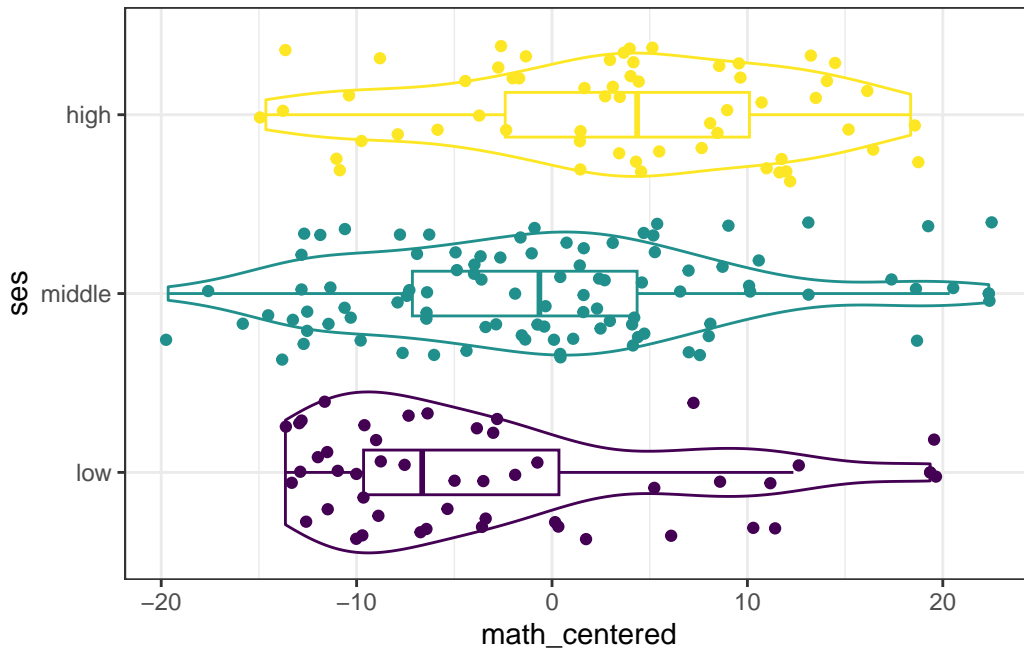


# Lab 9

## Part 1: Ordinal Models

Continuing from Tuesday's activity, we can also use the `hsb` dataset to fit ordinal regression models.

Consider using math scores to model the probability that a student's socioeconomic class is Low, Medium, or High.



We can use `MASS::polr` or `brms::brm` for ordinal regression we are modeling

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x$$

where  $Y \sim \text{Multinomial}(n, \pi)$  and  $\text{logit}[P(Y \leq J)] = \pi_1 + \dots + \pi_j$

```
ordinal_glm <- polr(ses ~ math_centered, data = hsb, method = 'logistic')
summary(ordinal_glm)
```

Re-fitting to get Hessian

Call:

```
polr(formula = ses ~ math_centered, data = hsb, method = "logistic")
```

Coefficients:

	Value	Std. Error	t value
math_centered	0.05736	0.01503	3.816

Intercepts:

	Value	Std. Error	t value
low middle	-1.2591	0.1728	-7.2872
middle high	0.9498	0.1611	5.8947

Residual Deviance: 405.8144

AIC: 411.8144

```
# ordinal_bayes <- brm(
#   ses ~ math_centered,
#   data = hsb,
#   family = cumulative(link = "logit"),
#   refresh = 0
# )
#
# print(ordinal_bayes)
```

- $\text{invlogit}(-1.25) = 0.2227001$  corresponds to the probability of the low class (at  $\text{math\_score} = 0$ )
- $\text{invlogit}(.94) = 0.7190997$  corresponds to the probability of the low or middle class (at  $\text{math\_score} = 0$ )
- with  $\beta$  positive (and having fairly small standard errors), an larger math scores imply higher probability of higher SES classes.

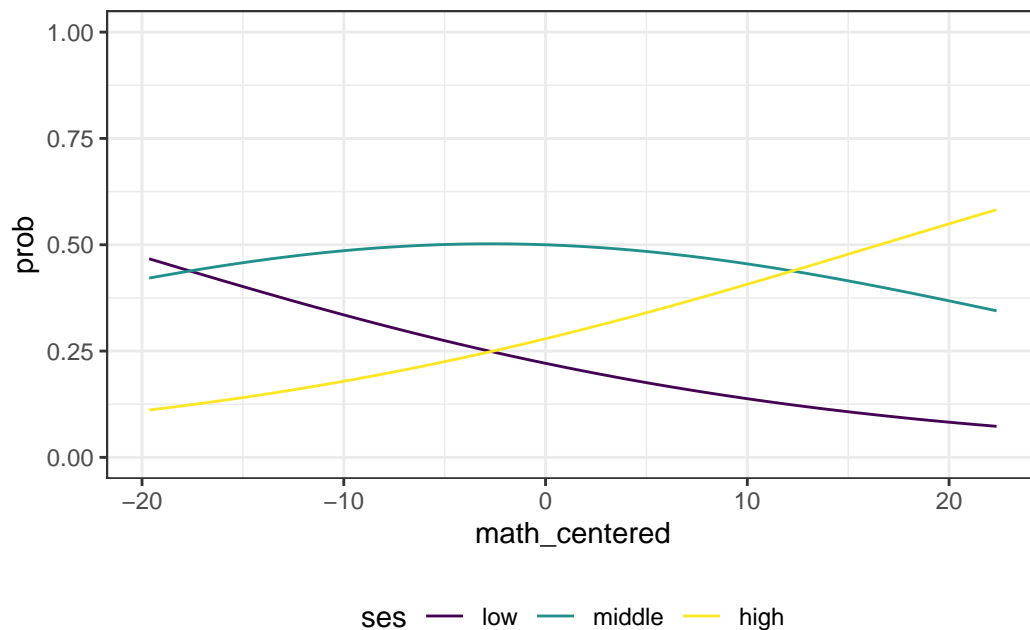
```

math_levels <- tibble(math_centered = seq(min(hsb$math_centered),
                                          max(hsb$math_centered), length.out = 50))

math_probs <- predict(ordinal_glm, newdata = math_levels, type = "probs")

melt(math_probs) |>
  bind_cols(math_centered = rep(math_levels$math_centered, 3)) |>
  mutate(ses = ordered(Var2, levels = c('low', 'middle', 'high'))) |>
  ggplot( aes(y = value, x = math_centered, color = ses )) + geom_line() +
  ylab('prob') +
  theme_bw() +
  theme(legend.position = 'bottom') +
  ylim(0,1)

```



## Lab 9 Questions

For today's lab, we will consider two variants of an ordinal regression model

1. An ordinal regression model with a linear relationship with a continuous predictor.
2. An ordinal regression model with a quadratic relationship with a continuous predictor.

### Question 1: Linear Ordinal

#### 1.1 (2 points)

Assume we fit the following model:

$$Y \sim Multinomial(n, \pi)$$
$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x$$
$$\text{logit}[P(Y \leq j)] = \pi_1 + \dots + \pi_j$$

Assume that  $\beta$  is positive, what do we expect for  $\pi_1$  and  $\pi_J$  in the limits ( $x \rightarrow \infty$ ) and ( $x \rightarrow -\infty$ )?

*In these situations, the probabilities will go to one and zero.*

#### 1.2 (4 points)

Assume the following characteristics and plot  $\pi_1 = Pr[Y \leq 1]$  against  $x$

- $x$  in range of  $[-5, 5]$
- $\alpha_{low|med} = -.5$
- $\alpha_{med|high} = 1$
- $\beta = .4$

```
n <- 200
x <- seq(-5, 5, length.out = n)
alpha_low_med <- -.5
alpha_med_high <- 1
beta <- .4
```

#### 1.3 (4 points)

Create two graphs (and include  $pi_{low}$  on both)

- $Pr[Y \leq med]$  against  $x$
- $\pi_{med}$  against  $x$

#### 1.4 (4 points)

Now add  $\pi_{hi}$  to the figure that includes (and include  $\pi_{low}$  and  $\pi_{med}$ )

#### 1.5 (4 points)

I've included code to simulate data from this framework. Fit the model and confirm that the estimated coefficients match what we'd expect. If not, discuss how they are different from what you'd expect.

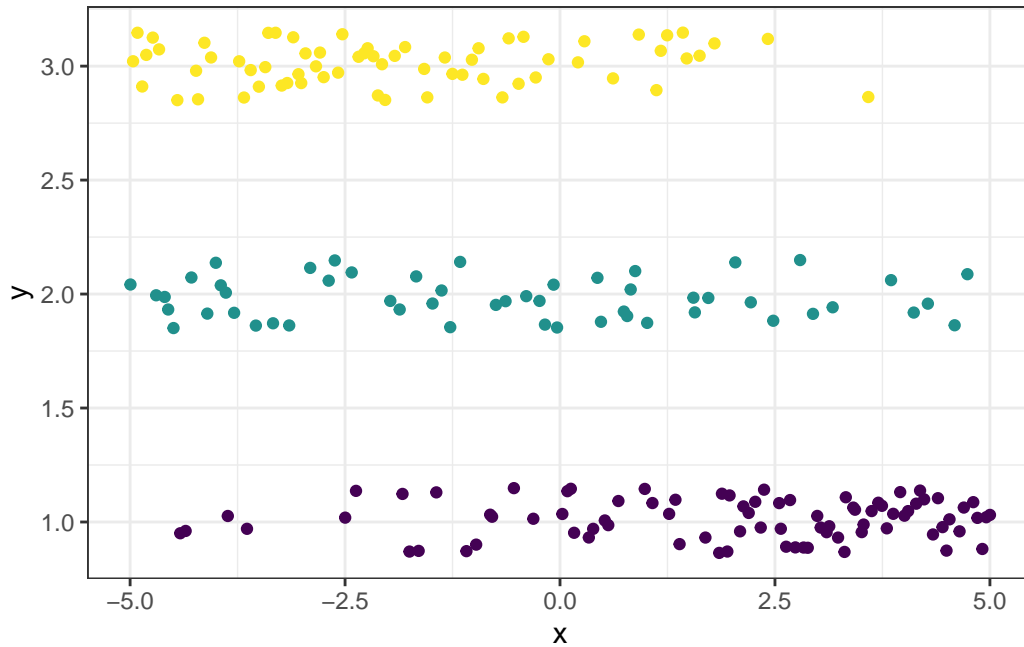
```
pi_values <- tibble(pi_1 = invlogit(alpha_low_med + x * beta ),
                   pi_2 = invlogit(alpha_med_high + x * beta ) -
                     invlogit(alpha_low_med + x * beta ),
                   pi_3 = 1 - invlogit(alpha_med_high + x * beta ),
                   x = x)

y_value <- rep('', n)

for (i in 1:n){
  y_value[i] <- sample(c('low','med','high'), size = 1, prob = pi_values[i,1:3], replace = F)
}

sim_data <- tibble(y_ord = y_value, x = x) |>
  mutate(y_cont = case_when(
    y_ord == 'low' ~ 1,
    y_ord == 'med' ~ 2,
    y_ord == 'high' ~ 3),
    y_ord = ordered(as.character(y_ord), levels = c('low','med','high')))

sim_data |>
  ggplot(aes(x = x, y = y_cont, color = y_ord)) +
  geom_jitter(height = .15) +
  theme_bw() +
  ylab('y') +
  theme(legend.position = 'none')
```



## Question 2: Quadratic Ordinal

### 2.1 (4 points)

Assume we fit the following model:

$$Y \sim \text{Multinomial}(n, \pi)$$

$$\text{logit}[P(Y \leq j)] = \alpha_j + \beta x + \gamma x^2$$

Assume the following characteristics and plot  $\pi_1 = Pr[Y \leq 1]$  against  $x$

- $x$  in range of  $[-5, 5]$
- $\alpha_{low|med} = -.5$
- $\alpha_{med|hi} = 1$
- $\beta = .4$
- $\gamma = -.3$

```
n <- 1000
x <- seq(-5, 5, length.out = n)
alpha_low_med <- -.5
alpha_med_high <- 1
beta <- .4
gamma <- -.3
```

## 2.2 (4 points)

Create two graphs (and include  $\pi_{low}$  on both)

- $Pr[Y \leq med]$  against  $x$
- $\pi_{med}$  against  $x$

## 2.3 (4 points)

Now add  $\pi_{high}$  to the figure that includes (and include  $\pi_{low}$  and  $\pi_{med}$ )