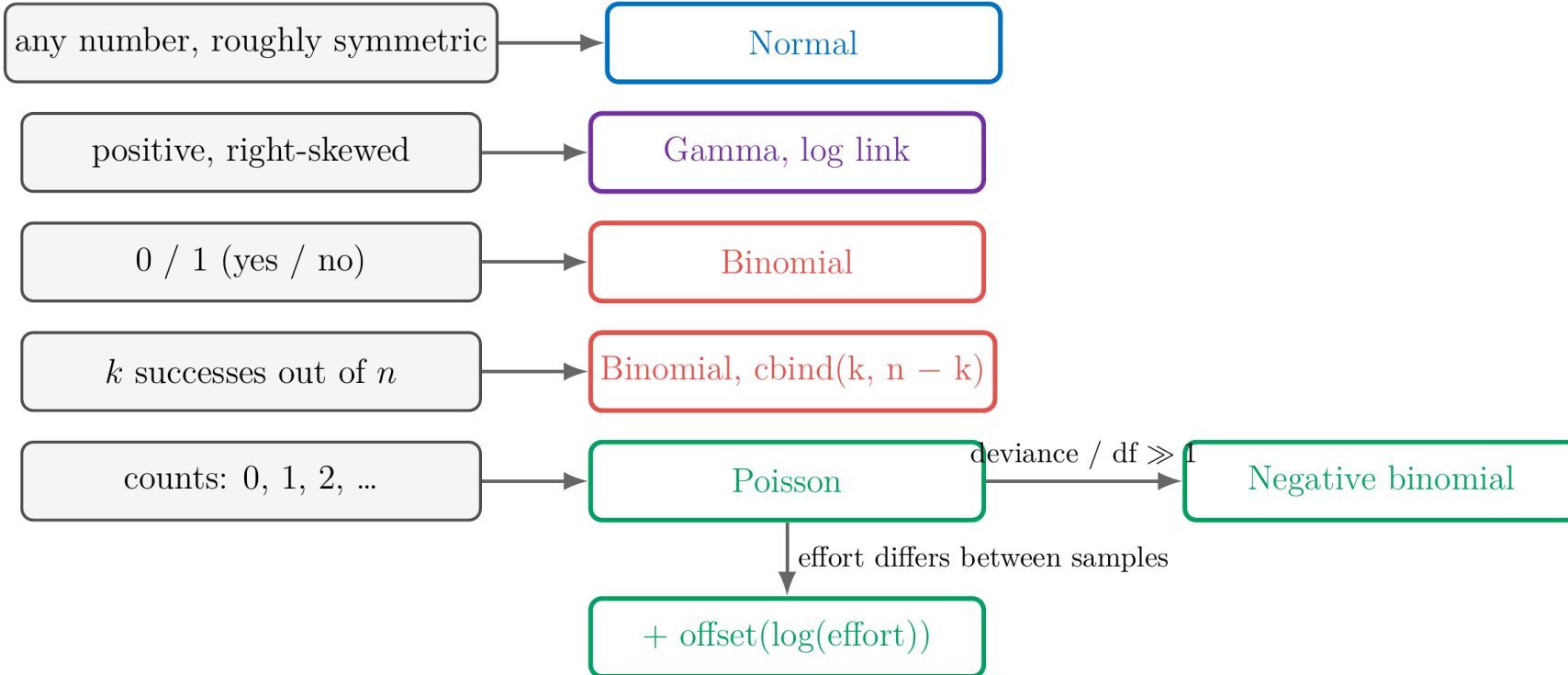
log $\mu = e^{\beta_0} \cdot e^{\beta_1 x}$ $x + 1$: mean $\times e^{\beta_1}$

logit odds $= e^{\beta_0} \cdot e^{\beta_1 x}$ $x + 1$: odds $\times e^{\beta_1}$

```
> new <- data.frame(nitrogen = c(50, 150), leaves = 1)
> predict(m2, new, type = "link")      # log-odds
      1      2
-3.7297544 -0.7483955
> predict(m2, new, type = "response")  # probability
      1      2
0.02343629 0.32117102
```

Which family?

y is ...



Cheat sheet

y looks like	family	link $g(\mu)$	spread $\text{Var}(y)$	in R	e^β is a ratio of
any number, symmetric	Normal	μ	σ^2	<code>lm(y ~ x)</code>	(no: β is a difference)
0 / 1	Binomial	$\log \frac{p}{1-p}$	$p(1-p)$	<code>glm(y ~ x, family = binomial)</code>	odds
k of n	Binomial	$\log \frac{p}{1-p}$	$np(1-p)$	<code>glm(cbind(k, n - k) ~ x, family = binomial)</code>	odds
counts	Poisson	$\log \mu$	μ	<code>glm(y ~ x, family = poisson)</code>	means (rates)
clumped counts	Negative binomial	$\log \mu$	$\mu + \mu^2/\theta$	<code>MASS::glm.nb(y ~ x)</code>	means (rates)
counts per effort	Poisson or NB	$\log \mu$	as above	<code>glm(y ~ x + offset(log(effort)), ...)</code>	rates per unit effort
positive, skewed	Gamma	$\log \mu$	μ^2/k	<code>glm(y ~ x, family = Gamma(link = "log"))</code>	means

Beyond these seven

ordered categories (none < mild < severe)

MASS::polr()

3+ unordered categories

nnet::multinom()

too many zeros

glmmTMB(..., ziformula = ~ 1)

proportions not from counts (% cover)

glmmTMB(..., family = beta_family())

plots within farms, repeated measures

lme4::glmer(..., family = ...)

Try it

[Interactive: every family](#)

pick a family, move β_0 and β_1 , watch both scales

[Code for every slide](#)

GLM_families.R

