Lecture 7

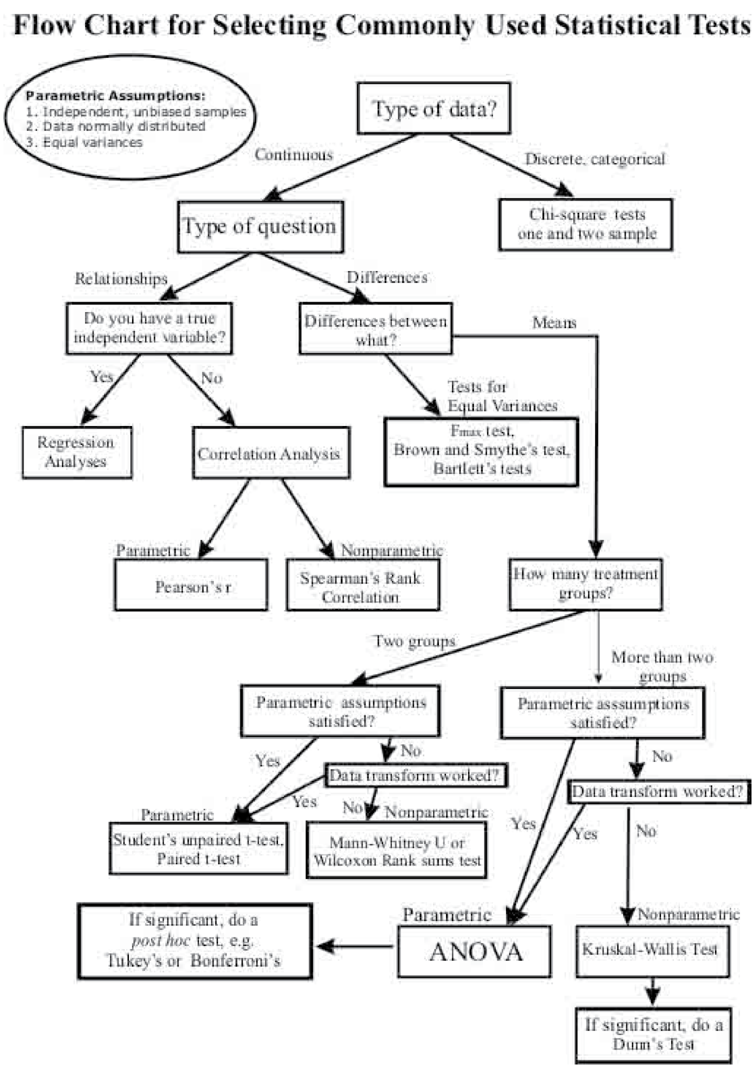
Generalized linear model review

WILD6900 (Spring 2021)

Readings

> Kéry & Schaub 48-55

Statistics cookbook



From linear models to GLMs

Linear models

$$response = deterministic\,part + stochastic\,part$$

$$\underbrace{\mu_i = \beta_0 + \beta_1 \times x_i}_{Deterministic}$$

$$\underbrace{y_i \sim normal(\mu_i, \tau)}_{Stochastic}$$

Linear models under the hood

Variations on the determinstic model

A simple linear model: the t-test

$$\mu_i = \beta_0 + \beta_1 \times x_i$$

$$x_i \in [0, 1]$$

Under the hood: The t-test

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ . \\ . \\ . \\ \mu_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ . & \\ . & \\ . & \\ 1 & x_N \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$\mu_1 = \beta_0 \times 1 + \beta_1 \times x_1$$
$$\mu_2 = \beta_0 \times 1 + \beta_1 \times x_2$$

Under the hood: The t-test

```
model.matrix(lm(mpg ~ am, data = mtcars))[c(3,1,5),]
```

```
##                   (Intercept) am1
## Datsun 710                  1   1
## Mazda RX4                   1   1
## Hornet Sportabout           1   0
```

β_0 is the mean mpg for automatic transmissions

β_1 is the *difference* between automatic and manual transmissions

Under the hood: The t-test

```
lm(mpg ~ am, data = mtcars)
```

term	estimate	std.error	statistic	p.value
(Intercept)	17.147	1.125	15.248	0e+00
am1	7.245	1.764	4.106	3e-04

Under the hood: The t-test

```
model.matrix(lm(mpg ~ as.factor(am) - 1, data = mtcars))[c(3,1,5),]
```

```
##                    am0 am1
## Datsun 710           0   1
## Mazda RX4            0   1
## Hornet Sportabout    1   0
```

β_0 is the mean mpg for automatic transmissions

β_1 is the mean mpg for manual transmissions

Under the hood: The t-test

```
lm(mpg ~ as.factor(am) - 1, data = mtcars)
```

term	estimate	std.error	statistic	p.value
am0	17.15	1.125	15.25	0
am1	24.39	1.360	17.94	0

The t-test becomes an ANOVA

$$y_i = \beta_0 + \beta_{1[j]} \times x_i$$

$$j \in [1, 2, 3, \ldots, J-1]$$

$$x_i \in [0, 1]$$

Under the hood: ANOVA

```
model.matrix(lm(mpg ~ as.factor(cyl), data = mtcars))[c(3,1,5),]
```

```
##                   (Intercept) cyl6 cyl8
## Datsun 710                  1    0    0
## Mazda RX4                   1    1    0
## Hornet Sportabout           1    0    1
```

β_0 is the mean mpg for 4-cylinders

$\beta_{1[6-cyl]}$ is the *difference* between 4-cyl & 6-cyl

$\beta_{1[8-cyl]}$ is the *difference* between 4-cyl & 8-cyl

Under the hood: ANOVA

Effects parameterization

```
lm(mpg ~ as.factor(cyl) , data = mtcars)
```

term	estimate	std.error	statistic	p.value
(Intercept)	26.664	0.9718	27.437	0e+00
cyl6	-6.921	1.5583	-4.441	1e-04
cyl8	-11.564	1.2986	-8.905	0e+00

Under the hood: ANOVA

Means parameterization

$$\mu_i = \beta_{0[j]}$$

```
lm(mpg ~ as.factor(cyl) - 1, data = mtcars)
```

term	estimate	std.error	statistic	p.value
cyl4	26.66	0.9718	27.44	0
cyl6	19.74	1.2182	16.21	0
cyl8	15.10	0.8614	17.53	0

The ANOVA becomes an ANCOVA

$$y_i = \beta_0 + \beta_{1[j]} \times \boldsymbol{x1}_i + \beta_2 \times \boldsymbol{x2}_i$$

$$\boldsymbol{x1}_i \in [\boldsymbol{0}, \boldsymbol{1}]$$

$$\boldsymbol{x2}_i \in [-\infty, \infty]$$

Under the hood: ANCOVA

```
model.matrix(lm(mpg ~ as.factor(cyl) + hp, data = mtcars))[c(3,1,5),]
```

```
##                    (Intercept) cyl6 cyl8  hp
## Datsun 710                   1    0    0  93
## Mazda RX4                    1    1    0 110
## Hornet Sportabout            1    0    1 175
```

β_0 is the mean mpg for 4-cylinders **@ 0hp**

$\beta_{1[6-cyl]}$ is the *difference* between 4-cyl & 6-cyl **@ 0hp**

$\beta_{1[8-cyl]}$ is the *difference* between 4-cyl & 8-cyl **@ 0hp**

β_2 is the effect of hp on mpg

Under the hood: ANCOVA

```
lm(mpg ~ as.factor(cyl) + hp, data = mtcars)
```

term	estimate	std.error	statistic	p.value
(Intercept)	28.650	1.5878	18.044	0.0000
cyl6	-5.968	1.6393	-3.640	0.0011
cyl8	-8.521	2.3261	-3.663	0.0010
hp	-0.024	0.0154	-1.560	0.1300

Under the hood: Interactions

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ . \\ . \\ . \\ \mu_N \end{bmatrix} = \begin{bmatrix} 1 & x1_1 & x2_1 & x1_1 * x2_1 \\ 1 & x1_2 & x2_2 & x1_2 * x2_2 \\ 1 & x1_3 & x2_3 & x1_3 * x2_3 \\ . & & & \\ . & & & \\ . & & & \\ 1 & x1_N & x2_N & x1_N * x2_N \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

Under the hood: Interactions

```
model.matrix(lm(mpg~as.factor(cyl)*hp, data = mtcars))[c(3,1,5),]
```

```
##                   (Intercept) cyl6 cyl8  hp cyl6:hp cyl8:hp
## Datsun 710                  1    0    0  93       0       0
## Mazda RX4                   1    1    0 110     110       0
## Hornet Sportabout           1    0    1 175       0     175
```

β_0 is the mean mpg for 4-cylinders **@ 0hp**

$\beta_{1[6-cyl]}$ is the *difference* between 4-cyl & 6-cyl **@ 0hp**

$\beta_{1[8-cyl]}$ is the *difference* between 4-cyl & 8-cyl **@ 0hp**

β_2 is the effect of hp on mpg for 4-cylinders

$\beta_{3[6-cyl]}$ is the *difference* between the effect of hp in 4-cyl vs 6-cyl

Under the hood: Interactions

```
lm(mpg ~ as.factor(cyl) * hp, data = mtcars)
```

term	estimate	std.error	statistic	p.value
(Intercept)	35.9830	3.8891	9.252	0.0000
cyl6	-15.3092	7.4346	-2.059	0.0496
cyl8	-17.9030	5.2596	-3.404	0.0022
hp	-0.1128	0.0457	-2.465	0.0206
cyl6:hp	0.1052	0.0685	1.536	0.1367
cyl8:hp	0.0985	0.0486	2.026	0.0531

Under the hood: Interactions

```
model.matrix(lm(mpg ~ as.factor(cyl) * hp - 1 - hp, data = mtcars))[c
```

```
##                    cyl4 cyl6 cyl8 cyl4:hp cyl6:hp cyl8:hp
## Datsun 710            1    0    0      93       0       0
## Mazda RX4             0    1    0       0     110       0
## Hornet Sportabout     0    0    1       0       0     175
```

$\beta_{0[4-cyl]}$ is the mean mpg for 4-cylinders **@ 0hp**

$\beta_{0[6-cyl]}$ is the mean mpg for 6-cylinders **@ 0hp**

$\beta_{0[8-cyl]}$ is the mean mpg for 8-cylinders **@ 0hp**

$\beta_{1[4-cyl]}$ is the effect of hp on mpg for 4-cylinders

$\beta_{1[6-cyl]}$ is the effect of hp on mpg for 6-cylinders

$\beta_{1[8-cyl]}$ is the effect of hp on mpg for 8-cylinders

Under the hood: Interactions

```
lm(mpg ~ as.factor(cyl) * hp - 1, data = mtcars)
```

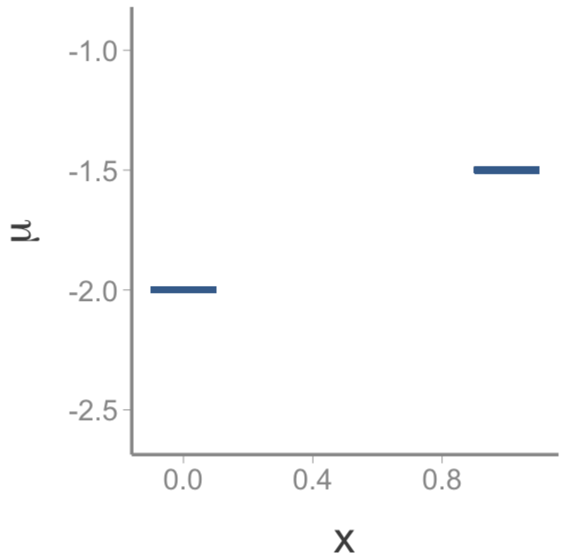
term	estimate	std.error	statistic	p.value
cyl4	35.9830	3.8891	9.2523	0.0000
cyl6	20.6739	6.3362	3.2628	0.0031
cyl8	18.0801	3.5410	5.1059	0.0000
cyl4:hp	-0.1128	0.0457	-2.4652	0.0206
cyl6:hp	-0.0076	0.0510	-0.1494	0.8824
cyl8:hp	-0.0142	0.0165	-0.8645	0.3952

Linear models under the hood

Variations on the stochastic model

Stochasticity in the linear model

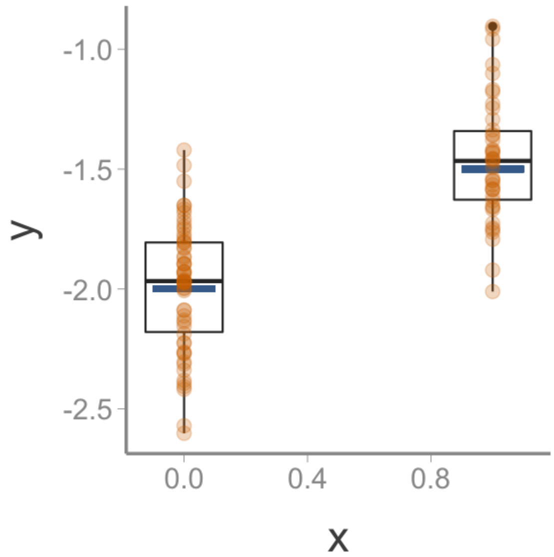
$$\underbrace{\mu_i = -2 + 0.5 \times x_i}_{Deterministic}$$



Stochasticity in the linear model

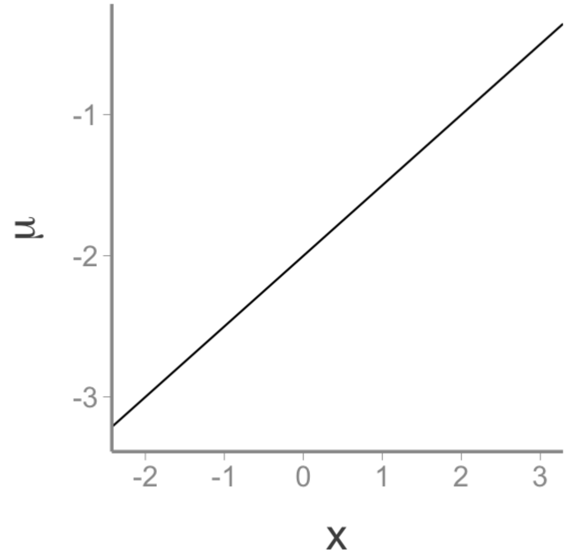
$$\mu_i = -2 + 0.5 \times x_i$$

$$\underbrace{y_i \sim normal(\mu_i, \tau)}_{Stochastic}$$



Stochasticity in the linear model

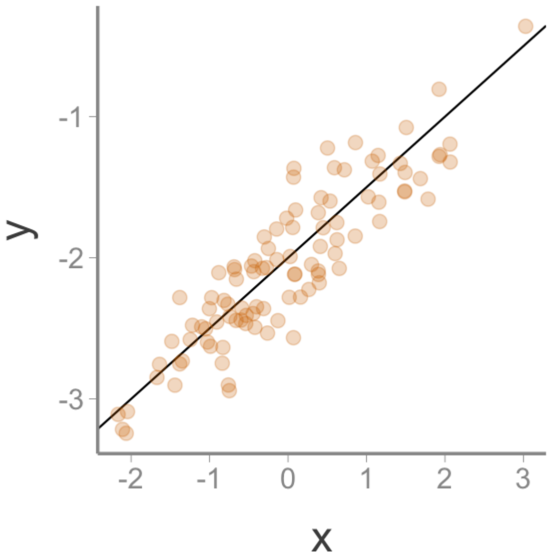
$$\mu_i = -2 + 0.5 \times x_i$$



Stochasticity in the linear model

$$\mu_i = -2 + 0.5 \times x_i$$

$$y_i \sim normal(\mu_i, \tau)$$



Components of the linear model

Components of the linear model

1) Distribution

$$y_i \sim normal(\mu_i, \tau)$$

2) Link function

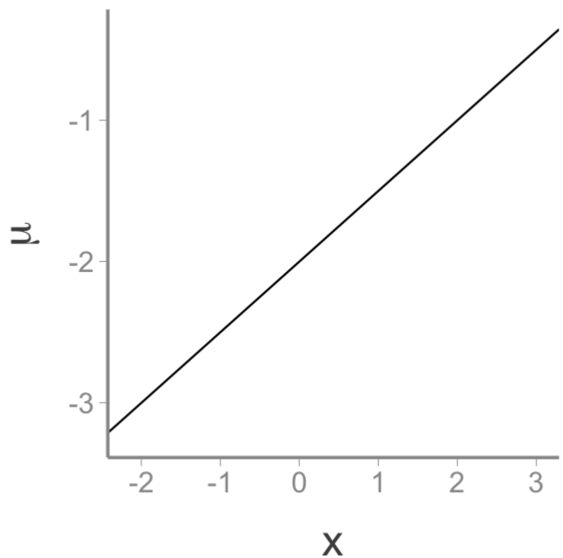
$$\mu_i = E(y_i) = linear\ predictor$$

3) Linear predictor

$$\beta_0 + \beta_1 \times x_i$$

Stochasticity in the linear model

What happens if $0 \leq y_i$?



Components of the generalized linear model

1) Distribution

$$y_i \sim normal(\mu_i, \tau)$$

Components of the generalized linear model

1) Distribution

$$y_i \sim Poisson(\lambda_i)$$

2) Link function

$$\lambda_i = E(y_i) = linear\ predictor$$

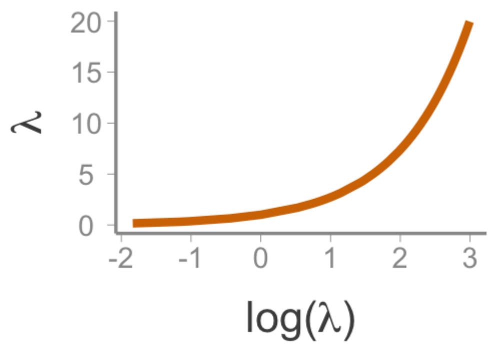
Components of the generalized linear model

1) Distribution

$$y_i \sim Poisson(\lambda_i)$$

2) Link function

$$log(\lambda_i) = log(E(y_i)) = linear\ predictor$$



Components of the generalized linear model

1) Distribution

$$y_i \sim Poisson(\lambda_i)$$

2) Link function

$$log(\lambda_i) = log(E(y_i)) = linear\ predictor$$

Components of the generalized linear model

1) Distribution

$$y_i \sim Poisson(\lambda_i)$$

2) Link function

$$log(\lambda_i) = log(E(y_i)) = linear\ predictor$$

3) Linear predictor

$$\beta_0 + \beta_1 \times x_i$$

Components of the generalized linear model

1) Distribution

$$y_i \sim Bernoulli(p_i)$$

2) Link function

$$logit(p_i) = log\left(\frac{p_i}{1-p_0}\right) = linear\ predictor$$

Components of the generalized linear model

1) Distribution

$$y_i \sim Bernoulli(p_i)$$

2) Link function

$$logit(p_i) = log\left(\frac{p_i}{1 - p_0}\right) = linear\ predictor$$

3) Linear predictor

$$\beta_0 + \beta_1 \times x_i$$

Components of the generalized linear model

1) Distribution

$$y_i \sim binomial(N, p_i)$$

2) Link function

$$logit(p_i) = log\left(\frac{p_i}{1 - p_0}\right) = linear\ predictor$$

3) Linear predictor

$$\beta_0 + \beta_1 \times x_i$$

Generalized linear models

- Flexible method to model observations arising from different probability distributions

- Link many classical tests under unified framework

- Underlie nearly all common ecological models