

労働経済学

川田恵介

Table of contents

1	労働市場の記述	3
1.1	本スライドの方針	3
1.2	実例: CPS1985	3
1.3	事例研究の課題	3
1.4	実例: histogram	4
1.5	実例: histogram	5
1.6	関心とする特徴の明示	5
1.7	Linear Model	5
1.8	“Long” linear model	6
1.9	記述モデルとしての利点	6
2	データの記述	6
2.1	OLS アルゴリズム	6
2.2	例: 賃金モデル (関数)	6
2.3	例: 賃金モデル (関数)	7
2.4	例: 賃金モデル (関数)	7
2.5	応用上のコツ	8
2.6	教科書的例: Y のモデル	9
2.7	実例: Y のモデル	9
2.8	実例: Y のモデル	10
2.9	実例: Y のモデル	10
2.10	応用上の問題点	11
2.11	OLS アルゴリズム (その2)	11
2.12	解釈	11
2.13	実例: $E[Y X]$ のモデル	11
2.14	実例: $E[Y X]$ の”長い”モデル	12
2.15	実例: $E[Y X]$ の重回帰	13
2.16	実例: $E[Y X]$ の重回帰	13
2.17	実例: Y が2値の場合	14

3	Estimand の定義	14
3.1	Sampling Uncertainly	15
3.2	概念: Estimator/Estimand	15
3.3	概念: 母集団	15
3.4	仮定: ランダムサンプリング	15
3.5	Estimand の定義	15
3.6	Estimand: Population OLS	16
3.7	Population OLS の魅力	16
3.8	例	16
3.9	仮定: No misspecification	16
3.10	例	17
3.11	まとめ	17
4	Estimand の推定	17
4.1	データから計算した OLS の性質	18
4.2	Sampling distribution	18
4.3	Asyptotic normality	18
4.4	Consistency	19
4.5	数値例: 母集団 (賃金 - 年齢)	20
4.6	Estimate: $N = 20$	20
4.7	Estimate: $N = 20000$	21
4.8	95% 信頼区間	21
4.9	例: 信頼区間	22
5	定式化の診断	22
5.1	決定係数	22
5.2	Binscatter	22
5.3	Binscatter	23
5.4	例: Wage	23
5.5	例: Gender	24
5.6	まとめ	24
6	補論: 伝統的な議論	24
6.1	確率モデル	24
6.2	$E[u \times X] = 0$	25
6.3	$E[u X] = 0$	25
6.4	$Var[u X] = Var[u]$	25
6.5	$u \sim N(0, \sigma)$	25
	Reference	26

1 労働市場の記述

- どのような家計/企業が、働いているか? /高い賃金を得ているか? /結婚しているか? /子供を持つか? /雇用を増やしているか? 等
 - 議論の出発点
 - データから回答

1.1 本スライドの方針

- “必要最低限”な仮定のみで得られる, non-trivial な手法を紹介
 - 必要最低限の仮定: 多重共線性がない、ランダムサンプリング、(+ regularity condition: 高次 Moment が有限)
 - 私見では、労働経済学において特に重要視される傾向
- 伝統的な教科書で採用される説明も紹介
 - 確率モデルの推定として定式化

1.2 実例: CPS1985

A tibble: 10 x 7

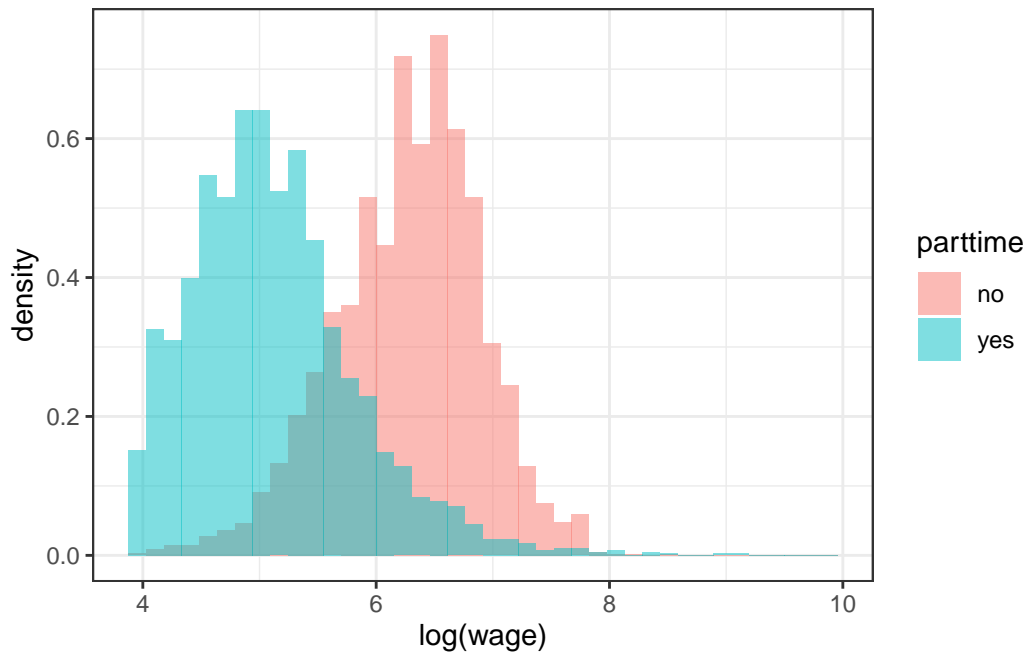
	wage	education	experience	ethnicity	smsa	region	parttime
	<dbl>	<int>	<int>	<fct>	<fct>	<fct>	<fct>
1	355.	7	45	cauc	yes	northeast	no
2	123.	12	1	cauc	yes	northeast	yes
3	370.	9	9	cauc	yes	northeast	no
4	755.	11	46	cauc	yes	northeast	no
5	594.	12	36	cauc	yes	northeast	no
6	377.	16	22	cauc	yes	northeast	no
7	285.	8	51	cauc	yes	northeast	no
8	561.	12	34	cauc	yes	northeast	no
9	264.	12	0	cauc	yes	northeast	no
10	1644.	14	18	cauc	yes	northeast	no

1.3 事例研究の課題

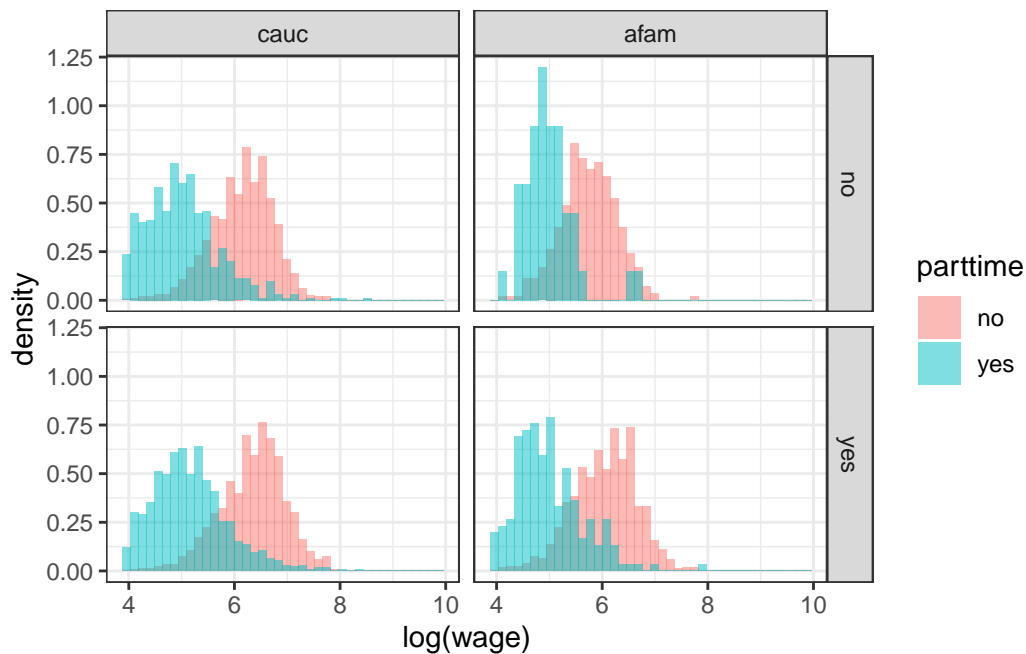
- 事例 (=データ) や社会の特徴を直接的に把握することは困難

- 大量の変数 (wage, education, ...) について、大量の事例が存在しており、人間の認知能力をも超えている
- “誰が見ても明らかな” 特徴” は、存在しない場合が多い

1.4 実例: histogram



1.5 実例: histogram



- 高賃金を得ている労働者の特徴は?

1.6 関心とする特徴の明示

- どのような特徴を分析対象とするのか、分析前に決定する
 - 例: 平均、分散、中央値
 - * データから社会の特徴を推論する上でも必須
- 労働経済学では、変数間の関係性把握に焦点が当たりがち
 - 良い出発点: Linear Model
 - * Y と X の関係性を簡潔に要約できる

1.7 Linear Model

- Y についての Linear Model $g_Y(X)$:

$$g_Y(X) = \beta_0 + \beta_1 X_1 + \dots + \beta_L X_L$$

- β_0, \dots, β_L = Parameters (データからは観察不可能であり、推定する必要がある)

– $X_1 \dots X_L$ = データから観察可能な Variable

* 例: $X = [\text{年齢}, \text{性別ダミー}]$

1.8 “Long” linear model

- Variables X についてではなく、Parameters β について Linear (足し算) であることが重要
- 短いモデル

$$g_Y(X) = \beta_0 + \beta_1 X$$

を”長くする”と

$$g_Y(X) = \beta_0 + \beta_1 \underbrace{X}_{\tilde{X}_1} + \beta_2 \underbrace{X^2}_{\tilde{X}_2}$$

1.9 記述モデルとしての利点

- パラメータの解釈が明確
 - β_1 = “モデル上”で X_1 が”1 単位”大きかった時に、 Y の値がどの程度大きいか?
 - モデルを超えた解釈も有する

2 データの記述

- Linear model の最も代表的な推定方法である OLS を紹介

2.1 OLS アルゴリズム

- 仮定: 多重共線性 ([wiki](#)) が無い
- 分析者が、モデル $g_Y(X) = \beta_0 + \dots + \beta_L X_L$ を設定
 - $\beta = [\beta_0, \dots, \beta_L]$ を二乗誤差の総和を最小にするように決定

$$\min \sum_i (Y_i - g_Y(X_i))^2$$

- $X := [X_1, \dots, X_L]$
- “ Y ” のモデル

2.2 例: 賃金モデル (関数)

- ミンサー型賃金モデル

$$\begin{aligned} \log(\text{wage}) \sim & \beta_0 + \beta_1 \times \text{EducationYear} \\ & + \beta_2 \times \text{Experience} + \beta_3 \text{Experience}^2 \end{aligned}$$

– 川口大司 (2011)

- β_1 = “Return to education”
- β_2/β_3 = “Return to experience”
 - “Retrun to human capital”

2.3 例: 賃金モデル (関数)

```
Fit = CPS1988 |>
  lm(log(wage) ~ education + experience + I(experience^2),
     data = _)

Fit$coefficients |>
  round(3)
```

(Intercept)	education	experience	I(experience^2)
4.278	0.087	0.078	-0.001

2.4 例: 賃金モデル (関数)

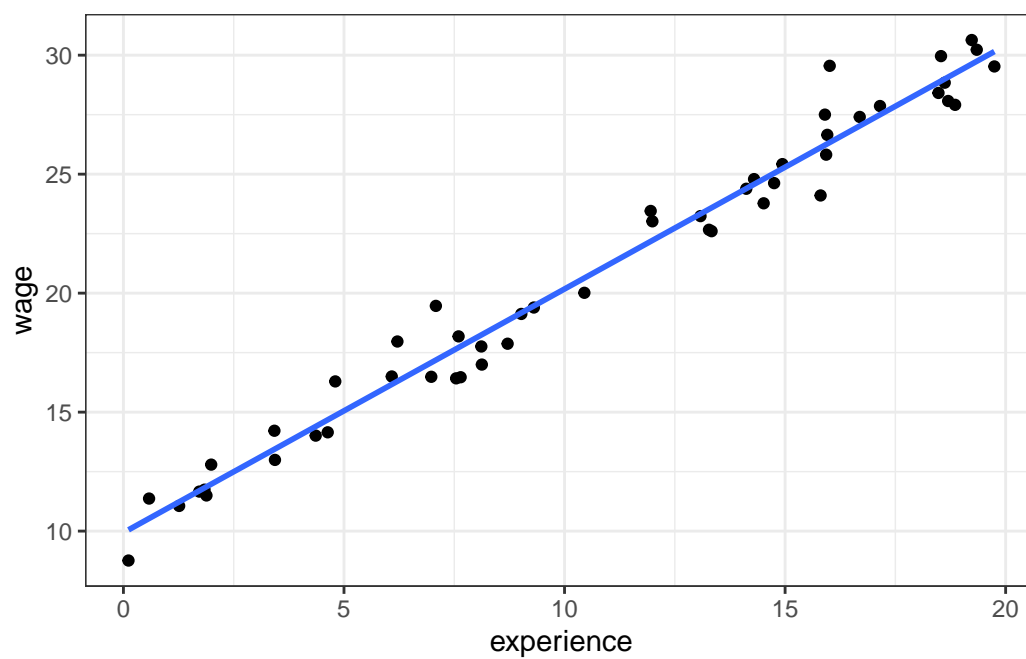
```
ggeffects::predict_response(
  Fit,
  "experience",
  vcov_fun = "vcovHC",
  vcov_type = "HC3" # Require sandwich
) |>
plot()
```



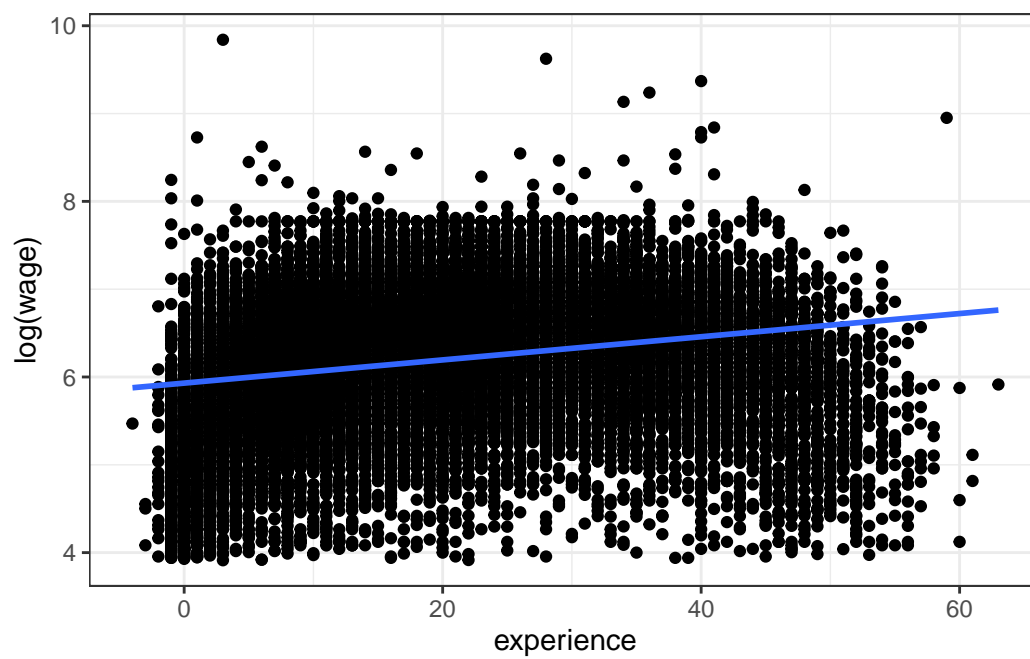
2.5 応用上のコツ

- しばしば OLS の推定結果を数値としてではなく、散布図上の線としてイメージ
 - － 機械学習などを理解する上でも有益

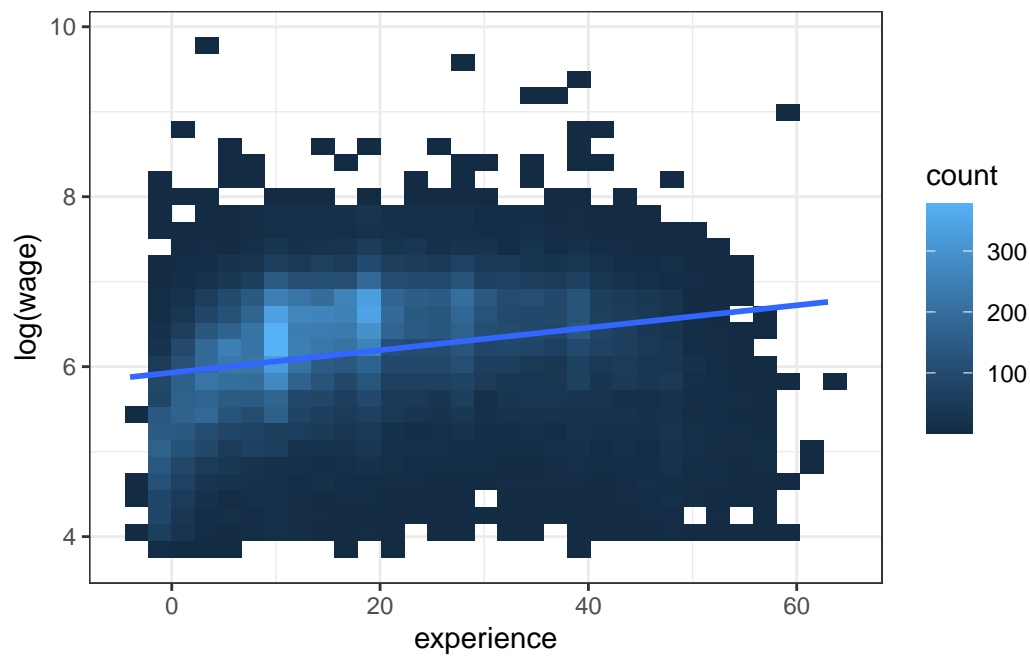
2.6 教科書的例: Y のモデル



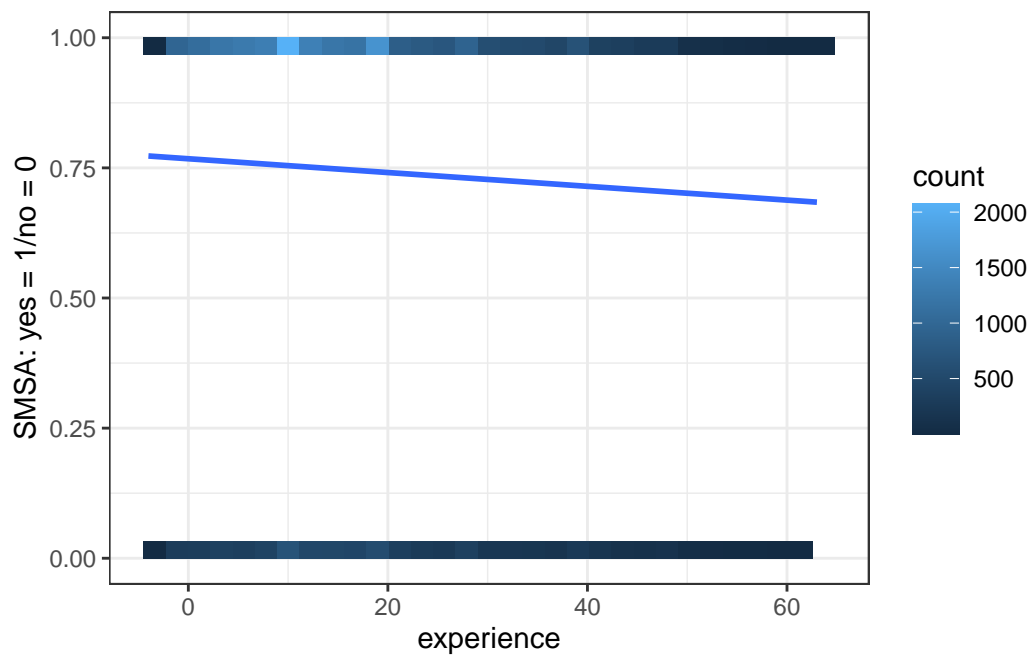
2.7 実例: Y のモデル



2.8 実例: Y のモデル



2.9 実例: Y のモデル



2.10 応用上の問題点

- 労働経済学における多くの応用では、「 X がよく似ていても Y の値が大きく異なる」
 - 賃金、就業、結婚等は非常に多くの要因が関わっている
 - * データから観察できる X は限られている
 - Y のモデルには見えない
- 別解釈が有効

2.11 OLS アルゴリズム (その2)

1. 分析者が、モデル $g_Y(X) = \beta_0 + \dots + \beta_L X_L$ を設定
2. $\beta = [\beta_0, \dots, \beta_L]$ を二乗誤差の総和を最小にするように決定

$$\min \sum_x [(Average(Y|X) - g_Y(X))^2 \times n(X)]$$

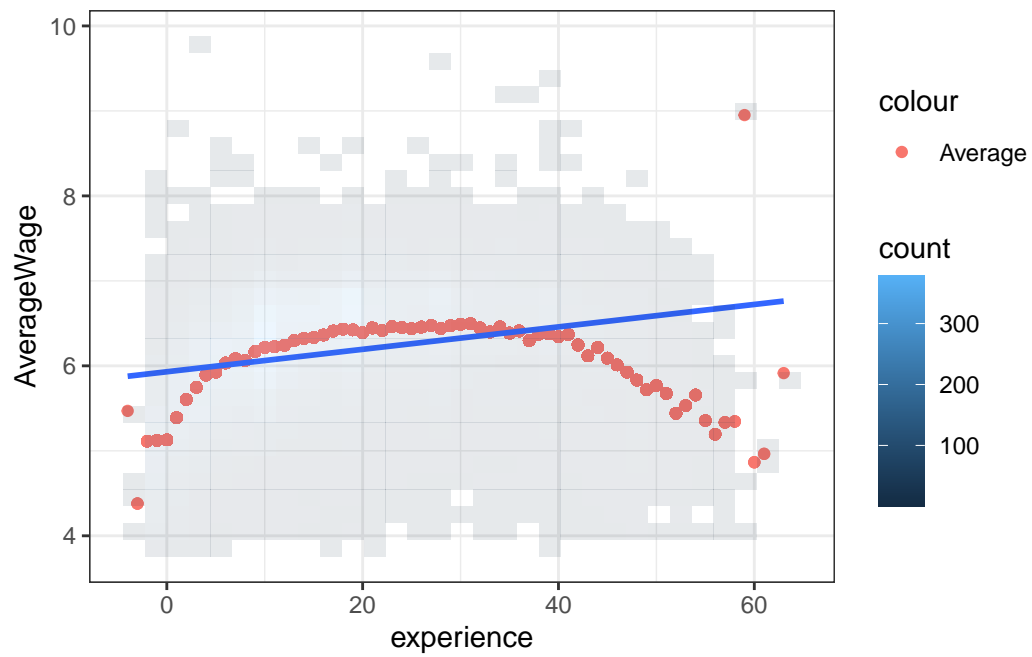
- $n(X)$ = 属性 X を持つ事例数
- $Average(Y|X)$ = 属性 X を持つ事例についての (データ上の) 平均値

2.12 解釈

- Y の平均値の近似モデル
 - 事例数が多い地点を重点的に近似

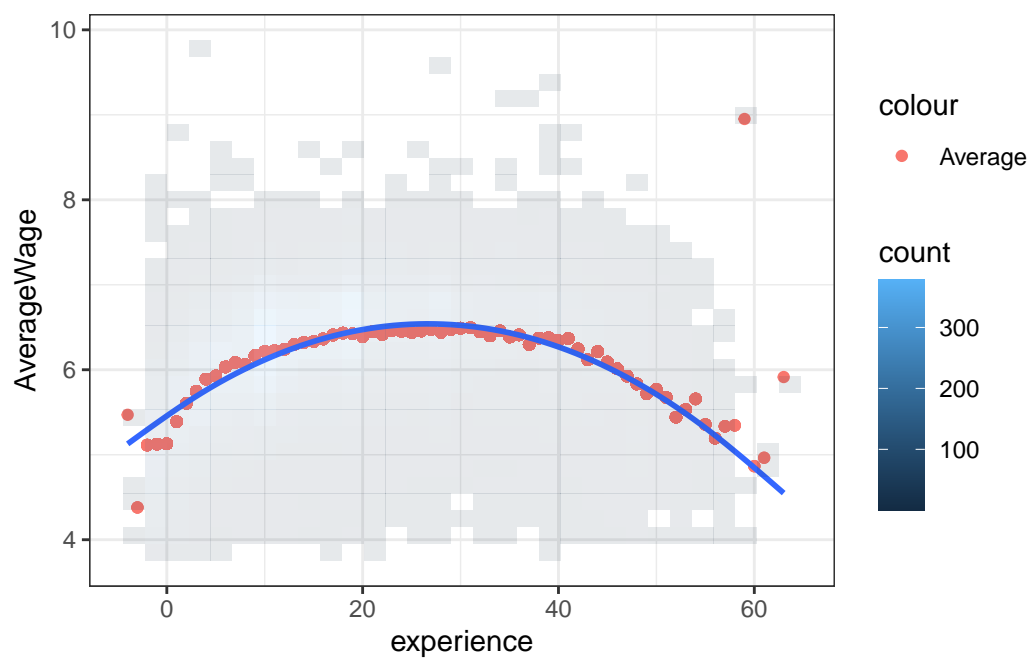
2.13 実例: $E[Y|X]$ のモデル

- $g_Y(X) = \beta_0 + \beta_1 Experience$



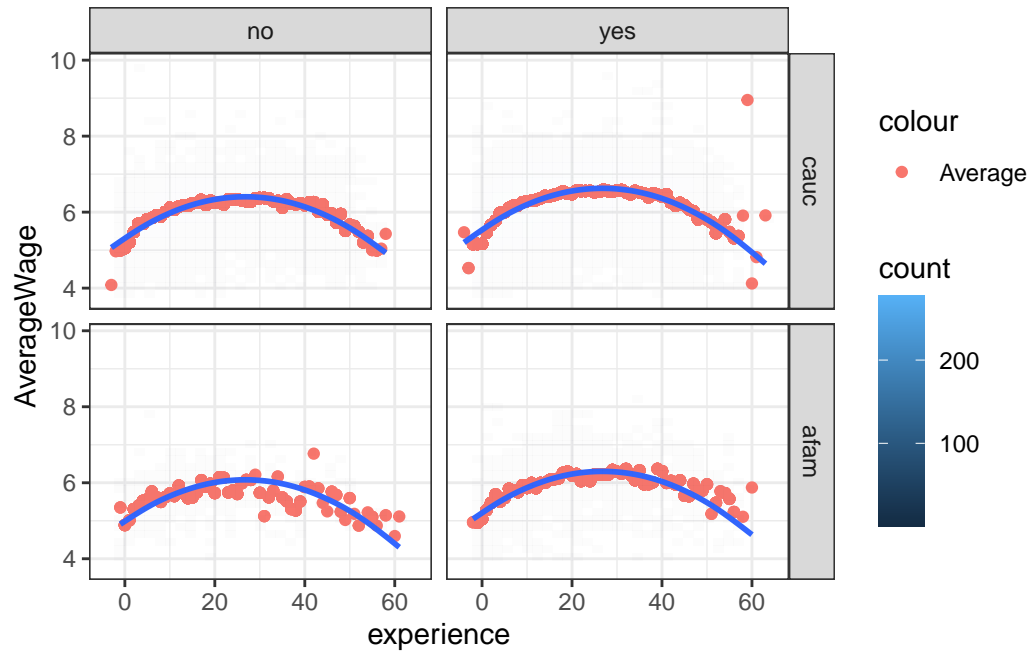
2.14 実例: $E[Y|X]$ の”長い”モデル

- $f_Y(X) = \beta_0 + \beta_1 Experience + \beta_2 Experience^2$



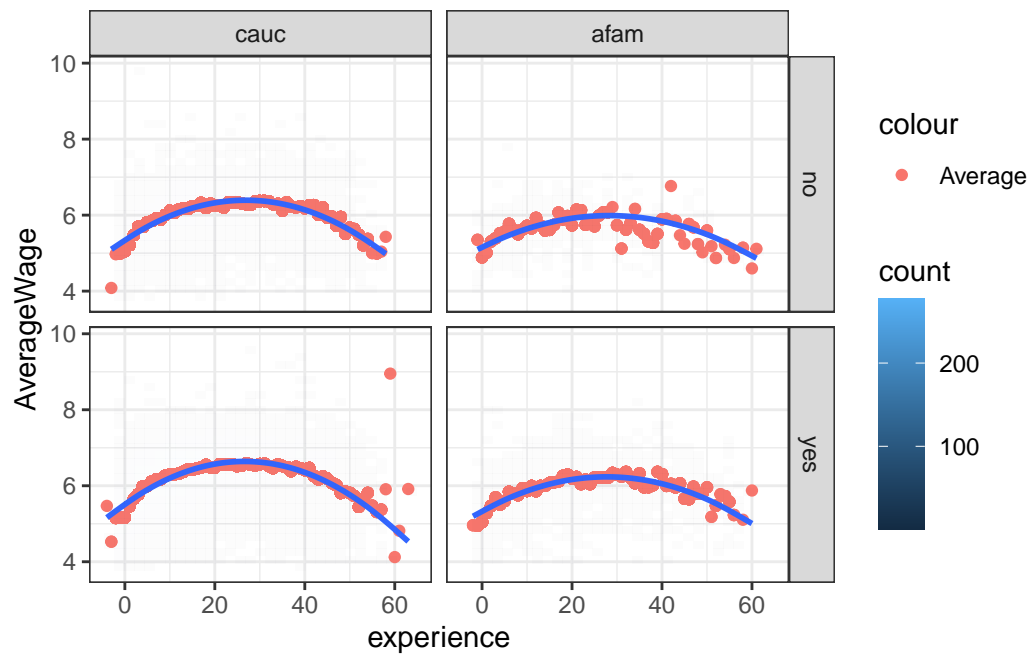
2.15 実例: $E[Y|X]$ の重回帰

- $f_Y(X) = \beta_0 + \beta_1 Experience + \beta_3 gender + \beta_4 married$

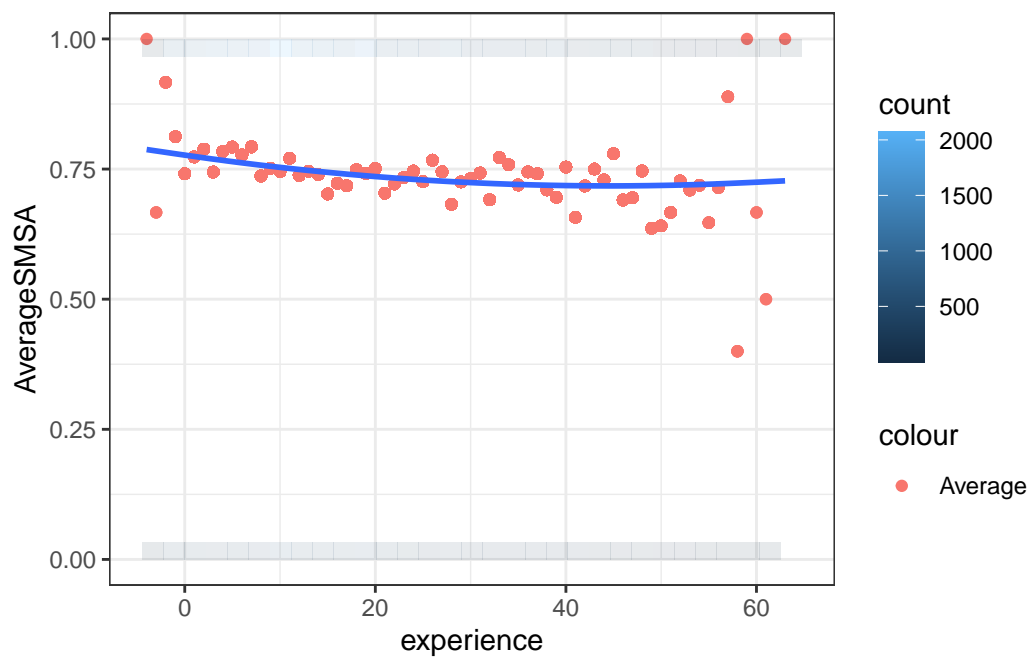


2.16 実例: $E[Y|X]$ の重回帰

- $f_Y(X) = \beta_0 + \beta_3 gender \times Experience + \beta_4 married \times Experience$



2.17 実例: Y が2値の場合



3 Estimand の定義

- データから労働市場についての含意を得たいが、非常に難しい

– OLS によって算出されるモデルは何を意味しているのか？

* 多くの人が合意可能な含意はあるのか？

3.1 Sampling Uncertainly

- 根本問題: 独立した研究者が、同時に同じ方法で調査/分析したとしても、同じ結果にならない
 - 例: 報道機関による世論調査
 - 調査する事例が異なるため

3.2 概念: Estimator/Estimand

- 対応策: 「全ての研究者で共通の真の答え」 (Estimand) と 「人によって異なるデータからの回答」 (Estimator) に分離する
 - 社会とデータの” 中間” 概念、**母集団**、を導入し、Estimand と Estimator を概念的に接続する

3.3 概念: 母集団

- (ラフに定義すると) “無限大の事例数” を持つデータ
 - 実際のデータは、母集団の一部
- 正式には、変数の同時分布 $f(Y, X)^*$ (母分布) として定義
- 分析目標の更新: データから **母集団** の特徴を理解できるか？

3.4 仮定: ランダムサンプリング

- データの事例は、母集団からランダムに選ばれる
 - 母分布に従って選ばれる

3.5 Estimand の定義

- 母集団 $f(Y, X)$ が直接観察できる (事例数無限大のランダムサンプルデータを持っている) 場合に、何を推定したいのか？
 - データと同様に、関心とする特徴の最初に明示する

3.6 Estimand: Population OLS

- 代表的な Estimand = “母集団で**仮想的に** OLS を行った結果”

$$g(Y)^* = \beta_0^* + \beta_1^* X_1 + \dots \beta_L^* X_L$$

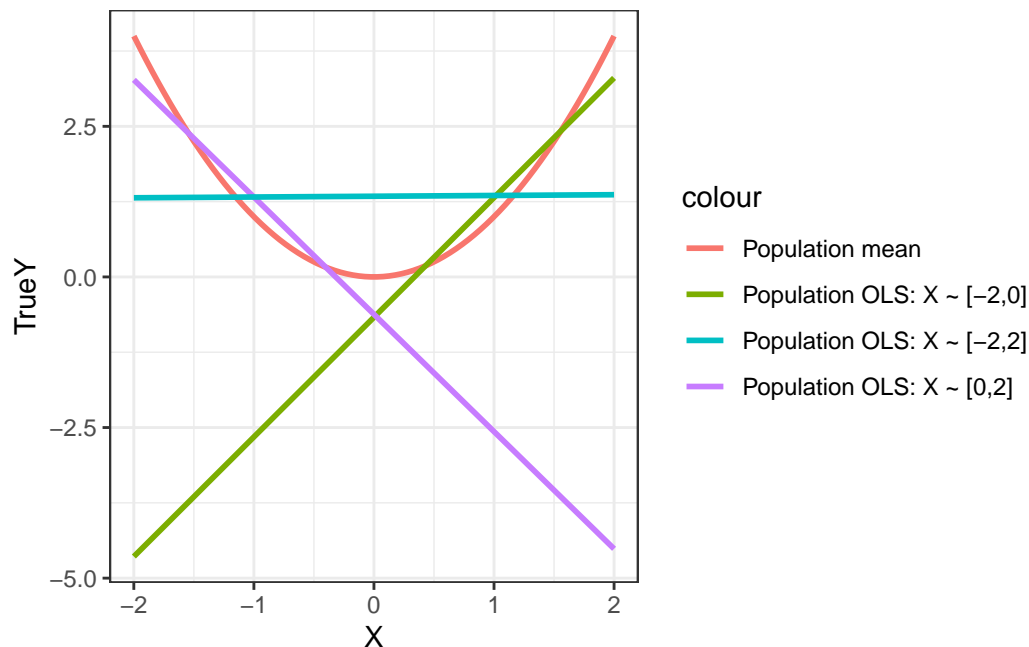
- Population mean $E[Y|X]$ を近似

3.7 Population OLS の魅力

1. ランダムサンプリングの**仮定のみ**で、データは有益な推定結果を提供できる

- Population OLS が推定できれば、
- 2. X と Y の母集団上での関係性について、Nontrivial な情報を提供
- 3. 条件付き母平均 $E[Y|X]$ を上手く近似できているかもしれない
- 注意点: $g(X)^*$ は、 $E[Y|X]$ のみならず、 X の分布に依存する

3.8 例

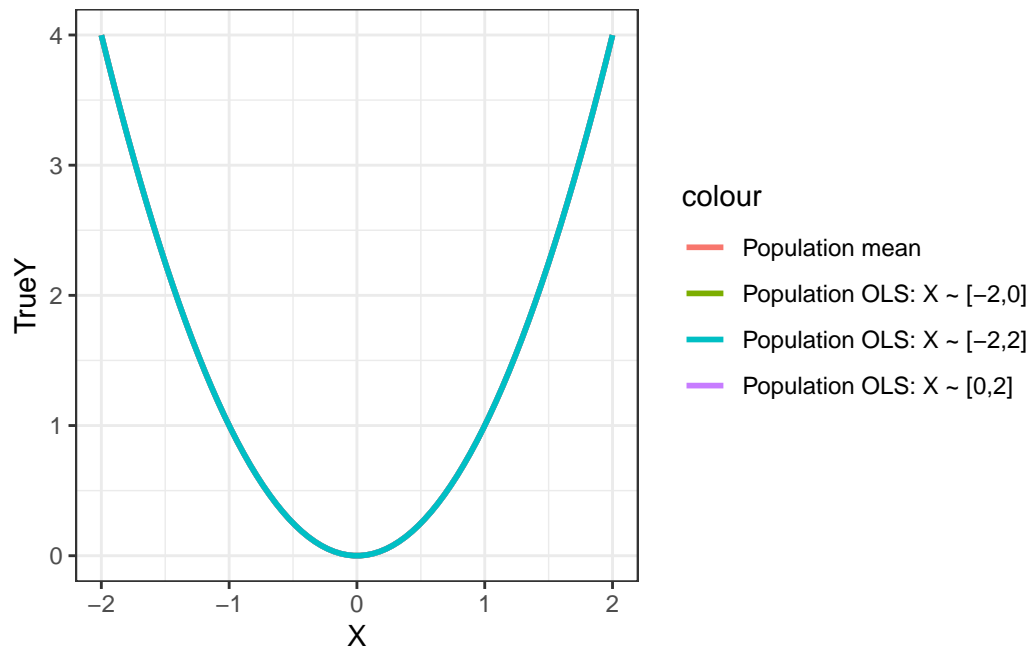


3.9 仮定: No misspecification

- Misspecification: β_0, \dots をどのように選んでも $g^*(X) \neq E[Y|X]$

- Misspecification があると Population OLS $g^*(X) \neq E[Y|X]$
- Misspecification がなければ、 X がどのような分布していても $g^*(X) = E[Y|X]$
- * 例外は X が分布を持たないケース

3.10 例



3.11 まとめ

- 同じ母集団/Estimand を対象とする独立した研究であれば、Estimand の値は共通
 - ランダムサンプルによって得られるデータが提供する回答は異なる
- 上記の問題構造に持ち込むために、母集団を導入

4 Estimand の推定

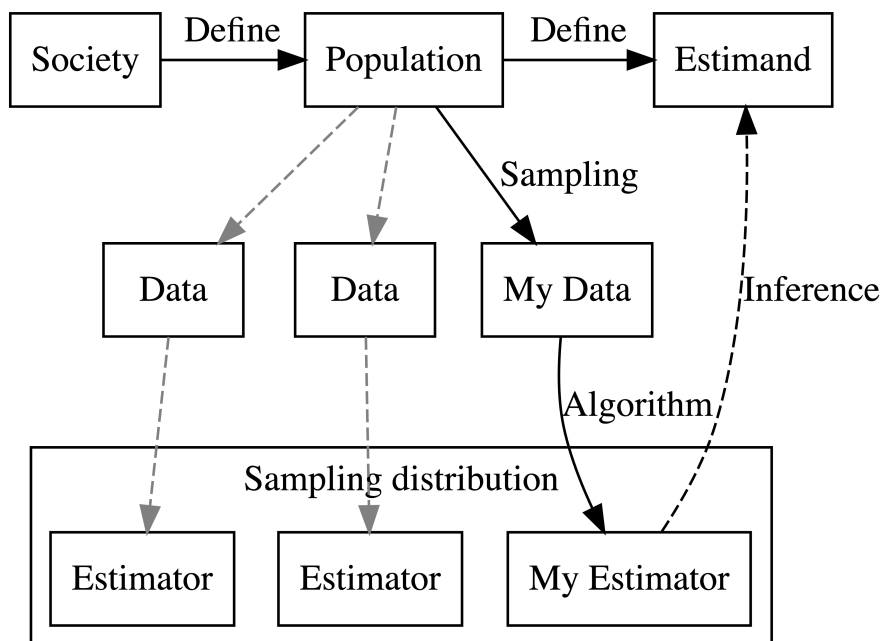
- Population は直接観察できないので、Estimand も観察できない
 - データから推論するしかない
 - * 推定結果 (Estimator)

4.1 データから計算した OLS の性質

- データが違うと異なるモデルが出てくる
 - モデル の分布は、事例数が増えたと、一定の規則性を持つ
- Random Sampling であれば、データが増えたと、Population OLS に近い Estimator が” 出やすくなる”
 - 正規分布で近似できる (中心極限定理)
 - 無限大の事例数の元で、Estimand (Population OLS) に収束する

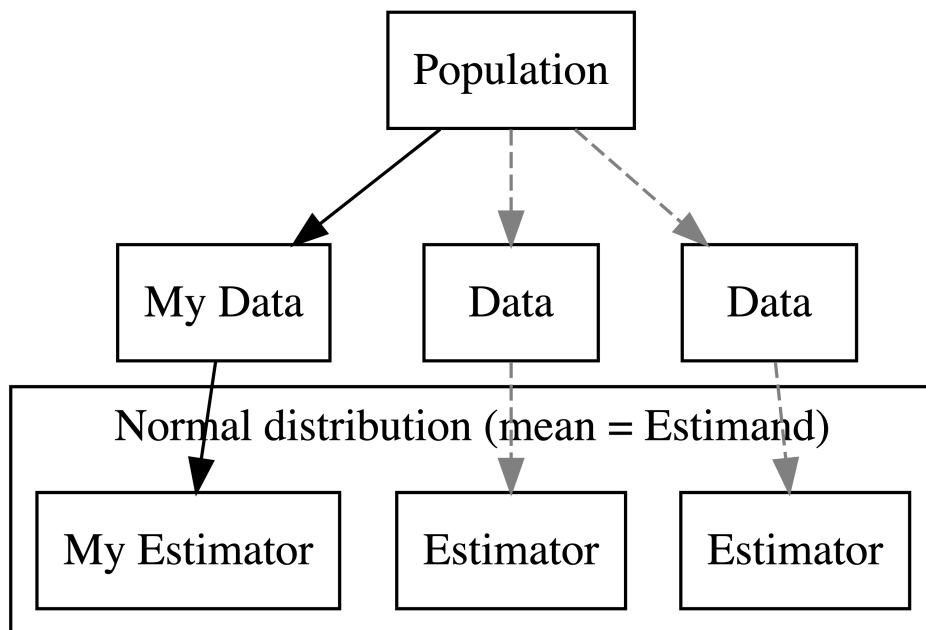
4.2 Sampling distribution

- 仮想的に生じる無数の結果の一つを観察する



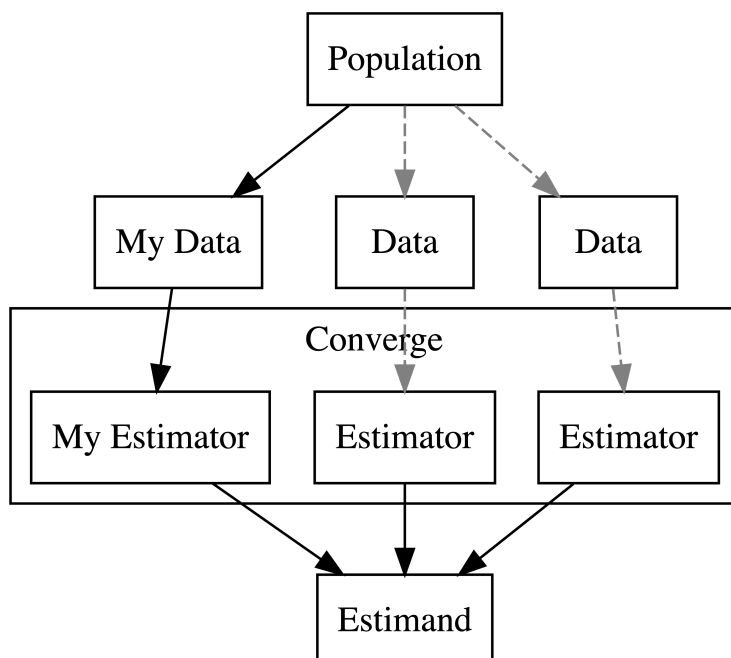
4.3 Asyptotic normality

- 事例数 N が増えたと、**Estimator** の分布に規則性を持つ

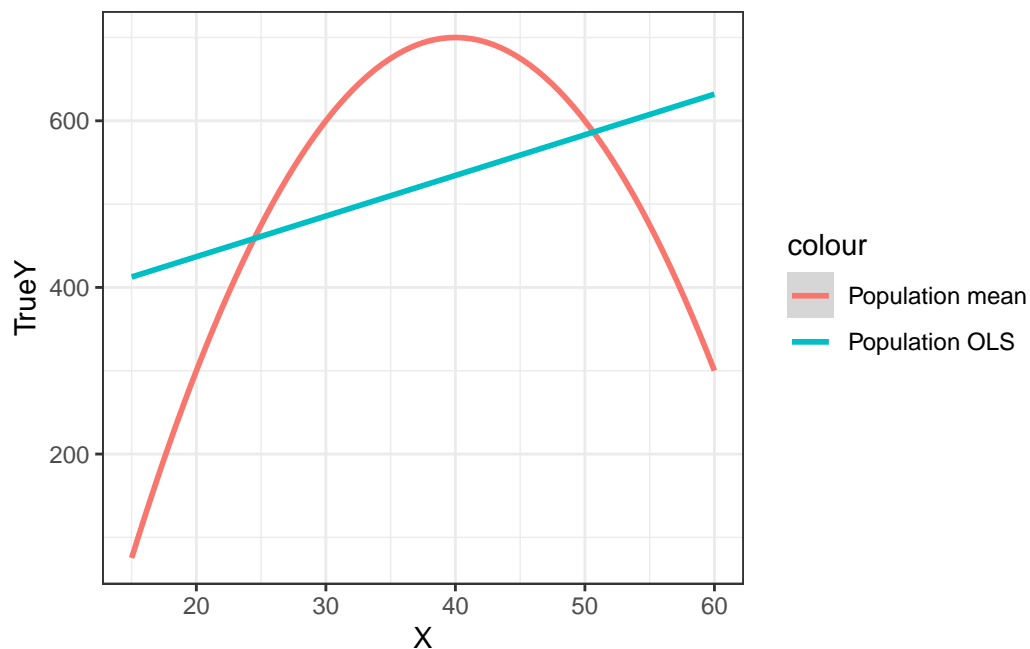


4.4 Consistency

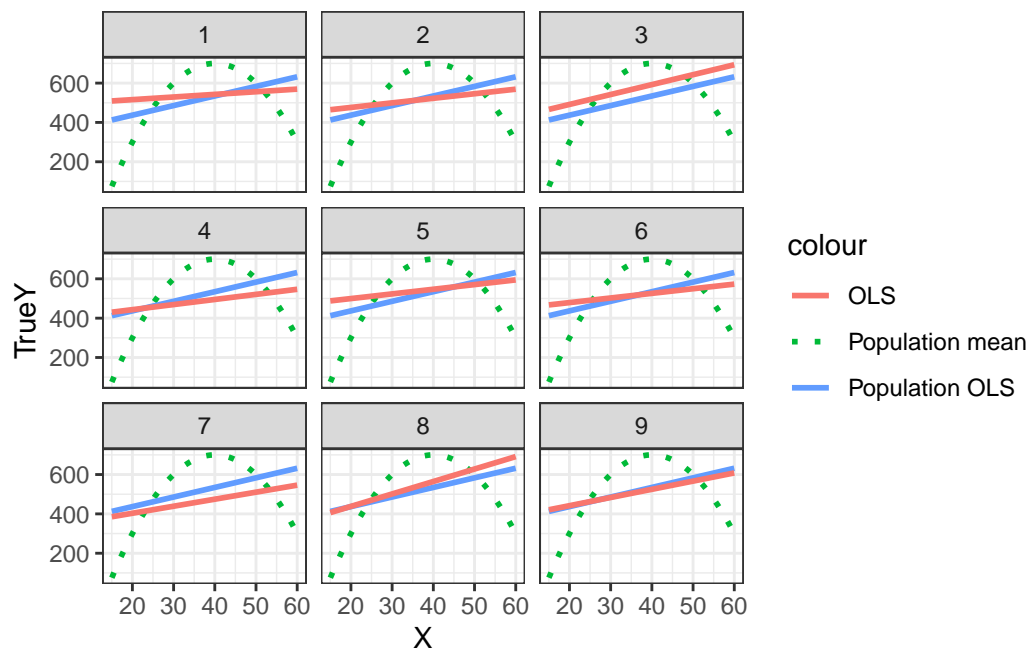
- 事例数が極めて大規模になると, \approx Estimand



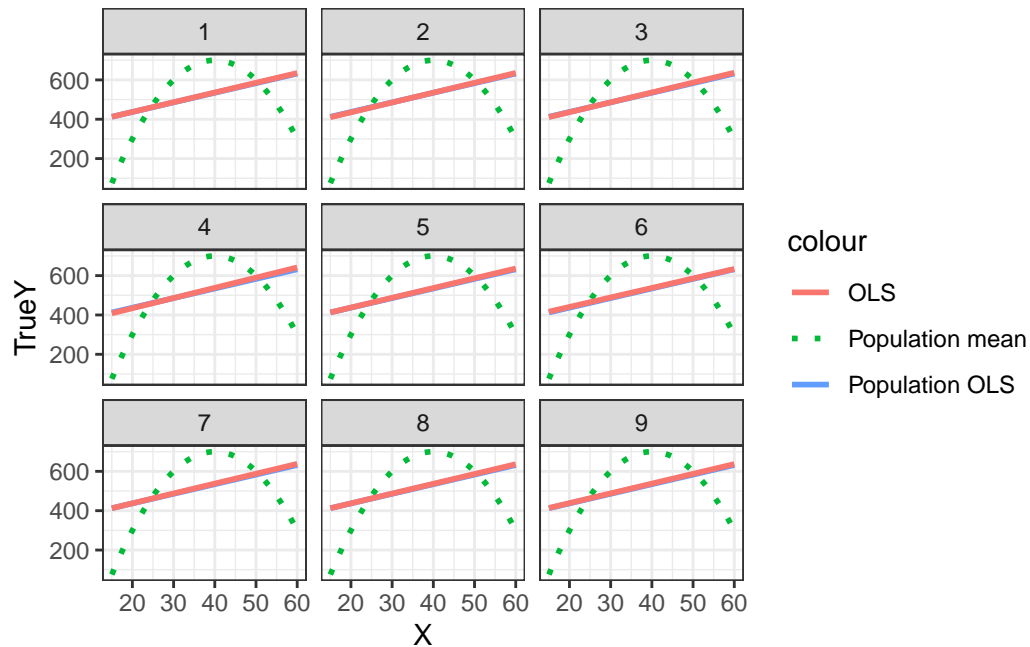
4.5 数值例: 母集団 (賃金 - 年齢)



4.6 Estimate: $N = 20$



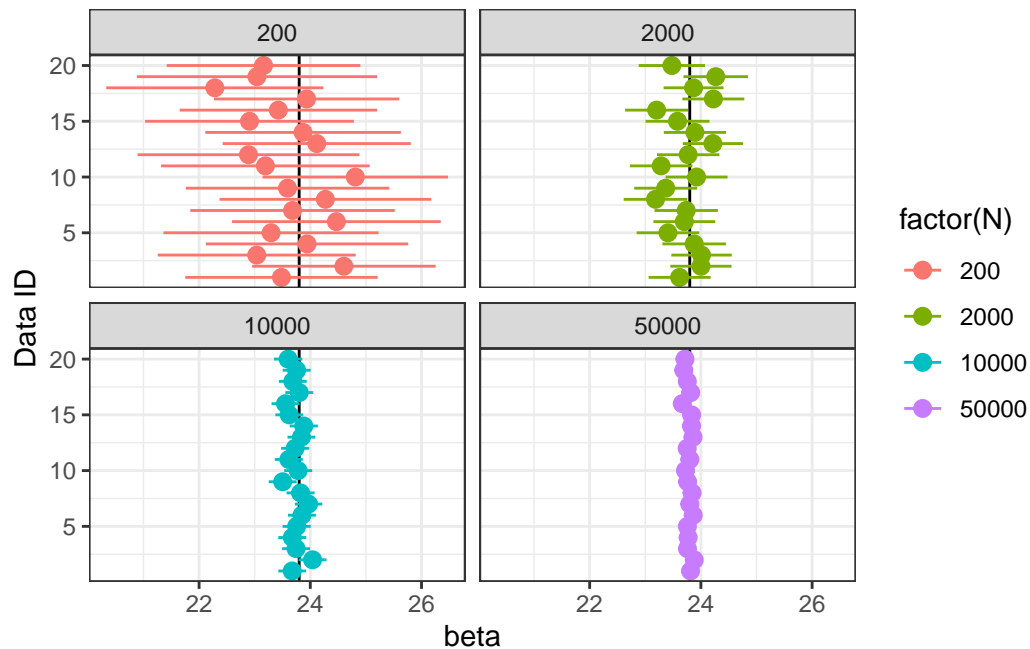
4.7 Estimate: $N = 20000$



4.8 95% 信頼区間

- Sampling uncertainty を直感的に要約
- Estimator ごとに計算される区間であり
 - 95 % の Estimator の信頼区間は、Estimand を含む
- Estimator が正規分布に従うと想定し、計算される
 - 中心極限定理から、ある程度の事例数があれば、近似的に正当化される
- 労働経済学において、結論を述べる最大の根拠

4.9 例: 信頼区間



5 定式化の診断

- Population OLS 以上の解釈には、厳密なテストが困難な仮定が必要
 - 推定後に診断し、結果とともに報告することが有益

5.1 決定係数

- 伝統的な指標であり、多くの関数が自動的に報告 ([wiki](#))
- モデルの Y への当てはまりの良さを測定する指標であり、多くの応用で非常に低い値となる
- 問題点
 - 過剰適合/過学習の結果として、非常に高い値が出てくる
 - Y の平均値のモデルの評価はできない

5.2 Binscatter

- 母平均とどの程度乖離しているのか?、可視的なやり方で診断

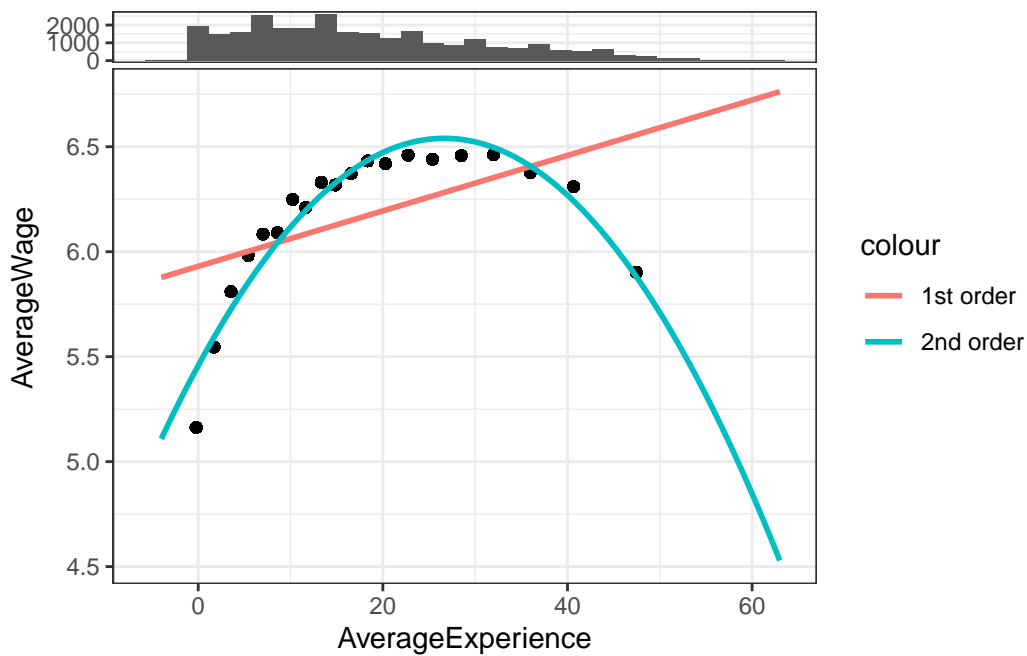
– データ上で OLS の結果と定式化に強く依存しない推定結果を比較

- Chetty, Looney, and Kroft (2009) , Chetty, Friedman, Olsen, et al. (2011) , Chetty, Friedman, Hilger, et al. (2011) などで採用された、人気の手法

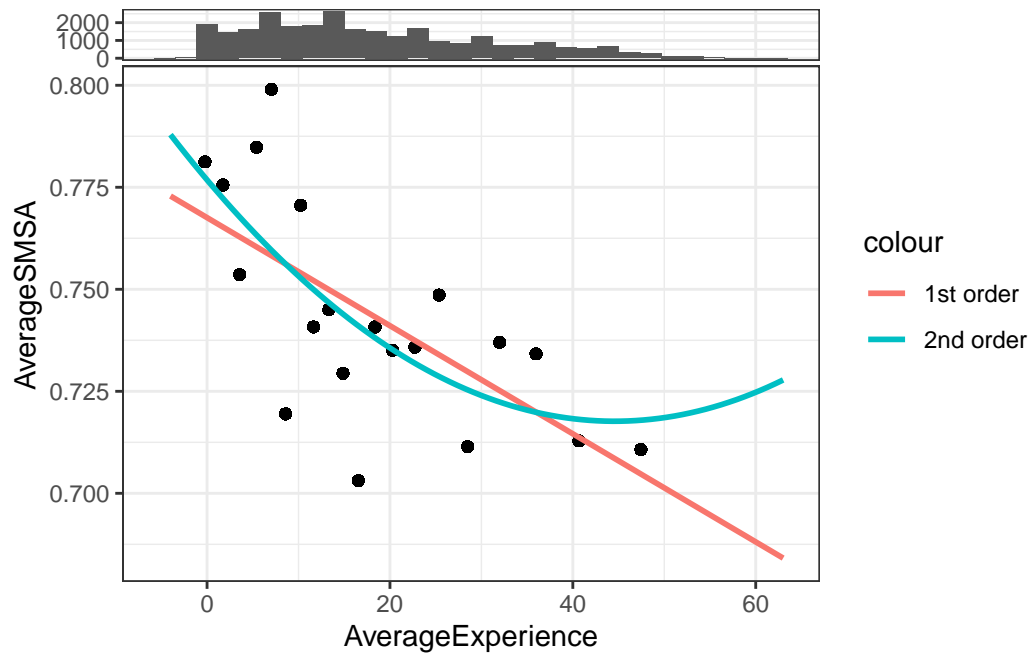
5.3 Binscatter

1. X を離散的カテゴリ (bin) に変換する (例) 年齢を 4 分割する
2. 各 bin 内で Y の平均値を計算する

5.4 例: Wage



5.5 例: Gender



5.6 まとめ

- Kuchibhotla, Brown, and Buja (2018)
- Aronow and Miller (2019)
- Ding (2024)

6 補論: 伝統的な議論

- 伝統的な入門書 (の最初の方の章) における議論と比較する
 - Wooldridge (2016), Stock and Watson (2020) など
- * 「正しい確率モデルのパラメタを推定する問題」に落とし込まれることが多い

6.1 確率モデル

- Y の正しい確率モデルを想定する:

$$Y = \underbrace{g(X)}_{\beta_0 + \beta_1 X_1 + \dots} + \underbrace{u}_{\text{誤差項} = Y - g(X)}$$

- 恒等式として成り立つ
- 誤差項の分布に“仮定”を追加することで β を”確定”させる

6.2 $E[u \times X] = 0$

•

$$Y = \underbrace{g(X)}_{\beta_0 + \beta_1 X_1 + \dots} + \underbrace{u}_{E[u \times X] = 0}$$

を母集団で満たす $\beta = \beta$ Population OLS

6.3 $E[u|X] = 0$

•

$$Y = \underbrace{g(X)}_{\beta_0 + \beta_1 X_1 + \dots} + \underbrace{u}_{E[u|X] = 0}$$

を母集団で満たす $\beta = E[Y|X] = g_Y(X)$ を達成

- 母平均について Mis-specification があれば、上記式を満たす β は**存在しない**
- $E[u|X] = 0$ が成り立てば、必ず $E[u \times X] = 0$ も成り立つ

6.4 $Var[u|X] = Var[u]$

- 分散均一
 - OLS は最善の不偏推定量 (BLUE) を提供
 - * ガウス・マルコフの定理 ([wiki](#))
 - 古典的な方法で標準誤差 (信頼区間) を計算可能

6.5 $u \sim N(0, \sigma)$

•

$$Y = \underbrace{g(X)}_{\beta_0 + \beta_1 X_1 + \dots} + \underbrace{u}_{\sim N(0, \sigma)}$$

- 古典的回帰モデル
- Estimand $f(Y|X)$ (Y の条件付き分布) の優れた Estimator (最尤法の推定値と一致)
 - 推定結果は、有限の事例数の元で、正規分布に従う
- 上記式を満たす β が存在しない (Mis-specification が存在する) 可能性が非常に高い

Reference

- Aronow, Peter M, and Benjamin T Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge University Press.
- Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.” *The Quarterly Journal of Economics* 126 (4): 1593–1660.
- Chetty, Raj, John N Friedman, Tore Olsen, and Luigi Pistaferri. 2011. “Adjustment Costs, Firm Responses, and Micro Vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records.” *The Quarterly Journal of Economics* 126 (2): 749–804.
- Chetty, Raj, Adam Looney, and Kory Kroft. 2009. “Salience and Taxation: Theory and Evidence.” *American Economic Review* 99 (4): 1145–77.
- Ding, Peng. 2024. “Linear Model and Extensions.” *arXiv Preprint arXiv:2401.00649*.
- Kuchibhotla, Arun K, Lawrence D Brown, and Andreas Buja. 2018. “Model-Free Study of Ordinary Least Squares Linear Regression.” *arXiv Preprint arXiv:1809.10538*.
- Stock, James H, and Mark W Watson. 2020. *Introduction to Econometrics*. Pearson.
- Wooldridge, Jeffrey M. 2016. *Introductory Econometrics*. Cengage AU.
- 川口大司. 2011. “ミンサー型賃金関数の日本の労働市場への適用.” *現代経済学の潮流*, 67–98.