Other Outcomes: Ordinal, Nominal, Count, Duration

Francisco Villamil

Applied Quantitative Methods II

MA in Social Sciences, Spring 2026

Logistics

- Presentations (April 16th)
  - → **5-minute presentations**, blocks of ≈5, random order, 15-min feedback after each block
    https://franvillamil.github.io/AQM2/logistics.html
  - → Send me **presentations** by email **in advance**
    3 slides max!: e.g. what, how, problems
- Rest of the course
  - → Computing session (April 23rd)
  - → Final session (April 30th), **exam** + review

Today's goals

- Understand the GLM framework as a unifying structure
- Model ordinal outcomes with ordered logit
- Model nominal (multi-category) outcomes with multinomial logit
- Model count data with Poisson and negative binomial regression
- Model time-to-event data with survival analysis and the Cox model
- Know which model to reach for and why

Roadmap

One framework, many models

Generalized Linear Models (GLMs) share three components:

Linear predictor	Link function	Distribution
$\eta_i = \mathbf{x}_i\boldsymbol{\beta}$	$g(\mu_i) = \eta_i$	$Y_i \sim F(\mu_i)$
A weighted sum of covariates	Connects mean to predictor	Data-generating process

- OLS is a GLM: identity link + normal distribution

One framework, many models

Generalized Linear Models (GLMs) share three components:

Linear predictor	Link function	Distribution
$\eta_i = \mathbf{x}_i\boldsymbol{\beta}$	$g(\mu_i) = \eta_i$	$Y_i \sim F(\mu_i)$
A weighted sum of covariates	Connects mean to predictor	Data-generating process

- OLS is a GLM: identity link + normal distribution
- Logit is a GLM: logit link + Bernoulli distribution

One framework, many models

Generalized Linear Models (GLMs) share three components:

Linear predictor	Link function	Distribution
$\eta_i = \mathbf{x}_i\boldsymbol{\beta}$	$g(\mu_i) = \eta_i$	$Y_i \sim F(\mu_i)$
A weighted sum of covariates	Connects mean to predictor	Data-generating process

- OLS is a GLM: identity link + normal distribution
- Logit is a GLM: logit link + Bernoulli distribution
- Choosing a model = choosing a link + distribution

Matching outcome type to model

Outcome type	**Distribution**	**Link**	**Model / Function**
Continuous	Normal	Identity	OLS (`lm()`)
Binary (0/1)	Bernoulli	Logit	Logit (`glm(...,binomial)`)
Ordinal	Prop. odds	Cumul. logit	Ordered logit (`polr()`)
Nominal (>2)	Multinomial	Log-odds	Multinom. logit (`multinom()`)
Count (≥ 0)	Poisson	Log	Poisson (`glm(...,poisson)`)
Count (overdisp)	Neg. Binomial	Log	NegBin (`glm.nb()`)
Duration (time)	—	—	Cox PH (`coxph()`)

First question: what is the data-generating process for your outcome?

Roadmap

Ordinal outcomes:

When categories have an order but not equal spacing

Ordinal outcomes are ordered categories

- Categories have a natural **ordering**, but not equal intervals

Ordinal outcomes are ordered categories

- Categories have a natural **ordering**, but not equal intervals
- Examples:

Ordinal outcomes are ordered categories

- Categories have a natural **ordering**, but not equal intervals
- Examples:
  - → Survey: "How satisfied are you?" ($1 =$ not at all, $5 =$ very)

Ordinal outcomes are ordered categories

- Categories have a natural **ordering**, but not equal intervals
- Examples:
  - → Survey: "How satisfied are you?" (1 = not at all, 5 = very)
  - → Party ID: Strong Dem → Weak Dem → Ind → Weak Rep → Strong Rep

Ordinal outcomes are ordered categories

- Categories have a natural **ordering**, but not equal intervals
- Examples:
  - → Survey: "How satisfied are you?" (1 = not at all, 5 = very)
  - → Party ID: Strong Dem → Weak Dem → Ind → Weak Rep → Strong Rep
  - → Democracy rating: Autocracy → Hybrid → Democracy

Ordinal outcomes are ordered categories

- Categories have a natural **ordering**, but not equal intervals
- Examples:
  - → Survey: "How satisfied are you?" (1 = not at all, 5 = very)
  - → Party ID: Strong Dem → Weak Dem → Ind → Weak Rep → Strong Rep
  - → Democracy rating: Autocracy → Hybrid → Democracy
  - → Conflict intensity: No conflict → Low → Medium → High

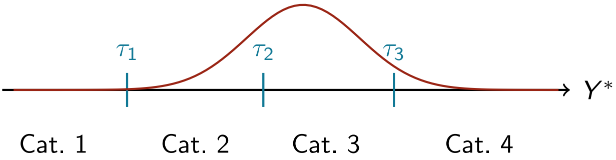
Ordinal outcomes are ordered categories

- Categories have a natural **ordering**, but not equal intervals
- Examples:
  - → Survey: "How satisfied are you?" (1 = not at all, 5 = very)
  - → Party ID: Strong Dem → Weak Dem → Ind → Weak Rep → Strong Rep
  - → Democracy rating: Autocracy → Hybrid → Democracy
  - → Conflict intensity: No conflict → Low → Medium → High
- The distance between "1" and "2" need not equal the distance between "2" and "3"

Ordinal outcomes are ordered categories

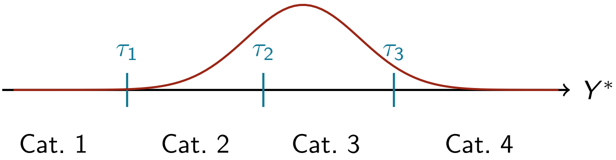
- Categories have a natural **ordering**, but not equal intervals
- Examples:
  - $\rightarrow$ Survey: "How satisfied are you?" (1 = not at all, 5 = very)
  - $\rightarrow$ Party ID: Strong Dem $\rightarrow$ Weak Dem $\rightarrow$ Ind $\rightarrow$ Weak Rep $\rightarrow$ Strong Rep
  - $\rightarrow$ Democracy rating: Autocracy $\rightarrow$ Hybrid $\rightarrow$ Democracy
  - $\rightarrow$ Conflict intensity: No conflict $\rightarrow$ Low $\rightarrow$ Medium $\rightarrow$ High
- The distance between "1" and "2" need not equal the distance between "2" and "3"
- OLS treats them as equally spaced — often wrong

The latent variable idea



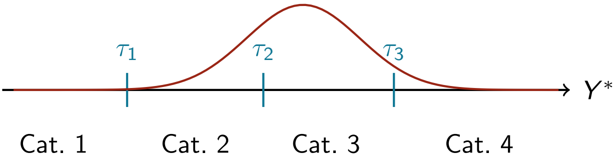
- Y^*: unobserved continuous variable (e.g., "true" satisfaction)

The latent variable idea



- Y^*: unobserved continuous variable (e.g., "true" satisfaction)
- We observe category j when $\tau_{j-1} < Y^* \leq \tau_j$

The latent variable idea



- Y^*: unobserved continuous variable (e.g., "true" satisfaction)
- We observe category j when $\tau_{j-1} < Y^* \leq \tau_j$
- Thresholds τ_j are **estimated** from the data

Ordered logit: the model

Cumulative link model (proportional odds)

$$\Pr(Y_i \leq j \mid \mathbf{x}_i) = \frac{1}{1 + \exp(-(\tau_j - \mathbf{x}_i \beta))}$$

- One set of β for all thresholds (**proportional odds assumption**)

Ordered logit: the model

Cumulative link model (proportional odds)

$$\Pr(Y_i \leq j \mid \mathbf{x}_i) = \frac{1}{1 + \exp(-(\tau_j - \mathbf{x}_i\boldsymbol{\beta}))}$$

- One set of $\boldsymbol{\beta}$ for all thresholds (**proportional odds assumption**)
- $\beta_k > 0$ shifts Y^* upward $\rightarrow$ **higher** categories become more likely

Ordered logit: the model

Cumulative link model (proportional odds)

$$\Pr(Y_i \leq j \mid \mathbf{x}_i) = \frac{1}{1 + \exp(-(\tau_j - \mathbf{x}_i \boldsymbol{\beta}))}$$

- One set of $\boldsymbol{\beta}$ for all thresholds (**proportional odds assumption**)
- $\beta_k > 0$ shifts Y^* upward $\rightarrow$ **higher** categories become more likely
- $\Pr(Y_i = j) = \Pr(Y_i \leq j) - \Pr(Y_i \leq j - 1)$: probability for each category by differencing cumulative probabilities

Ordered logit in R

```
library(MASS); library(marginaleffects)
m_ord = polr(factor(satisfaction) ~ age + educ,
     data = mydata, Hess = TRUE)
avg_slopes(m_ord) # AMEs per predictor per category
```

- `Hess = TRUE`: stores the Hessian — required for SEs

Ordered logit in R

```
library(MASS); library(marginaleffects)
m_ord = polr(factor(satisfaction) ~ age + educ,
     data = mydata, Hess = TRUE)
avg_slopes(m_ord) # AMEs per predictor per category
```

- `Hess = TRUE`: stores the Hessian — required for SEs
- Sign warning: raw `polr()` coefficients use a reversed sign convention — always interpret via `avg_slopes()`

Ordered logit: R output

Using `housing` data from MASS: satisfaction with housing conditions

```
> m_ord = polr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing, Hess = TRUE)
> summary(m_ord)

Coefficients:
                Value Std. Error t value
InflMedium     0.5664    0.10465   5.412
InflHigh       1.2888    0.12716  10.136
TypeApartment -0.5724    0.11924  -4.800
TypeAtrium    -0.3662    0.15517  -2.360
TypeTerrace   -1.0910    0.15149  -7.202
ContHigh       0.3603    0.09554   3.771

Intercepts:
            Value   Std. Error t value
Low|Medium  -0.4961  0.1248    -3.9739
Medium|High  0.6907  0.1255     5.5049

Residual Deviance: 3479.149
AIC: 3495.149
```

- **Coefficients**: log-odds scale, not directly interpretable
- **Intercepts**: estimated thresholds τ_j between categories
- No *p*-values — use *t*-values > 2 as rough guide

Ordered logit: marginal effects

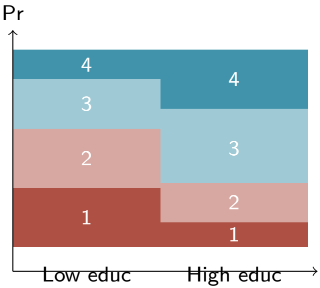
`avg_slopes(m_ord)` — average marginal effects on probability scale

```
> avg_slopes(m_ord)

 Term  Group          Contrast Estimate Std. Error      z Pr(>|z|)
 Cont  Low    High - Low       -0.0727     0.0193  -3.77   <0.001
 Cont  Medium High - Low       -0.0064     0.0022  -2.93    0.003
 Cont  High   High - Low        0.0791     0.0206   3.83   <0.001
 Infl  Low    Medium - Low     -0.1287     0.0236  -5.46   <0.001
 Infl  Low    High - Low       -0.2619     0.0237 -11.04   <0.001
 Infl  High   Medium - Low      0.1202     0.0220   5.46   <0.001
 Infl  High   High - Low        0.2909     0.0279  10.42   <0.001
 Type  Low    Terrace - Tower   0.2239     0.0312   7.17   <0.001
 Type  High   Terrace - Tower  -0.2393     0.0316  -7.59   <0.001
```

- One row per predictor $\times$ outcome category
- **Group**: which satisfaction level (Low, Medium, High)
- E.g., high influence $\rightarrow$ +29pp probability of "High" satisfaction
- AMEs across categories **sum to zero** for each predictor

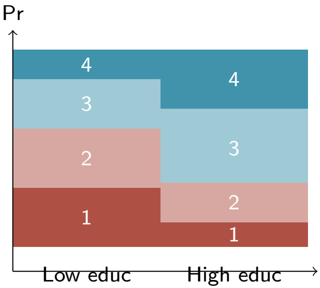
Interpreting ordered logit

- Raw coefficients: log-odds — not intuitive



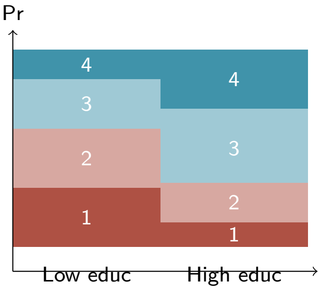
Interpreting ordered logit

- Raw coefficients: log-odds — not intuitive
- **Better**: average marginal effects



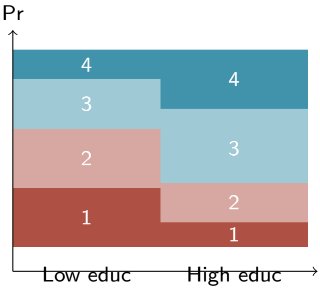
Interpreting ordered logit

- Raw coefficients: log-odds — not intuitive
- **Better**: average marginal effects
  - → One AME per predictor *per category*



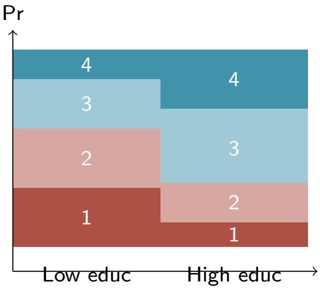
Interpreting ordered logit

- Raw coefficients: log-odds — not intuitive
- **Better**: average marginal effects
  - → One AME per predictor *per category*
  - → "A one-unit increase in education changes the probability of category j by ΔP_j"



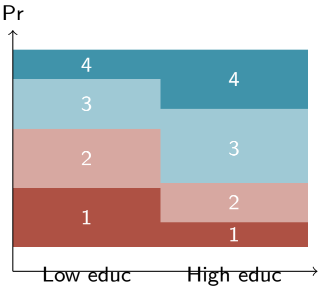
Interpreting ordered logit

- Raw coefficients: log-odds — not intuitive
- **Better**: average marginal effects
  - → One AME per predictor *per category*
  - → "A one-unit increase in education changes the probability of category j by ΔP_j"
- **Best**: predicted probability plots



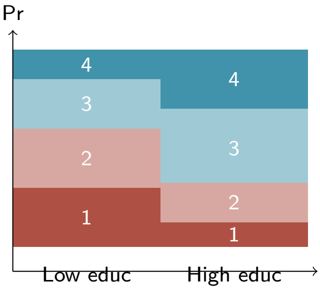
Interpreting ordered logit

- Raw coefficients: log-odds — not intuitive
- **Better**: average marginal effects
  - → One AME per predictor *per category*
  - → "A one-unit increase in education changes the probability of category j by ΔP_j"
- **Best**: predicted probability plots
  - → How does $\Pr(Y = j)$ change across X?



Interpreting ordered logit

- Raw coefficients: log-odds — not intuitive
- **Better**: average marginal effects
  - → One AME per predictor *per category*
  - → "A one-unit increase in education changes the probability of category j by ΔP_j"
- **Best**: predicted probability plots
  - → How does $\Pr(Y = j)$ change across X?
  - → Use `plot_predictions(m_ord, condition = "educ")`



Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - $\rightarrow$ Democracy scores (autocracy $\rightarrow$ anocracy $\rightarrow$ democracy)

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → . . .)

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → . . .)
  - → Health status (healthy → mild → severe → critical)

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → . . .)
  - → Health status (healthy → mild → severe → critical)
- **Idea**: given current state S_t, model the probability of transitioning to each state at S_{t+1} using ordinal logit

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → ...)
  - → Health status (healthy → mild → severe → critical)
- **Idea**: given current state S_t, model the probability of transitioning to each state at S_{t+1} using ordinal logit
- The ordinal structure constrains transitions to respect the natural ordering — large jumps (e.g., autocracy → full democracy) are possible but predicted to be rare

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → ...)
  - → Health status (healthy → mild → severe → critical)
- **Idea**: given current state S_t, model the probability of transitioning to each state at S_{t+1} using ordinal logit
- The ordinal structure constrains transitions to respect the natural ordering — large jumps (e.g., autocracy → full democracy) are possible but predicted to be rare
- Covariates shift the latent propensity, making upward or downward transitions more likely

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - $\rightarrow$ Democracy scores (autocracy $\rightarrow$ anocracy $\rightarrow$ democracy)
  - $\rightarrow$ Credit ratings (AAA $\rightarrow$ AA $\rightarrow$ A $\rightarrow \ldots$)
  - $\rightarrow$ Health status (healthy $\rightarrow$ mild $\rightarrow$ severe $\rightarrow$ critical)
- **Idea**: given current state S_t, model the probability of transitioning to each state at S_{t+1} using ordinal logit
- The ordinal structure constrains transitions to respect the natural ordering — large jumps (e.g., autocracy $\rightarrow$ full democracy) are possible but predicted to be rare
- Covariates shift the latent propensity, making upward or downward transitions more likely
  - $\rightarrow$ E.g., economic crisis $\rightarrow$ higher probability of moving down the scale

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → ...)
  - → Health status (healthy → mild → severe → critical)
- **Idea**: given current state S_t, model the probability of transitioning to each state at S_{t+1} using ordinal logit
- The ordinal structure constrains transitions to respect the natural ordering — large jumps (e.g., autocracy → full democracy) are possible but predicted to be rare
- Covariates shift the latent propensity, making upward or downward transitions more likely
  - → E.g., economic crisis → higher probability of moving down the scale

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → . . .)
  - → Health status (healthy → mild → severe → critical)
- **Idea**: given current state S_t, model the probability of transitioning to each state at S_{t+1} using ordinal logit
- The ordinal structure constrains transitions to respect the natural ordering — large jumps (e.g., autocracy → full democracy) are possible but predicted to be rare
- Covariates shift the latent propensity, making upward or downward transitions more likely
  - → E.g., economic crisis → higher probability of moving down the scale

Application: Markov transition models

- Ordinal models can be used to study **transitions between ordered states**
  - → Democracy scores (autocracy → anocracy → democracy)
  - → Credit ratings (AAA → AA → A → . . .)
  - → Health status (healthy → mild → severe → critical)
- **Idea**: given current state S_t, model the probability of transitioning to each state at S_{t+1} using ordinal logit
- The ordinal structure constrains transitions to respect the natural ordering — large jumps (e.g., autocracy → full democracy) are possible but predicted to be rare
- Covariates shift the latent propensity, making upward or downward transitions more likely
  - → E.g., economic crisis → higher probability of moving down the scale

Roadmap

Nominal outcomes:

Multiple unordered categories — vote choice, transport mode, conflict type

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking
- Examples:

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking
- Examples:
  - → Vote choice in a multi-party election (PSOE, PP, Vox, Sumar, ...)

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking
- Examples:
  - → Vote choice in a multi-party election (PSOE, PP, Vox, Sumar, ...)
  - → Mode of transport (car, bus, train, bike, walk)

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking
- Examples:
  - → Vote choice in a multi-party election (PSOE, PP, Vox, Sumar, . . .)
  - → Mode of transport (car, bus, train, bike, walk)
  - → Type of conflict (interstate, civil war, non-state)

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking
- Examples:
  - → Vote choice in a multi-party election (PSOE, PP, Vox, Sumar, . . .)
  - → Mode of transport (car, bus, train, bike, walk)
  - → Type of conflict (interstate, civil war, non-state)
  - → Occupation category (manager, professional, manual, service)

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking
- Examples:
  - → Vote choice in a multi-party election (PSOE, PP, Vox, Sumar, . . .)
  - → Mode of transport (car, bus, train, bike, walk)
  - → Type of conflict (interstate, civil war, non-state)
  - → Occupation category (manager, professional, manual, service)
- Ordered logit is wrong here — it would impose an ordering

When categories have no natural order

- **Nominal outcomes**: categories, but no ranking
- Examples:
  - → Vote choice in a multi-party election (PSOE, PP, Vox, Sumar, ...)
  - → Mode of transport (car, bus, train, bike, walk)
  - → Type of conflict (interstate, civil war, non-state)
  - → Occupation category (manager, professional, manual, service)
- Ordered logit is wrong here — it would impose an ordering
- We need a model that treats all categories symmetrically

Multinomial logit: the model

Softmax probabilities

$$\Pr(Y_i = j \mid \mathbf{x}_i) = \frac{\exp(\beta_j' \mathbf{x}_i)}{\sum_{k=1}^{J} \exp(\beta_k' \mathbf{x}_i)}$$

- J categories $\rightarrow$ $J - 1$ sets of coefficients (one category is the **reference**)

Multinomial logit: the model

Softmax probabilities

$$\Pr(Y_i = j \mid \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}_j' \mathbf{x}_i)}{\sum_{k=1}^{J} \exp(\boldsymbol{\beta}_k' \mathbf{x}_i)}$$

- J categories $\rightarrow$ $J - 1$ sets of coefficients (one category is the **reference**)
- Reference category: $\boldsymbol{\beta}_1 = \mathbf{0}$ by convention

Multinomial logit: the model

Softmax probabilities

$$\Pr(Y_i = j \mid \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}_j' \mathbf{x}_i)}{\sum_{k=1}^{J} \exp(\boldsymbol{\beta}_k' \mathbf{x}_i)}$$

- J categories $\to$ $J - 1$ sets of coefficients (one category is the **reference**)
- Reference category: $\boldsymbol{\beta}_1 = \mathbf{0}$ by convention
- Each β_{jk}: effect of x_k on log-odds of category j **vs. reference**

Multinomial logit: the model

Softmax probabilities

$$\Pr(Y_i = j \mid \mathbf{x}_i) = \frac{\exp(\boldsymbol{\beta}_j' \mathbf{x}_i)}{\sum_{k=1}^{J} \exp(\boldsymbol{\beta}_k' \mathbf{x}_i)}$$

- J categories $\rightarrow$ $J - 1$ sets of coefficients (one category is the **reference**)
- Reference category: $\boldsymbol{\beta}_1 = \mathbf{0}$ by convention
- Each β_{jk}: effect of x_k on log-odds of category j **vs. reference**
- All predicted probabilities sum to 1: $\sum_{j=1}^{J} \Pr(Y_i = j) = 1$

Multinomial logit in R

```
library(nnet)

m_mnom = multinom(vote ~ age + ideology + income,
      data = beps)

# AMEs (one per predictor per category)

avg_slopes(m_mnom)
```

- First category = reference; coefficients = log-odds vs. reference

Multinomial logit in R

```
library(nnet)

m_mnom = multinom(vote ~ age + ideology + income,
      data = beps)

# AMEs (one per predictor per category)

avg_slopes(m_mnom)
```

- First category = reference; coefficients = log-odds vs. reference
- Change reference: `relevel(factor(vote), ref = "Labour")`

Multinomial logit in R

```
library(nnet)

m_mnom = multinom(vote ~ age + ideology + income,
      data = beps)

# AMEs (one per predictor per category)

avg_slopes(m_mnom)
```

- First category = reference; coefficients = log-odds vs. reference
- Change reference: `relevel(factor(vote), ref = "Labour")`
- Interpret via `avg_slopes()` or `plot_predictions()`

Multinomial logit: R output

Using `iris` data: predicting species from flower measurements

```
> m_mnom = multinom(Species ~ Sepal.Length + Sepal.Width + Petal.Length,
+                   data = iris, trace = FALSE)
> summary(m_mnom)

Coefficients:
           (Intercept) Sepal.Length Sepal.Width Petal.Length
versicolor    15.84477    -5.106783   -9.373084     15.67647
virginica    -22.68156    -8.980237   -9.993958     28.90429

Std. Errors:
           (Intercept) Sepal.Length Sepal.Width Petal.Length
versicolor    38.86805     104.8927    164.5751     79.52894
virginica     39.08951     104.9167    164.5911     79.55610

Residual Deviance: 23.77288
AIC: 39.77288
```

- One row of coefficients **per non-reference category**
- Reference = *setosa*; each row = log-odds vs. reference
- Coefficients are log-odds → hard to interpret directly
- Use `avg_slopes()` or `plot_predictions()` instead

Multinomial logit: marginal effects

`avg_slopes(m_mnom)` — average marginal effects on probability scale

```
> avg_slopes(m_mnom)

         Term      Group    Estimate Std. Error       z Pr(>|z|)
 Petal.Length setosa     -0.00003018   0.001756  -0.017   0.9863
 Petal.Length versicolor -0.31811938   0.023102 -13.770   <0.001
 Petal.Length virginica   0.31814956   0.023035  13.812   <0.001
 Sepal.Length setosa      0.00000983   0.000578   0.017   0.9864
 Sepal.Length versicolor  0.09315278   0.032003   2.911   0.0036
 Sepal.Length virginica  -0.09316261   0.031998  -2.912   0.0036
 Sepal.Width  setosa      0.00001805   0.001106   0.016   0.9870
 Sepal.Width  versicolor  0.01491495   0.054905   0.272   0.7859
 Sepal.Width  virginica  -0.01493299   0.054891  -0.272   0.7856
```

- One row per predictor $\times$ category (like ordered logit)
- 1cm longer petal $\rightarrow$ -32pp *versicolor*, $+32$pp *virginica*
- AMEs across categories **sum to zero** for each predictor

The IIA assumption

Independence of Irrelevant Alternatives (IIA): odds ratio btw any two alternatives **unaffected** by presence of other alternatives

$$\frac{\Pr(Y = j)}{\Pr(Y = k)} = \exp\big((\boldsymbol{\beta}_j - \boldsymbol{\beta}_k)'\mathbf{x}\big)$$

- The famous **red bus / blue bus** problem:

The IIA assumption

Independence of Irrelevant Alternatives (IIA): odds ratio btw any two alternatives **unaffected** by presence of other alternatives

$$\frac{\Pr(Y = j)}{\Pr(Y = k)} = \exp\big((\boldsymbol{\beta}_j - \boldsymbol{\beta}_k)'\mathbf{x}\big)$$

- The famous **red bus / blue bus** problem:
  - → 50% car, 50% red bus — now introduce a blue bus

The IIA assumption

Independence of Irrelevant Alternatives (IIA): odds ratio btw any two alternatives **unaffected** by presence of other alternatives

$$\frac{\Pr(Y=j)}{\Pr(Y=k)} = \exp\big((\boldsymbol{\beta}_j - \boldsymbol{\beta}_k)'\mathbf{x}\big)$$

- The famous **red bus / blue bus** problem:
  - → 50% car, 50% red bus — now introduce a blue bus
  - → IIA implies: 33% car, 33% red bus, 33% blue bus

The IIA assumption

Independence of Irrelevant Alternatives (IIA): odds ratio btw any two alternatives **unaffected** by presence of other alternatives

$$\frac{\Pr(Y=j)}{\Pr(Y=k)} = \exp\big((\boldsymbol{\beta}_j - \boldsymbol{\beta}_k)'\mathbf{x}\big)$$

- The famous **red bus / blue bus** problem:
  - → 50% car, 50% red bus — now introduce a blue bus
  - → IIA implies: 33% car, 33% red bus, 33% blue bus
  - → Reality: 50% car, 25% red bus, 25% blue bus

The IIA assumption

Independence of Irrelevant Alternatives (IIA): odds ratio btw any two alternatives **unaffected** by presence of other alternatives

$$\frac{\Pr(Y=j)}{\Pr(Y=k)} = \exp\big((\boldsymbol{\beta}_j - \boldsymbol{\beta}_k)'\mathbf{x}\big)$$

- The famous **red bus / blue bus** problem:
  - → 50% car, 50% red bus — now introduce a blue bus
  - → IIA implies: 33% car, 33% red bus, 33% blue bus
  - → Reality: 50% car, 25% red bus, 25% blue bus
- IIA holds when alternatives are **genuinely distinct**

The IIA assumption

Independence of Irrelevant Alternatives (IIA): odds ratio btw any two alternatives **unaffected** by presence of other alternatives

$$\frac{\Pr(Y = j)}{\Pr(Y = k)} = \exp\big((\boldsymbol{\beta}_j - \boldsymbol{\beta}_k)'\mathbf{x}\big)$$

- The famous **red bus / blue bus** problem:
  - → 50% car, 50% red bus — now introduce a blue bus
  - → IIA implies: 33% car, 33% red bus, 33% blue bus
  - → Reality: 50% car, 25% red bus, 25% blue bus
- IIA holds when alternatives are **genuinely distinct**
- Violations: similar alternatives that substitute for each other

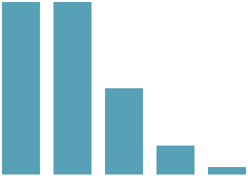
Roadmap

Count data:

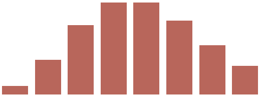
Non-negative integers — conflicts, protests, bills, articles

Count outcomes

- Outcomes that are non-negative integers: 0, 1, 2, 3, . . .



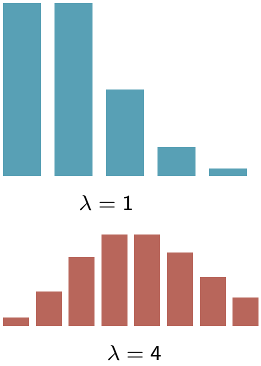
$\lambda = 1$



$\lambda = 4$

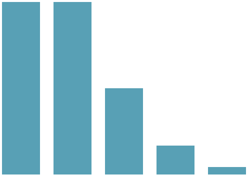
Count outcomes

- Outcomes that are non-negative integers: 0, 1, 2, 3, . . .
- Examples: conflict events, bills, protests, publications, attacks

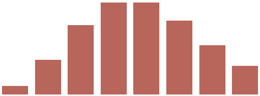


Count outcomes

- Outcomes that are non-negative integers: 0, 1, 2, 3, . . .
- Examples: conflict events, bills, protests, publications, attacks
- **Why not OLS?**



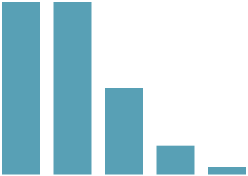
$\lambda = 1$



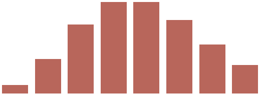
$\lambda = 4$

Count outcomes

- Outcomes that are non-negative integers: 0, 1, 2, 3, . . .
- Examples: conflict events, bills, protests, publications, attacks
- **Why not OLS?**
  - → Counts ≥ 0: OLS can predict negative values



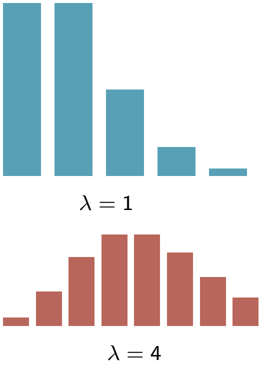
$\lambda = 1$



$\lambda = 4$

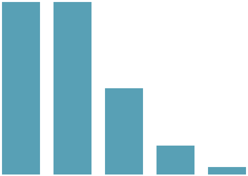
Count outcomes

- Outcomes that are non-negative integers: 0, 1, 2, 3, . . .
- Examples: conflict events, bills, protests, publications, attacks
- **Why not OLS?**
  - → Counts ≥ 0: OLS can predict negative values
  - → Variance increases with the mean

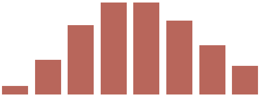


Count outcomes

- Outcomes that are non-negative integers: 0, 1, 2, 3, ...
- Examples: conflict events, bills, protests, publications, attacks
- **Why not OLS?**
  - → Counts ≥ 0: OLS can predict negative values
  - → Variance increases with the mean
  - → Strongly right-skewed; often many zeros



$\lambda = 1$



$\lambda = 4$

The Poisson model

Poisson regression

$$Y_i \sim \text{Poisson}(\mu_i), \quad \mu_i = \exp(\mathbf{x}_i\boldsymbol{\beta})$$

$$\Leftrightarrow \quad \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$$

- $\mu_i = E[Y_i \mid \mathbf{x}_i]$: expected count (always > 0)

The Poisson model

Poisson regression

$$Y_i \sim \text{Poisson}(\mu_i), \quad \mu_i = \exp(\mathbf{x}_i \boldsymbol{\beta})$$

$$\Leftrightarrow \quad \log(\mu_i) = \mathbf{x}_i \boldsymbol{\beta}$$

- $\mu_i = E[Y_i \mid \mathbf{x}_i]$: expected count (always > 0)
- The **log link** ensures predictions are always non-negative

The Poisson model

Poisson regression

$$Y_i \sim \text{Poisson}(\mu_i), \quad \mu_i = \exp(\mathbf{x}_i\boldsymbol{\beta})$$

$$\Leftrightarrow \quad \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$$

- $\mu_i = E[Y_i \mid \mathbf{x}_i]$: expected count (always > 0)
- The **log link** ensures predictions are always non-negative
- Coefficient interpretation: $\exp(\beta_k) =$ **multiplicative** change in expected count

The Poisson model

Poisson regression

$$Y_i \sim \text{Poisson}(\mu_i), \quad \mu_i = \exp(\mathbf{x}_i \boldsymbol{\beta})$$

$$\Leftrightarrow \quad \log(\mu_i) = \mathbf{x}_i \boldsymbol{\beta}$$

- $\mu_i = E[Y_i \mid \mathbf{x}_i]$: expected count (always > 0)
- The **log link** ensures predictions are always non-negative
- Coefficient interpretation: $\exp(\beta_k) =$ **multiplicative** change in expected count
  - $\rightarrow$ $\beta_k = 0.3 \Rightarrow \exp(0.3) \approx 1.35$: 35% increase in expected count

The Poisson model

Poisson regression

$$Y_i \sim \text{Poisson}(\mu_i), \quad \mu_i = \exp(\mathbf{x}_i \boldsymbol{\beta})$$

$$\Leftrightarrow \quad \log(\mu_i) = \mathbf{x}_i \boldsymbol{\beta}$$

- $\mu_i = E[Y_i \mid \mathbf{x}_i]$: expected count (always > 0)
- The **log link** ensures predictions are always non-negative
- Coefficient interpretation: $\exp(\beta_k) =$ **multiplicative** change in expected count
  - $\rightarrow$ $\beta_k = 0.3 \Rightarrow \exp(0.3) \approx 1.35$: 35% increase in expected count
- Poisson assumption: $\text{Var}(Y_i) = E[Y_i]$ (mean $=$ variance)

Poisson in R and interpretation

```
# Fit Poisson model
m_pois = glm(n_conflicts ~ gdp_pc + pop_l,
     family = "poisson", data = mydata)

summary(m_pois)
exp(coef(m_pois)) # incidence rate ratios
```

Raw coefficient	**Exponentiated**
$\hat{\beta}_{\text{gdp\_pc}} = -0.24$	$\exp(-0.24) \approx 0.79$
Hard to interpret directly	$\rightarrow$ 21% fewer conflicts per unit GDP

Also: `avg_slopes(m_pois)` gives AMEs on count scale

Poisson: R output

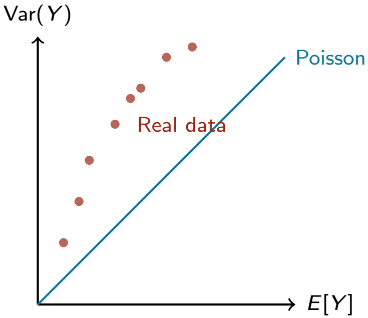
Using `warpbreaks`: number of breaks in yarn by wool type and tension

```
> m_pois = glm(breaks ~ wool + tension, family = "poisson", data = warpbreaks)
> summary(m_pois)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.69196    0.04541  81.302  < 2e-16 ***
woolB       -0.20599    0.05157  -3.994 6.49e-05 ***
tensionM    -0.32132    0.06027  -5.332 9.73e-08 ***
tensionH    -0.51849    0.06396  -8.107 5.21e-16 ***

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 297.37  on 53  degrees of freedom
Residual deviance: 210.39  on 50  degrees of freedom
AIC: 493.06

> exp(coef(m_pois))
(Intercept)       woolB    tensionM    tensionH
     40.124       0.814       0.725       0.595
```

- Coefficients on **log scale** $\rightarrow$ exponentiate for IRRs
- `woolB`: $\exp(-0.21) = 0.81 \rightarrow 19\%$ fewer breaks
- Red flag: residual deviance (210) $\gg$ df (50) $\rightarrow$ overdispersion!

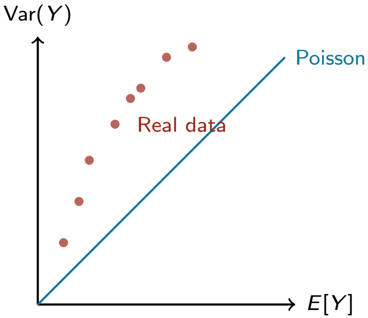
The overdispersion problem

- Poisson requires: $\text{Var}(Y) = E[Y]$



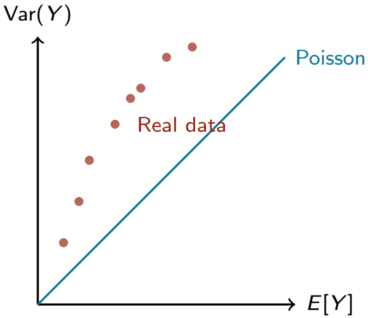
The overdispersion problem

- Poisson requires: $\text{Var}(Y) = E[Y]$
- Real data almost always: $\text{Var}(Y) \gg E[Y]$



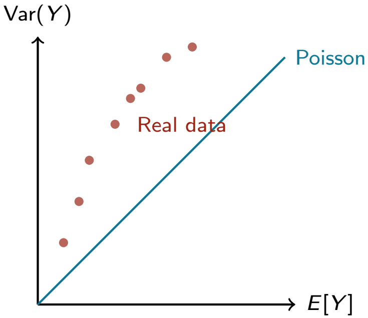
The overdispersion problem

- Poisson requires: $\text{Var}(Y) = E[Y]$
- Real data almost always: $\text{Var}(Y) \gg E[Y]$
- Consequences of ignoring overdispersion:



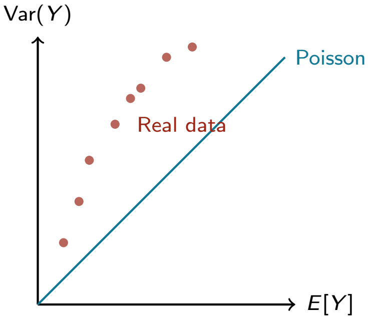
The overdispersion problem

- Poisson requires: $\text{Var}(Y) = E[Y]$
- Real data almost always: $\text{Var}(Y) \gg E[Y]$
- Consequences of ignoring overdispersion:
  - $\rightarrow$ Standard errors too **small**



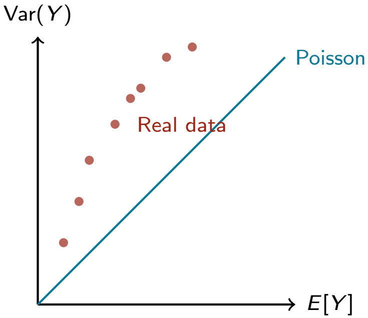
The overdispersion problem

- Poisson requires: $\text{Var}(Y) = E[Y]$
- Real data almost always: $\text{Var}(Y) \gg E[Y]$
- Consequences of ignoring overdispersion:
  - → Standard errors too **small**
  - → *p*-values too **small**



The overdispersion problem

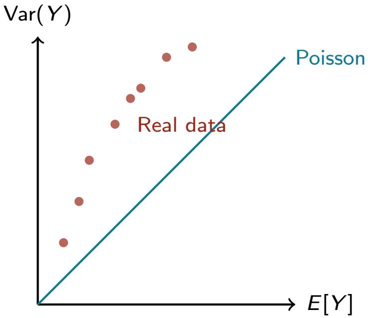
- Poisson requires: $\text{Var}(Y) = E[Y]$
- Real data almost always:
  $\text{Var}(Y) \gg E[Y]$
- Consequences of ignoring overdispersion:
  - $\rightarrow$ Standard errors too **small**
  - $\rightarrow$ *p*-values too **small**
  - $\rightarrow$ False positives



The overdispersion problem

- Poisson requires: $\text{Var}(Y) = E[Y]$
- Real data almost always: $\text{Var}(Y) \gg E[Y]$
- Consequences of ignoring overdispersion:
  - $\rightarrow$ Standard errors too **small**
  - $\rightarrow$ p-values too **small**
  - $\rightarrow$ False positives
- Quick check:

  Residual deviance $\gg$ df



Negative binomial regression

Negative Binomial model

$$Y_i \sim \text{NegBin}(\mu_i, \theta), \quad \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$$

$$\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

- $\theta > 0$: dispersion parameter estimated from the data

Negative binomial regression

Negative Binomial model

$$Y_i \sim \text{NegBin}(\mu_i, \theta), \quad \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$$

$$\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

- $\theta > 0$: dispersion parameter estimated from the data
  - → $\theta \to \infty$: approaches Poisson (no overdispersion)

Negative binomial regression

Negative Binomial model

$$Y_i \sim \text{NegBin}(\mu_i, \theta), \quad \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$$

$$\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

- $\theta > 0$: dispersion parameter estimated from the data
  - → $\theta \to \infty$: approaches Poisson (no overdispersion)
  - → Small θ: high overdispersion

Negative binomial regression

Negative Binomial model

$$Y_i \sim \text{NegBin}(\mu_i, \theta), \quad \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$$

$$\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

- $\theta > 0$: dispersion parameter estimated from the data
  - → $\theta \to \infty$: approaches Poisson (no overdispersion)
  - → Small θ: high overdispersion
- Same interpretation as Poisson: $\exp(\beta_k)$ = incidence rate ratio

Negative binomial regression

Negative Binomial model

$$Y_i \sim \text{NegBin}(\mu_i, \theta), \quad \log(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}$$

$$\text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

- $\theta > 0$: dispersion parameter estimated from the data
  - → $\theta \to \infty$: approaches Poisson (no overdispersion)
  - → Small θ: high overdispersion
- Same interpretation as Poisson: $\exp(\beta_k)$ = incidence rate ratio
- In R: `MASS::glm.nb(y ~ x1 + x2)`

Negative binomial: R output

Same warpbreaks data — compare with Poisson output

```
> m_nb = glm.nb(breaks ~ wool + tension, data = warpbreaks)
> summary(m_nb)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.6734     0.0979  37.520  < 2e-16 ***
woolB        -0.1862     0.1010  -1.844   0.0651 .
tensionM     -0.2992     0.1217  -2.458   0.0140 *
tensionH     -0.5114     0.1237  -4.133 3.58e-05 ***

(Dispersion parameter for Negative Binomial(9.9444) family taken to be 1)

    Null deviance: 75.464  on 53  degrees of freedom
Residual deviance: 53.723  on 50  degrees of freedom
AIC: 408.76
              Theta:  9.94
          Std. Err.:  2.56

> exp(coef(m_nb))
(Intercept)       woolB    tensionM    tensionH
     39.384       0.830       0.741       0.600
```

- **SEs are larger** than Poisson (e.g., `woolB`: 0.10 vs. 0.05)
- `woolB` now only marginal ($p = 0.065$) — was "highly significant"
- Deviance/df $\approx$ 1.07 — overdispersion is handled

Count models: a summary

Model	**R function**	**When to use**
Poisson	`glm(...,poisson)`	Equidispersed counts
Negative binomial	`MASS::glm.nb()`	Overdispersed counts
Zero-inflated Poisson	`pscl::zeroinfl()`	Excess zeros, equidisp.
Zero-inflated NegBin	`pscl::zeroinfl()`	Excess zeros, overdisp.

- **Default**: start Poisson, test overdispersion, switch to NegBin if needed (**in practice**: often go ahead to NB)

Count models: a summary

Model	R function	When to use
Poisson	`glm(...,poisson)`	Equidispersed counts
Negative binomial	`MASS::glm.nb()`	Overdispersed counts
Zero-inflated Poisson	`pscl::zeroinfl()`	Excess zeros, equidisp.
Zero-inflated NegBin	`pscl::zeroinfl()`	Excess zeros, overdisp.

- **Default**: start Poisson, test overdispersion, switch to NegBin if needed (**in practice**: often go ahead to NB)
- Zero-inflation: two-process model (structural vs. count zeros)

Count models: a summary

Model	R function	When to use
Poisson	`glm(...,poisson)`	Equidispersed counts
Negative binomial	`MASS::glm.nb()`	Overdispersed counts
Zero-inflated Poisson	`pscl::zeroinfl()`	Excess zeros, equidisp.
Zero-inflated NegBin	`pscl::zeroinfl()`	Excess zeros, overdisp.

- **Default**: start Poisson, test overdispersion, switch to NegBin if needed (**in practice**: often go ahead to NB)
- Zero-inflation: two-process model (structural vs. count zeros)
- All share the log link: $\exp(\hat{\beta})$ = incidence rate ratio (IRR)

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate

- **Two-process model**:

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate
- **Two-process model**:
  - → Process 1 (logit): is the unit "at risk" of a positive count?

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate
- **Two-process model**:
  - → Process 1 (logit): is the unit "at risk" of a positive count?
  - → Process 2 (Poisson/NegBin): if at risk, how many events?

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate
- **Two-process model**:
  - → Process 1 (logit): is the unit "at risk" of a positive count?
  - → Process 2 (Poisson/NegBin): if at risk, how many events?
- Distinguishes **structural zeros** (unit can't have the event) from **sampling zeros** (could happen, just didn't)

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate
- **Two-process model**:
  - → Process 1 (logit): is the unit "at risk" of a positive count?
  - → Process 2 (Poisson/NegBin): if at risk, how many events?
- Distinguishes **structural zeros** (unit can't have the event) from **sampling zeros** (could happen, just didn't)
- In R: `pscl::zeroinfl(y ~ x | z, dist = "negbin")`

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate
- **Two-process model**:
  - → Process 1 (logit): is the unit "at risk" of a positive count?
  - → Process 2 (Poisson/NegBin): if at risk, how many events?
- Distinguishes **structural zeros** (unit can't have the event) from **sampling zeros** (could happen, just didn't)
- In R: `pscl::zeroinfl(y ~ x | z, dist = "negbin")`
  - → `x`: predictors for the count process

Zero-inflated models

- Sometimes data have **more zeros than any count distribution predicts**
  - → Conflict events: many country-years at peace (structural zeros)
  - → Patent counts: many firms never innovate
- **Two-process model**:
  - → Process 1 (logit): is the unit "at risk" of a positive count?
  - → Process 2 (Poisson/NegBin): if at risk, how many events?
- Distinguishes **structural zeros** (unit can't have the event) from **sampling zeros** (could happen, just didn't)
- In R: `pscl::zeroinfl(y ~ x | z, dist = "negbin")`
  - → `x`: predictors for the count process
  - → `z`: predictors for the zero-inflation process

Roadmap

Duration analysis:

Time until an event — war onset, cabinet survival, regime collapse

Time-to-event outcomes

- Outcomes that record **how long until something happens**

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)
  - → Duration of a coalition government in office

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)
  - → Duration of a coalition government in office
  - → Time to re-arrest after prison release

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)
  - → Duration of a coalition government in office
  - → Time to re-arrest after prison release
  - → How long an authoritarian regime survives

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)
  - → Duration of a coalition government in office
  - → Time to re-arrest after prison release
  - → How long an authoritarian regime survives
  - → Time until a bill is passed (from introduction)

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)
  - → Duration of a coalition government in office
  - → Time to re-arrest after prison release
  - → How long an authoritarian regime survives
  - → Time until a bill is passed (from introduction)
- Why not OLS on duration?

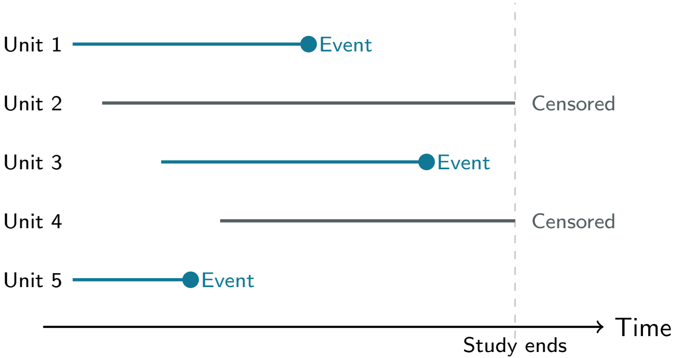
Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)
  - → Duration of a coalition government in office
  - → Time to re-arrest after prison release
  - → How long an authoritarian regime survives
  - → Time until a bill is passed (from introduction)
- Why not OLS on duration?
  - → Durations are ≥ 0 and right-skewed

Time-to-event outcomes

- Outcomes that record **how long until something happens**
- Examples:
  - → Time until a war begins (from peace)
  - → Duration of a coalition government in office
  - → Time to re-arrest after prison release
  - → How long an authoritarian regime survives
  - → Time until a bill is passed (from introduction)
- Why not OLS on duration?
  - → Durations are ≥ 0 and right-skewed
  - → **Censoring**: many observations end without the event occurring

The censoring problem



Censored units: event did not occur during observation window

What does survival data look like?

Unit	Spell start	Duration	Event	Covariates ...
Country A	1975	12 years	1 (breakdown)	GDP, region, ...
Country B	1991	30 years	0 (censored)	GDP, region, ...
Country C	1985	8 years	1 (breakdown)	GDP, region, ...
Country D	2002	19 years	0 (censored)	GDP, region, ...
Country E	1960	3 years	1 (breakdown)	GDP, region, ...

- Each row is a **spell**: one unit observed from entry to event or censoring

What does survival data look like?

Unit	**Spell start**	**Duration**	**Event**	**Covariates ...**
Country A	1975	12 years	1 (breakdown)	GDP, region, ...
Country B	1991	30 years	0 (censored)	GDP, region, ...
Country C	1985	8 years	1 (breakdown)	GDP, region, ...
Country D	2002	19 years	0 (censored)	GDP, region, ...
Country E	1960	3 years	1 (breakdown)	GDP, region, ...

- Each row is a **spell**: one unit observed from entry to event or censoring
- Two outcome variables: **duration** (T) and **event indicator** (d: 1 = event, 0 = censored)

What does survival data look like?

Unit	Spell start	Duration	Event	Covariates ...
Country A	1975	12 years	1 (breakdown)	GDP, region, ...
Country B	1991	30 years	0 (censored)	GDP, region, ...
Country C	1985	8 years	1 (breakdown)	GDP, region, ...
Country D	2002	19 years	0 (censored)	GDP, region, ...
Country E	1960	3 years	1 (breakdown)	GDP, region, ...

- Each row is a **spell**: one unit observed from entry to event or censoring
- Two outcome variables: **duration** (T) and **event indicator** (d: 1 = event, 0 = censored)
- In R: combined into a `Surv(time, event)` object

Key concepts: survival and hazard

Survival function $S(t)$

$S(t) = \Pr(T > t)$

Probability of surviving past time t

$S(0) = 1$, $S(\infty) = 0$

Monotonically non-increasing

Hazard function $h(t)$

$$h(t) = \frac{f(t)}{S(t)}$$

Instantaneous rate of event at t,

conditional on survival to t

"Danger per unit time at time t"

- High $h(t)$: event is likely soon

Key concepts: survival and hazard

Survival function $S(t)$

$S(t) = \Pr(T > t)$

Probability of surviving past time t

$S(0) = 1$, $S(\infty) = 0$

Monotonically non-increasing

Hazard function $h(t)$

$$h(t) = \frac{f(t)}{S(t)}$$

Instantaneous rate of event at t,

conditional on survival to t

"Danger per unit time at time t"

- High $h(t)$: event is likely soon
- $S(t)$ and $h(t)$ are two views of the same object

Key concepts: survival and hazard

Survival function $S(t)$

$S(t) = \Pr(T > t)$

Probability of surviving past time t

$S(0) = 1$, $S(\infty) = 0$

Monotonically non-increasing

Hazard function $h(t)$

$$h(t) = \frac{f(t)}{S(t)}$$

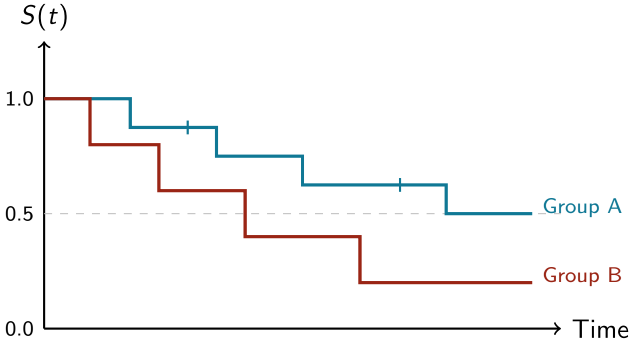
Instantaneous rate of event at t,

conditional on survival to t

"Danger per unit time at time t"

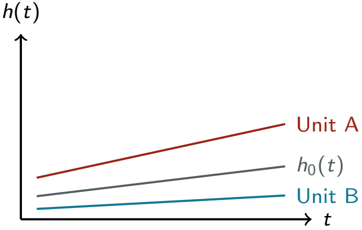
- High $h(t)$: event is likely soon
- $S(t)$ and $h(t)$ are two views of the same object
- Covariates shift $h(t)$ — that is what regression models

Kaplan-Meier: visualizing survival



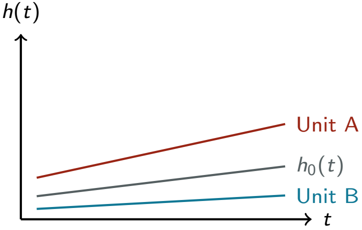
Kaplan-Meier estimator: non-parametric, no covariates, visual comparison of groups

The proportional hazards idea



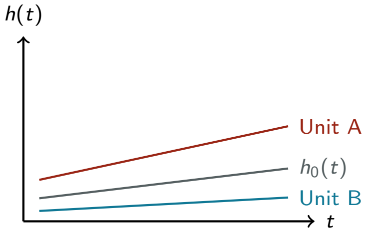
- **Key idea**: each unit has its own hazard, but all are proportional shifts of the same baseline $h_0(t)$

The proportional hazards idea



- **Key idea**: each unit has its own hazard, but all are proportional shifts of the same baseline $h_0(t)$
- The ratio $h_i(t)/h_j(t)$ is **constant over time** — curves never cross

The proportional hazards idea



- **Key idea**: each unit has its own hazard, but all are proportional shifts of the same baseline $h_0(t)$
- The ratio $h_i(t)/h_j(t)$ is **constant over time** — curves never cross
- Violation: if the gap between lines closes or reverses over time $\Rightarrow$ PH assumption fails

Cox proportional hazards model

Cox PH model

$$h_i(t) = h_0(t) \cdot \exp(\mathbf{x}_i \boldsymbol{\beta})$$

- $h_0(t)$: **baseline hazard** — left completely unspecified (semiparametric)

Cox proportional hazards model

Cox PH model

$$h_i(t) = h_0(t) \cdot \exp(\mathbf{x}_i \boldsymbol{\beta})$$

- $h_0(t)$: **baseline hazard** — left completely unspecified (semiparametric)
- $\exp(\mathbf{x}_i \boldsymbol{\beta})$: multiplicative shift of the baseline

Cox proportional hazards model

Cox PH model

$$h_i(t) = h_0(t) \cdot \exp(\mathbf{x}_i \boldsymbol{\beta})$$

- $h_0(t)$: **baseline hazard** — left completely unspecified (semiparametric)
- $\exp(\mathbf{x}_i \boldsymbol{\beta})$: multiplicative shift of the baseline
- **Proportional hazards assumption**: the hazard ratio between any two units is *constant over time*

Cox proportional hazards model

Cox PH model

$$h_i(t) = h_0(t) \cdot \exp(\mathbf{x}_i \boldsymbol{\beta})$$

- $h_0(t)$: **baseline hazard** — left completely unspecified (semiparametric)
- $\exp(\mathbf{x}_i \boldsymbol{\beta})$: multiplicative shift of the baseline
- **Proportional hazards assumption**: the hazard ratio between any two units is *constant over time*
- $\exp(\beta_k)$: **hazard ratio** for a one-unit increase in x_k

Cox proportional hazards model

Cox PH model

$$h_i(t) = h_0(t) \cdot \exp(\mathbf{x}_i \boldsymbol{\beta})$$

- $h_0(t)$: **baseline hazard** — left completely unspecified (semiparametric)
- $\exp(\mathbf{x}_i \boldsymbol{\beta})$: multiplicative shift of the baseline
- **Proportional hazards assumption**: the hazard ratio between any two units is *constant over time*
- $\exp(\beta_k)$: **hazard ratio** for a one-unit increase in x_k
  - $\rightarrow$ $\exp(\beta_k) > 1$: higher hazard (event happens sooner)

Cox proportional hazards model

Cox PH model

$$h_i(t) = h_0(t) \cdot \exp(\mathbf{x}_i \boldsymbol{\beta})$$

- $h_0(t)$: **baseline hazard** — left completely unspecified (semiparametric)
- $\exp(\mathbf{x}_i \boldsymbol{\beta})$: multiplicative shift of the baseline
- **Proportional hazards assumption**: the hazard ratio between any two units is *constant over time*
- $\exp(\beta_k)$: **hazard ratio** for a one-unit increase in x_k
  - $\rightarrow$ $\exp(\beta_k) > 1$: higher hazard (event happens sooner)
  - $\rightarrow$ $\exp(\beta_k) < 1$: lower hazard (event happens later)

Cox model in R

```
library(survival); library(broom); library(ggplot2)

# Kaplan-Meier estimator

km_fit = survfit(Surv(time, event) ~ group, data = d)

ggplot(tidy(km_fit), aes(time, estimate, color =
strata)) +
  geom_step() + geom_ribbon(aes(ymin = conf.low,
  ymax = conf.high, fill = strata), alpha = .2)

# Cox PH model

cox_fit = coxph(Surv(time, event) ~ x1 + x2, data = d)

summary(cox_fit) # shows hazard ratios

exp(coef(cox_fit))
```

Cox PH: R output

Using `lung` data (survival pkg): survival of lung cancer patients

```
> cox_fit = coxph(Surv(time, status) ~ age + sex + ph.ecog, data = lung)
> summary(cox_fit)

  n= 227, number of events= 164

             coef exp(coef) se(coef)      z Pr(>|z|)
age      0.011067  1.011128 0.009267  1.194  0.23242
sex     -0.552612  0.575445 0.167739 -3.294  0.00099 ***
ph.ecog  0.463728  1.589991 0.113577  4.083 4.45e-05 ***

        exp(coef) exp(-coef) lower .95 upper .95
age        1.0111     0.9890    0.9929    1.0297
sex        0.5754     1.7378    0.4142    0.7994
ph.ecog    1.5900     0.6289    1.2727    1.9864

Concordance= 0.637  (se = 0.025 )
Likelihood ratio test= 30.5  on 3 df,   p=1e-06
Wald test            = 29.93  on 3 df,   p=1e-06
Score (logrank) test = 30.5  on 3 df,   p=1e-06
```

- `exp(coef)` = **hazard ratios** — the key quantity to report
- `sex` (2 = female): HR = 0.58 → 42% *lower* hazard (longer survival)
- `ph.ecog`: HR = 1.59 → 59% *higher* hazard per unit increase
- Concordance = 0.64: moderate predictive accuracy

Interpreting the Cox model

	Coef	Exp(coef)	SE	p-value
gdp_pc	-0.31	0.73	0.08	< 0.001
military	0.64	1.90	0.15	0.001
log_pop	0.18	1.20	0.10	0.07

- $\exp(\hat{\beta}_{\texttt{gdp\_pc}}) = 0.73$: richer countries have 27% **lower** hazard of event

Interpreting the Cox model

	Coef	Exp(coef)	SE	p-value
`gdp_pc`	-0.31	0.73	0.08	< 0.001
`military`	0.64	1.90	0.15	0.001
`log_pop`	0.18	1.20	0.10	0.07

- $\exp(\hat{\beta}_{\texttt{gdp\_pc}}) = 0.73$: richer countries have 27% **lower** hazard of event
- $\exp(\hat{\beta}_{\texttt{military}}) = 1.90$: military regimes face 90% **higher** hazard

Interpreting the Cox model

	Coef	Exp(coef)	SE	*p*-value
gdp_pc	−0.31	0.73	0.08	< 0.001
military	0.64	1.90	0.15	0.001
log_pop	0.18	1.20	0.10	0.07

- $\exp(\hat{\beta}_{\texttt{gdp\_pc}}) = 0.73$: richer countries have 27% **lower** hazard of event
- $\exp(\hat{\beta}_{\texttt{military}}) = 1.90$: military regimes face 90% **higher** hazard
- Hazard ratio < 1: covariate *reduces* event rate $\rightarrow$ *increases* survival

Interpreting the Cox model

	Coef	Exp(coef)	SE	p-value
gdp_pc	-0.31	0.73	0.08	< 0.001
military	0.64	1.90	0.15	0.001
log_pop	0.18	1.20	0.10	0.07

- $\exp(\hat{\beta}_{\texttt{gdp\_pc}}) = 0.73$: richer countries have 27% **lower** hazard of event
- $\exp(\hat{\beta}_{\texttt{military}}) = 1.90$: military regimes face 90% **higher** hazard
- Hazard ratio < 1: covariate *reduces* event rate $\rightarrow$ *increases* survival
- Hazard ratio > 1: covariate *increases* event rate $\rightarrow$ *decreases* survival

What if proportional hazards fails?

- Test with `cox.zph()`: significant p-value $\Rightarrow$ effect changes over time

What if proportional hazards fails?

- Test with `cox.zph()`: significant p-value $\Rightarrow$ effect changes over time
- **Solutions**:

What if proportional hazards fails?

- Test with `cox.zph()`: significant p-value $\Rightarrow$ effect changes over time
- **Solutions**:
  - $\rightarrow$ **Stratify**: `coxph(Surv(t,d) ~ x + strata(z))`
    Fits separate baseline hazard per group; useful for control variables

What if proportional hazards fails?

- Test with `cox.zph()`: significant p-value $\Rightarrow$ effect changes over time

- **Solutions**:
  - $\rightarrow$ **Stratify**: `coxph(Surv(t,d) ~ x + strata(z))`
    Fits separate baseline hazard per group; useful for control variables
  - $\rightarrow$ **Time interaction**: let β vary with time
    Models the changing effect explicitly

What if proportional hazards fails?

- Test with `cox.zph()`: significant p-value $\Rightarrow$ effect changes over time
- **Solutions**:
  - $\rightarrow$ **Stratify**: `coxph(Surv(t,d) ~ x + strata(z))`
    Fits separate baseline hazard per group; useful for control variables
  - $\rightarrow$ **Time interaction**: let β vary with time
    Models the changing effect explicitly
  - $\rightarrow$ **Split follow-up**: divide time into periods, fit separate models
    Simple but choice of cutpoint can be arbitrary

What if proportional hazards fails?

- Test with `cox.zph()`: significant p-value $\Rightarrow$ effect changes over time

- **Solutions**:
  - → **Stratify**: `coxph(Surv(t,d) ~ x + strata(z))`
    Fits separate baseline hazard per group; useful for control variables
  - → **Time interaction**: let β vary with time
    Models the changing effect explicitly
  - → **Split follow-up**: divide time into periods, fit separate models
    Simple but choice of cutpoint can be arbitrary
  - → **Different model**: parametric (Weibull) or accelerated failure time
    Do not require proportional hazards

Roadmap

Pulling it together:

Which model for which data?

Choosing the right model

If your outcome is...	Use	Key check
Continuous	OLS	Linearity, homoskedasticity
Binary (0/1)	Logit / LPM	Rare events $\rightarrow$ prefer logit
Ordered categories	Ordered logit	Test PO assumption (Brant)
Unordered categories	Multinomial logit	IIA plausible?
Non-negative integer	Poisson	Test overdispersion
Overdispersed count	Neg. Binomial	Often the safe default
Many zeros	Zero-inflated	Two-process data?
Time until event	Cox PH	Test PH assumption

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)
- **Step 3**: Estimate the model in R

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)
- **Step 3**: Estimate the model in R
- **Step 4**: Check assumptions

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)
- **Step 3**: Estimate the model in R
- **Step 4**: Check assumptions
  - $\rightarrow$ Counts: overdispersion test $\rightarrow$ switch to NegBin if needed

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)
- **Step 3**: Estimate the model in R
- **Step 4**: Check assumptions
  - $\rightarrow$ Counts: overdispersion test $\rightarrow$ switch to NegBin if needed
  - $\rightarrow$ Ordinal: Brant test for proportional odds

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)
- **Step 3**: Estimate the model in R
- **Step 4**: Check assumptions
  - → Counts: overdispersion test → switch to NegBin if needed
  - → Ordinal: Brant test for proportional odds
  - → Nominal: IIA plausibility (substantive reasoning)

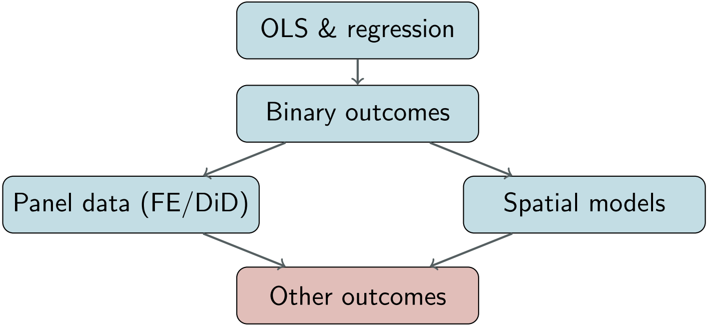
Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)
- **Step 3**: Estimate the model in R
- **Step 4**: Check assumptions
  - $\rightarrow$ Counts: overdispersion test $\rightarrow$ switch to NegBin if needed
  - $\rightarrow$ Ordinal: Brant test for proportional odds
  - $\rightarrow$ Nominal: IIA plausibility (substantive reasoning)
  - $\rightarrow$ Duration: Schoenfeld residuals for PH assumption

Practical checklist

- **Step 1**: What is the DGP? Identify the outcome type
- **Step 2**: Choose model family accordingly (see table)
- **Step 3**: Estimate the model in R
- **Step 4**: Check assumptions
  - $\rightarrow$ Counts: overdispersion test $\rightarrow$ switch to NegBin if needed
  - $\rightarrow$ Ordinal: Brant test for proportional odds
  - $\rightarrow$ Nominal: IIA plausibility (substantive reasoning)
  - $\rightarrow$ Duration: Schoenfeld residuals for PH assumption
- **Steps 5–6**: Interpret via AMEs / predicted probabilities (**never raw coefficients alone**); visualize results

This semester



Today: ordinal, nominal, count, duration

(Computing session on April 23rd)

For the exam

- Be able to **identify** the right model given an outcome description

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)
  - → Equidispersion (Poisson)

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)
  - → Equidispersion (Poisson)
  - → Proportional hazards (Cox)

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)
  - → Equidispersion (Poisson)
  - → Proportional hazards (Cox)
- Interpret coefficients correctly:

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)
  - → Equidispersion (Poisson)
  - → Proportional hazards (Cox)
- Interpret coefficients correctly:
  - → Ordered/multinomial logit: AMEs or predicted probabilities

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)
  - → Equidispersion (Poisson)
  - → Proportional hazards (Cox)
- Interpret coefficients correctly:
  - → Ordered/multinomial logit: AMEs or predicted probabilities
  - → Poisson/NegBin: $\exp(\hat{\beta})$ = incidence rate ratio

For the exam

- Be able to **identify** the right model given an outcome description
- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)
  - → Equidispersion (Poisson)
  - → Proportional hazards (Cox)
- Interpret coefficients correctly:
  - → Ordered/multinomial logit: AMEs or predicted probabilities
  - → Poisson/NegBin: $\exp(\hat{\beta})$ = incidence rate ratio
  - → Cox PH: $\exp(\hat{\beta})$ = hazard ratio

For the exam

- Be able to **identify** the right model given an outcome description

- Know the key **assumptions** and how to check them:
  - → Proportional odds (ordered logit)
  - → IIA (multinomial logit)
  - → Equidispersion (Poisson)
  - → Proportional hazards (Cox)

- Interpret coefficients correctly:
  - → Ordered/multinomial logit: AMEs or predicted probabilities
  - → Poisson/NegBin: $\exp(\hat{\beta})$ = incidence rate ratio
  - → Cox PH: $\exp(\hat{\beta})$ = hazard ratio

- Know the relevant **R functions**

Questions?