

Biostatistics Resources



Biostatistics Collaboration Center (BCC)

- Supports **non-cancer** research at NU
- Provides investigators an initial 1-2 hour consultation subsidized by the FSM Office of Research

Biostatistics Research Core (BRC)

- Supports **Lurie Children's Hospital** affiliates
- **Stanley Manne Research Institute** at Lurie Children's

Quantitative Data Sciences Core (QDSC)

- Supports all **cancer-related** research at NU
- Provides free support to all Cancer Center members subsidized by RHLCCC
- Grant

Northwestern University Data Analysis and Coordinating Center (NUDACC)

- Supports prospective, multicenter research
- Spans the full life cycle of research
- Grant

Contact Information

Non-cancer

Cancer

Lurie Children's

Data Coordinating

- Biostatistics Collaboration Center (BCC)
 - Website: <http://www.feinberg.northwestern.edu/sites/bcc/index.html>
 - Email: bcc@northwestern.edu
 - Phone: 312.503.2288
- Quantitative Data Sciences Core (QDSC)
 - Website: <https://www.cancer.northwestern.edu/research/shared-resources/quantitative-data-sciences.html>
 - Email: qdsc_rhlccc@northwestern.edu
 - Phone: 312.503.2288
- Biostatistics Research Core (BRC)
 - Website: <https://www.luriechildrens.org/en/research/research-areas/clinical-research/biostatistics-research-core/>
 - Email: merreed@luriechildrens.org
 - Phone: 773.755.6328
- Northwestern University Data Analysis and Coordinating Center (NUDACC)
 - Website: <https://www.feinberg.northwestern.edu/sites/nudacc/>
 - Email: nudacc@northwestern.edu



Outline

- Basics
- Common Mistakes
- Related Topics
- Alternatives

The Basics

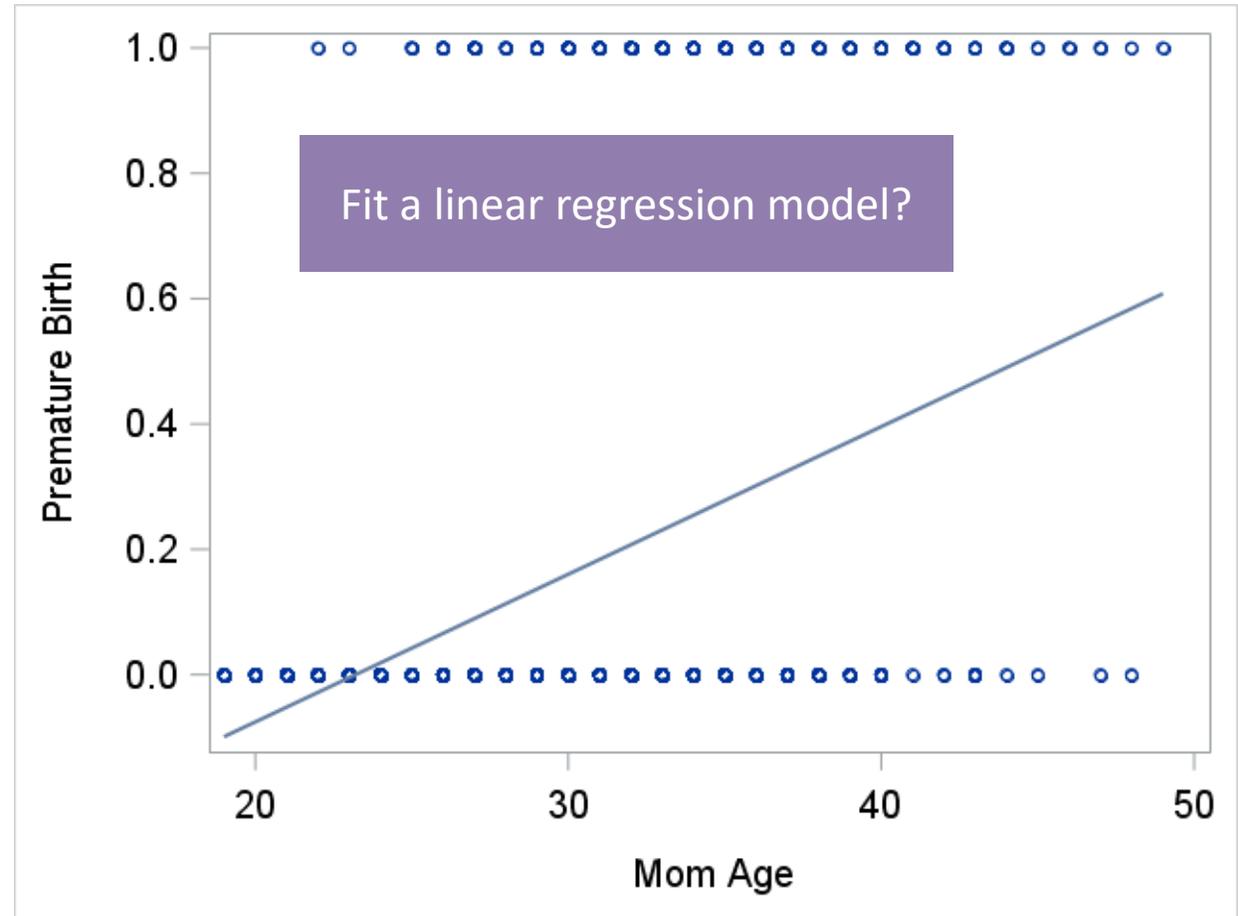
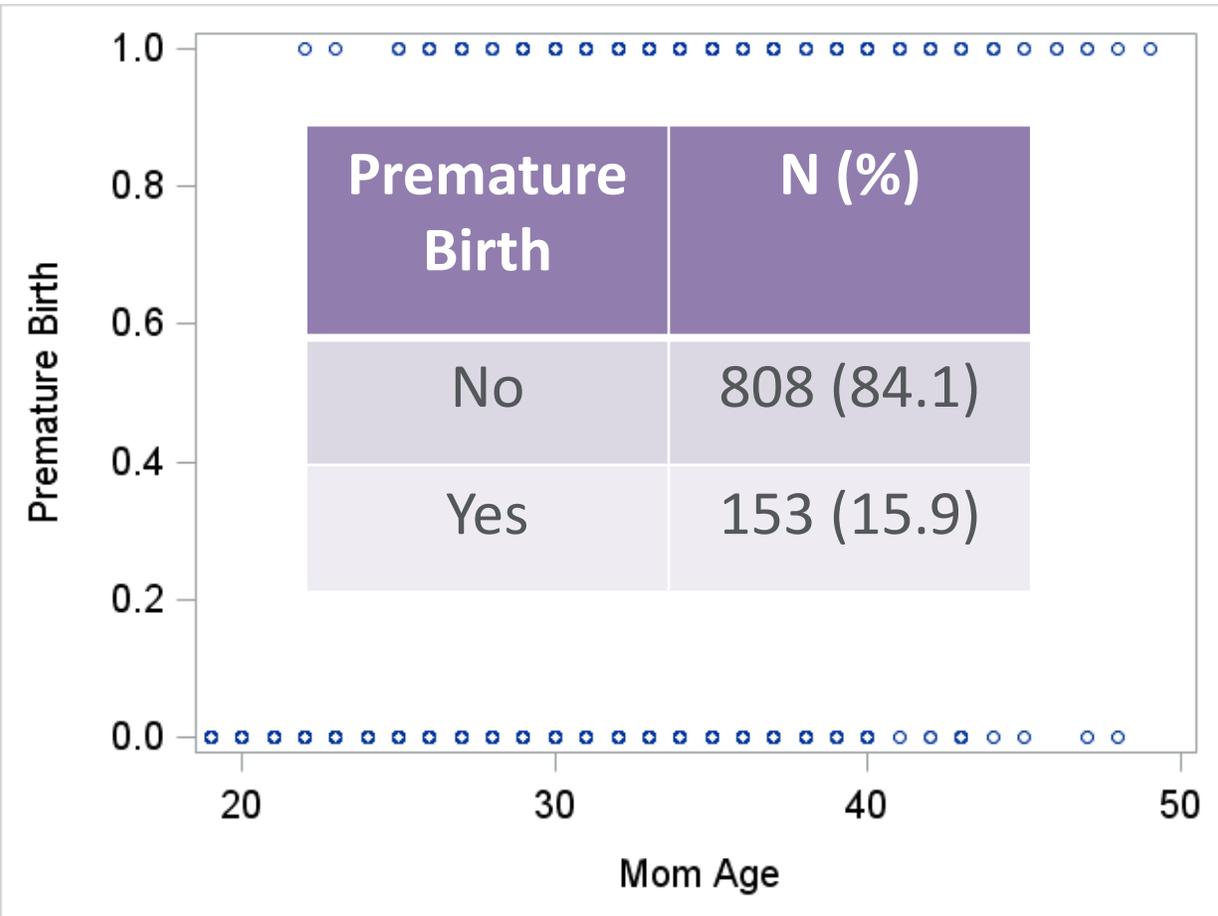
When do we use logistic regression?

- Binary (class) outcome
- One or more independent variables
 - Categorical or continuous

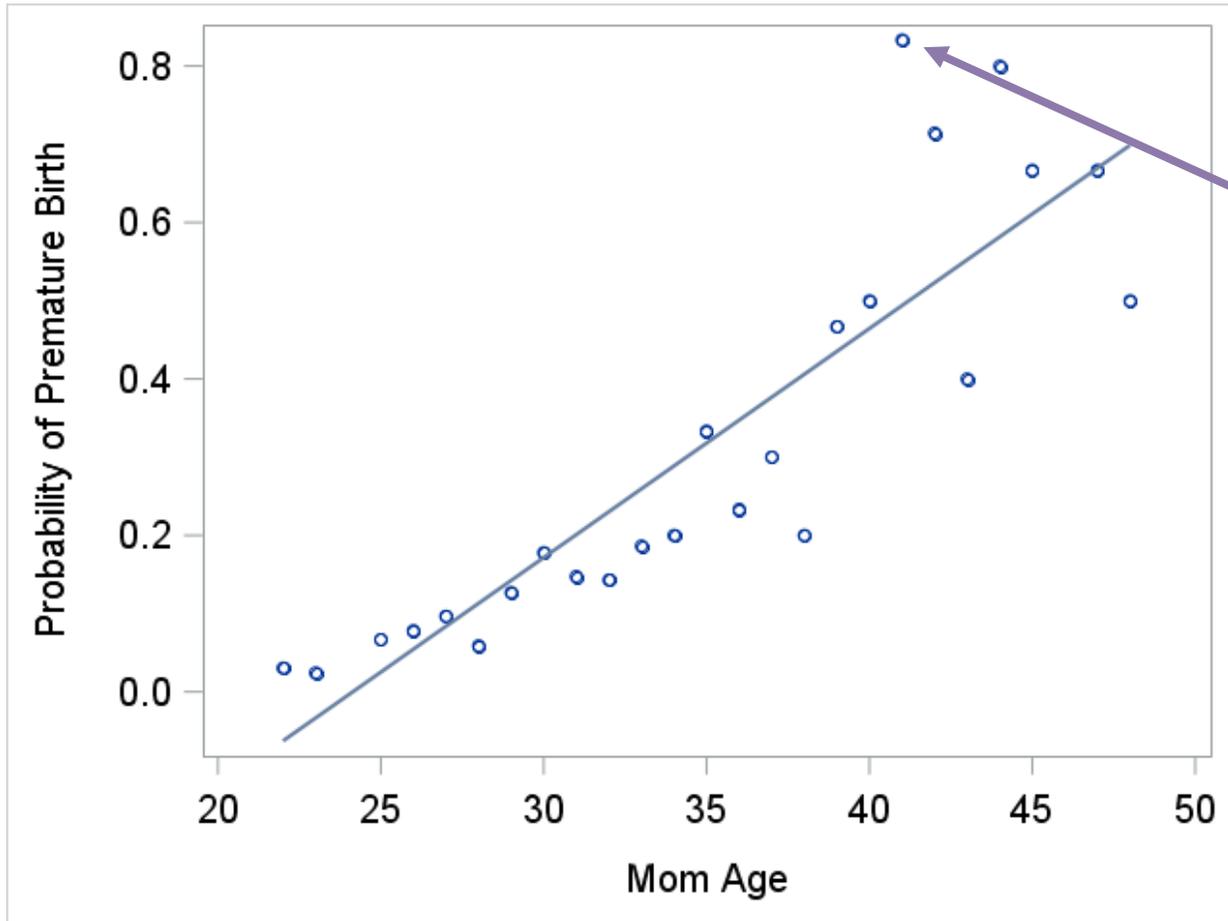
Is smoking associated with premature birth?

Is maternal pre-pregnancy BMI associated with childhood obesity?

Why the “logit”?



Why the “logit”?

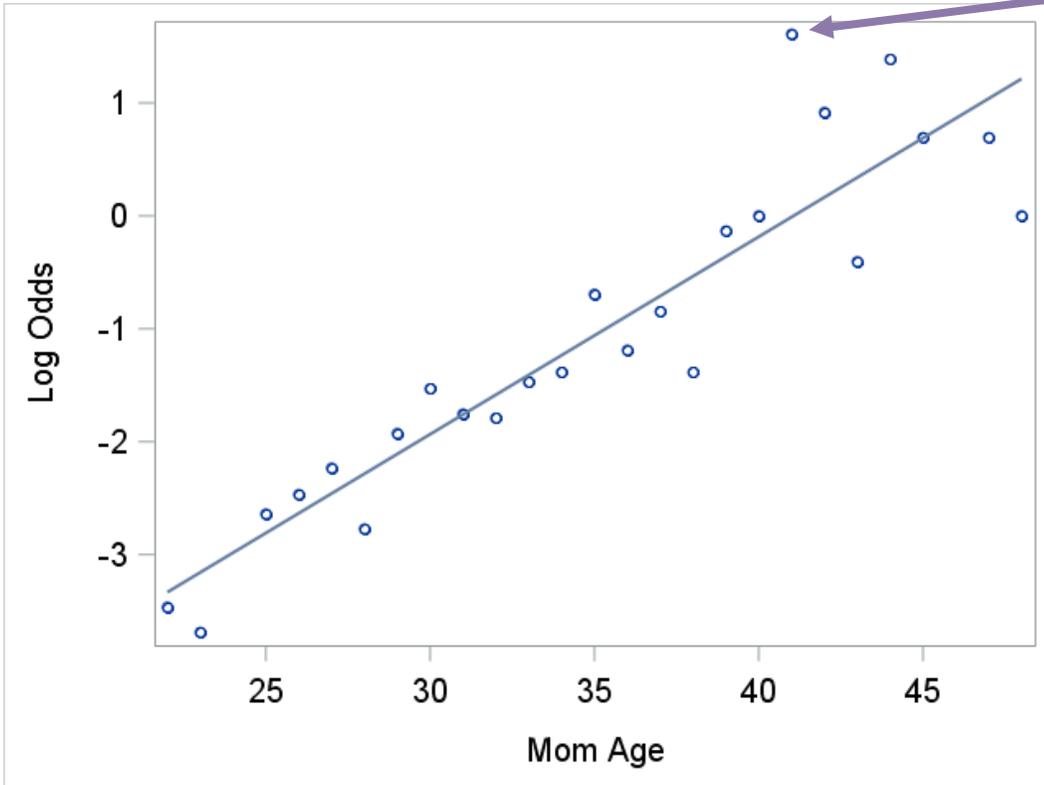


	Premature	Not Premature
41 Years	5	1

Probability = 5/6

Why the “logit”?

	Premature	Not Premature
41 Years	5	1



Odds = 5/1



Odds of rolling a 2 = 1/5 = 0.2

Probability of rolling a 2 = 1/6 = 0.17

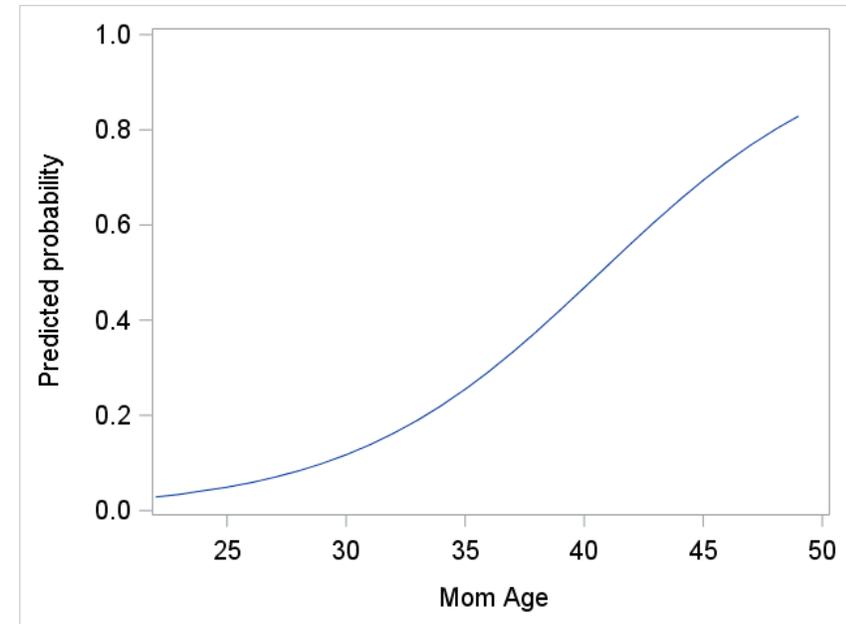
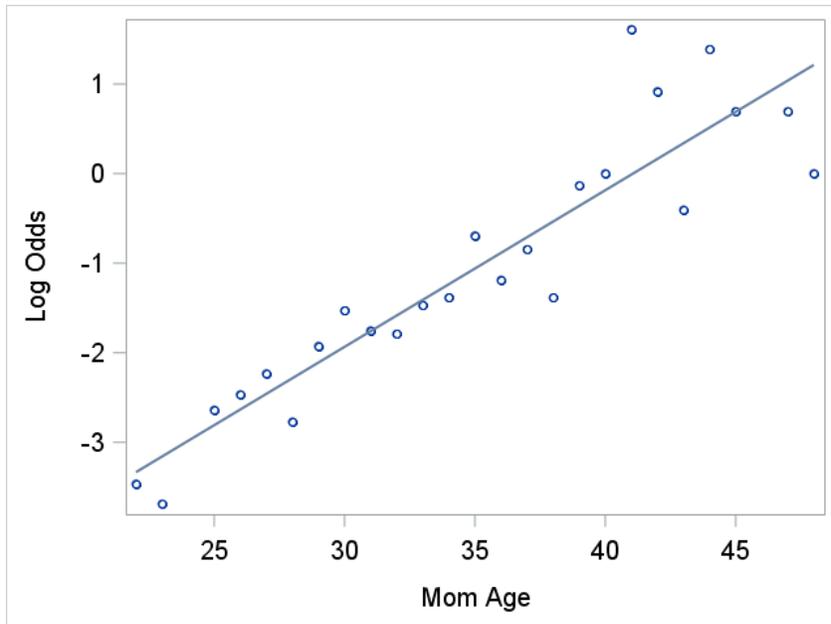
Why the “logit”?

Binary
Outcome



Independent
Variables

Logit



Guarantees the probability is bounded by 0 and 1

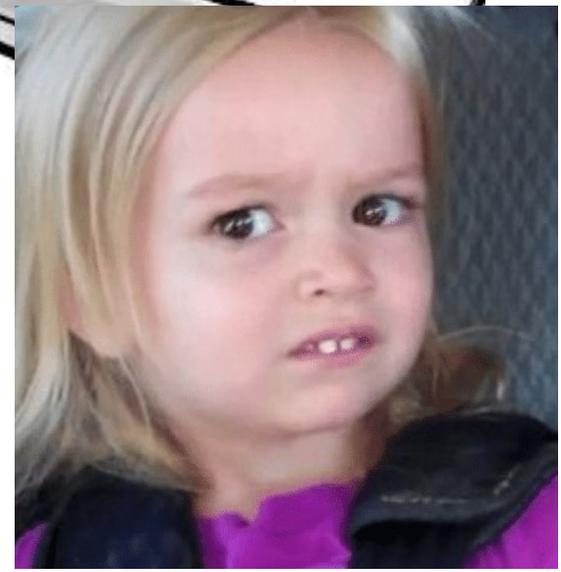
How do we interpret?

Data

ID	Outcome	Age
1	0	30
2	0	35
3	1	40



Parameter	Estimate	P-value
Intercept	-7.6984	<.0001
Mom Age	0.1893	<.0001



A one year increase in mom age is associated with an increase in the log odds of premature birth of 0.19

How do we interpret?

Log Odds

Parameter	Estimate	P-value
Intercept	-7.6984	<.0001
Mom Age	0.1893	<.0001

Exponentiate



Odds Scale

$$\text{Exp}(0.19)=1.21$$

$$\frac{\text{Odds of Premature Birth}_{\text{maternal age}=X+1}}{\text{Odds of Premature Birth}_{\text{maternal age}=X}}$$

A one year increase in mom age is associated with a 21% increase in the odds of premature birth.

How do we interpret?

- Depends on the reference category/level for the predictor
- Depends on the modeling of the outcome

	Premature Birth	
	Yes	No
Smoker	80	135
Non-Smoker	73	673

$$\text{OR Premature Birth (smoker/non-smoker)} = \frac{80/135}{73/673} = 5.46$$

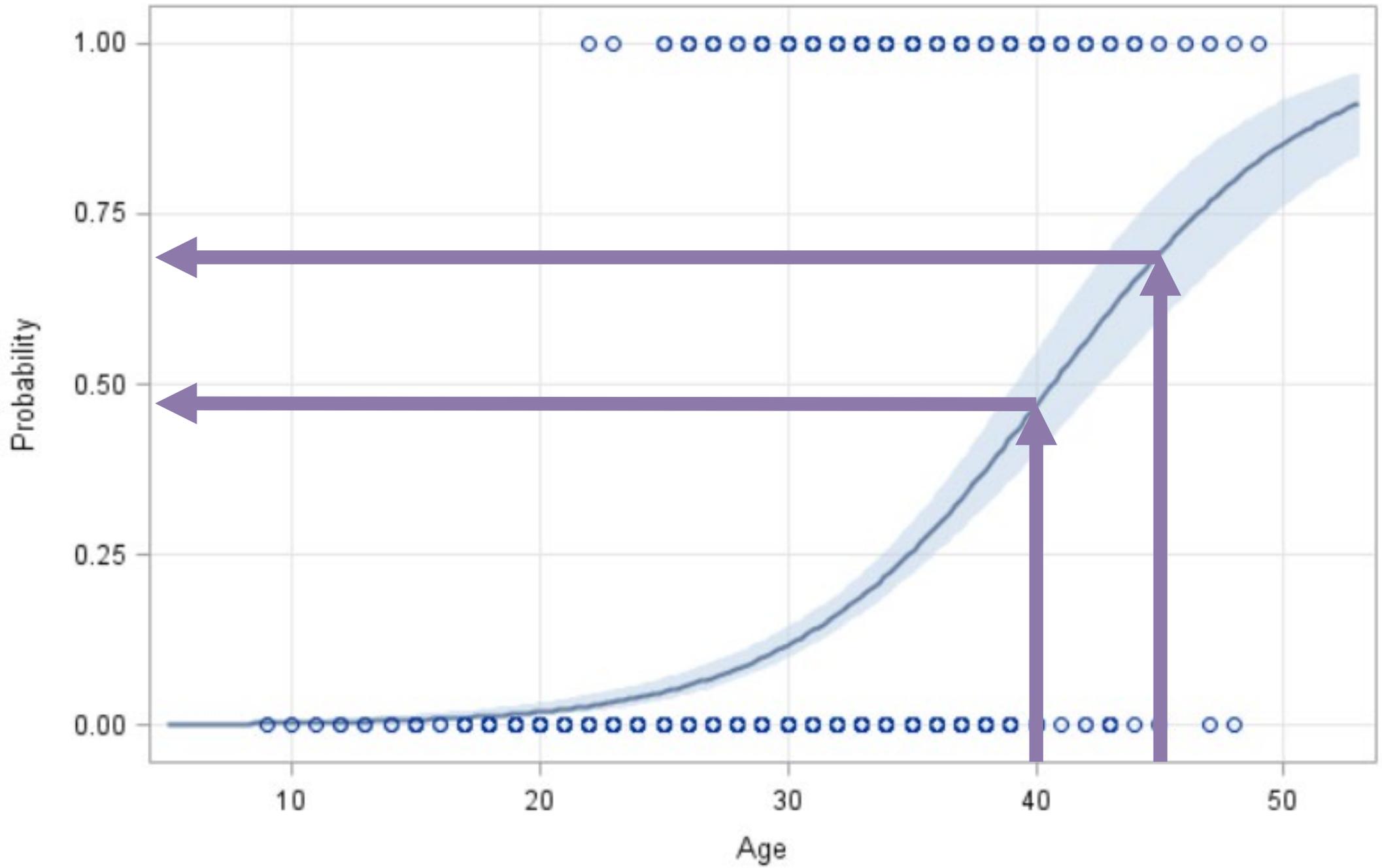
$$\text{OR Premature Birth (non-smoker/smoker)} = \frac{73/673}{80/135} = 0.18$$

$$\text{OR Not Premature Birth (non-smoker/smoker)} = \frac{673/73}{135/80} = 5.46$$

$$\text{OR Not Premature Birth (smoker/non-smoker)} = \frac{135/80}{673/73} = 0.18$$

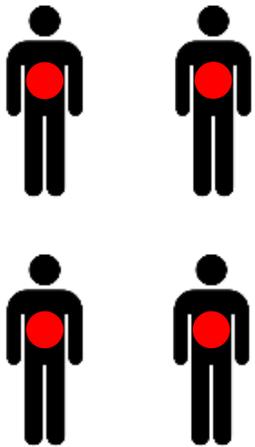
Increased odds

Decreased odds

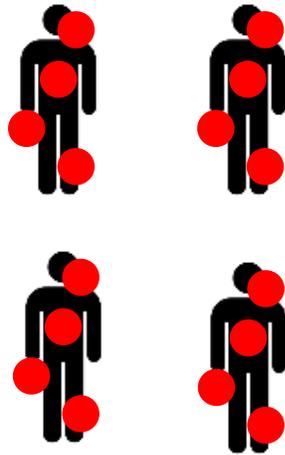


Model Assumptions

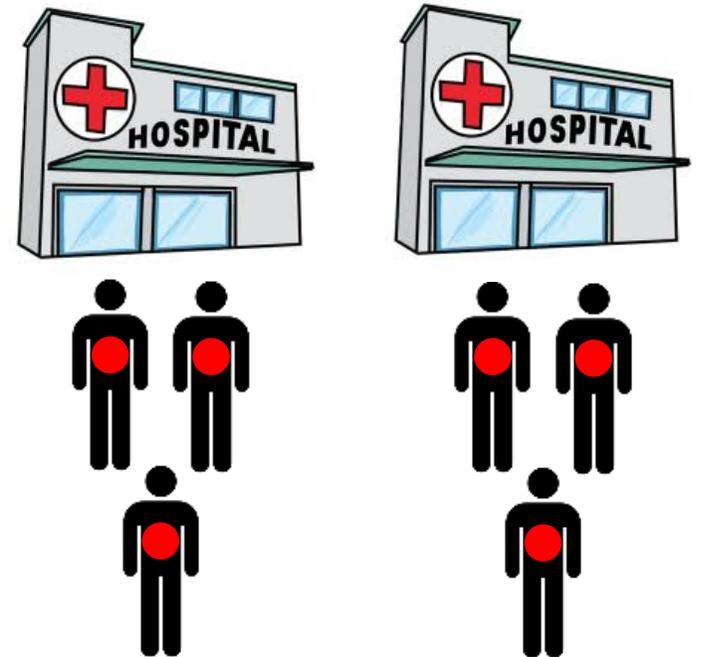
Independence



Independent observations



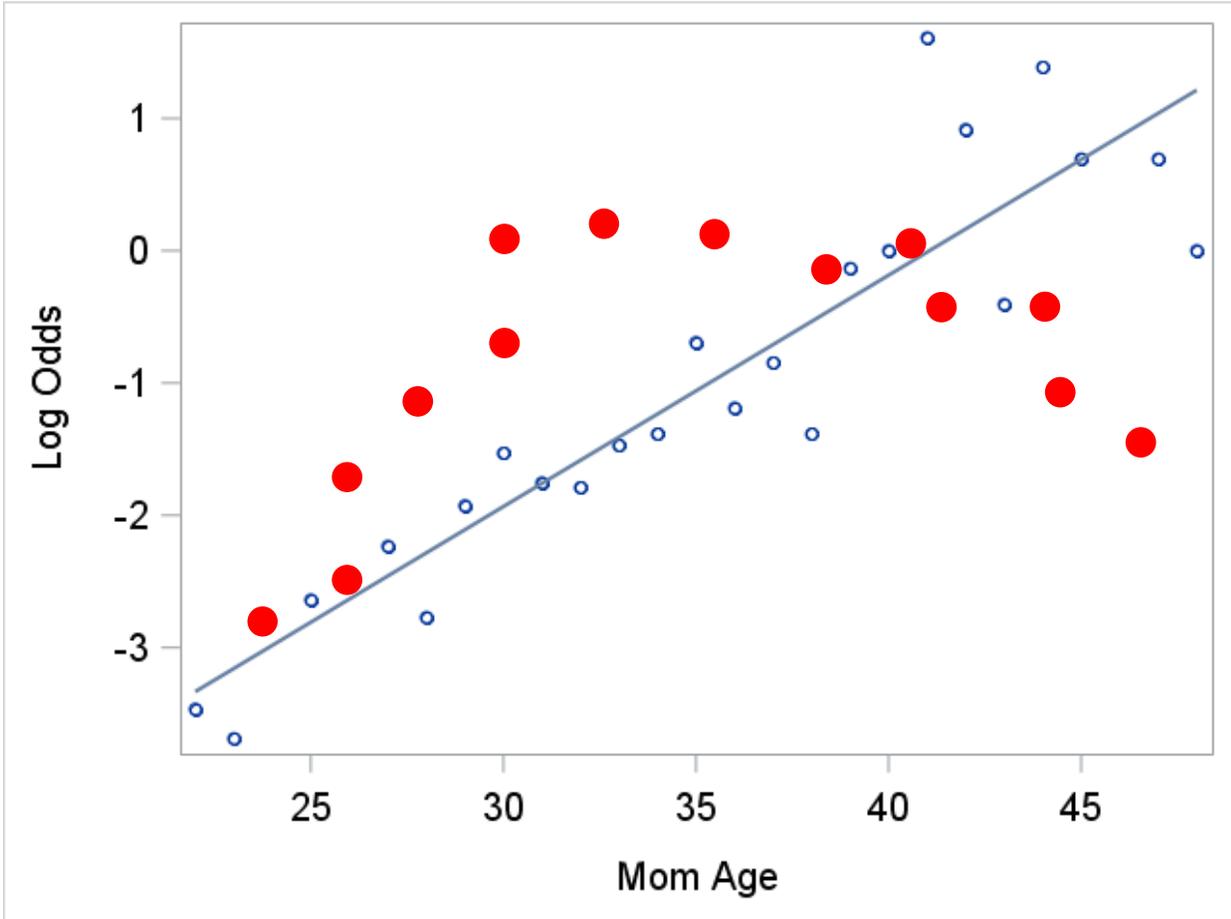
Repeated Measures



Clustered Observations

Model Assumptions

Linearity



Model Assumptions

Absence of multicollinearity

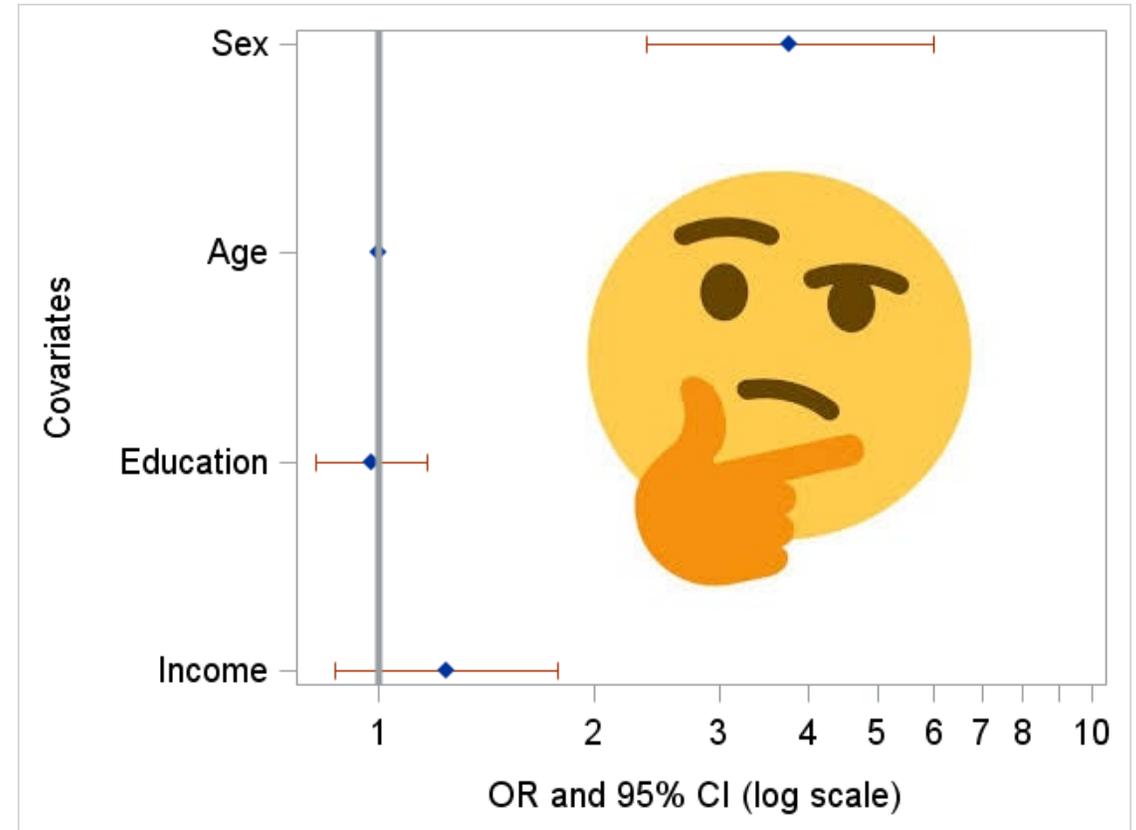
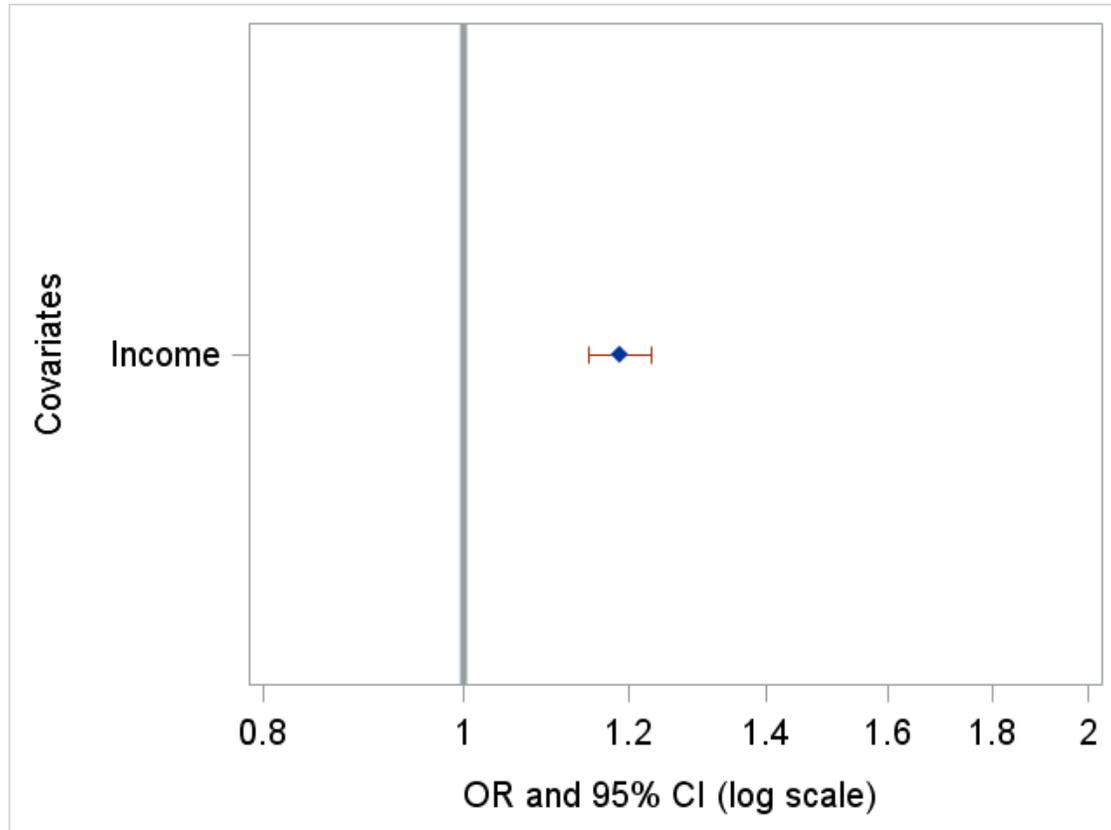
Depression

Income

Sex

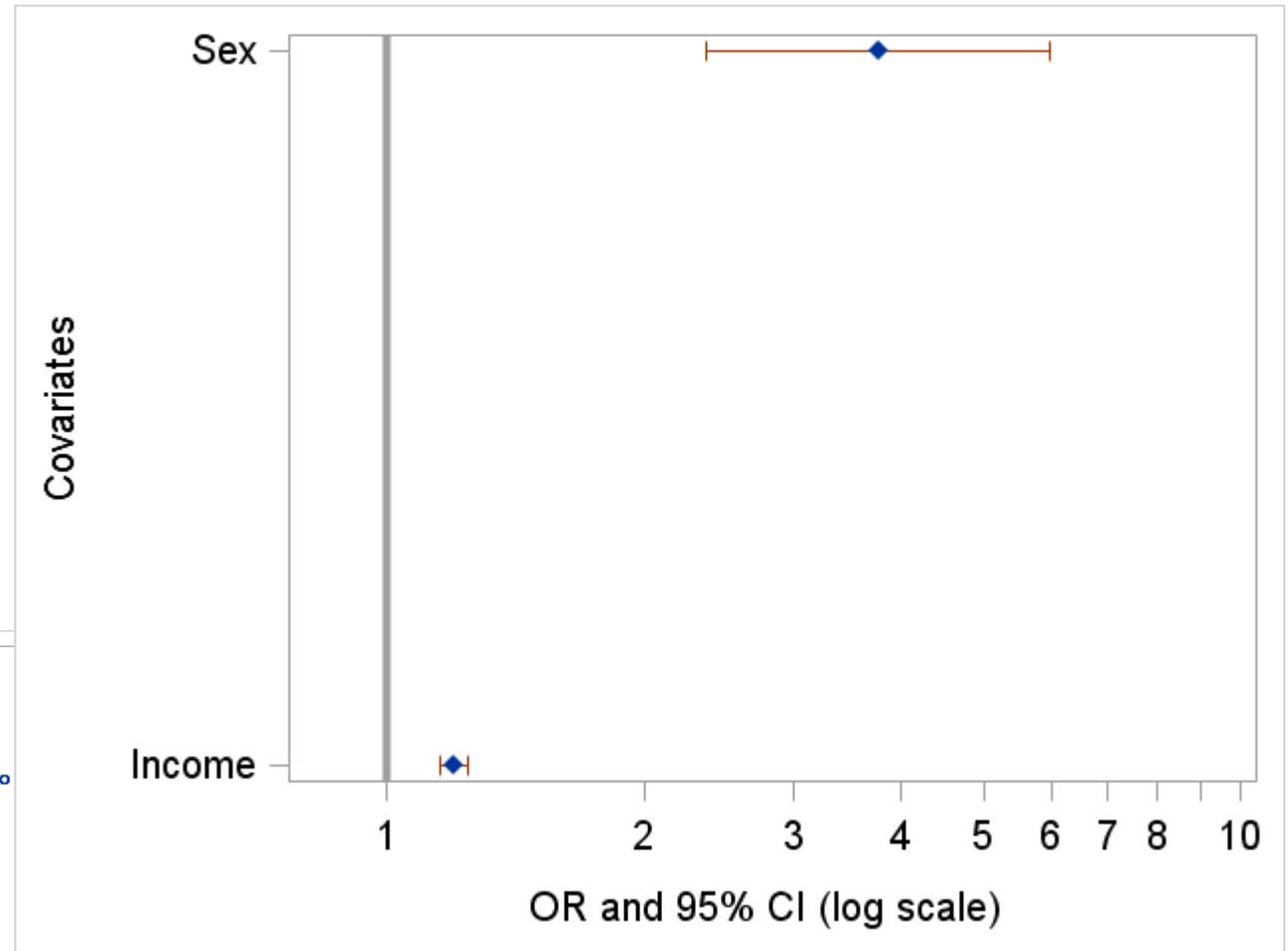
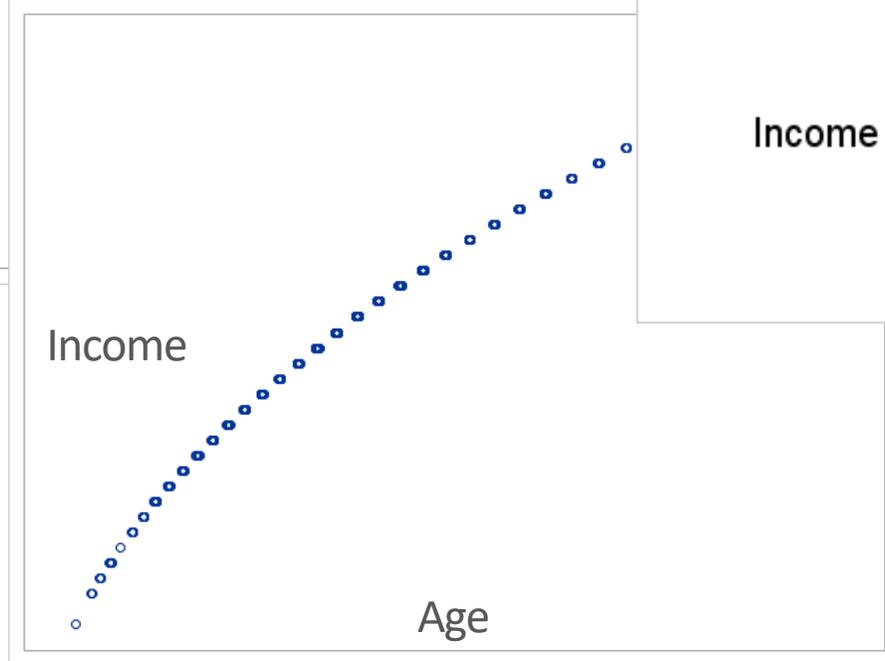
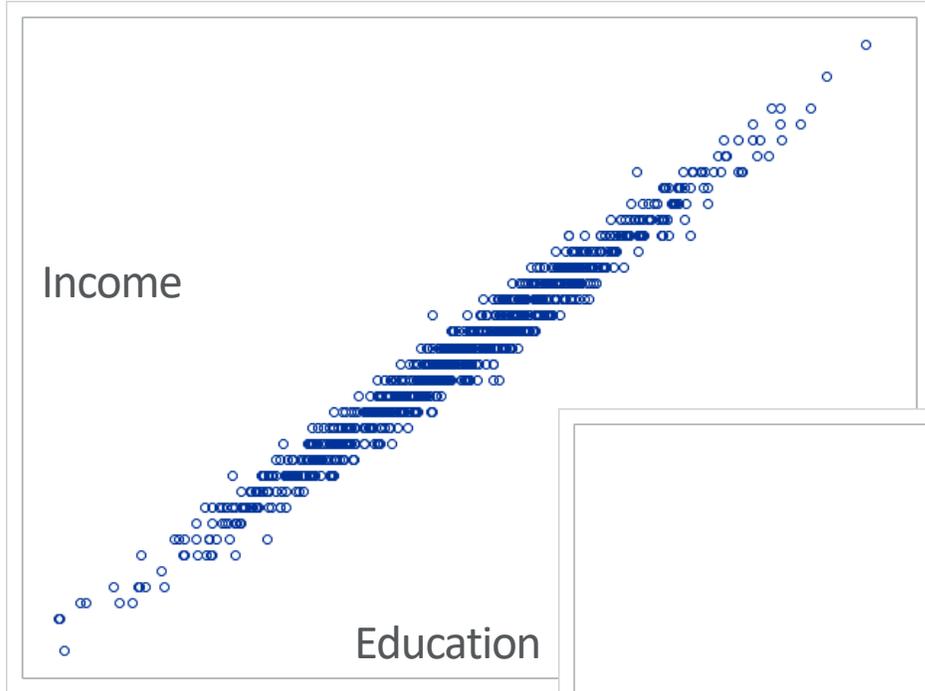
Education Level

Age



Model Assumptions

Absence of multicollinearity



Common Mistakes

Common Mistake #1: Interpreting Odds Ratio as 'Risk'

- Odds ratios \neq relative risk (risk ratio)
 - Except for rare outcomes (ORs approximate RRs)
 - When outcome is common, $OR > RR$
- Odds ratio is ratio of two odds



Odds of rolling a 2 = $1/5 = 0.2$
Risk of rolling a 2 = $1/6 = 0.17$

$$\frac{\text{Odds of Premature Birth}_{\text{maternal age}=X+1}}{\text{Odds of Premature Birth}_{\text{maternal age}=X}}$$

- Relative Risk is ratio of two probabilities
 - Risk of an event is the probability of an event

Common Mistake #1: Interpreting Odds Ratio as 'Risk'

- Hypothetical trial (Ranganathan et al. 2015)

	Death	Survival	Total
Treatment A	18	46	64
Treatment B	29	36	65
Total	47	82	129

RISK

Treatment A	0.28
Treatment B	0.44
Ratio	$0.28/0.44 = 0.64$

Example from Ranganathan et al. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640017/>

Common Mistake #1: Interpreting Odds Ratio as 'Risk'

- Hypothetical trial (Ranganathan et al. 2015)

	Death	Survival	Total
Treatment A	18	46	64
Treatment B	29	36	65
Total	47	82	129

RISK

Treatment A	0.28
Treatment B	0.44
Ratio	$0.28/0.44 = 0.64$

ODDS

Treatment A	0.39
Treatment B	0.81
Ratio	$0.39/0.81 = 0.48$

Example from Ranganathan et al. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640017/>

Common Mistake #1: Interpreting Odds Ratio as 'Risk'

- Hypothetical trial (Ranganathan et al. 2015)

	Death	Survival	Total
Treatment A	5	59	64
Treatment B	9	56	65
Total	14	115	129

RISK

Treatment A	0.08
Treatment B	0.14
Ratio	$0.08/0.14 = 0.56$

ODDS

Treatment A	0.08
Treatment B	0.16
Ratio	$0.08/0.16 = 0.53$

Common Mistake #1: Interpreting Odds Ratio as 'Risk'

- Hypothetical trial (Ranganathan et al. 2015)

	Death	Survival	Total
Treatment A	1	63	64
Treatment B	3	62	65
Total	4	125	129

RISK

Treatment A	0.02
Treatment B	0.05
Ratio	$0.02/0.05 = 0.34$

ODDS

Treatment A	0.02
Treatment B	0.05
Ratio	$0.02/0.05 = 0.33$

Common Mistake #2: Confusing Odds and Probabilities

- Odds \neq Probability
- A large odds ratio can represent a small probability and vice-versa

5 year survival

Treatment

→ OR = 5.2

Wow! There's a really big difference in the probability of 5 year survival between the treatment and control

	Death	Survival
Treatment	132	68
Control	182	18

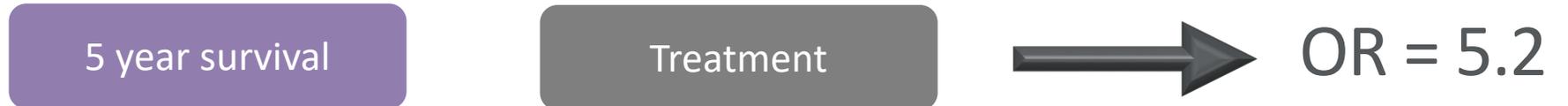
→ $0.34 - 0.09 = \mathbf{0.25}$

	Death	Survival
Treatment	950	50
Control	990	10

→ $0.05 - 0.04 = \mathbf{0.01}$

Common Mistake #2: Confusing Odds and Probabilities

- Odds \neq Probability
- A large odds ratio can represent a small probability and vice-versa



$$\frac{68/132}{18/182} = 5.2$$

	Death	Survival
Treatment	132	68
Control	182	18

$$0.34 - 0.09 = 0.25$$

$$\frac{50/950}{10/990} = 5.2$$

	Death	Survival
Treatment	950	50
Control	990	10

$$0.05 - 0.04 = 0.01$$

Common Mistake #2: Confusing Odds and Probabilities

- **Odds ratio** represents the constant effect of a predictor
- The effect of the predictor on the **probability** of the outcome depends on the value of the predictor



1.07

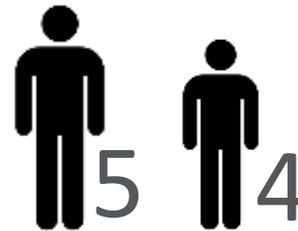


1.07

Odds Ratio

Common Mistake #2: Confusing Odds and Probabilities

- Odds ratio represents the constant effect of a predictor
- The effect of the predictor on the probability of the outcome depends on the value of the predictor



Odds Ratio

1.07

1.07

Predicted Probability

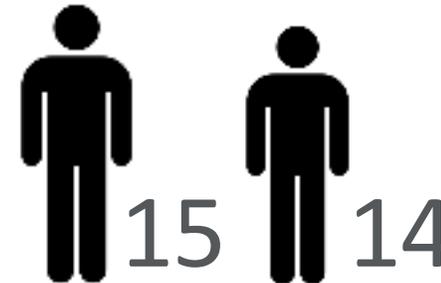
0.108 vs 0.102

Difference in Probability

0.006

Common Mistake #2: Confusing Odds and Probabilities

- Odds ratio represents the constant effect of a predictor
- The effect of the predictor on the probability of the outcome depends on the value of the predictor



Odds Ratio	1.07	1.07
Predicted Probability	0.108 vs 0.102	0.195 vs 0.185
Difference in Probability	0.006	0.01

Common Mistake #3: Confusing Marginal and Conditional Odds Ratios

- Marginal (“Average”) OR
 - Average exposure/treatment effect on the whole population
 - “Population-average” OR

Odds of 5 year survival

Treatment

- Conditional (“Subject-specific”) OR
 - Exposure/treatment effect for particular characteristics
 - “Adjusted” OR

Odds of 5 year survival

Treatment

Sex

Common Mistake #3: Confusing Marginal and Conditional Odds Ratios

	Failure	Success
Control	110	90
Treatment	90	110

OR Treatment vs Control = 1.49

Males (n=200)		
	Failure	Success
Control	80	20
Treatment	70	30

OR Treatment vs Control = 1.71

Females (n=200)		
	Failure	Success
Control	30	70
Treatment	20	80

OR Treatment vs Control = 1.71

Common Mistake #3: Confusing Marginal and Conditional Odds Ratios

- Marginal OR does not always equal Conditional OR, even in absence of confounding
- Marginal OR
 - Unadjusted/simple logistic regression model
 - Randomized trial
 - Propensity score methods
 - Average effect
- Conditional OR
 - Multivariable logistic regression model
 - Customizing treatment for a specific individual with specific characteristics

Common Mistake #4: Interpreting main effects in presence of interaction

EX: Interested in the effect of a treatment on the odds of 5 year survival. Investigator believes the treatment effect will be different for males and females.

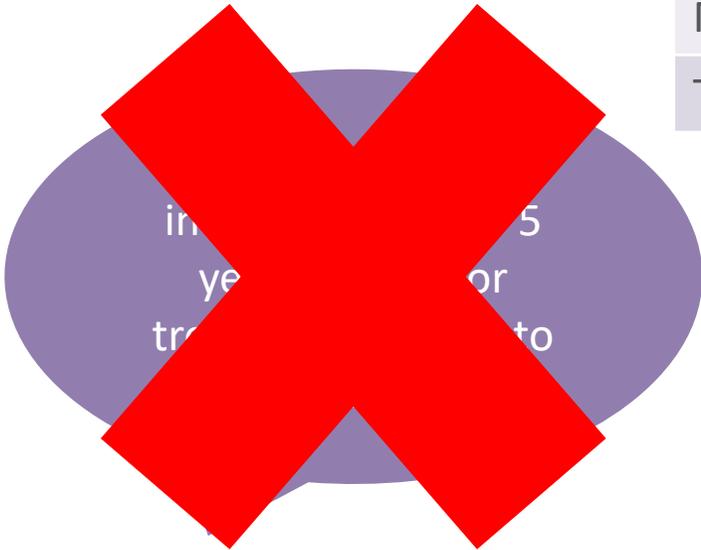
	Odds Ratio (95% CI)
Treatment	1.65 (1.35, 1.95)
Male	1.22 (1.10, 1.34)
Treatment*Male	1.11 (1.07, 1.15)

There is a 65% increased odds of 5 year survival for treated compared to controls

Common Mistake #4: Interpreting main effects in presence of interaction

EX: Interested in the effect of a treatment on the odds of 5 year survival. Investigator believes the treatment effect will be different for males and females.

	Odds Ratio (95% CI)
Treatment	1.65 (1.35, 1.95)
Male	1.22 (1.10, 1.34)
Treatment*Male	1.11 (1.07, 1.15)



Common Mistake #4: Interpreting main effects in presence of interaction

EX: Interested in the effect of a treatment on the odds of 5 year survival. Investigator believes the treatment effect will be different for males and females.

	Odds Ratio (95% CI)
Treatment	1.65 (1.35, 1.95)
Male	1.22 (1.10, 1.34)
Treatment*Male	1.11 (1.07, 1.15)

- For females, there is a 65% increased odds of 5 year survival in the treatment vs control
- For males, there is an 82% increased odds of 5 year survival in the treatment vs control

Common Mistake #5: Overfitting a model with a small sample size

Myocardial
Infarction

BMI Category

Smoking Status

	No MI (n=31)	MI (n=10)
BMI		
Normal	16	1
Overweight	9	2
Obese	6	7
Smoke		
Non-smoker	21	2
Smoker	10	8

Common Mistake #5: Overfitting a model with a small sample size

Effect	OR	95% CI	
Overweight vs Normal	3.000	0.199	45.242
Obese vs Normal	>999	<0.001	>999.999
Smoker vs Non-Smoker	>999	<0.001	>999.999

	Non-Smoker	
	No MI	MI
BMI		
Normal	10	0
Overweight	5	0
Obese	6	2

	Smoker	
	No MI	MI
BMI		
Normal	6	1
Overweight	4	2
Obese	0	5

Common Mistake #5: Overfitting a model with a small sample size

	No MI (n=31)	MI (n=10)
BMI		
Normal	16	1
Overweight	9	2
Obese	6	7

	No MI (n=310)	MI (n=100)
BMI		
Normal	160	10
Overweight	90	20
Obese	60	70

Effect	OR	95% CI	
Overweight vs Normal	3.55	0.27	44.88
Obese vs Normal	18.67	1.88	185.39

Effect	OR	95% CI	
Overweight vs Normal	3.55	1.60	7.93
Obese vs Normal	18.67	9.03	38.58

Common Mistake #5: Overfitting a model with a small sample size

	No MI (n=31)	MI (n=10)
BMI		
Normal	16	1
Overweight	9	2
Obese	6	7

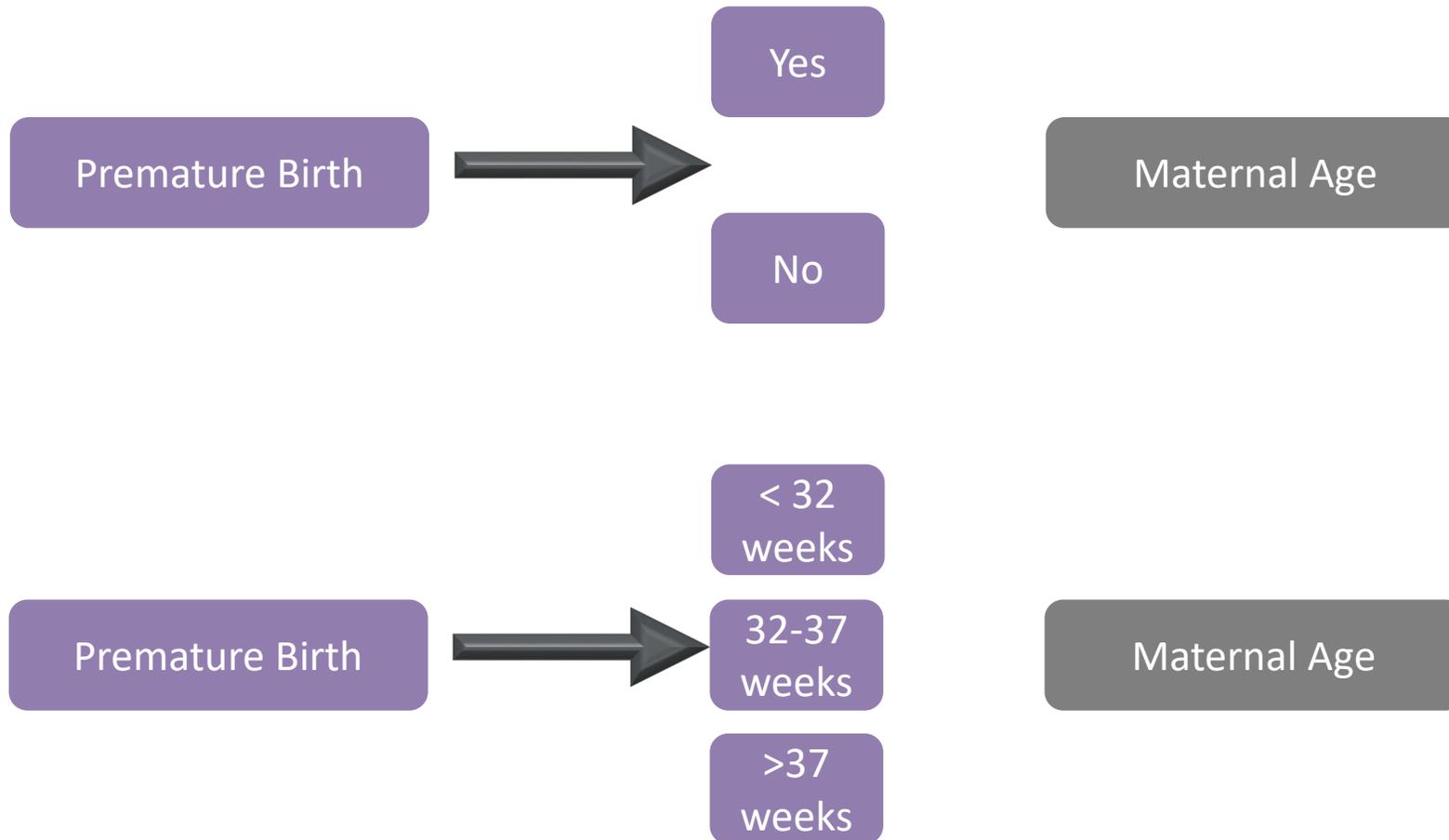
	No MI (n=290)	MI (n=120)
BMI		
Normal	150	20
Overweight	80	30
Obese	60	70

Effect	OR	95% CI	
Overweight vs Normal	3.55	0.27	44.88
Obese vs Normal	18.67	1.88	185.39

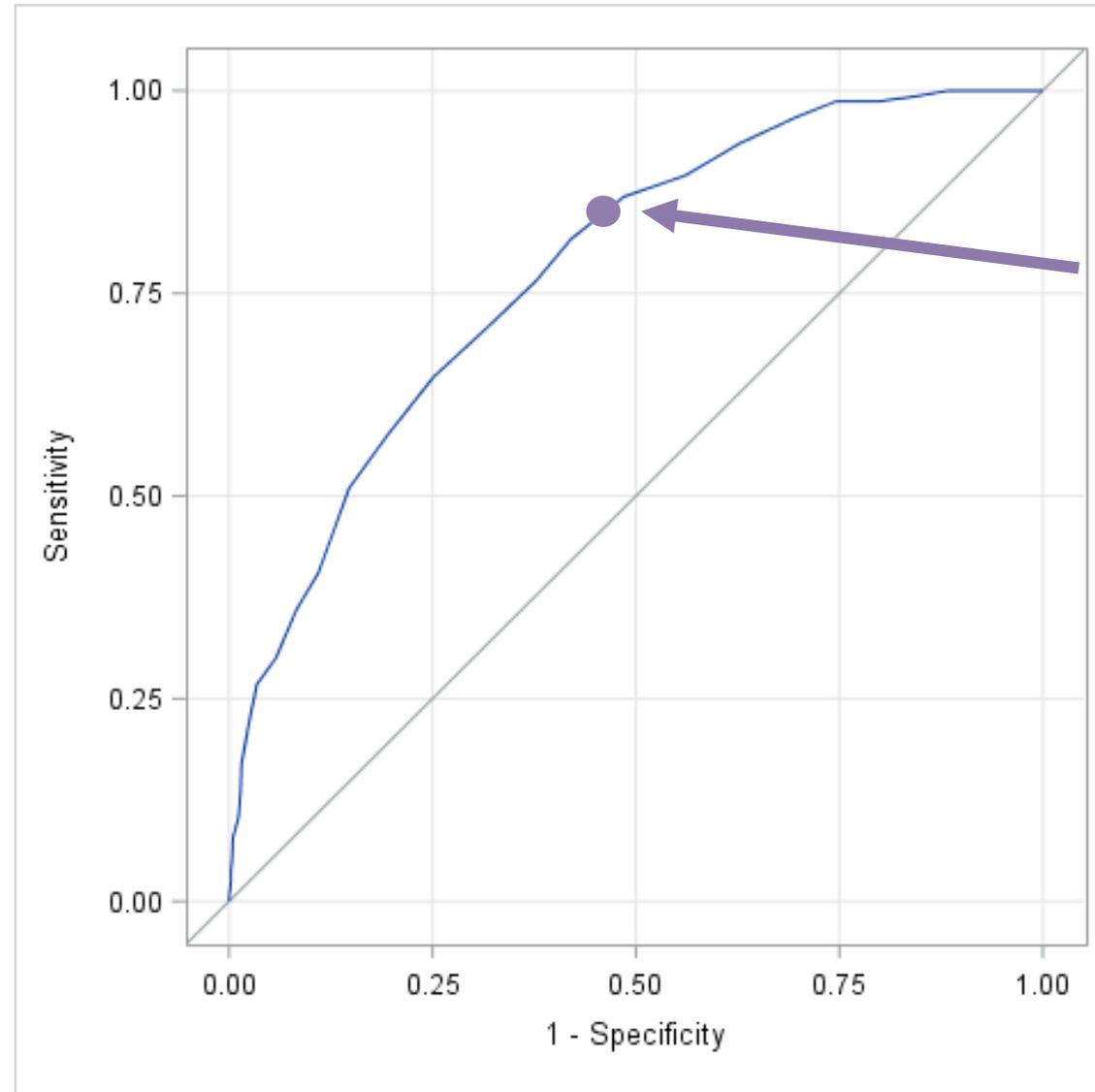
Effect	OR	95% CI	
Overweight vs Normal	2.81	1.50	5.27
Obese vs Normal	8.74	4.90	15.63

Related Topics

Related Topics: Ordinal/Multinomial



Related Topics: ROC Curve



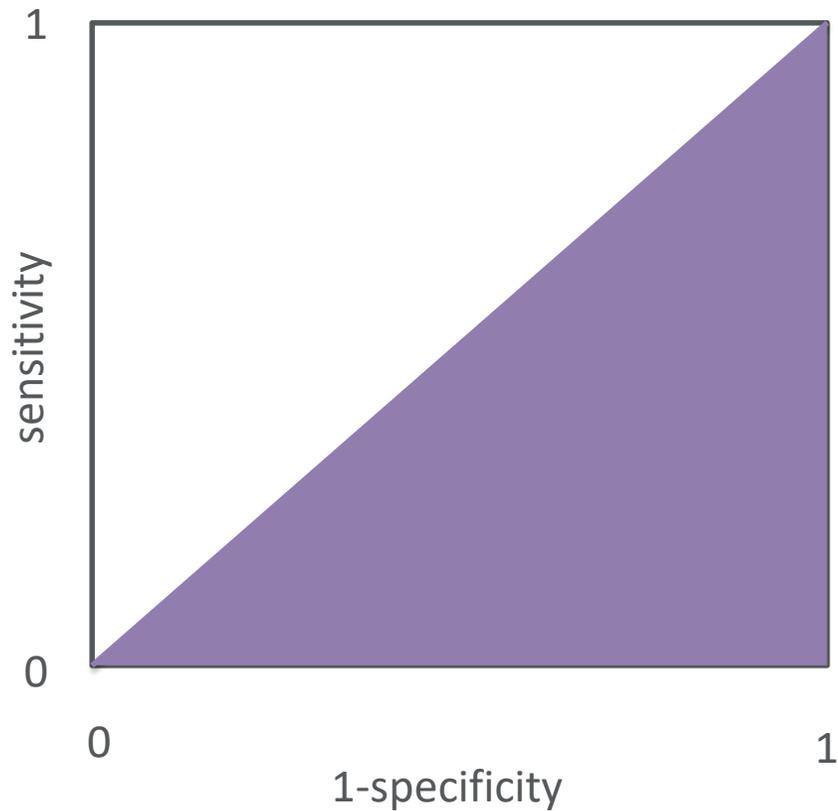
	Term	Preterm
Age <30	491	28
Age 30+	356	125

Sensitivity = 0.82

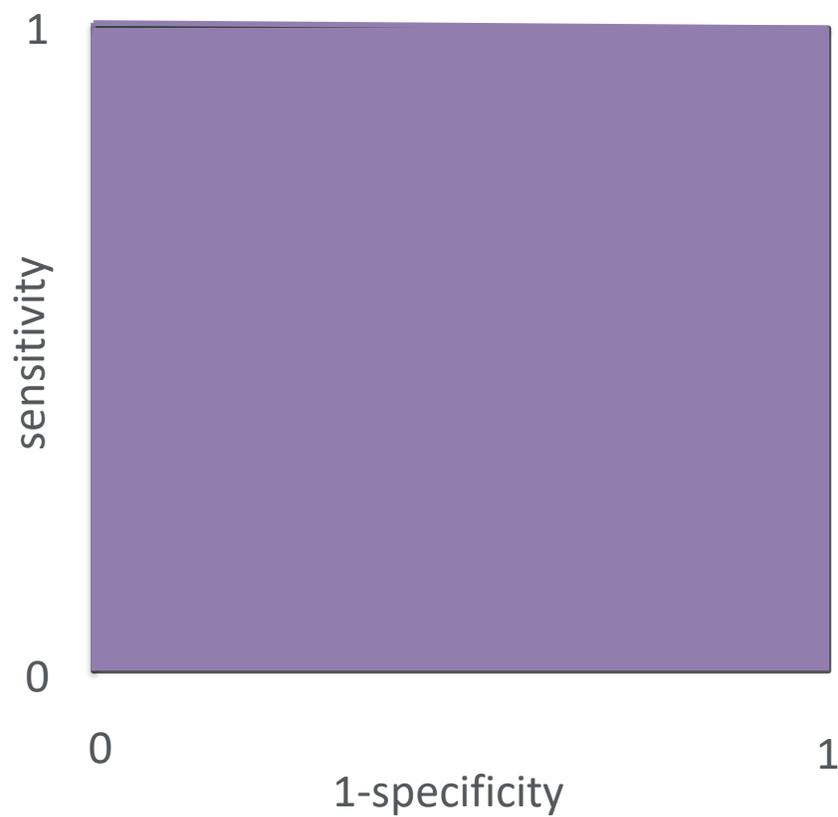
Specificity = 0.58

Related Topics: ROC Curve

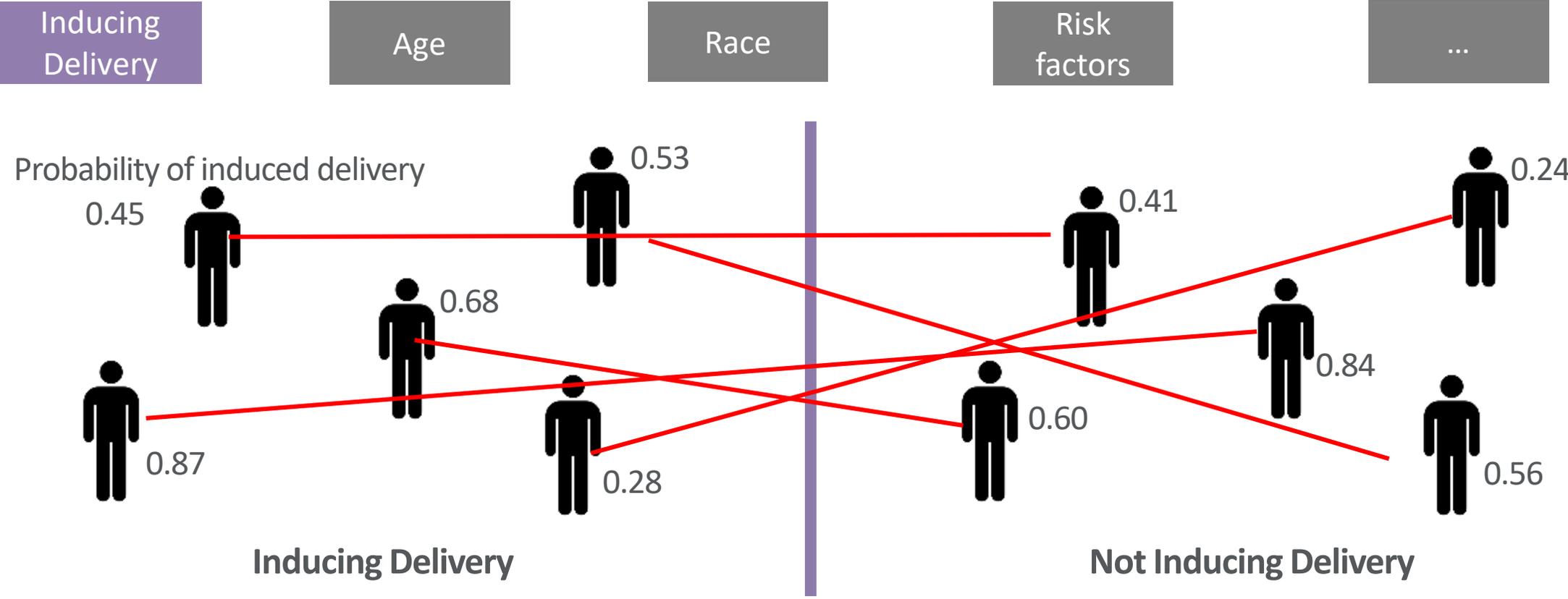
'Coin toss', AUC=0.5



Perfectly accurate, AUC=1



Related Topics: Propensity Scores



Alternative Methods for Binary Outcome

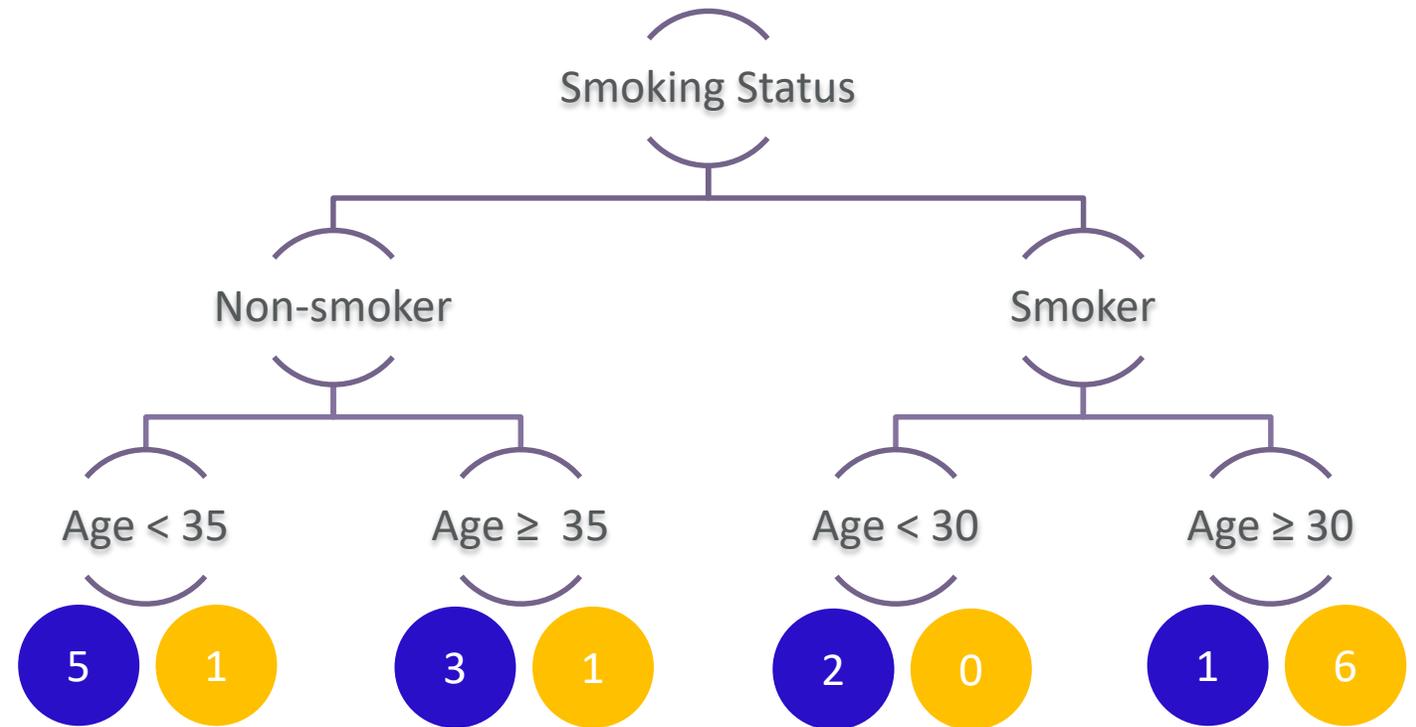
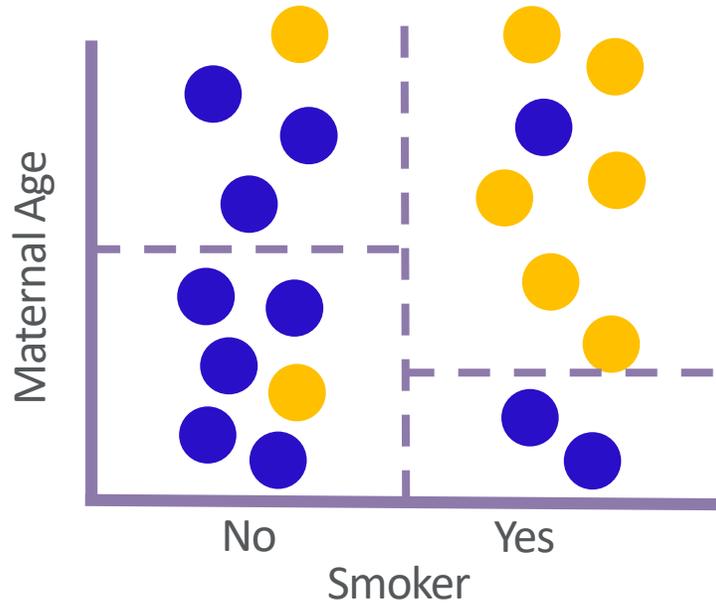
Relative Risk Models

- Log-binomial model
 - Binomial distribution
 - Log link
- Poisson model
 - Poisson distribution
 - Log link
 - Robust standard errors

A one year increase in mom age is associated with a 21% increase in the odds of premature birth. **OR=1.21**

A one year increase in mom age is associated with a 14% increase in the risk of premature birth. **RR=1.14**

Classification and Regression Tree (CART)



- Premature Birth
- Not Premature Birth

May the odds be ever in your favor

Questions?

References

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3936971/>
- <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.7170>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640017/>