statistically speaking

Interactions in Poisson and Logistic Regression

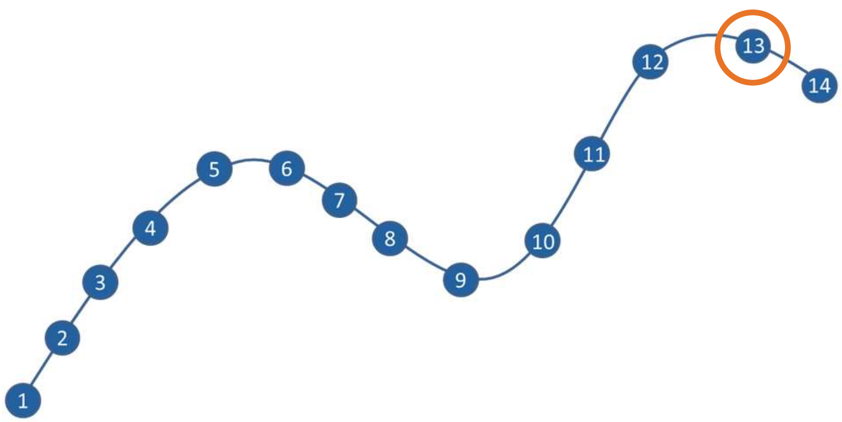
Kim Love

Where This Fits

Stage
3 – Extensions of Linear Models

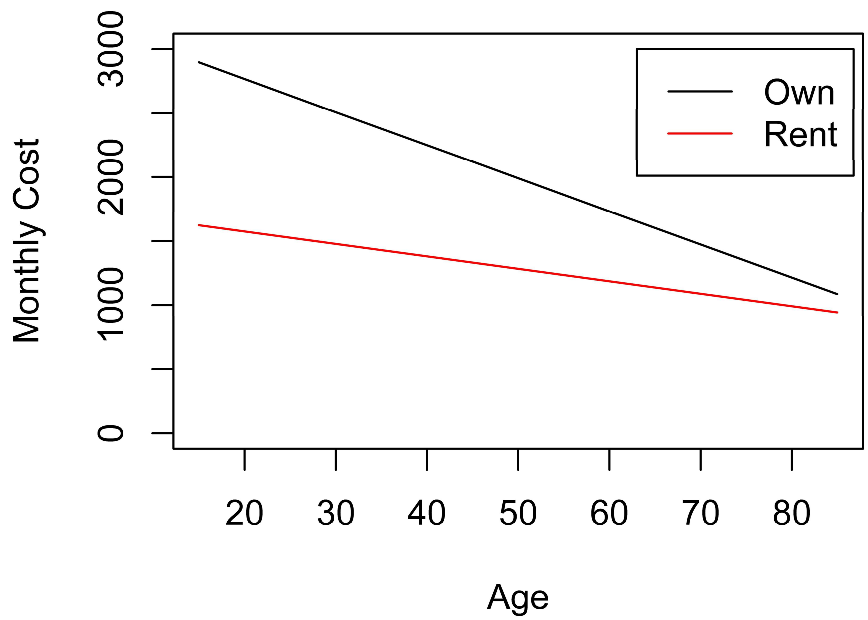
Component
Data Analysis Skill

Step
13 – Interpret Results



Introduction

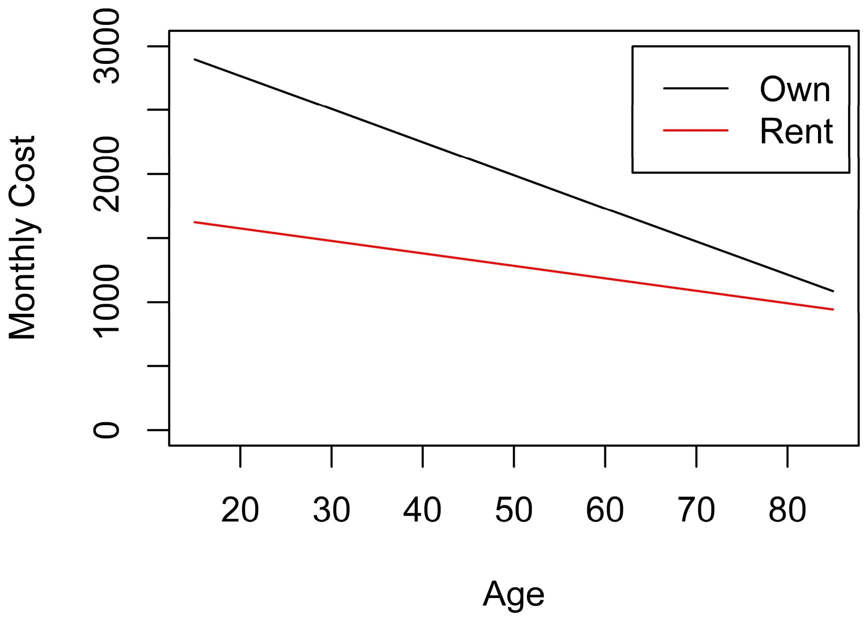
Statistical interactions are tricky to understand



Introduction

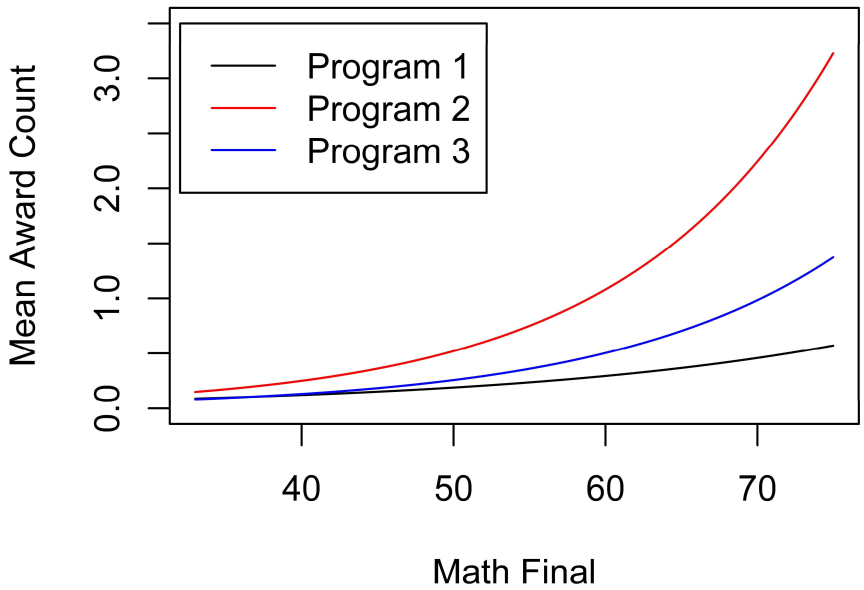
Statistical interactions are tricky to understand

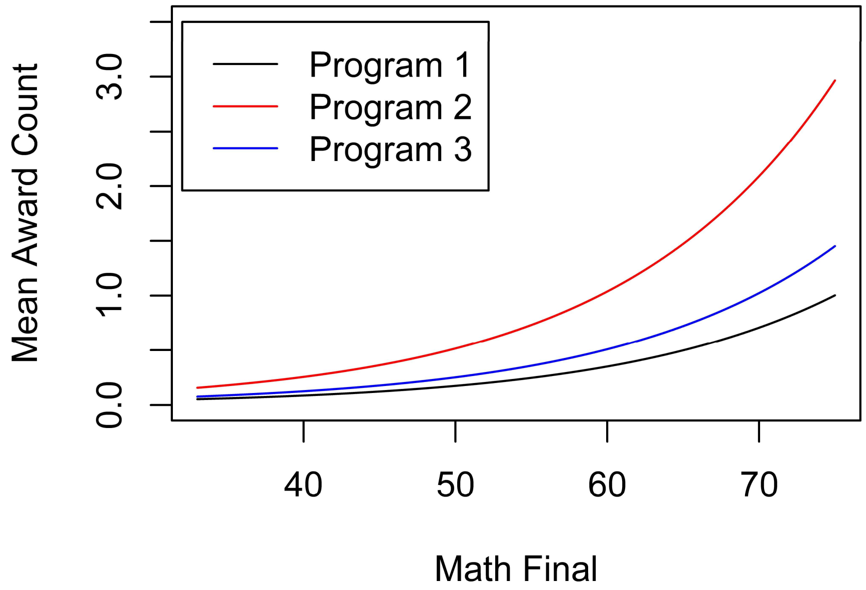
Variable	Category	Coefficient
Intercept		1769.3
Tenure	Own	1513.1
	Rent	0
Age		-9.7
Tenure x Age	Own	-16.1
	Rent	0



Introduction

Even more difficult when relationships are not linear!





Goals

Two goals:

- Learn to interpret interactions in count and binary outcome models
- Gain more general comfort with these models

What You'll Learn Today:

1. Review of Interactions in Linear Models
2. Overview of Poisson Regression Equations
3. Interactions in Poisson Regression
4. Overview of Logistic Regression Equations
5. Interactions in Logistic Regression

1. Review of Interactions in Linear Models

Interactions in Linear Models

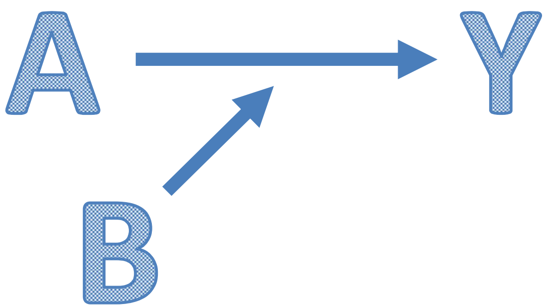
Example data set: 2021 American Housing Survey

Tenure	Age	Monthly Cost
Own	74	779
Own	55	2360
Rent	80	570
Own	60	1082
Own	38	739
Own	49	6134
	…	

Interactions in Linear Models

Interaction:

- Effect of variable A depends on value of variable B
  - Or vice versa
- Example with tenure status and age:
  - Difference between renting and owning depends on age
  - Effect of age depends on tenure status



Interactions in Linear Models

Example (tenure and age):

- NO interaction:

Variable	Category	Coefficient
Intercept		2168.9
Tenure	Own	663.7
	Rent	0
Age		-17.9

$$\hat{y} = 2168.9 + 663.7 * \text{Own} - 17.9 * \text{Age}$$

Interactions in Linear Models

Some estimated monthly costs:

$$\hat{y} = 2168.9 + 663.7 * \text{Own} - 17.9 * \text{Age}$$

- A 60-year-old homeowner has $\hat{y}$ of 2168.9 + 663.7 – 17.9*60 = $1758.60
- A 60-year-old renter has $\hat{y}$ of 2168.9 – 17.9*60 = $1094.90
- A 40-year-old homeowner has $\hat{y}$ of 2168.9 + 663.7 – 17.9*40 = $2116.60
- A 40-year-old renter has $\hat{y}$ of 2168.9 – 17.9*40 = $1452.90

Interactions in Linear Models

These are called *marginal means*

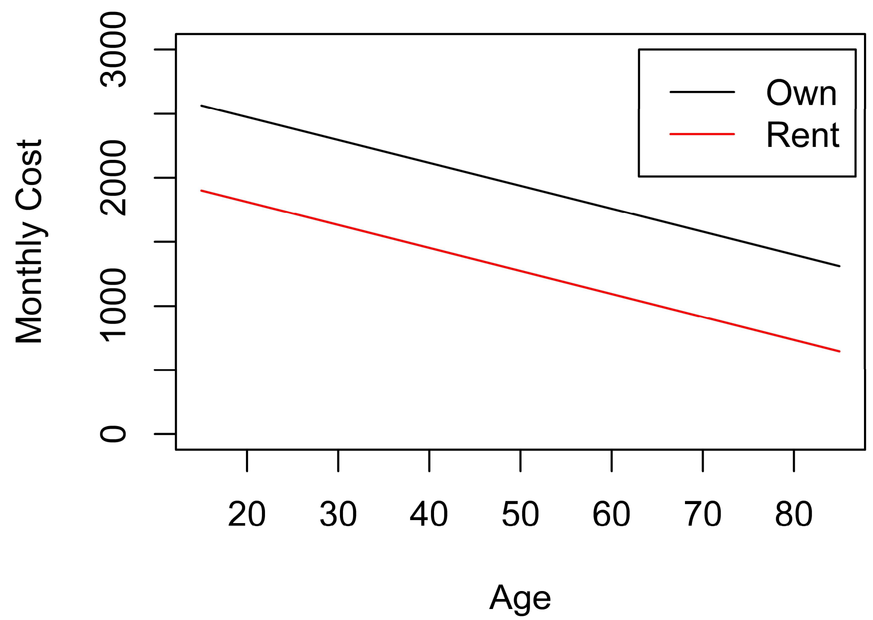
Age	Tenure	Est. Costs
60	Own	1758.60
60	Rent	1094.90
40	Own	2116.60
40	Rent	1452.90

Interactions in Linear Models

Example (tenure and age):

- NO interaction:

Variable	Category	Coefficient
Intercept		2168.9
Tenure	Own	663.7
	Rent	0
Age		-17.9



Interactions in Linear Models

Example (tenure and age):

- WITH interaction:

Variable	Category	Coefficient
Intercept		1769.3
Tenure	Own	1514.1
	Rent	0
Age		-9.7
Tenure x Age	Own	-16.1
	Rent	0

$$\hat{y} = 1769.3 + 1514.1 * \text{Own} - 9.7 * \text{Age} - 16.1 * \text{Own} * \text{Age}$$

Interactions in Linear Models

Some estimated monthly costs:

$$\hat{y} = 1769.3 + 1514.1 * \text{Own} - 9.7 * \text{Age} - 16.1 * \text{Own} * \text{Age}$$

- A 60-year-old homeowner has $\hat{y}$ of 1769.3 + 1514.1 – (9.7+16.1)*60 = $1758.60
- A 60-year-old renter has $\hat{y}$ of 1769.3 – 9.7*60 = $1758.60
- A 40-year-old homeowner has $\hat{y}$ of 1769.3 + 1514.1 – (9.7+16.1)*40 = $1758.60
- A 40-year-old renter has $\hat{y}$ of 1769.3 – 9.7*40 = $1758.60

Interactions in Linear Models

Marginal means with interaction

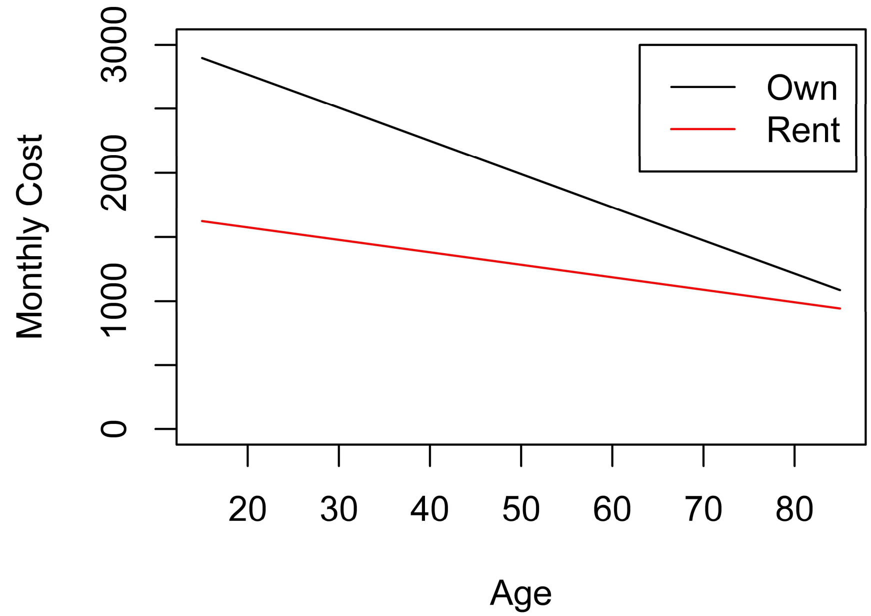
Age	Tenure	Est. Costs
60	Own	1735.4
60	Rent	1187.3
40	Own	2251.4
40	Rent	1381.3

Interactions in Linear Models

Example (tenure and age):

- WITH interaction:

Variable	Category	Coefficient
Intercept		1769.3
Tenure	Own	1514.1
	Rent	0
Age		-9.7
Tenure x Age	Own	-16.1
	Rent	0



Interactions in Linear Models

Example (tenure and age):

- Comparison:

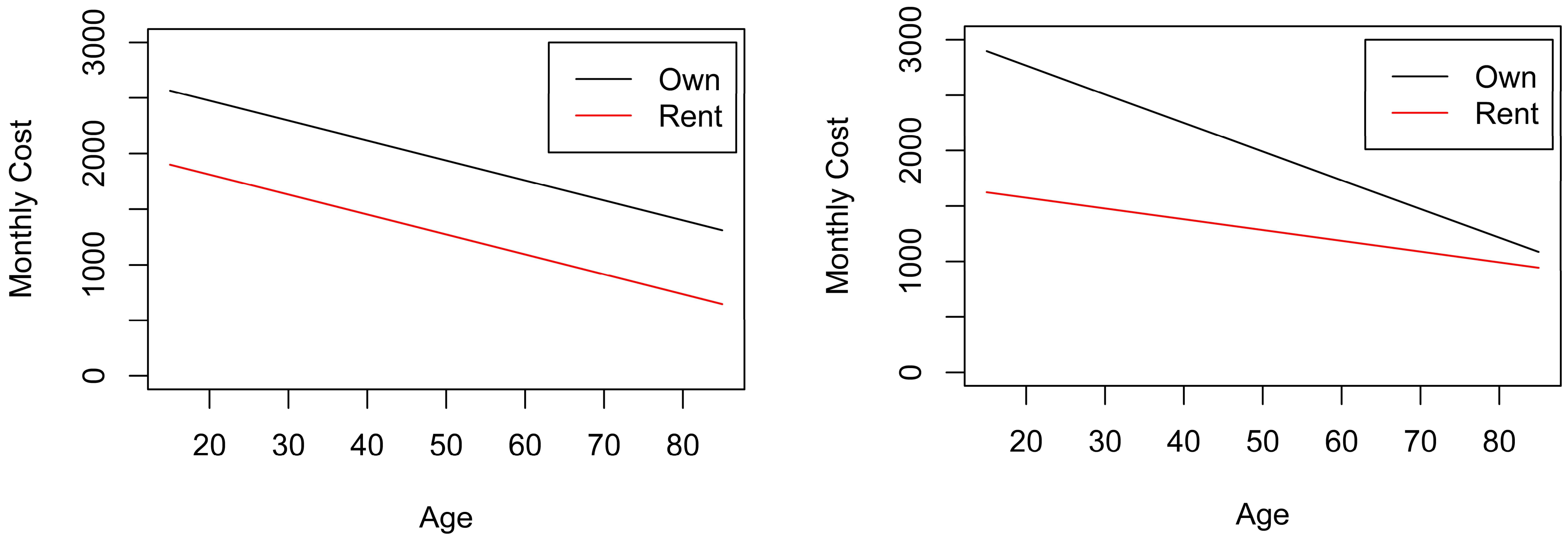
Variable	Category	Coefficient
Intercept		2168.9
Tenure	Own	663.7
	Rent	0
Age		-17.9

Variable	Category	Coefficient
Intercept		1769.3
Tenure	Own	1514.1
	Rent	0
Age		-9.7
Tenure x Age	Own	-16.1
	Rent	0

Interactions in Linear Models

Example (tenure and age):

- Comparison:



2. Overview of Poisson Regression Equations

Poisson Regression Equations

Purpose:

- Used when counts have a Poisson distribution
- Predict or estimate a mean count
- Determine *what* variables change the mean count
- Determine *how* variables change the mean count

Poisson Regression Equations

Example data set: Academic Awards

Program	MathFinal	NumAwards
3	40	0
3	33	0
2	48	0
2	41	1
2	43	1
2	46	0
…		

Poisson Regression Equations

General formula:

$$\ln(\lambda) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \dots + \beta_k x_k$$

- λ is the *mean count*
- *Link function* on the left side relates λ to xs
  - This is a *natural log* link

Poisson Regression Equations

Or, we can write

$$\lambda = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots + \beta_k x_k)$$

Poisson Regression Equations

Example (program and math final):

Variable	Category	Coefficient
Intercept		-4.88
Program	1	-0.37
	2	0.71
	3	0
MathFinal		0.07

$$\ln(\lambda) = -4.88 - 0.37 * \text{Program1} + 0.71 * \text{Program2} + 0.07 * \text{MathFinal}$$

Poisson Regression Equations

Math score of 60 in program 1:

$$\ln(\lambda) = -4.88 - 0.37 * \text{Program1} + 0.71 * \text{Program2} + 0.07 * \text{MathFinal}$$

- Someone in Program 1 with a math final score of 60 has ln(λ) of -4.88 – 0.370+ 0.07*60 = -1.05

Poisson Regression Equations

Math score of 60 in program 1:

$$\lambda = \exp(-1.05) = 0.35$$

- Someone in Program 1 with a math final score of 60 earns an average of 0.35 awards

Poisson Regression Equations

Marginal means in $\ln(\lambda)$ scale

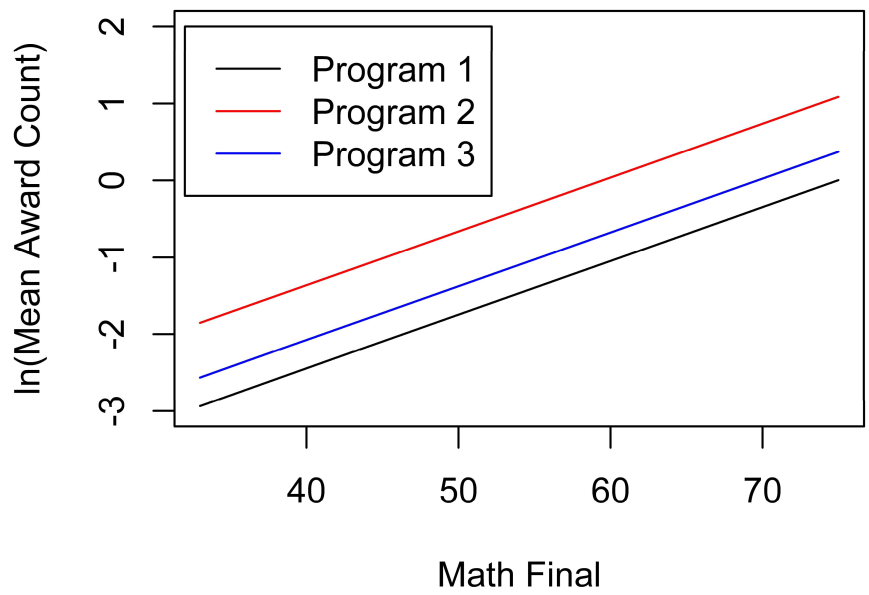
Math Final	Program	ln(λ)
60	1	-1.05
60	3	-0.68
40	1	-2.45
40	3	-2.08

Variable	Category	Coefficient
Intercept		-4.88
Program	1	-0.37
	2	0.71
	3	0
MathFinal		0.07

- Same group difference in $\ln(\lambda)$ for all math scores

Poisson Regression Equations

Same group difference in $\ln(\lambda)$ for all math scores



Poisson Regression Equations

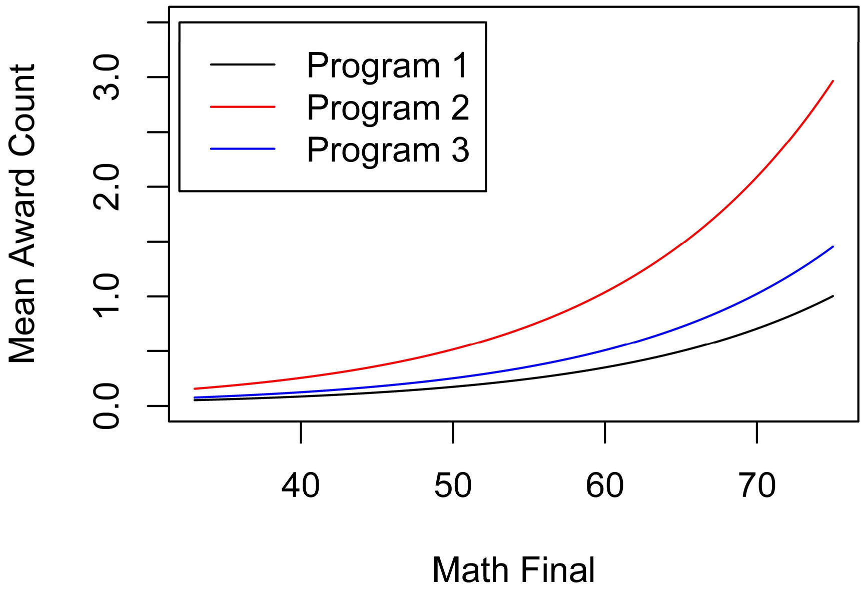
Compare program effect across math scores:

Math Final	Program	ln(λ)	λ
60	1	-1.05	0.35
60	3	-0.68	0.51
40	1	-2.45	0.09
40	3	-2.08	0.13

- *Not* same group difference in λ for all math scores

Poisson Regression Equations

Not same group difference in λ for all math scores



Poisson Regression Equations

How do we talk about this?

- Difference based on log mean counts
- Difference in log mean counts not really meaningful

Poisson Regression Equations

Discuss as *incidence rate ratios* (IRR)

Math Final	Program	ln(λ)	λ
60	1	-1.05	0.35
60	3	-0.68	0.51
40	1	-2.45	0.09
40	3	-2.08	0.13

- For math final of 60, $\exp(-1.05 - -0.68) = \frac{\exp(-1.05)}{\exp(-0.68)} = \frac{0.35}{0.51} = \frac{\lambda_1}{\lambda_3} = 0.69$
- For math final of 40, $\exp(-2.45 - -2.08) = \frac{\exp(-2.45)}{\exp(-2.08)} = \frac{0.09}{0.12} = \frac{\lambda_1}{\lambda_3} = 0.69$

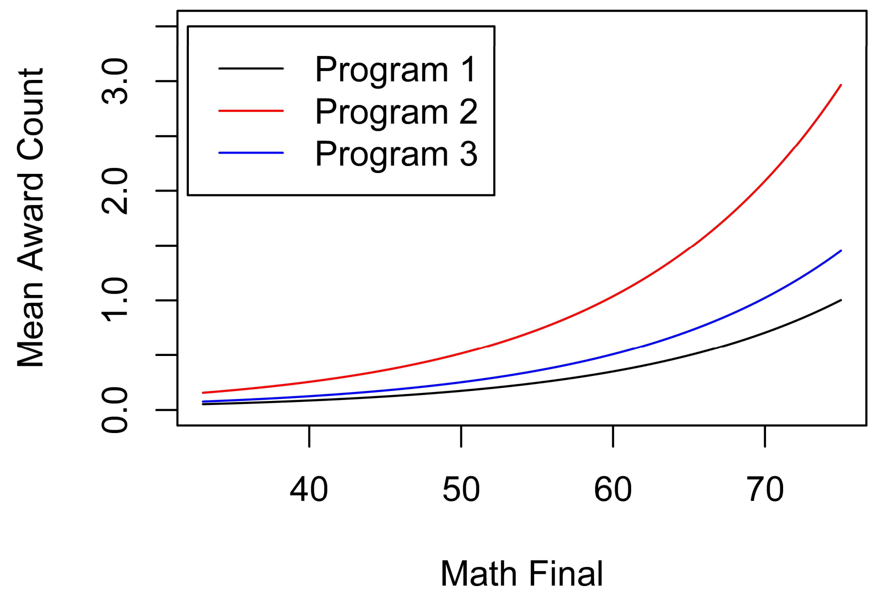
Poisson Regression Equations

More generally, exp(coeff) = IRR

Variable	Category	Coefficient	IRR
Intercept		-4.87	0.01
Program	1	-0.37	0.69
	2	0.71	2.04
	3	0	1
MathFinal		0.07	1.07

Poisson Regression Equations

Variable	Category	Coefficient	IRR
Intercept		-4.88	0.01
Program	1	-0.37	0.69
	2	0.71	2.04
	3	0	1
MathFinal		0.07	1.07



3. Interactions in Poisson Regression

Interactions in Poisson Regression

Now incorporate an interaction:

Variable	Category	Coefficient	IRR
(Intercept)		-4.71	0.01
Program	1	0.85	2.33
	2	0.40	1.50
	3	0	1
MathFinal		0.07	1.07
Program x MathFinal	1	-0.02	0.98
	2	0.01	1.01
	3	0	1

Interactions in Poisson Regression

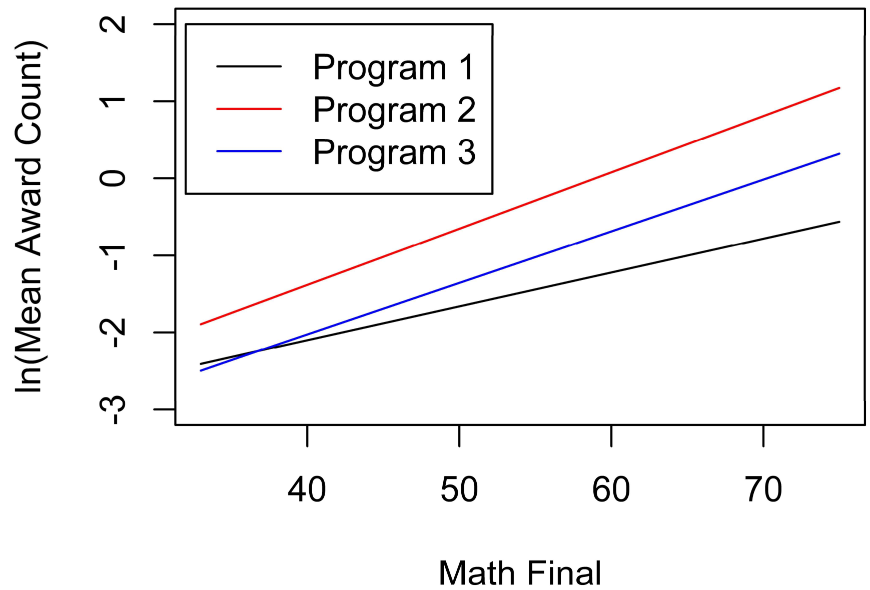
The equation is

$$\ln(\lambda) = -4.71 - 0.85 * \text{Program1} + 0.40 * \text{Program2} + 0.07 * \text{MathFinal} - 0.02 * \text{Program1} * \text{MathFinal} + 0.01 * \text{Program2} * \text{MathFinal}$$

- For program 1, $\ln(\lambda) = -4.71 - 0.85 + (0.07 - 0.02) * \text{MathFinal}$
- For program 2, $\ln(\lambda) = -4.71 + 0.40 + (0.07 + 0.01) * \text{MathFinal}$
- For program 3, $\ln(\lambda) = -4.71 + 0.07 * \text{MathFinal}$

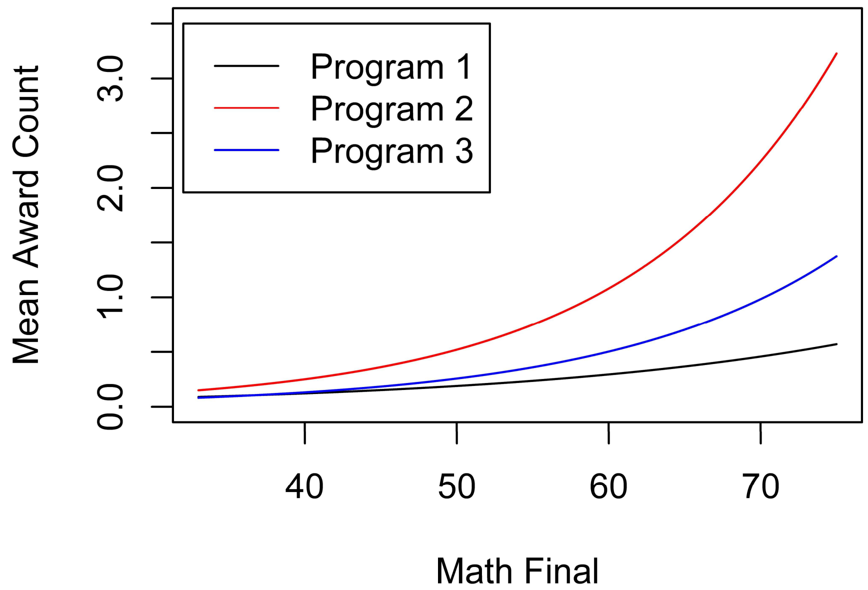
Interactions in Poisson Regression

Interaction in $\ln(\lambda)$ scale:



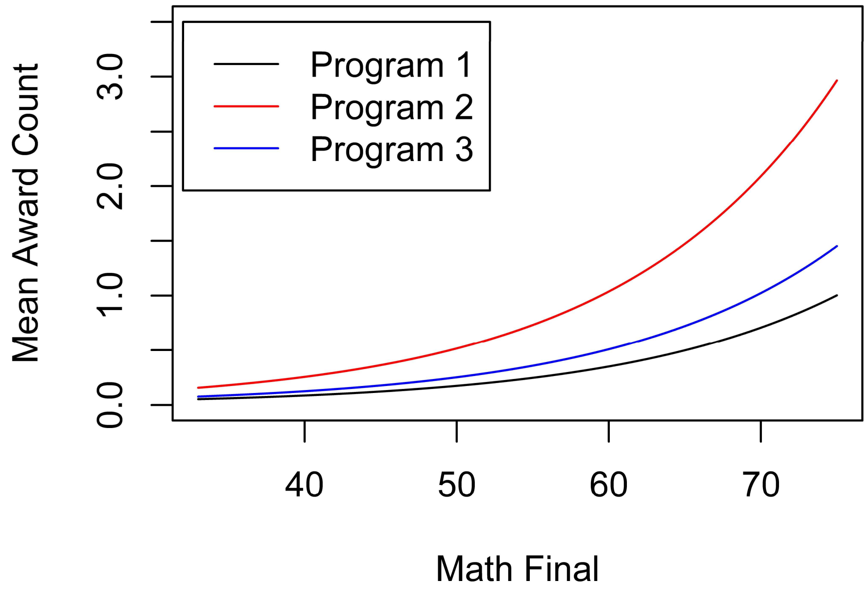
Interactions in Poisson Regression

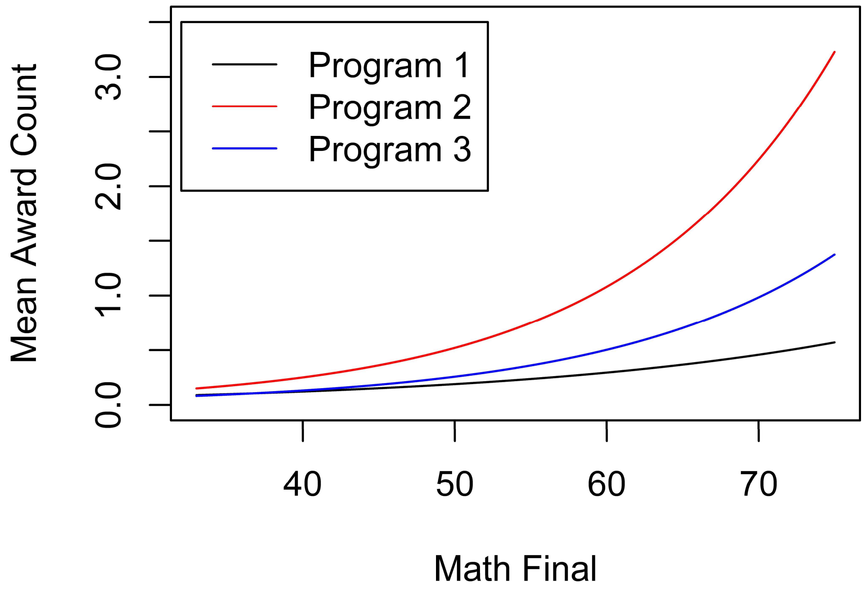
Interaction in λ scale:



Interactions in Poisson Regression

Interaction v. none in λ scale:





Interactions in Poisson Regression

How do we talk about this interaction?

- The interaction multiplies the IRR

Interactions in Poisson Regression

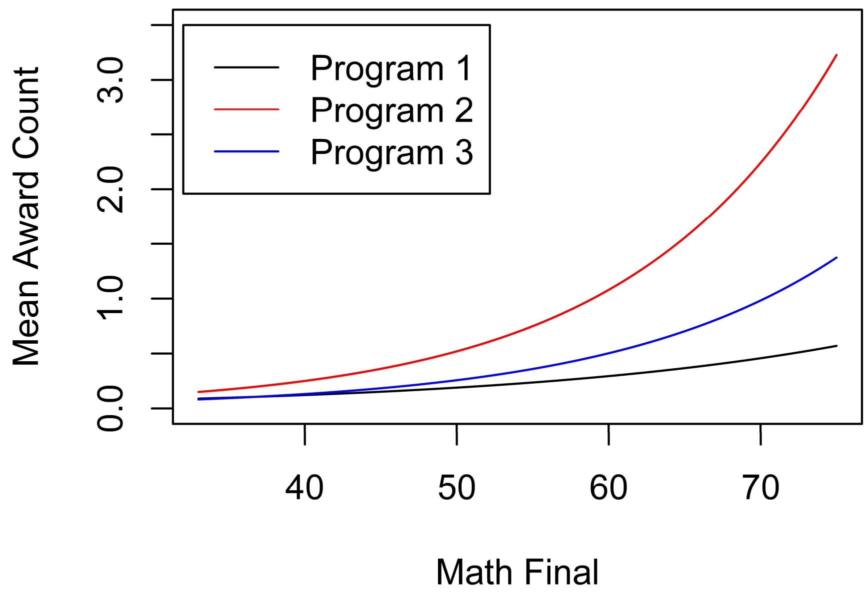
Variable	Category	Coefficient	IRR
(Intercept)		-4.71	0.01
Program	1	0.85	2.33
	2	0.40	1.50
	3	0	1
MathFinal		0.07	1.07
Program x MathFinal	1	-0.02	0.98
	2	0.01	1.01
	3	0	1

- For each additional point on the math final, the IRR of program 1 (versus 3) is multiplied by 0.98
- For program 1, the IRR of the math final is multiplied by 0.98 relative to program 3

Interactions in Poisson Regression

- For each additional point on the math final, the IRR of program 1 (versus 3) is multiplied by 0.98

- For program 1, the IRR of the math final is multiplied by 0.98 relative to program 3



4. Overview of Logistic Regression Equations

Logistic Regression Equations

Purpose:

- Used for binary outcomes
- Predict or estimate probability (or proportion) of “events”
- Determine *what* variables change this
- Determine *how* variables change this

Logistic Regression Equations

Example data set: Task Completion

Group	Age	Complete
1	25	1
1	22	1
1	20	0
1	18	1
1	24	1
1	25	1
…		

Logistic Regression Equations

General formula:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots + \beta_k x_k$$

- p is the *probability* of the event occurring
- *Link function* on the left side relates p to xs
  - This is a *logit link*
  - Logit is also called *log odds*

Logistic Regression Equations

Or, we can write

$$p = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \dots + \beta_k x_k\ ))}$$

Logistic Regression Equations

Example (group and age):

Variable	Category	Coefficient
Intercept		-7.42
Group	1	-0.10
	2	-2.10
	3	0
Age		0.48

$$\ln\left(\frac{p}{1-p}\right) = -7.42 - 0.10 * \text{Group1} - 2.10 * \text{Group2} + 0.48 * \text{Age}$$

Logistic Regression Equations

18-year-old member of group 2:

$$\ln\left(\frac{p}{1-p}\right) = -7.42 - 0.10 * \text{Group1} - 2.10 * \text{Group2} + 0.48 * \text{Age}$$

- An 18-year-old member of group 2 has logit -7.42 - 2.10 + 0.48*18 = -0.92

Logistic Regression Equations

18-year-old member of group 2:

$$p = \frac{1}{1 + \exp(-(-0.92))} = 0.285$$

- An 18-year-old member of group 2 has a 28.5% probability of completing the task

Logistic Regression Equations

Compare group effect across ages, logit scale:

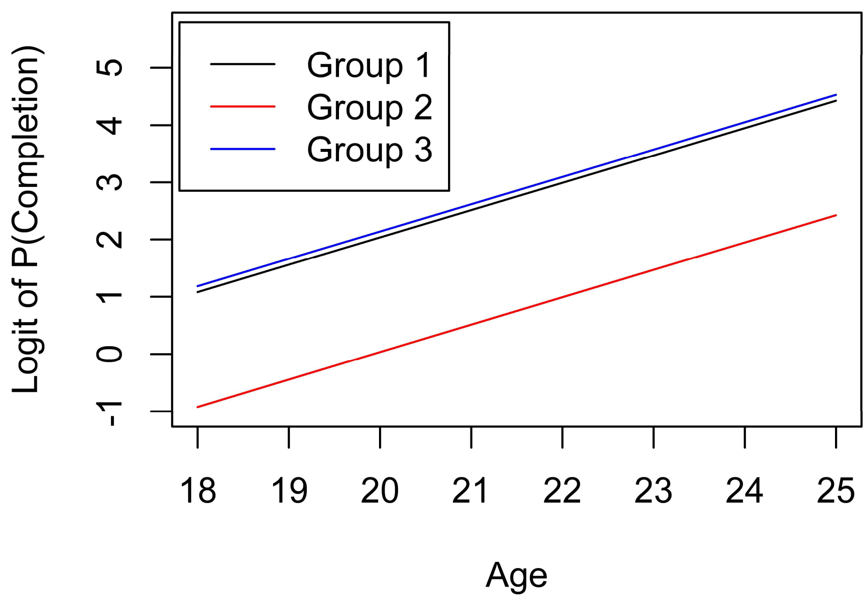
Age	Group	logit
18	2	-0.92
18	3	1.18
24	2	1.95
24	3	4.05

Variable	Category	Coefficient
Intercept		-7.42
Group	1	-0.10
	2	-2.10
	3	0
Age		0.48

- Same group difference in logit for all ages

Logistic Regression Equations

Logit same group difference for all ages



Logistic Regression Equations

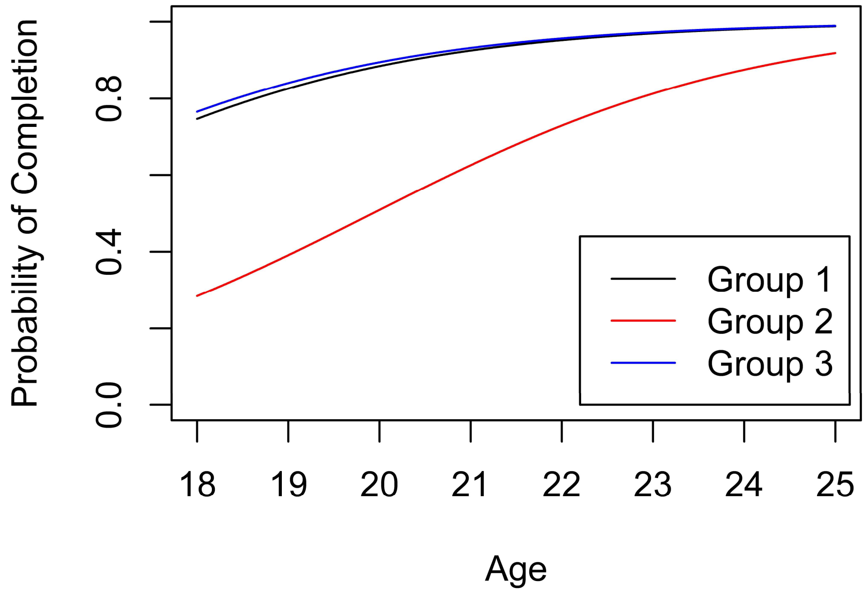
Compare group effect across ages, probability scale:

Age	Group	logit	probability
18	2	-0.92	0.285
18	3	1.18	0.766
24	2	1.95	0.875
24	3	4.05	0.983

- *Not* same group difference in probability for all ages

Logistic Regression Equations

Not same group difference in probability for all ages



Logistic Regression Equations

How do we talk about this?

- Difference in logit scale not really meaningful
- Can we talk about ratios of p? Like IRR?
  - Group 2 v Group 3, age 18: $\frac{0.285}{0.766} = 0.372$
  - Group 2 v Group 3, age 24: $\frac{0.875}{0.983} = 0.890$
- Hmm…

Logistic Regression Equations

Discuss as *odds ratios*

- Logit is $\ln\left(\frac{p}{1-p}\right)$

Age	Group	logit	probability	odds
18	2	-0.92	0.285	0.40
18	3	1.18	0.766	3.27
24	2	1.95	0.875	7.01
24	3	4.05	0.983	57.51

- For an 18-year-old, $\exp(-0.92 - 1.18) = \frac{\exp(-0.92)}{\exp(1.18)} = \frac{0.40}{3.27} = \frac{\frac{p_2}{1-p_2}}{\frac{p_3}{1-p_3}} = 0.122$
- For a 24-year-old, $\exp(1.95 - 4.05) = \frac{\exp(1.95)}{\exp(4.05)} = \frac{7.01}{57.51} = \frac{\frac{p_2}{1-p_2}}{\frac{p_3}{1-p_3}} = 0.122$

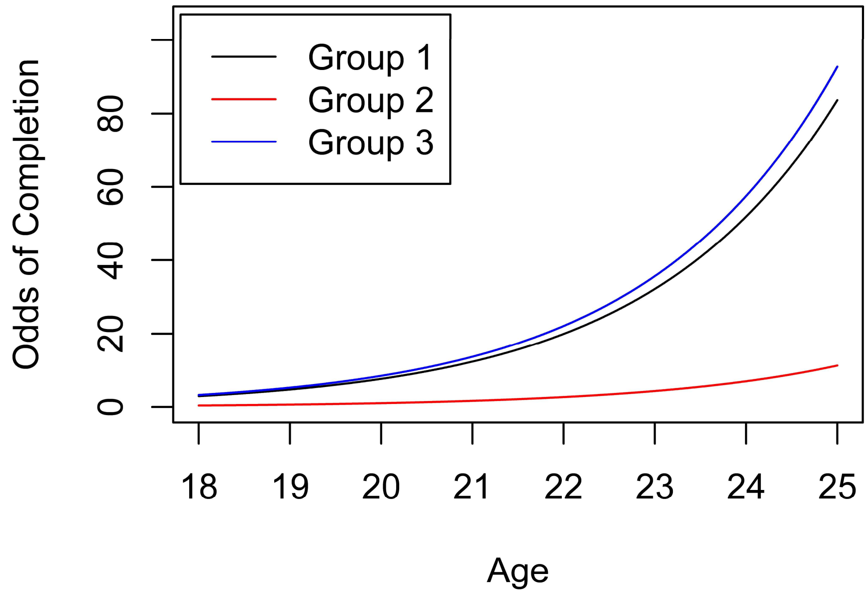
Logistic Regression Equations

More generally, exp(coeff) = OR

Variable	Category	Coefficient	OR
Intercept		-7.42	0.001
Group	1	-0.10	0.902
	2	-2.10	0.122
	3	0	1
Age		0.48	1.613

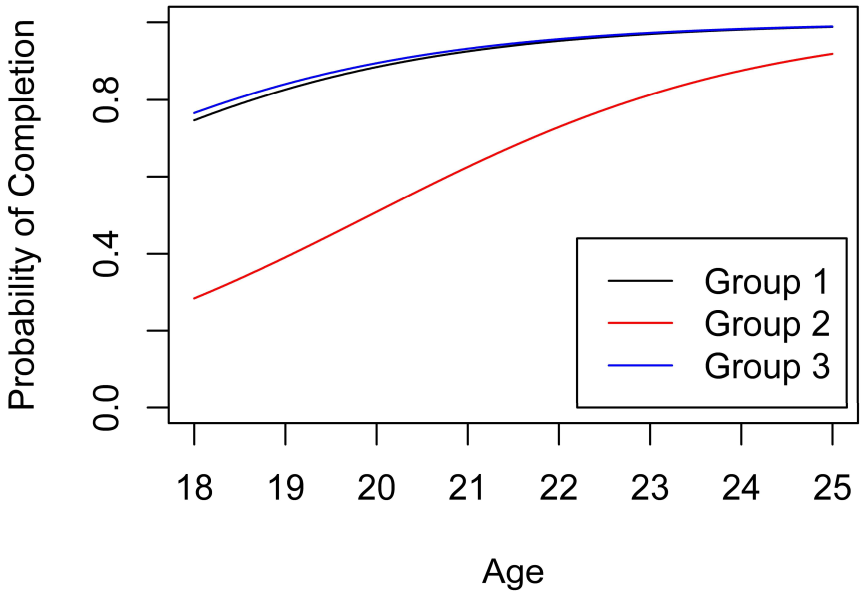
Logistic Regression Equations

Variable	Category	Coefficient	OR
Intercept		-7.42	0.001
Group	1	-0.10	0.902
	2	-2.10	0.122
	3	0	1
Age		0.48	1.613



Logistic Regression Equations

Variable	Category	Coefficient	OR
Intercept		-7.42	0.001
Group	1	-0.10	0.902
	2	-2.10	0.122
	3	0	1
Age		0.47	1.613



5. Interactions in Logistic Regression

Interactions in Logistic Regression

Now incorporate an interaction:

Variable	Category	Coefficient	OR
Intercept		-0.86	0.421
Group	1	-10.36	0.001
	2	-10.45	0.001
	3	0	1
Age		0.16	1.171
Group x Age	1	0.51	1.667
	2	0.42	1.501
	3	0	1

Interactions in Logistic Regression

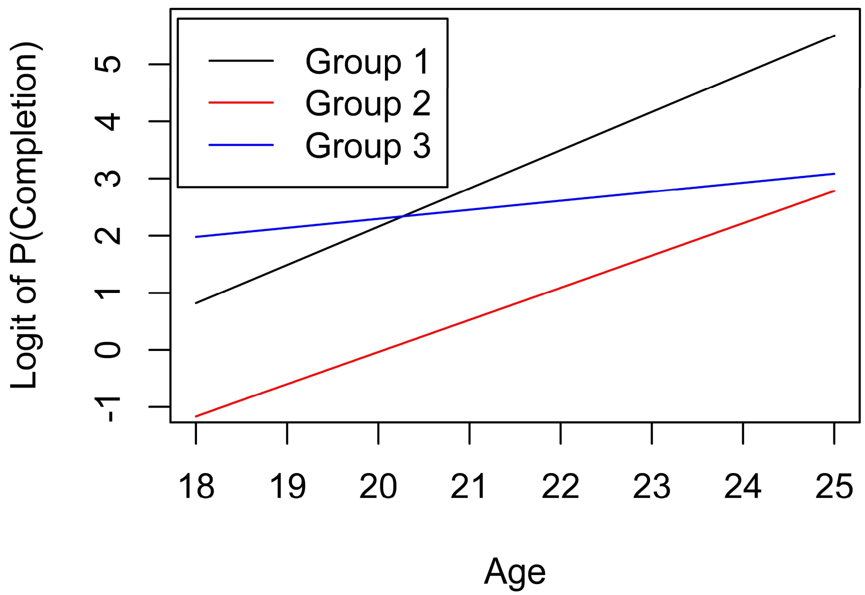
The equation is

$$\ln\left(\frac{p}{1-p}\right) = -0.86 - 10.36 * \text{Group1} - 10.45 * \text{Group2} + 0.16 * \text{Age} + 0.51 * \text{Group1} * \text{Age} + 0.41 * \text{Group2} * \text{Age}$$

- For group 1, $\ln\left(\frac{p}{1-p}\right) = -0.86 - 10.36 + (0.16 + 0.51) * \text{Age}$
- For group 2, $\ln\left(\frac{p}{1-p}\right) = -0.86 - 10.45 + (0.16 + 0.41) * \text{Age}$
- For group 3, $\ln\left(\frac{p}{1-p}\right) = -0.86 + 0.16 * \text{Age}$

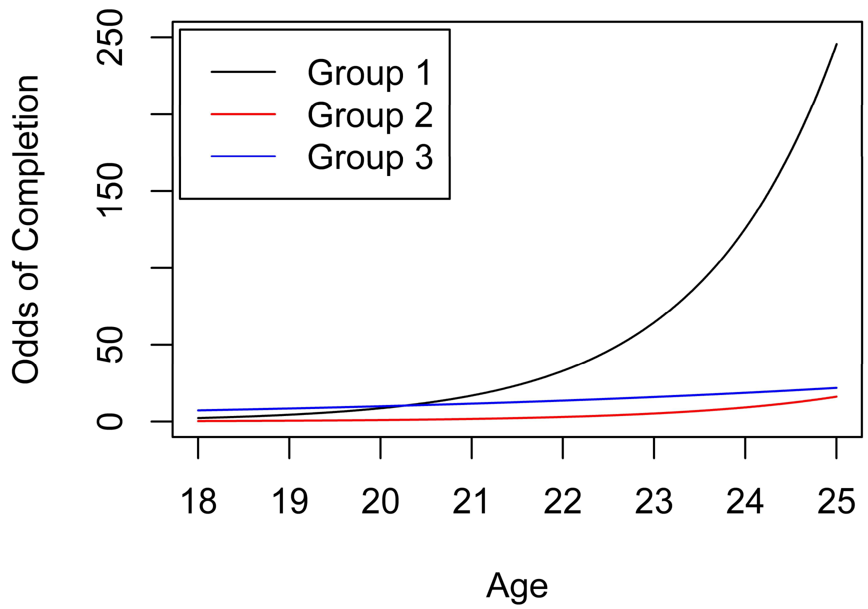
Interactions in Logistic Regression

Interaction in $\ln\left(\frac{p}{1-p}\right)$ (logit) scale:



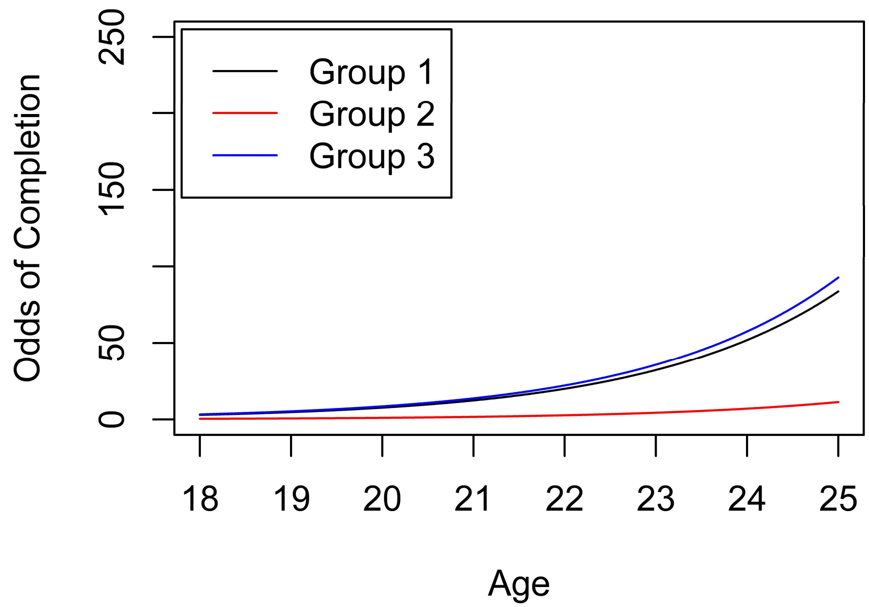
Interactions in Logistic Regression

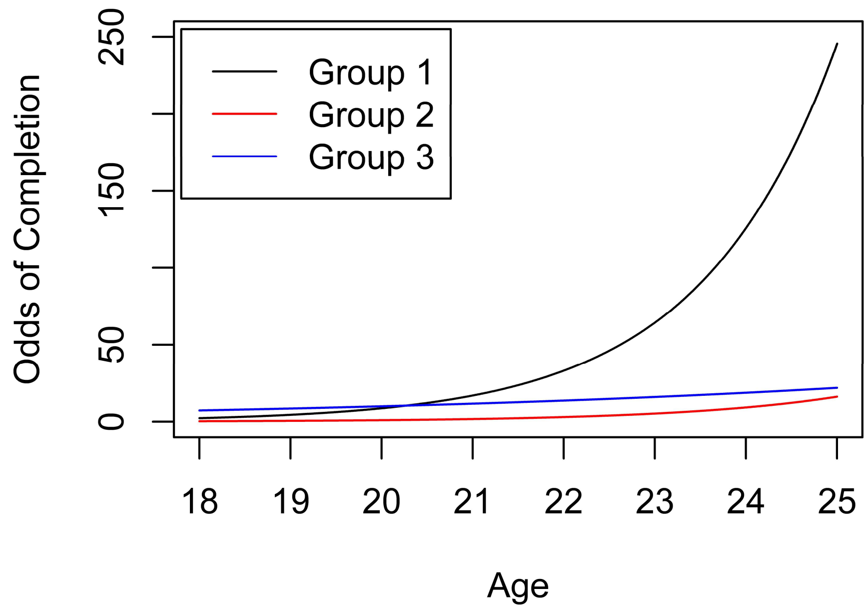
Interaction in $\frac{p}{1-p}$ (odds) scale:



Interactions in Logistic Regression

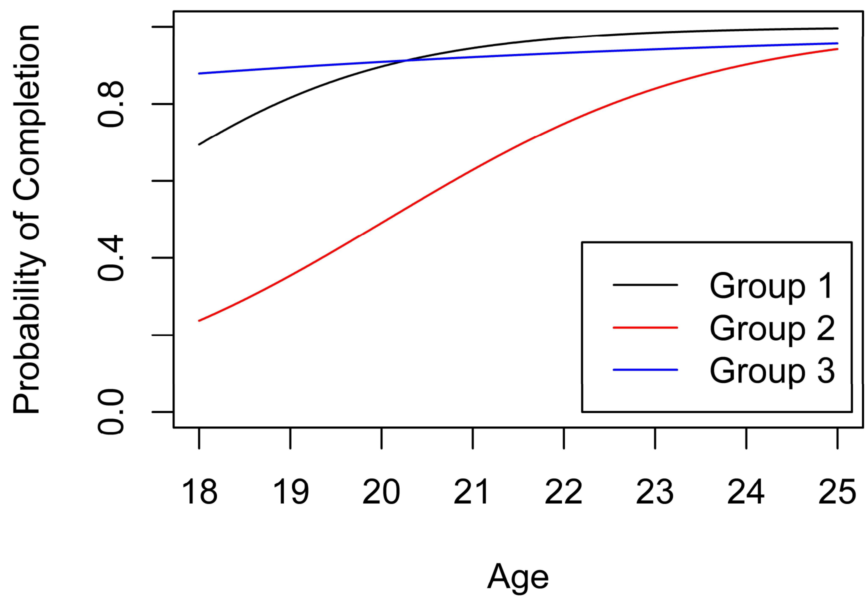
Interaction v. none in odds scale:





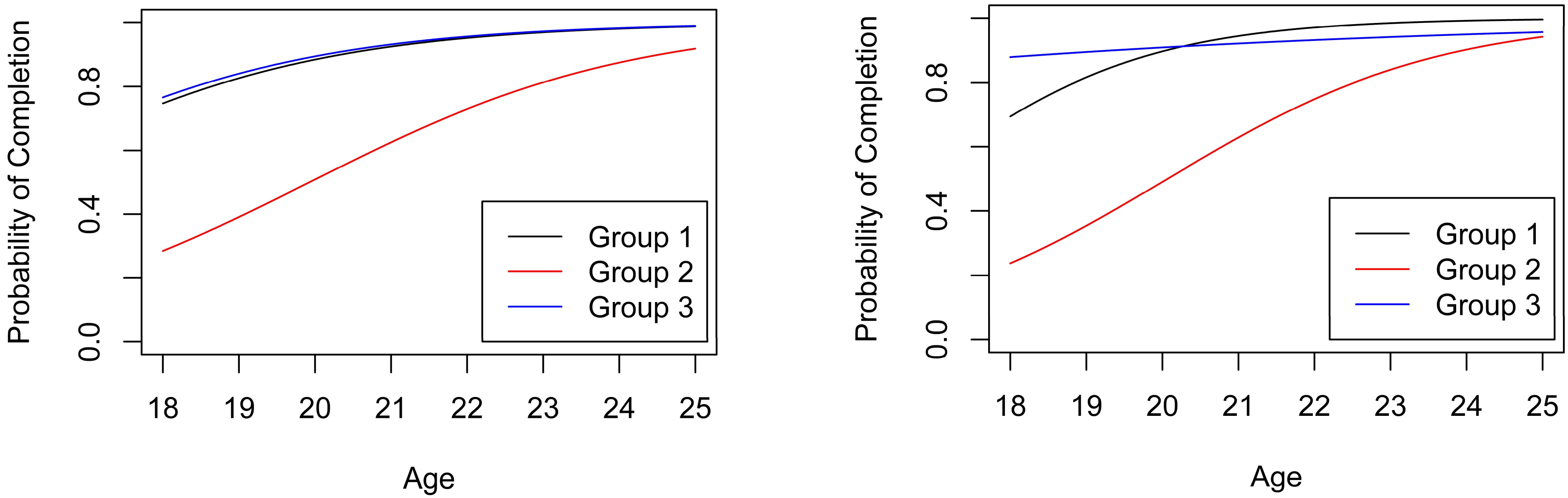
Interactions in Logistic Regression

Interaction in p (probability) scale:



Interactions in Logistic Regression

Interaction v. none in probability scale:



Interactions in Logistic Regression

How do we talk about this interaction?

- Like with IRR, the interaction multiplies the OR

Interactions in Logistic Regression

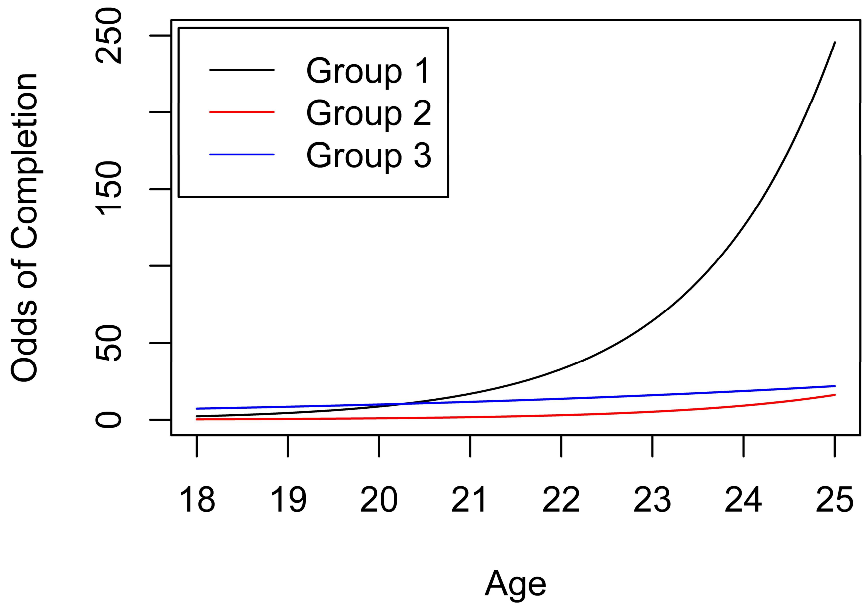
Variable	Category	Coefficient	OR
Intercept		-0.86	0.421
Group	1	-10.36	0.001
	2	-10.45	0.001
	3	0	1
Age		0.16	1.171
Group x Age	1	0.51	1.667
	2	0.41	1.501
	3	0	1

- For each additional year of age, the OR of program 2 (versus 3) is multiplied by 1.667
- For program 2, the OR of the math final is multiplied by 1.667 relative to program 3

Interactions in Logistic Regression

- For each additional year of age, the OR of program 2 (versus 3) is multiplied by 1.67

- For program 2, the OR of the math final is multiplied by 1.67 relative to program 3



6. Additional Comments

Poisson v Negative Binomial

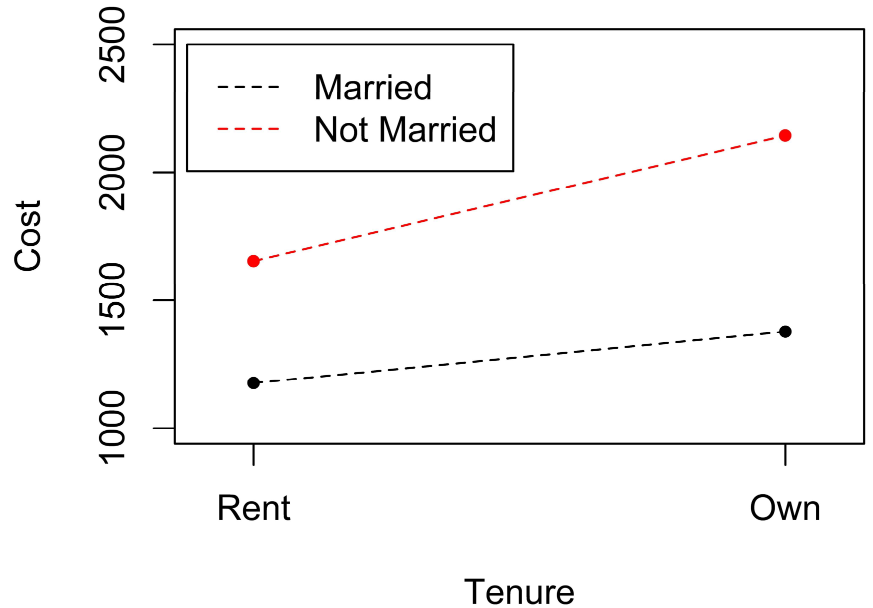
One more reminder:

- All Poisson information applies to negative binomial
  - Same link function
  - Different variability of outcome distribution (unimportant to discussion!)
- In general, interpretation applies based on link functions

Other Types of Interactions

This training focused on categorical by numeric interactions

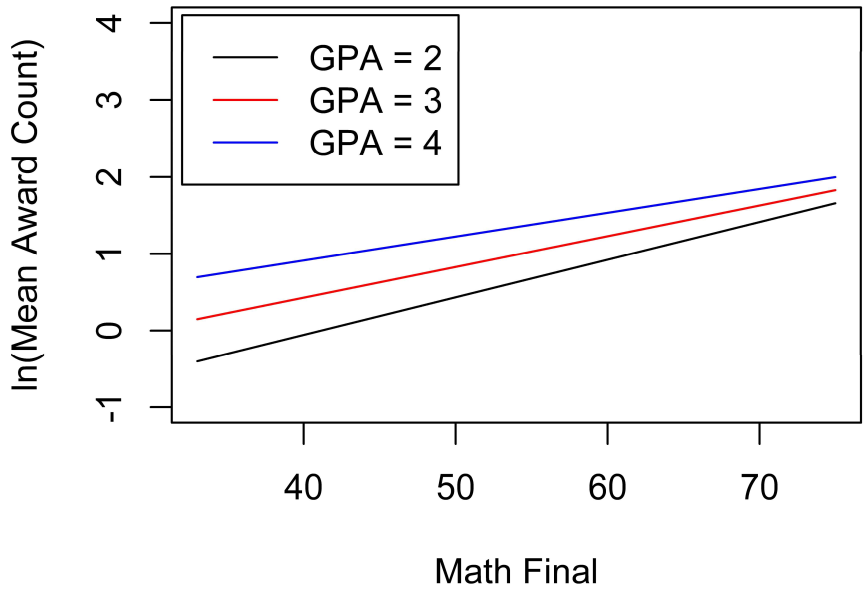
- Categorical by categorical
  - E.g., difference between owning and renting depends on marital status



Other Types of Interactions

This training focused on categorical by numeric interactions

- Numeric by numeric
  - E.g., effect of math final score depends on GPA



Questions?

Resources

Stats Amore webinars:

- https://programs.theanalysisfactor.com/statistically-speaking/trainings/statistical-contrasts/
- https://programs.theanalysisfactor.com/statistically-speaking/trainings/marginal-means-your-new-best-friend/
- https://programs.theanalysisfactor.com/statistically-speaking/trainings/dummy-and-effect-coding/

Resources

Stats Amore webinars, cont'd:

- https://programs.theanalysisfactor.com/statistically-speaking/trainings/interactions-in-anova-and-regression-models/
- https://programs.theanalysisfactor.com/statistically-speaking/trainings/interactions-in-anova-and-regression-models-part-2/
- https://programs.theanalysisfactor.com/statistically-speaking/trainings/explaining-logistic-regression-results-to-non-researchers/
- https://programs.theanalysisfactor.com/statistically-speaking/trainings/count-models/

Resources

Other resources:

- Article on Interpreting Linear Regression Coefficients: https://www.theanalysisfactor.com/interpreting-regression-coefficients/
- Craft of Statistical Analysis webinar on Interpreting Linear Regression Coefficients: https://thecraftofstatisticalanalysis.com/webinar-recording-signup/?cosid=303

References

The American Housing Survey:

https://www.census.gov/programs-surveys/ahs.html

Awards data set (simulated by UCLA):

https://stats.oarc.ucla.edu/r/dae/poisson-regression/