POLSCI 9592

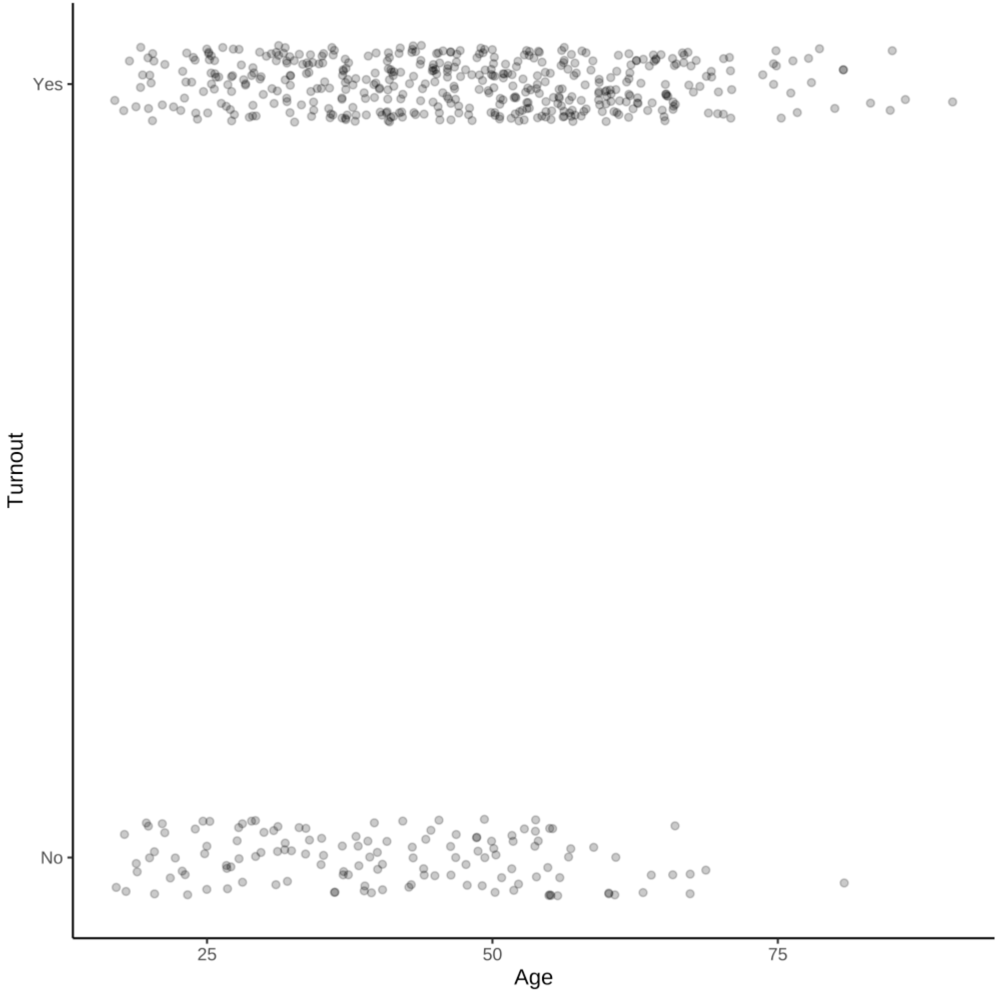
Lecture 2: Bianry Dependent Variable Models

Dave Armstrong

Goals for This Session

1. Develop and Evaluate the Linear Probability Model
2. Describe the Generalized Linear Model Framework
3. Estimate GLMs for Binary Dependent Variables
4. Consider Different Methods of Describing Effects.
5. What Should You Present?

What are we up to?



What's the best way to model the relationship between these variables?

- Straight line?
- S-shaped (sigmoid) curve?
- Step
- Something else?

Models for Binary Data: Preliminaries

I will refer to $b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k$ as $\mathbf{X}\beta$ and "the linear predictor", which in GLM notation is often referred to as η (the Greek letter "eta").

In the linear model, we are modeling $E(y|\mathbf{X})$ (the expected value of y given $\mathbf{X}$) as:

$$y = \mathbf{X}\beta + \varepsilon$$

or

$$E(y|\mathbf{X}) = \mathbf{X}\beta$$

Linear Probability Model

The definition of an expectation is:

$$E(y) = \sum y \times Pr(y)$$

For a binary variable where $Y = \{0, 1\}$, we get

$$\begin{aligned} E(y) &= 1 \times Pr(y = 1) + 0 \times Pr(y = 0) \\ &= 1 \times Pr(y = 1) + 0 \times (1 - Pr(y = 1)) \\ &= Pr(y = 1) \end{aligned}$$

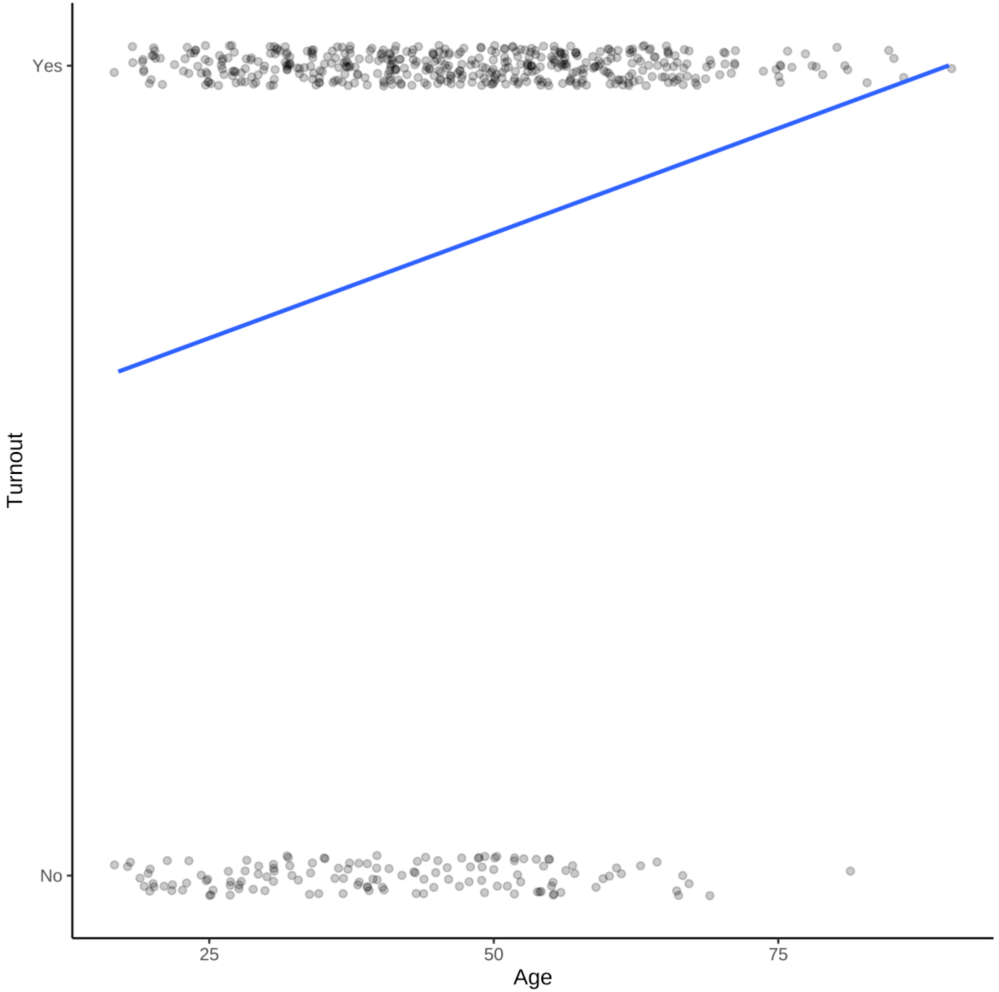
If y is binary and we model it with the linear model, then we are asserting:

$$Pr(y = 1|\mathbf{X}) = E(y|\mathbf{X}) = \mathbf{X}\beta$$

Voting Example

```r
mod <- lm(voted ~ age,
   data=dat)
summary(mod)
```

```
##
## Call:
## lm(formula = voted ~ age, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95374  0.07087  0.18871  0.25864  0.37778
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.534159   0.055181   9.680  < 2e-16 ***
## age         0.005180   0.001149   4.507 7.97e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.414 on 582 degrees of freedom
## Multiple R-squared:  0.03372,    Adjusted R-squared:  0.03206
## F-statistic: 20.31 on 1 and 582 DF,  p-value: 7.967e-06
```



A Better Specified LPM

```r
dat$ideo_strength <- abs(dat$leftright-5)
mod <- lm(voted ~ age + educ + income +
            ideo_strength + female + race,
          data=dat)
```

```r
summary(mod)
```

```
##
## Call:
## lm(formula = voted ~ age + educ + income + ideo_strength + female +
##     race, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.98154 -0.10438  0.09751  0.26112  0.68097
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.391614   0.103560  -3.782 0.000172 ***
## age            0.004557   0.001076   4.234 2.67e-05 ***
## educ           0.041008   0.007101   5.775 1.26e-08 ***
## income         0.011400   0.003198   3.565 0.000394 ***
## ideo_strength  0.032798   0.009483   3.459 0.000583 ***
## female         0.049907   0.031653   1.577 0.115416
## raceWhite      0.128341   0.048707   2.635 0.008642 **
## raceBlack      0.300102   0.057220   5.245 2.20e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3767 on 576 degrees of freedom
## Multiple R-squared:  0.2084,    Adjusted R-squared:  0.1988
## F-statistic: 21.66 on 7 and 576 DF,  p-value: < 2.2e-16
```

Model Interpretation

- Model fit looks reasonable
- The probability of the oldest person in the dataset voting is 33% more than the probability of the youngest person voting
- The probability of the most educated person voting is about 69% higher than the probability of the least educated person voting.
- Those with the highest income have probability of voting about 27% higher than those with the lowest income.
- Those at the extremes of the ideological spectrum have probabilities of voting 16% higher than those in the middle of the ideological spectrum.
- Females are more likely to vote than men, though not significantly so.
- Whites and black are both more likely to vote than those in the "other" category. Further, blacks have a signiêcantly higher probability of voting than do whites.

Does the Model Make Sense?

The substantive conclusions seem mostly reasonable if not a bit exaggerated for education and perhaps age, but

- Is the linear functional form right? Probably not. Further, the model imposes (at least in this case) a constant marginal effect. That means regardless of where you start, the variable always gives the smae change in predicted probabilities.
- Do the errors have the same variance? No - the errors will almost certainly be heteroskedastic.
- Are the errors normally distributed? No - the errors will likely be bimodal.
- Also, it is possible that $\hat{y} > 1$ or that $\hat{y} < 0$, which is theoretically not possible for a probability.

Summary

The problems could be solved if we coul;d make the outcome variable:

1. Continuous, and
2. Unbounded

Then we could model:

$$\widehat{\text{Outcome}} = b_0 + b_1 x_1 + \ldots + b_k x_k$$

The Solution

We start by wanting to predict a probability: $\Pr(Y = 1|\mathbf{X})$, which is continuous but bounded in $[0, 1]$. We could think of some transformations that may produce the right result:

1. Odds: $\frac{\Pr(Y=1|\mathbf{X})}{\Pr(Y=0|\mathbf{X})}$, which is still bounded in $[0, \infty]$
2. The log of the odds: $\log\left(\frac{\Pr(Y=1|\mathbf{X})}{\Pr(Y=0|\mathbf{X})}\right)$ is continuous and unbounded.

So, we could model:

$$\log\left(\frac{\Pr(Y = 1|\mathbf{X})}{\Pr(Y = 0|\mathbf{X})}\right) = b_0 + b_1 x_1 + \ldots + b_k x_k$$

This is the logistic regression model.

Log-odds to Probabilities

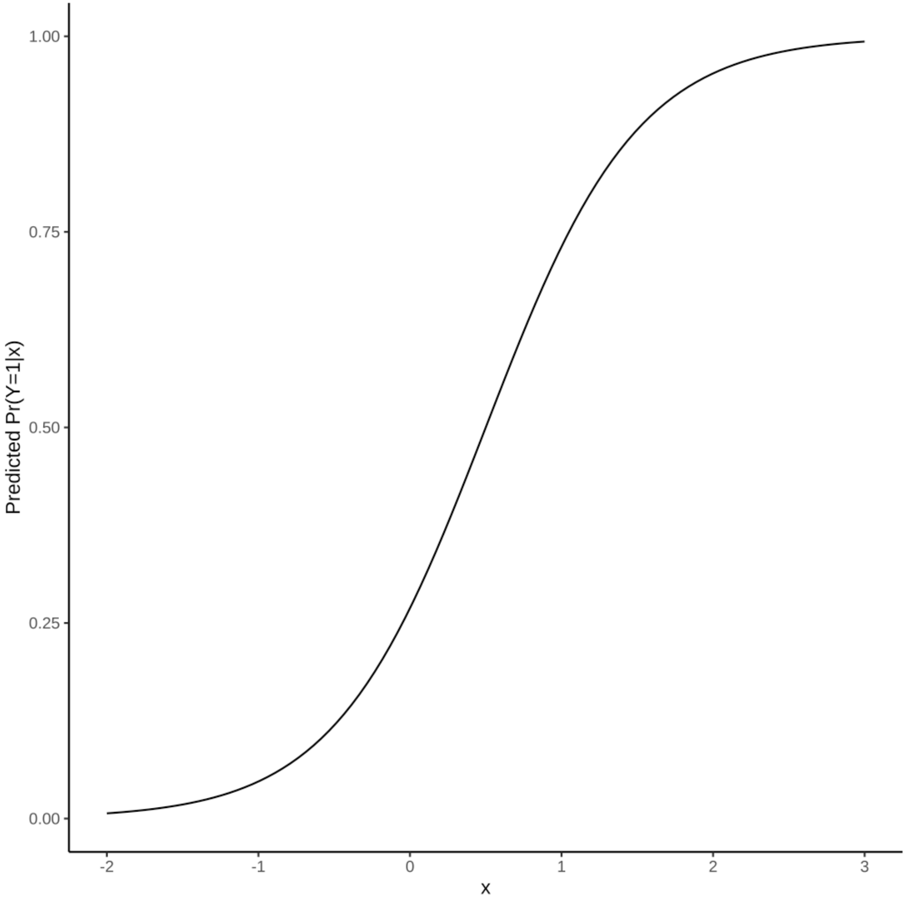
If we think of a simple model:

$$\log\left(\frac{\Pr(Y=1|\mathbf{X})}{\Pr(Y=0|\mathbf{X})}\right) = -1 + 2x$$

where x is in the range $[-2, 3]$, we could unwind the transformation to get the probabilities:

$$\Pr(Y=1|\mathbf{X}) = \frac{e^{-1+2x}}{1+e^{-1+2x}}$$

This produces a function that is non-linear in x.



We did it!

Great, so we have a solution, can we use linear regression now

- Nope - because we don't actually observe the log of the odds - only wheter $y = \{0, 1\}$.
- We have to use MLE
  - We want to make $\Pr(Y = 1|\mathbf{X})$ as big as possible for the observations where $Y = 1$ and,
  - as small as possible for the observations where $Y = 0$.

What Distribution?

The dependent variable is whether a respondent voted or not $y = \{0, 1\}$, so what distribution could we use?

- There aren't a lot of two-point distributions, but the Bernoulli is a common one. Its PMF is:

$$f(y_i|p_i) = p_i^{y_i}(1 - p_i)^{1-y_i}$$

Likelihood Function

First, let's consider the likelihood function:

$$L_i = \prod_i \hat{p}_i^{y_i}(1 - \hat{p}_i)^{(1-y_i)}$$
$$\log L_i = \sum_i y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i)$$

where $p_i = Pr(y_i = 1|\mathbf{X})$. For our purposes, most of the time:

$$Pr(y_i = 1|\mathbf{X}) = \Lambda(\mathbf{Xb}) \qquad \text{(Logit)}$$
$$Pr(y_i = 1|\mathbf{X}) = \Phi(\mathbf{Xb}) \qquad \text{(Probit)}$$

$\Phi(\cdot)$ and $\Lambda(\cdot)$ are the CDFs for the normal and logistic distributions, respectively.

Logit

We will often use logistic regression because the interpretation becomes a bit easier. Here:

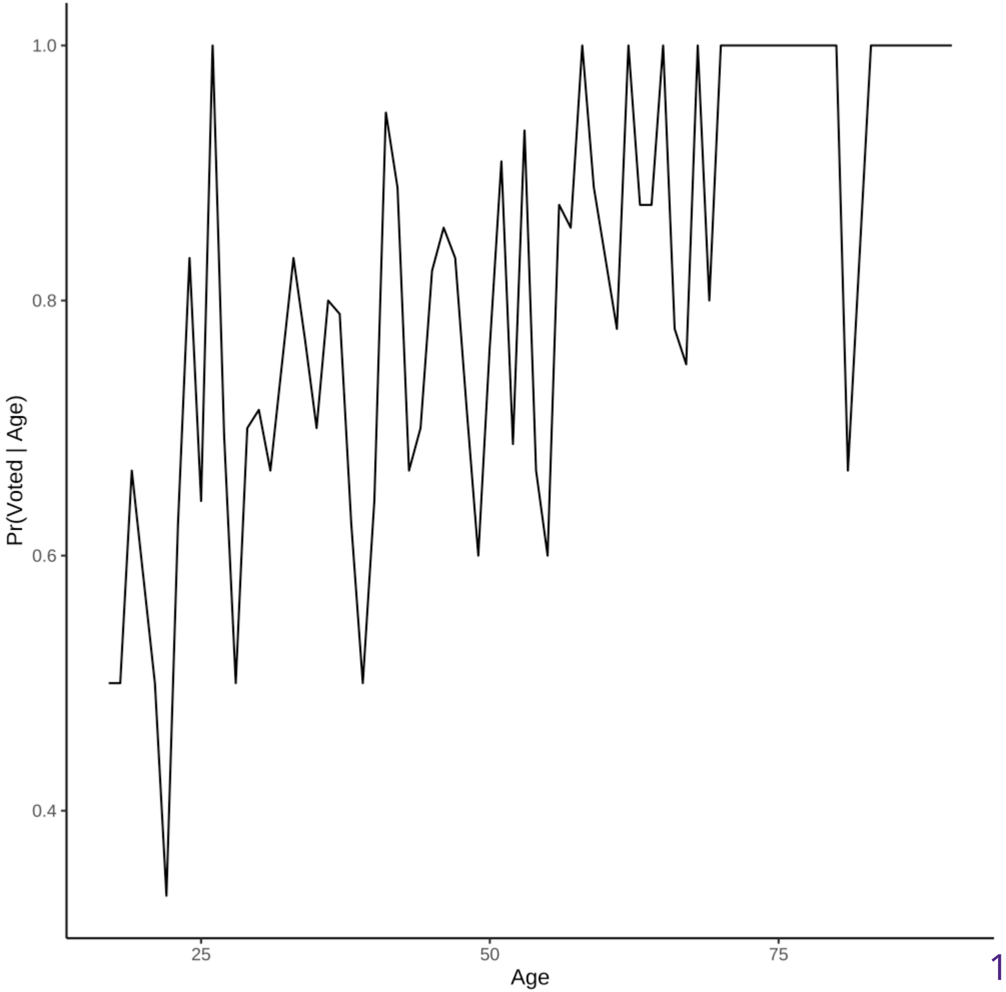
$$
\begin{aligned}
Pr(y_i = 1|\mathbf{X}) &= \Lambda(\mathbf{Xb}) \\
&= \frac{e^{\mathbf{Xb}}}{1 + e^{\mathbf{Xb}}}
\end{aligned}
$$

Figuring out the predicted probability "by hand" here doesn't require integration (as it does in the case of the probit model).

Simple Example: Age and Turnout

We might want to know how age affects turnout. One possibility would be to calculate turnout for each individual value of age.

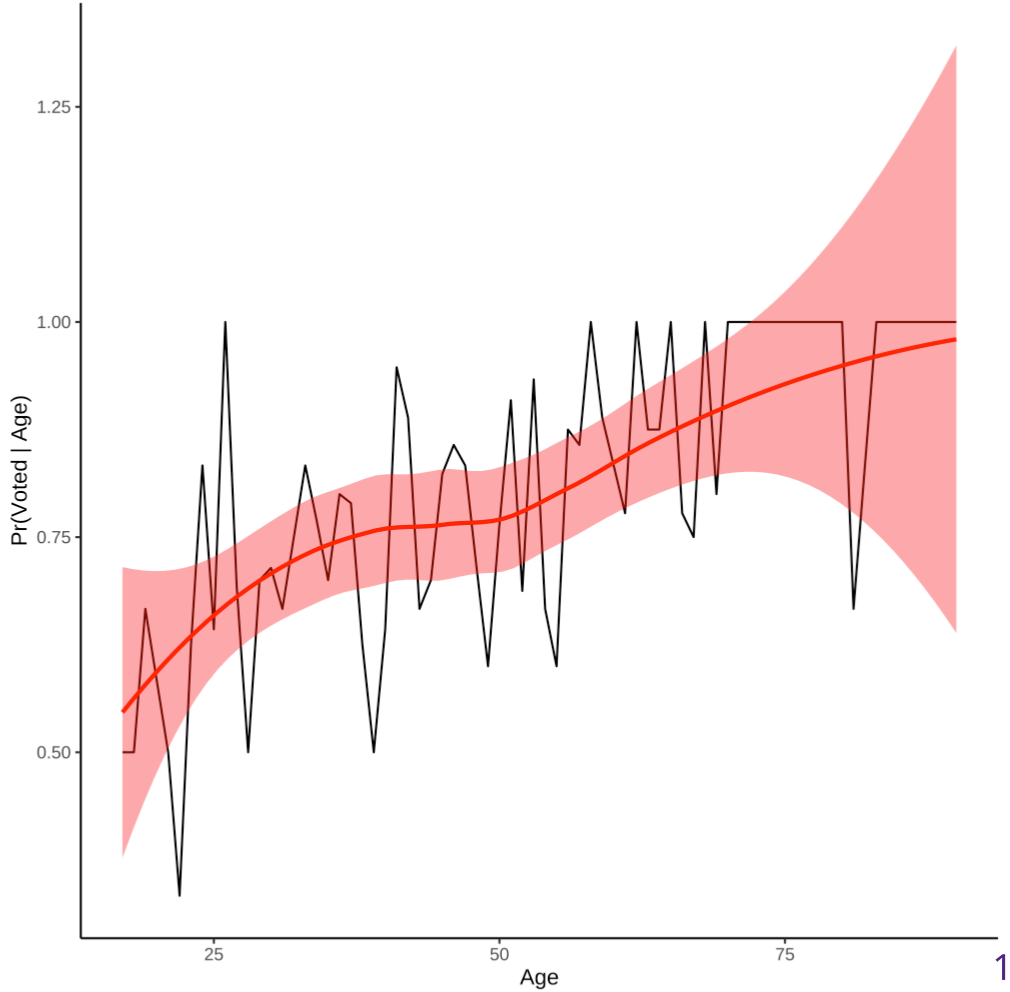
```
dat_ag <- dat %>%
  group_by(age) %>%
  summarise(turnout = mean(voted, na.rm=TRUE))
ggplot(dat_ag, aes(x=age, y=turnout)) +
  geom_line(col="black") +
  theme_classic() +
  labs(x="Age", y="Pr(Voted | Age)")
```



Smoothing out the Predictions

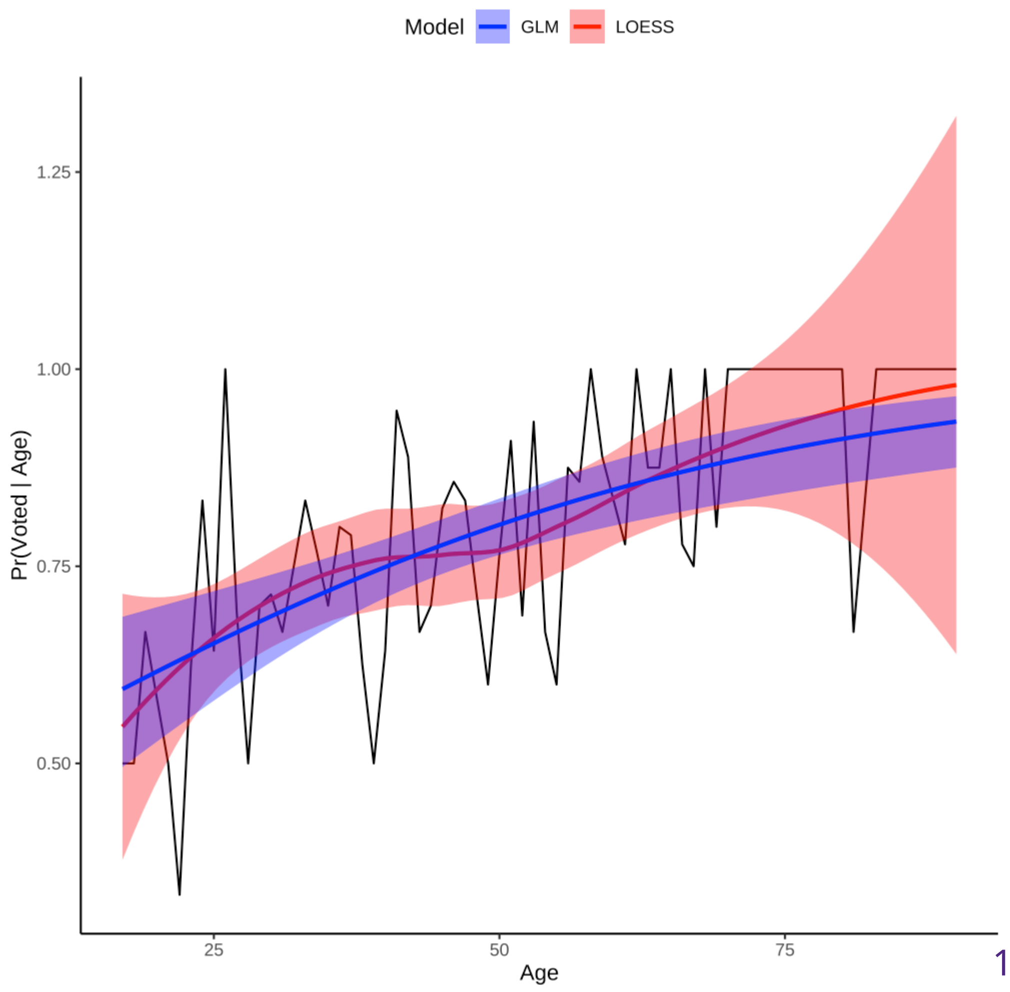
We could smooth out the predictions with a local polynomial regression.

```
ggplot() +
  geom_line(data = dat_ag,
            aes(x=age, y=turnout),
            col="black") +
  geom_smooth(data = dat,
              aes(x=age, y=voted),
              method="loess",
              se=TRUE,
              fill="red",
              color="red") +
  theme_classic() +
  labs(x="Age", y="Pr(Voted | Age)")
```



Logit Model

```r
ggplot() +
  geom_line(data = dat_ag,
            aes(x=age, y=turnout),
            col="black") +
  geom_smooth(data = dat,
              aes(x=age, y=voted,
                  fill="LOESS",
                  color="LOESS"),
              method="loess",
              se=TRUE) +
  geom_smooth(data = dat,
              aes(x=age, y=voted,
                  fill="GLM",
                  color="GLM"),
              method="glm",
              se=TRUE,
              method.args=list(family=binomial)) +
  scale_fill_manual(values=c("blue", "red")) +
  scale_colour_manual(values=c("blue", "red")) +
  theme_classic() +
  theme(legend.position="top") +
  labs(x="Age",
       y="Pr(Voted | Age)",
       colour="Model",
       fill="Model")
```

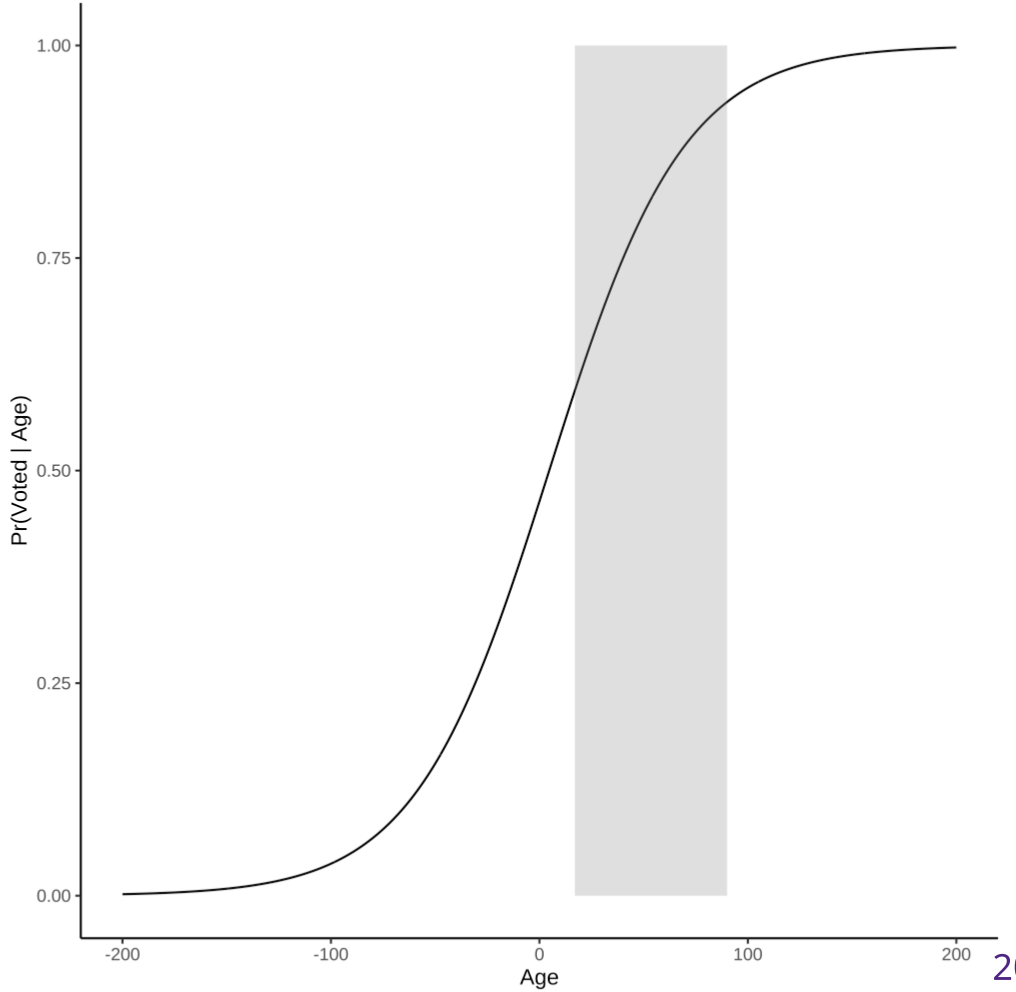


Where is the S-shaped Curve?

You might wonder, why isn't the effect shaped like an "S" as we discussed before?

- because the range of age only covers a small range of the s-shaped curve.

```
mod <- glm(voted ~ age, data=dat, family=binomial)
b <- mod$coef
s <- seq(-200, 200, length = 1000)
p <- plogis(b[1] + b[2] * s)
ggplot(mapping=aes(x=s, y=p)) +
  geom_line() +
  geom_polygon(mapping=aes(x=c(17,90,90,17,17),
                           y=c(0,0,1,1,0)),
               fill="gray50",
               alpha=.25) +
  theme_classic() +
  labs(x="Age", y="Pr(Voted | Age)")
```



Estimation

Estimation could be done in a number of ways, easiest is with the generalized linear model. In our normal linear model:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$
$$E(\mathbf{Y}) = \eta = \mathbf{X}\beta$$

The generalization, is:

$$g(\mu) = \eta = \mathbf{X}\beta$$

where $g(\cdot)$ is a "link function" that transforms the unbounded linear predictor into the response space of $\mathbf{Y}$.

GLMs

GLMs have 4 components:

1. Stochastic component: $\mathbf{Y}$ is a random or stochastic component that we expect to change from sample to sample in the frequentist thought experiment.
2. Systematic component: $\theta = \mathbf{X}\beta$
3. Link function: The stochastic and systematic components are linked through a function which "tricks" the model into thinking that it is still acting on normally distributed outcomes.
4. Residuals: The residuals can be computed the same way as in the linear model, but there are other, perhaps more useful options here, too.

Model of Voting

In our model of voting:

$$\mathbf{X}\beta = b_0 + b_1\text{Age} + b_2\text{Race=Black} + b_3\text{Race=Other}$$

$$log\left(\frac{Pr(\text{Voted}|\mathbf{X})}{1 - Pr(\text{Voted}|\mathbf{X})}\right) = \mathbf{X}\beta$$

$$Pr\widehat{(\text{Voted})} = \frac{e^{\mathbf{X}\beta}}{1 + e^{\mathbf{X}\beta}}$$

Estimated Voting Model

```r
mod <- glm(voted ~ age + race,
           data=dat,
           family=binomial(link="logit"))
```

```r
summary(mod)
```

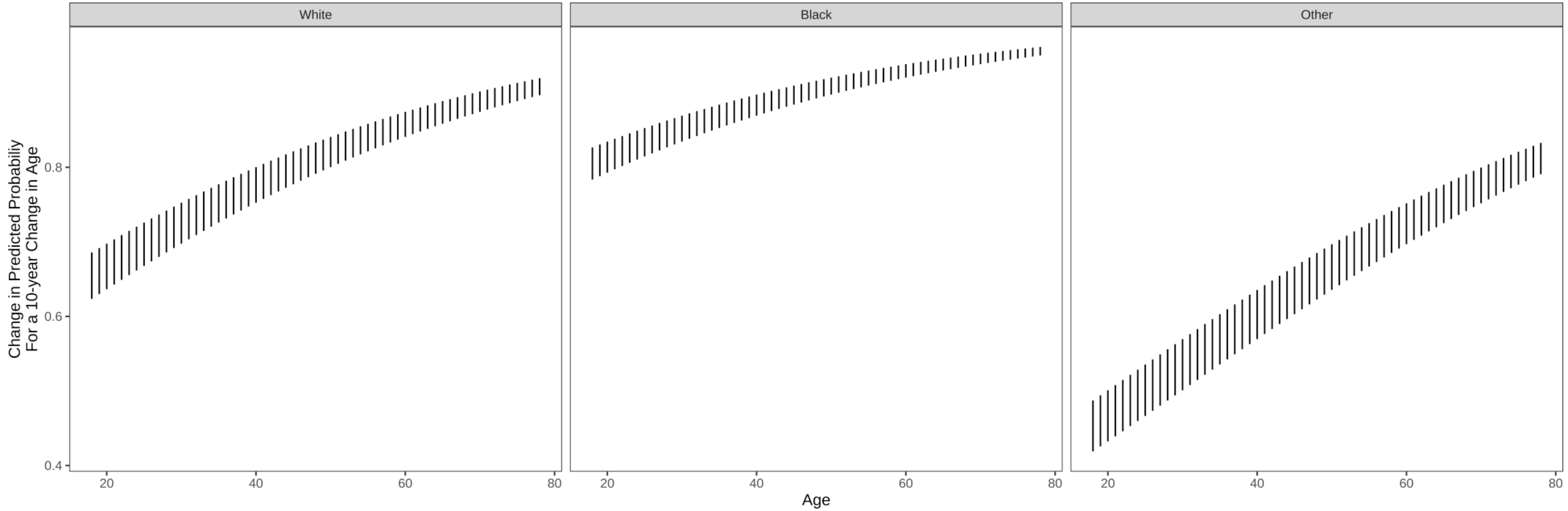
```
## 
## Call:
## glm(formula = voted ~ age + race, family = binomial(link = "logit"), 
##     data = dat)
## 
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)    
## (Intercept) -0.824407   0.365715  -2.254 0.024181 *  
## age          0.027611   0.007294   3.785 0.000154 ***
## raceWhite    0.831845   0.268921   3.093 0.001980 ** 
## raceBlack    1.613868   0.372594   4.331 1.48e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 629.10  on 583  degrees of freedom
## Residual deviance: 588.32  on 580  degrees of freedom
## AIC: 596.32
## 
## Number of Fisher Scoring iterations: 4
```

Interpreting Coefficients

We could interpret coefficients the same as in the linear model, but with a different dependent variable.

- For every one-unit change in age, the log of the odds of the probability of voting goes up by 0.028 holding race constant. *How enlightening!*
- This is often not an intuitive metric for your readers (or yourselves).
- And, it doesn't really tell us anything about the actual probability of voting.

Visual Display of Log-Odds $\rightarrow$ Probabilities



For all of these changes, the log odds ratio is the same:

$$\log\left[\left(\frac{\Pr(y=1|\text{Age}=a_0+10,\text{Race})}{1-\Pr(y=1|\text{Age}=a_0+10,\text{Race})}\right)\bigg/\left(\frac{\Pr(y=1|\text{Age}=a_0,\text{Race})}{1-\Pr(y=1|\text{Age}=a_0,\text{Race})}\right)\right]=0.27$$

What do OLS Coefficients Mean?

In the OLS context (assuming a continuous x), we could think about the the coefficients in two different ways.

$$y = b_0 + b_1 x + b_2 z + e$$

- Marginal Effect:

$$\frac{\partial E(y|x,z)}{\partial x} = b_1$$

- First Difference:

$$E(y|x = x_0 + 1, z = z_0) - E(y|x = x_0, z_0) = b_1$$

- Marginal Effect = First Difference
- Marginal Effect and First Difference are constant.

Marginal Effects and First Differences in the Logit Model

$$\log\left(\frac{\Pr(y=1|x,z)}{1-\Pr(y=1|x,z)}\right) = b_0 + b_1 x + b_2 z$$

Marginal Effect:

$$\frac{\partial E(y|x)}{\partial x} = b_1 f(b_0 + b_1 x + b_2 z)$$

First Difference:

$$\Delta_x = F(y|x = x_0 + 1, z = z) - F(y|x = x_0, z = z)$$

- Marginal Effect doesn't necessarily equal First Difference
- Neither Marginal Effect nor First Difference is constant.

Marginal Effects vs. First Differences: Age, Race Turnout.

Consider white respondents and the effect of a 10-year change age when Age is 25 vs when it is 85:

$$ME_{25} = 0.0276 \times 10 \times f(0.0074 + 0.0276 \times 25) = 0.06125$$

$$FD_{25} = F(0.0074 + 0.0276 \times 30) - F(0.0074 + 0.0276 \times 20) = .06118$$

$$ME_{85} = 0.0276 \times 10 \times f(0.0074 + 0.0276 \times 85) = 0.02188$$

$$FD_{85} = F(0.0074 + 0.0276 \times 90) - F(0.0074 + 0.0276 \times 80) = 0.02191$$

Taking Stock

What dowe know so far?

- The coefficients don't have a nice intuitive interpretation like they do in OLS.
- The coefficients do not necessarily indicate anything in particular about the absolute value of the predicted probability.
- Different one-unit changes can have different effects.

So, how do we characterize the *effect* of a variable in a single number?

Approaches to Identifying the Effect

- Hold all other variables constant at their means and calculate the first difference or marginal effect of x. [First Difference or Marginal Effect at Means]
- Hold all other variables constant at reasonable/representative and calculate the first difference or marginal effect of x. [First Difference or Marginal Effects at Reasonable values]
- Calculate the marginal effect or first difference for all observations and then average over all observations. [Average First Difference or Marginal Effects]

FD at Reasonable Values

Age:

```
library(marginaleffects)
comparisons(mod, newdata=datagrid(age = 40, race="White"), variables=list(age=10))
```

```
##
##  Term Contrast age  race Estimate Std. Error    z Pr(>|z|)    S  2.5 % 97.5 %
##   age      +10  40 White   0.0478     0.0127 3.75   <0.001 12.5 0.0228 0.0728
##
## Columns: rowid, term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high, age, race, predicted_lo, predicted_hi, pr
## Type:  response
```

Race:

```
comparisons(mod, newdata=datagrid(age=45), variables=list(race="pairwise"))
```

```
##
##  Term       Contrast age Estimate Std. Error    z Pr(>|z|)    S  2.5 % 97.5 %
##  race Black - Other  45    0.281     0.0644 4.37  < 0.001 16.3 0.1549  0.407
##  race Black - White  45    0.107     0.0363 2.94  0.00329  8.2 0.0356  0.178
##  race White - Other  45    0.174     0.0612 2.85  0.00439  7.8 0.0544  0.294
##   race
##  White
##  White
##  White
##
```

ME at Reasonable Values

```
comparisons(mod, newdata=datagrid(age = 40, race="White"), variables="age", comparison = "dydx")
```

```
##
##  Term age  race Estimate Std. Error    z Pr(>|z|)    S   2.5 % 97.5 %
##   age  40 White  0.00514    0.00146 3.53   <0.001 11.2 0.00228  0.008
##
## Columns: rowid, term, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high, age, race, predicted_lo, predicted_hi, predicted, v
## Type:  response
```

Marginal effects (first derivatives) only really make sense for continuous variables, so we shouldn't try to do that for `race`.

Average First Differences

Age:

```
avg_comparisons(mod, variables = list(age=10), comparison="difference")
```

```
##
##  Term  Contrast Estimate Std. Error    z Pr(>|z|)    S  2.5 % 97.5 %
##   age mean(+10)   0.0423     0.0102 4.17   <0.001 15.0 0.0224 0.0622
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high, predicted_lo, predicted_hi, predicted
## Type:  response
```

Race:

```
avg_comparisons(mod, variables = list(race="pairwise"))
```

```
##
##  Term                  Contrast Estimate Std. Error    z Pr(>|z|)    S  2.5 %
##  race mean(Black) - mean(Other)    0.276     0.0630 4.39  < 0.001 16.4 0.1530
##  race mean(Black) - mean(White)    0.107     0.0366 2.92  0.00348  8.2 0.0352
##  race mean(White) - mean(Other)    0.169     0.0594 2.85  0.00435  7.8 0.0530
##  97.5 %
##   0.400
##   0.179
##   0.286
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high, predicted_lo, predicted_hi, predicted
```

Average Marginal Effects

Age:

```r
avg_comparisons(mod, variables = "age", comparison="dydx")
```

```
##
##  Term    Contrast Estimate Std. Error    z Pr(>|z|)    S   2.5 % 97.5 %
##   age mean(dY/dX)  0.00453    0.00116 3.91   <0.001 13.4 0.00226 0.0068
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low, conf.high, predicted_lo, predicted_hi, predicted
## Type:  response
```

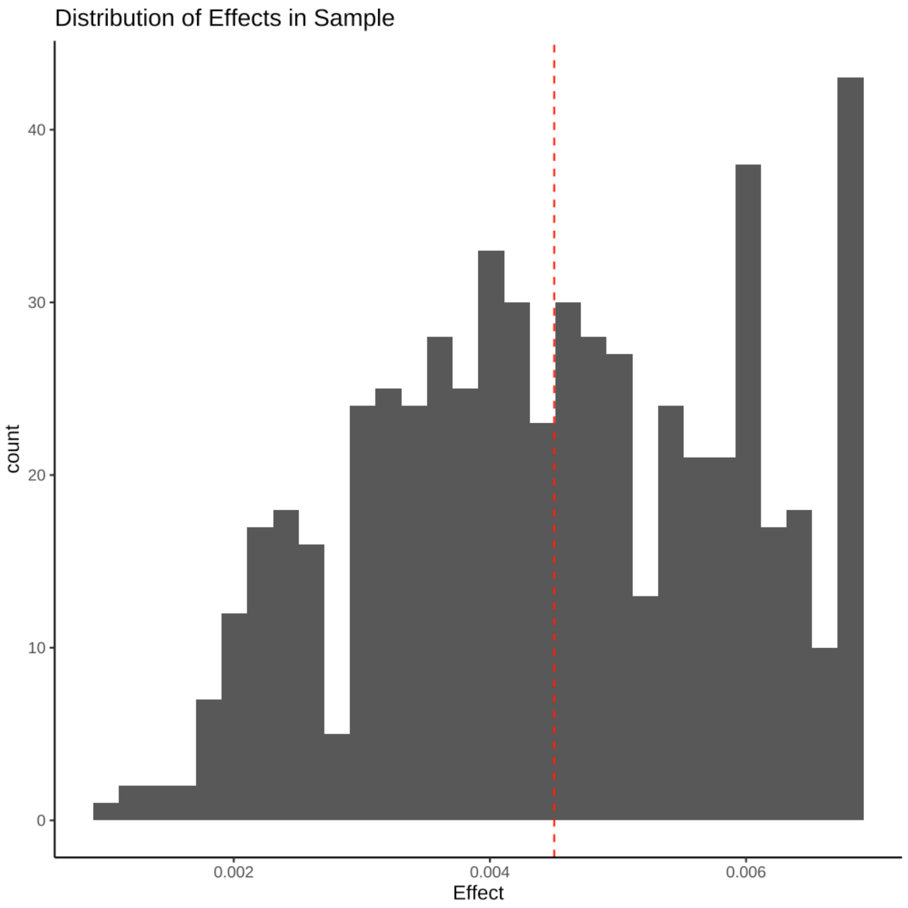
Distribution of Marginal Effects

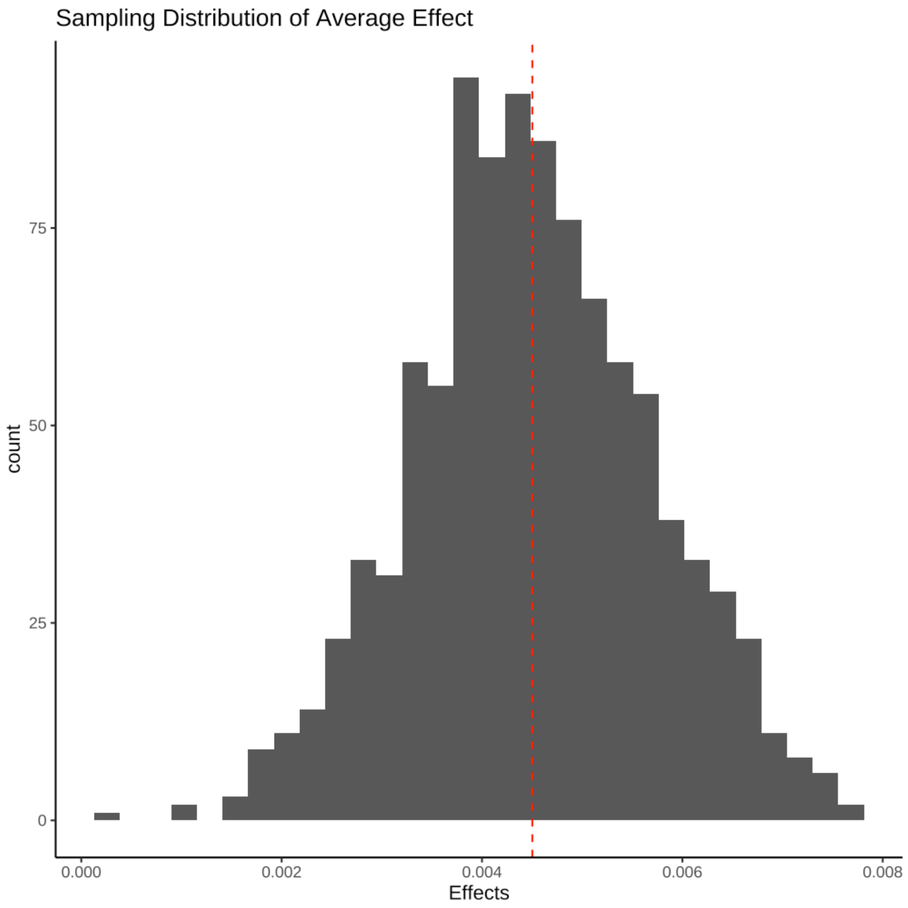
We might be interested in the distribution of marginal effects in two different senses.

- How much does the average marginal effect (or average discrete change) change as a function of sampling variability?
- How are marginal effects or discrete changes distributed in the sample?

The answers to these will be necessarily different.

Effect Distributions



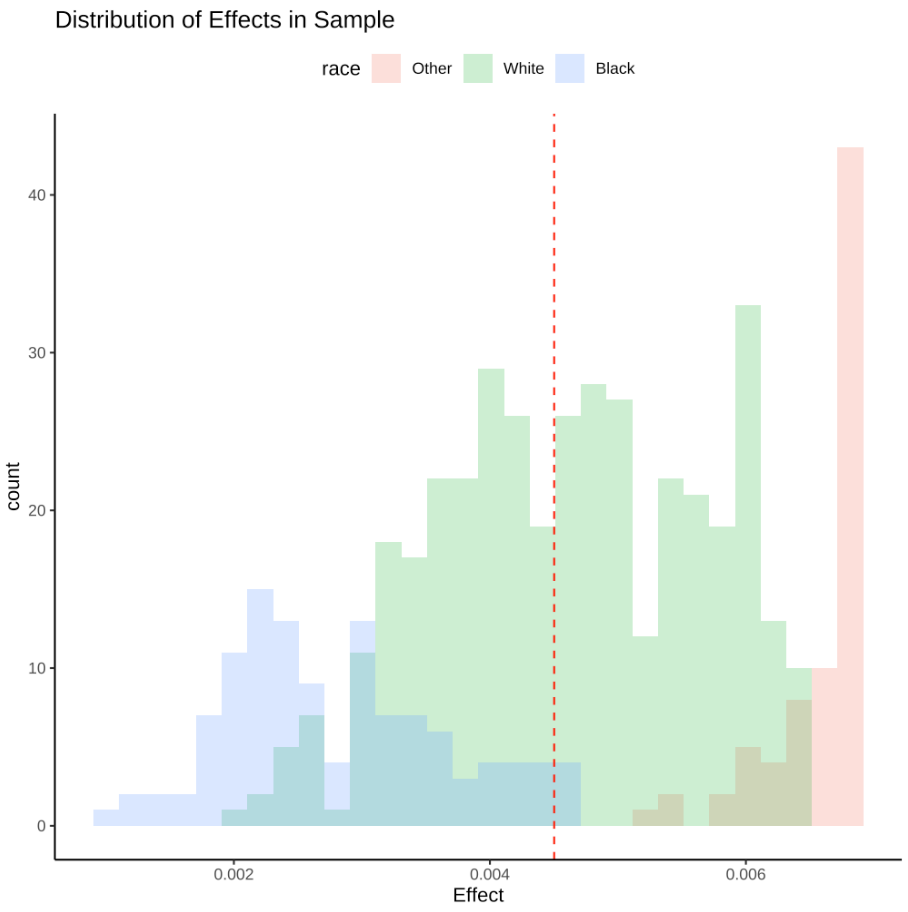


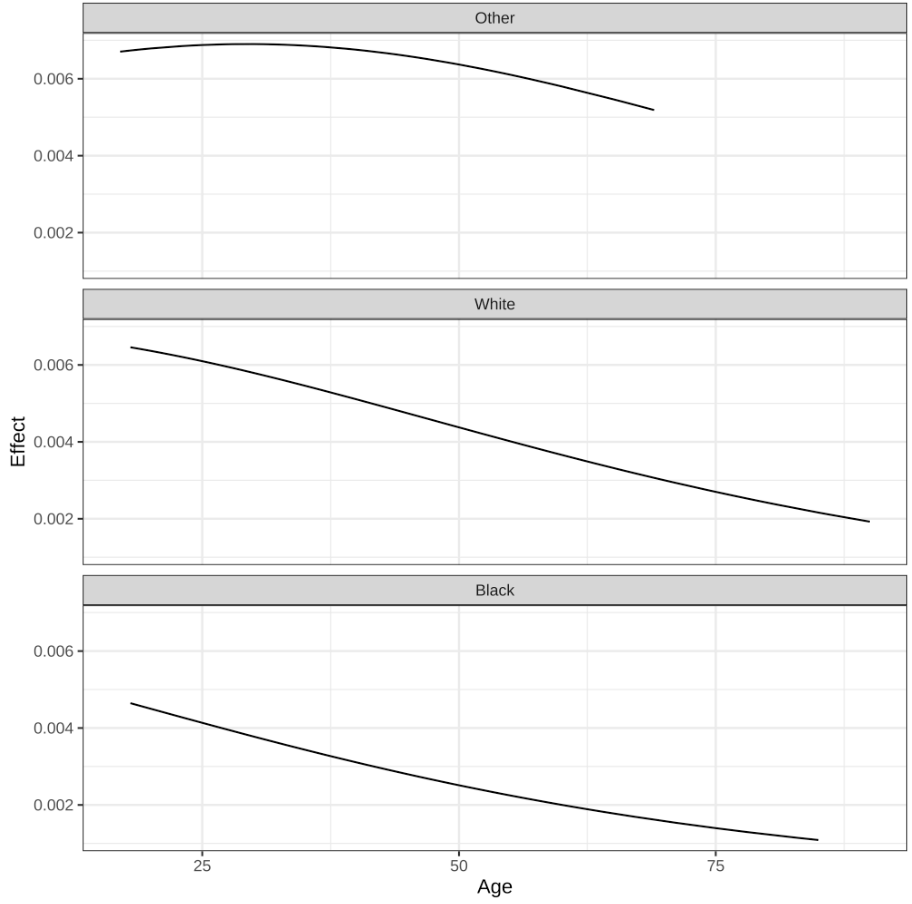
Considering the two Distributions

Note that the two distributions are quite different.

- The sampling distribution is, not surprisingly, approximately normal.
- The distribution in the sample looks skewed negative or maybe multimodal.

Effect Distributions by Race





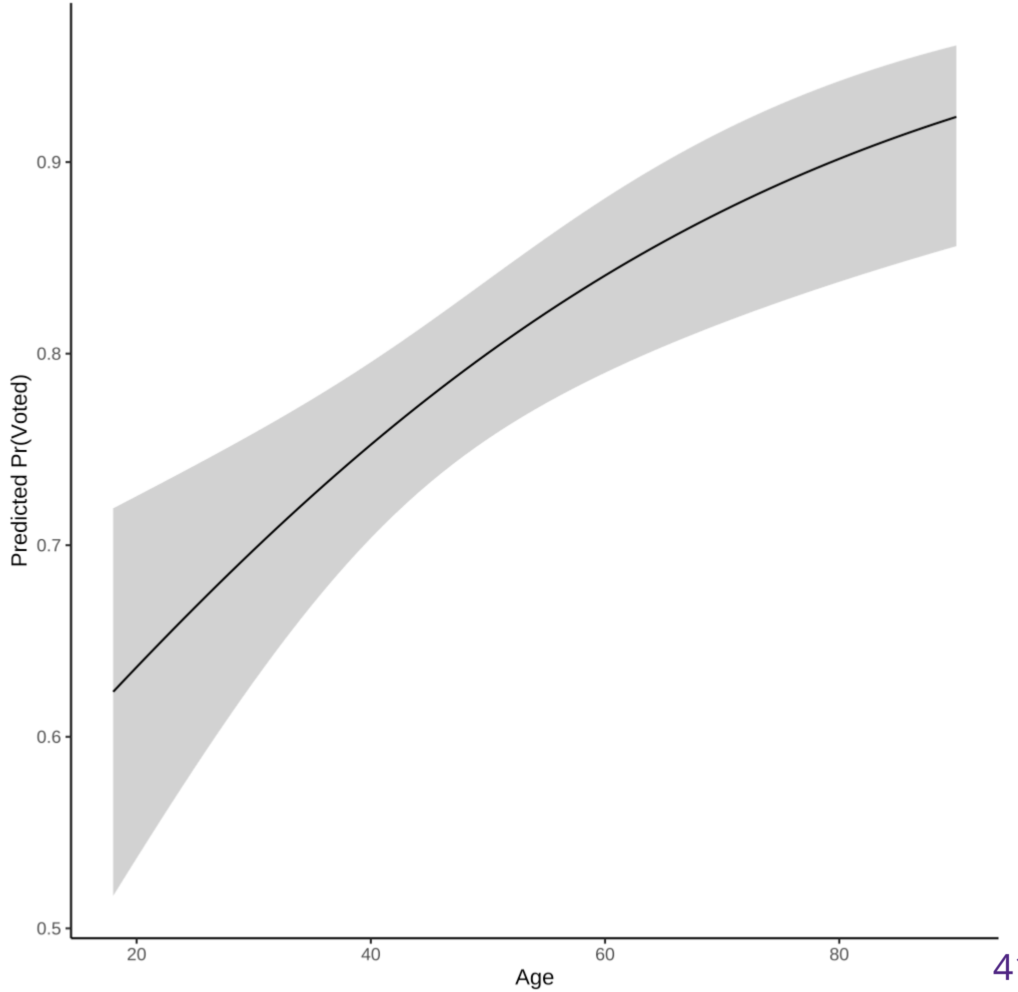
Effect Plot

Effect plots are a way of showing how the probability of the outcome of interest changes as you vary one variable over its range.

- You can use either an average effect or an effect at reasonable values approach.

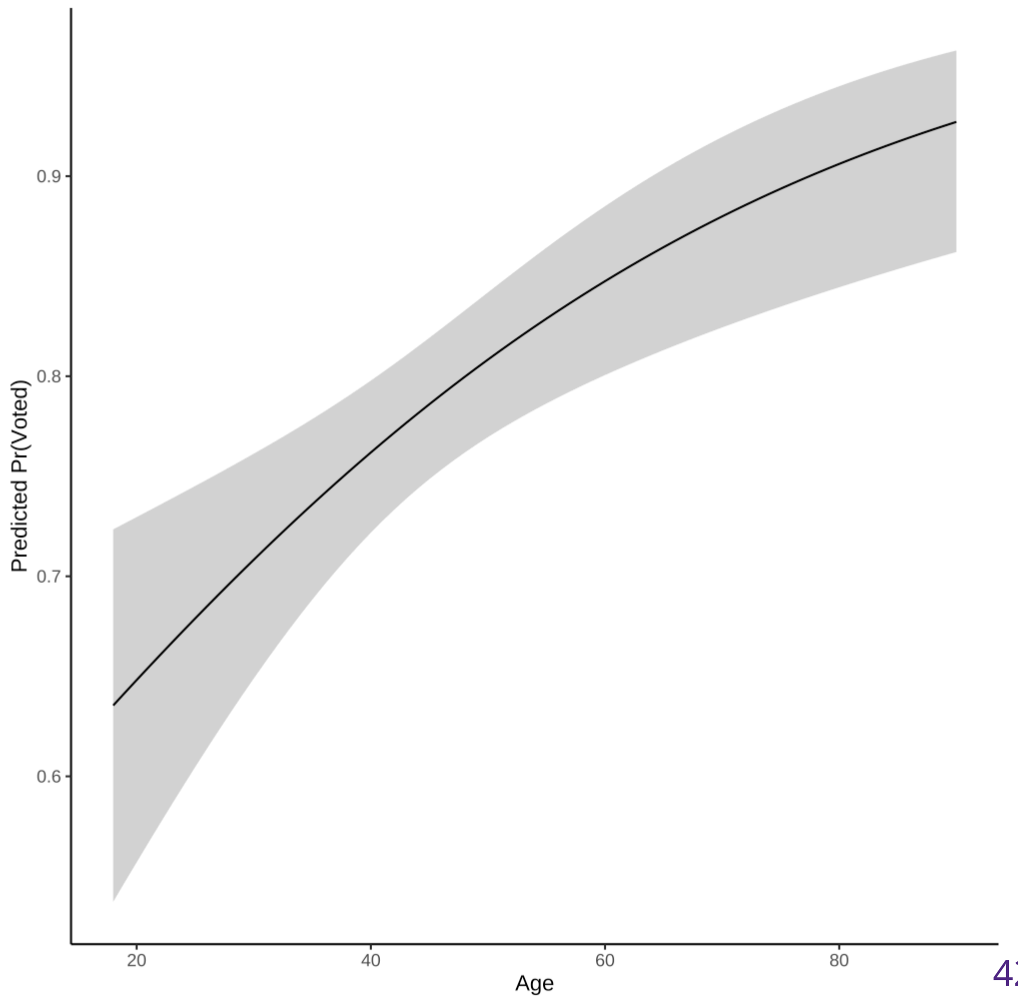
Effect Plot: RV Approach

```
p_age <- predictions(mod,
          newdata=datagrid(age=18:90))

ggplot(p_age, aes(x=age, y=estimate,
                  ymax=conf.high,
                  ymin=conf.low)) +
  geom_ribbon(alpha=.25) +
  geom_line() +
  theme_classic() +
  labs(x="Age", y="Predicted Pr(Voted)")
```



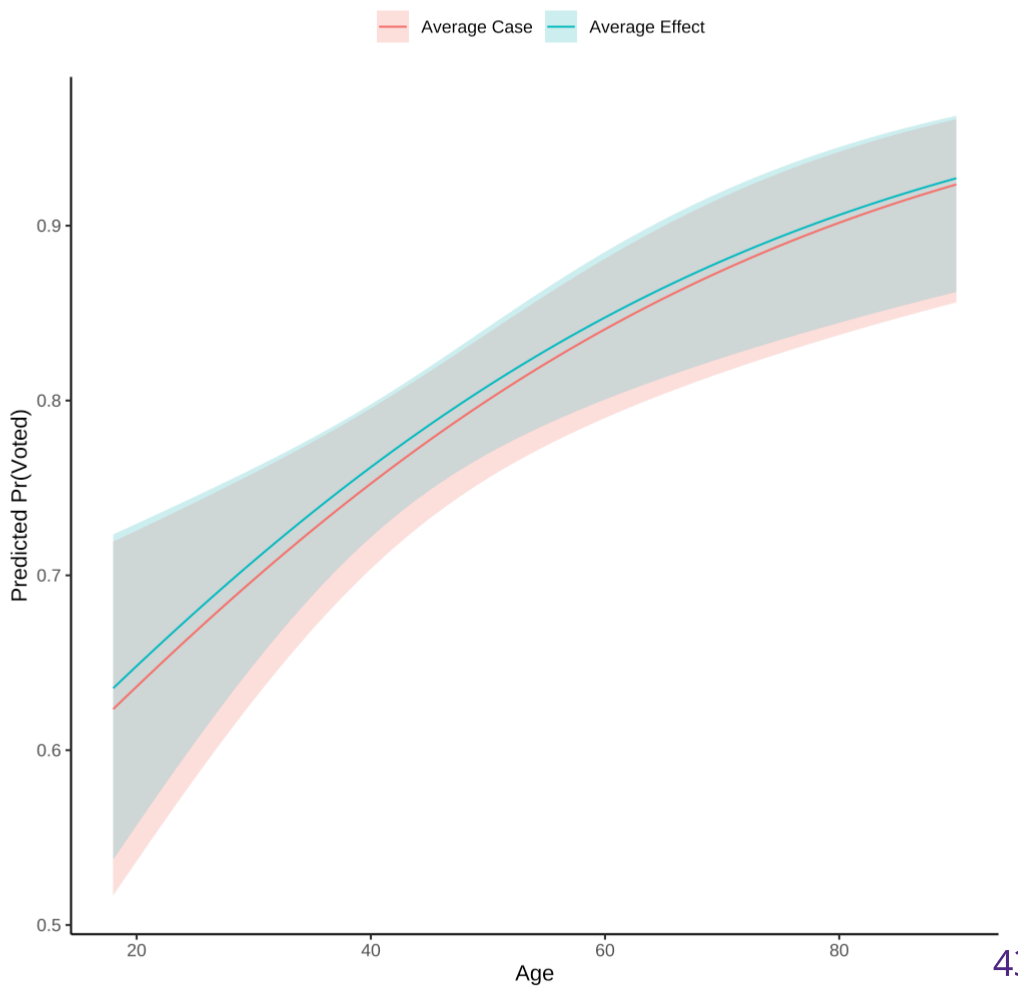
Effect Plot: AE Approach

```r
ap_age <- avg_predictions(mod,
          variables = list(age=18:90))

ggplot(ap_age, aes(x=age, y=estimate,
                  ymax=conf.high,
                  ymin=conf.low)) +
  geom_ribbon(alpha=.25) +
  geom_line() +
  theme_classic() +
  labs(x="Age", y="Predicted Pr(Voted)")
```



Comparison

```
age_both <- p_age %>%
  as.data.frame() %>%
  mutate(type="Average Case") %>%
  bind_rows(ap_age %>%
              as.data.frame() %>%
              mutate(type="Average Effect"))

ggplot(age_both, aes(x=age, y=estimate,
                     ymax=conf.high,
                     ymin=conf.low)) +
  geom_ribbon(aes(fill=type), alpha=.25) +
  geom_line(aes(color=type)) +
  theme_classic() +
  theme(legend.position="top") +
  labs(x="Age", y="Predicted Pr(Voted)",
       color="", fill="")
```



What Should You Present

- For variables of theoretical interest
  - A graph of predicted probabilities if the variable is continuous or a table or graph if it is categorical.
  - In the discussion, you can talk about the distribution of discrete changes if you like and identify important groups in that distribution.
- For variables that have theoretical importance to others.
  - You might want to present a first difference or two for variables others will find really interesting, particularly if you can show your theoretically important variable is more important.
- Make sure to present a table of descriptive statistics either in the paper or an appendix so people could calculate MEMs or MERs if they wanted to on their own.

Example Table

```r
tidy.comparisons <- function(x, ...){
  comps %>% select(term, estimate, std.error,
                   p.value, conf.low, conf.high)
}
registerS3method("tidy", "comparisons", tidy.comparisons)
comps <- avg_comparisons(mod) %>%
  mutate(term =c("age", "raceBlack", "raceWhite"))
f <- function(x) format(round(x, 3), big.mark=",")
gm <- list(
  list("raw" = "nobs", "clean" = "N", "fmt" = f),
  list("raw" = "logLik", "clean" = "LL", "fmt" = f),
  list("raw" = "aic", "clean" = "AIC", "fmt" = f),
  list("raw" = "bic", "clean" = "BIC", "fmt" = f))

modelsummary(
  list("GLM" = mod,
       "FD" = comps),
  estimate = c("{estimate}{stars}",
               "{estimate}"),
  stars = c("*" = .05),
  coef_map = c("age" = "Age",
               "raceBlack" = "Race: Black",
               "raceOther" = "Race: Other",
               "raceWhite" = "Race: White",
               "(Intercept)" = "Constant"),
  gof_map = gm,
  notes = "* p < 0.05 (two-tailed)",
  output = "flextable"
)
```

	GLM	FD
Age	0.028*	0.005
	(0.007)	(0.001)
Race: Black	1.614*	0.276
	(0.373)	(0.063)
Race: White	0.832*	0.169
	(0.269)	(0.059)
Constant	-0.824*	
	(0.366)	
N	584	
LL	-294.162	
AIC	596.323	
BIC	613.803	

* p < 0.05 (two-tailed)

Review

We covered the following topics:

1. Develop and Evaluate the Linear Probability Model
2. Describe the Generalized Linear Model Framework
3. Estimate GLMs for Binary Dependent Variables
4. Consider Different Methods of Describing Effects.
5. What Should You Present?

Exercises

If you load the file `ces0419.rda`, it will put an object in your workspace called `ces0419`, which has selected variables form the Canadian Election Study for the years 2004-2019. The variables in the data are described below. Estimate a model that predicts support for the Liberal party (you'll have to make a dummy variable indicating liberal vote vs vote for another party, first). Use whatever variables you like. Then answer the following questions:

1. Which variables are statistically significant in the model?
2. Pick 2 variables and find the first differences using both the First Differences at Reasonable Values and Average First Differences approaches.
3. Pick 2 variables and make effects graphs for them, again using both the Reasonable Values and Average Effects approaches.
4. What can does your model say about the likelihood of voting for the Liberal Candidate?