Lecture 12: Gelman Hill Ch 6

Generalized Linear Models

Generalized linear models are defined by three characteristics:

1. A probability distribution
2. A link function
3. Linear combination of predictors.

Two previous examples

- Normal Linear model:

$$\begin{aligned} y_i &\sim N(\mu_i, \sigma^2) \\ \mu_i &= X_i\underline{\beta} \end{aligned}$$

- Logistic Regression

$$\begin{aligned} y_i &\sim Binomial(n_i, p_i) \\ logit(p_i) &= X_i\underline{\beta} \end{aligned}$$

Other Examples

- Poisson Regression

$$\begin{aligned} y_i &\sim Poisson(\mu_i) \\ log(\mu_i) &= X_i\underline{\beta} \end{aligned}$$

- t Linear model:

$$\begin{aligned} y_i &\sim t(\mu_i, \sigma^2, \nu) \\ \mu_i &= X_i\underline{\beta} \end{aligned}$$

- Multinomial Regression (Ordinal Regression)

$$\begin{aligned} y_i &\sim Multinomial(1, \underline{p_i}) \\ \underline{p_i} &= \dots \end{aligned}$$

The takeaway is that we can use regression principles to model any type of data, assuming we can specify a probability distribution and a link function associated with a linear combination of predictors.

Overdispersion

Overdispersion, (too much dispersion), implies that the data has additional variance beyond what our model can capture.

Overdispersion is a common issue with GLMs, particularly when using a Poisson regression model for count data.

Sometimes this can be remedied by directly modeling additional correlation between observations using hierarchical models (mixed models), spatialtemporal structure, or multivariate methods.

Other times, a different data distribution is necessary to capture the variance in the data.

Count Regression

Poisson regression, assuming overdispersion is not an issue, provides a fairly intuitive interpretation of the coefficients.

The coefficients associated with continuous variables correspond to the expected difference in logarithms *or* the exponential of the coefficient corresponds to multiplicative increase in the response.

With overdispersion, set the family to be "quasipoisson" or as is typically done in a Bayesian framework, use a negative binomial - (this allows another term for the variance).

Latent Regression: Binary and Categorical

For binary data, consider a latent (unknown) continuous variable that is mapped to a binary output.

add sketch

If the latent variable, z is greater than 0, then $y = 1$; otherwise, $y = 0$.

This model can be fit using a probit link:

$$Pr[y_i = 1|x] = \Phi(X_i\underline{\beta})$$

, where Φ is the CDF of a standard normal.

The same idea can be used for ordinal regression, where the cutoffs are model parameters.

add sketch

Count Regression Demo

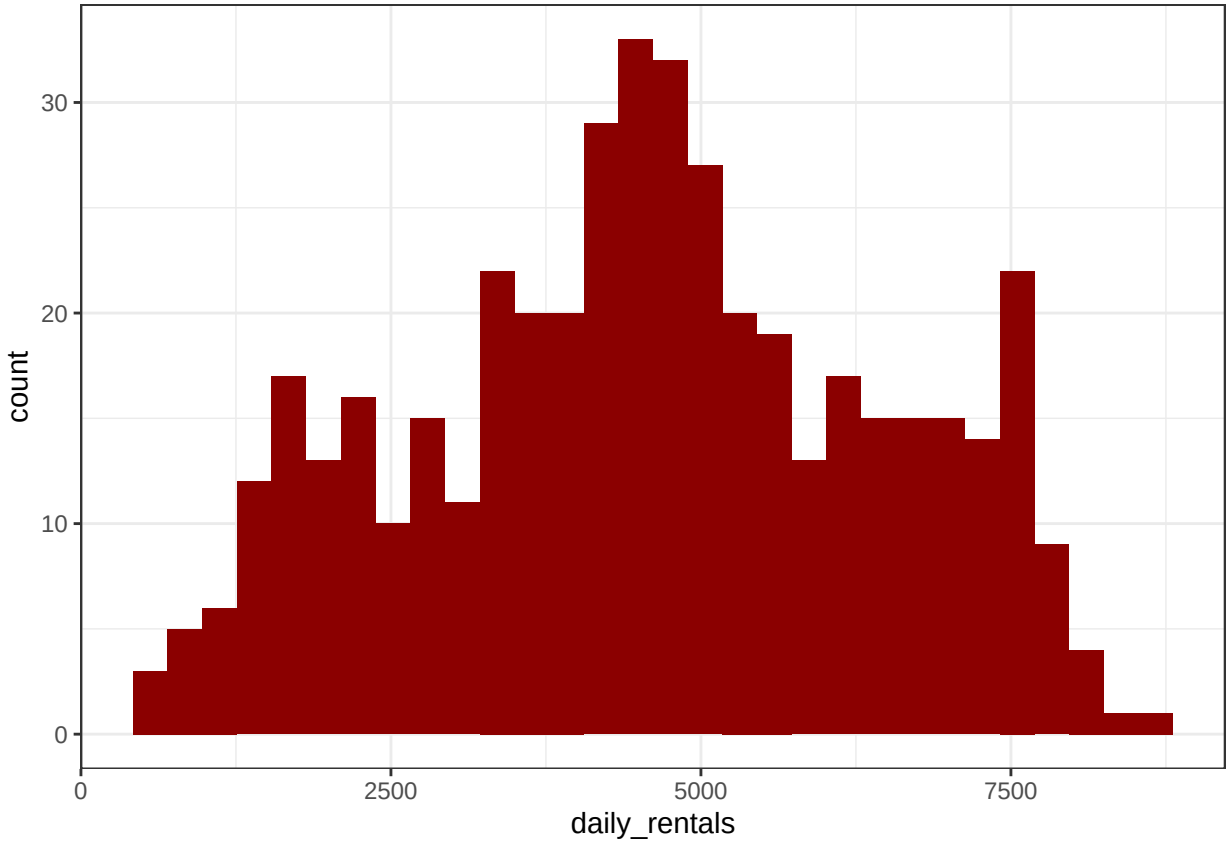
Model Formulation

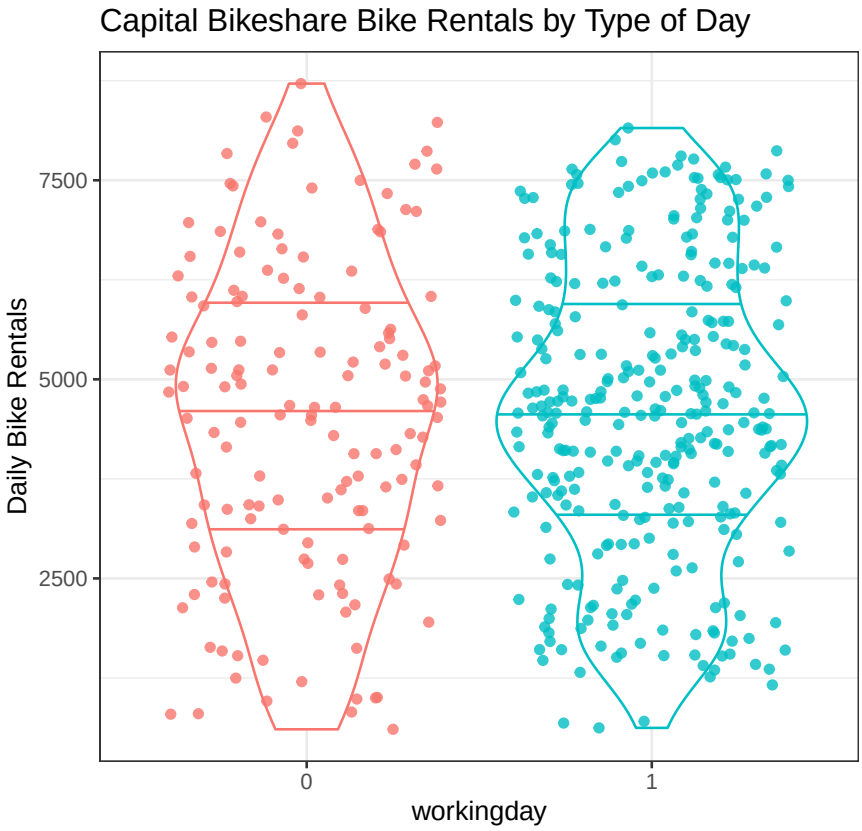
- Poisson Regression

$$\begin{aligned} y_i &\sim Poisson(\mu_i) \\ \log(\mu_i) &= X_i\underline{\beta} \\ \mu_i &= \exp(X_i\underline{\beta}) \end{aligned}$$

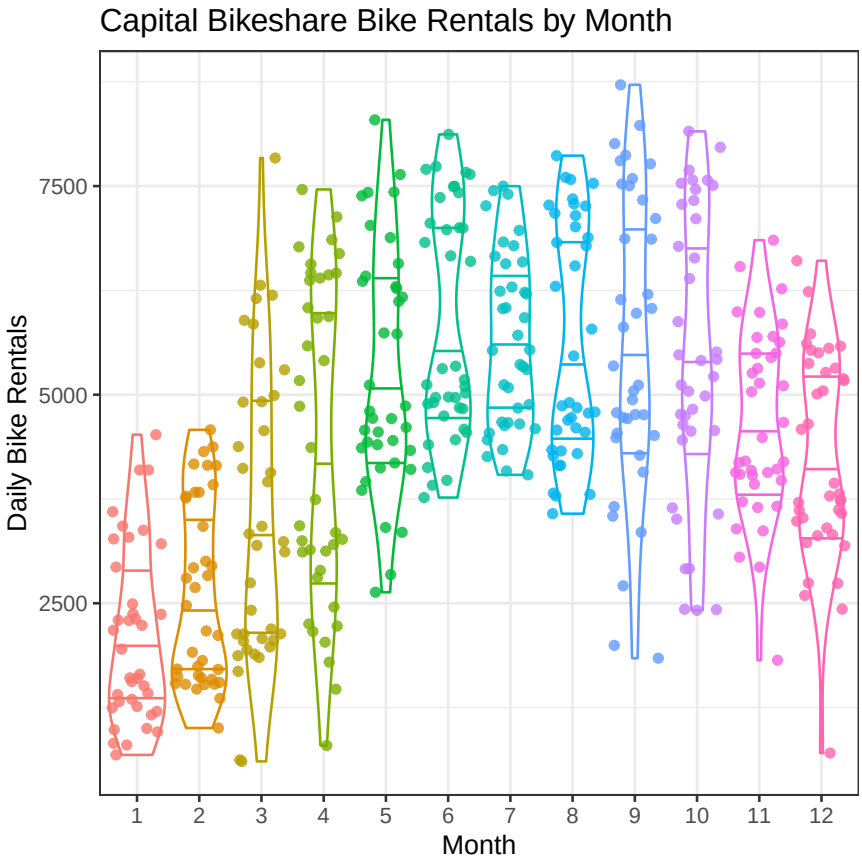
- workingday: potentially useful, not sure if more people would rent bikes to commute to work during the week or get around and sitesee on weekends
- max_temp: there is almost certainly a positive relationship between max_temp and bike rentals. However, it is possible that it might get too hot and bike rental would go down again.
- wind_speed: could be a proxy for inclement weather where high windspeeds results in lower bike rental counts
- month: almost certainly a seasonal pattern in bike rentals corresponding to month.
- interactions: it is possible that month may interact with other variables. For instance a warm day in spring/winter might lead to a different (slope) relationship for bike rentals than in the summer/fall.

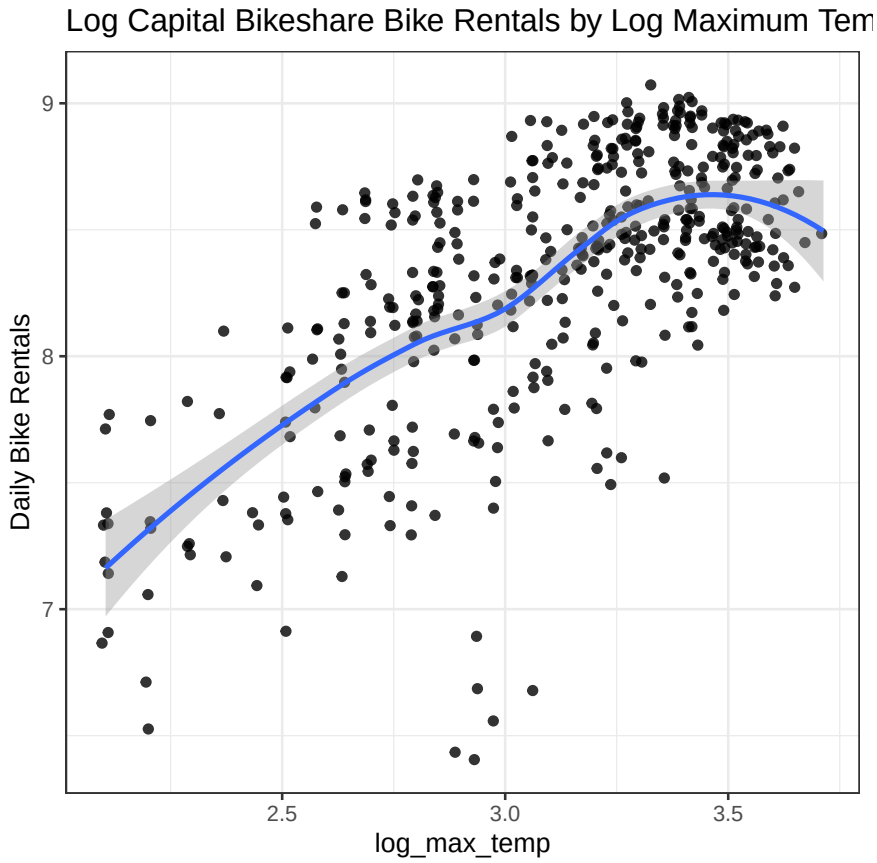
Data Viz

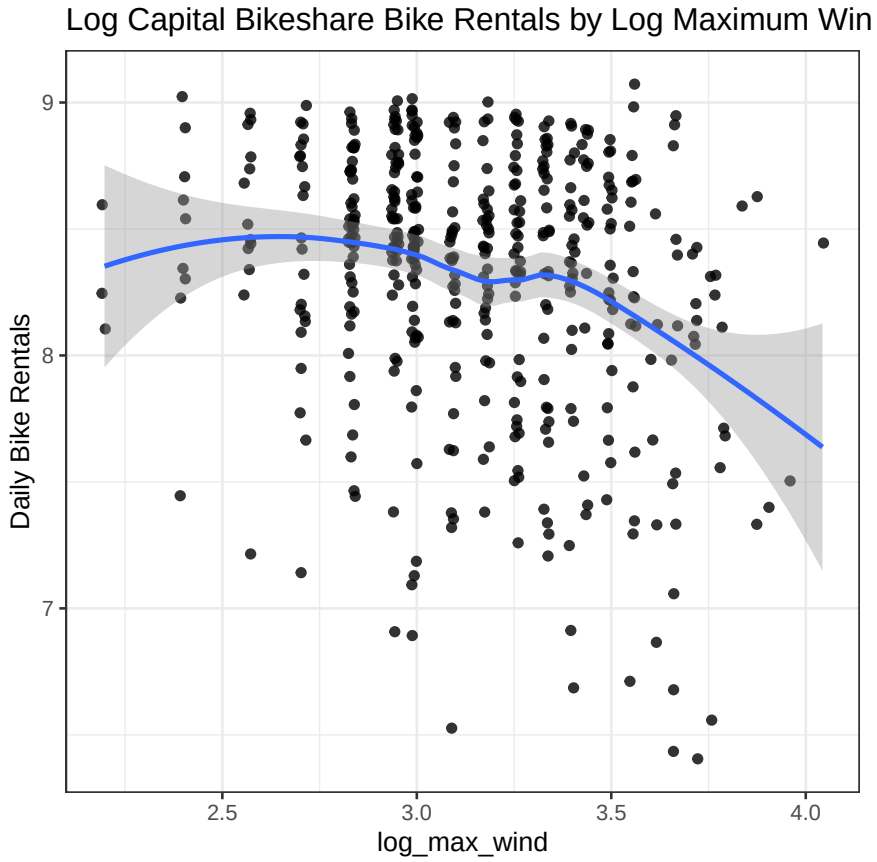




workingday = 1 implies a workday




Log Capital Bikeshare Bike Rentals by Log Maximum Tem
Daily Bike Rentals
9
8
7
2.5
3.0
3.5
log_max_temp


Log Capital Bikeshare Bike Rentals by Log Maximum Win
Daily Bike Rentals
9
8
7
2.5
3.0
3.5
4.0
log_max_wind

Model Fitting

Deviance is a summary of model fit

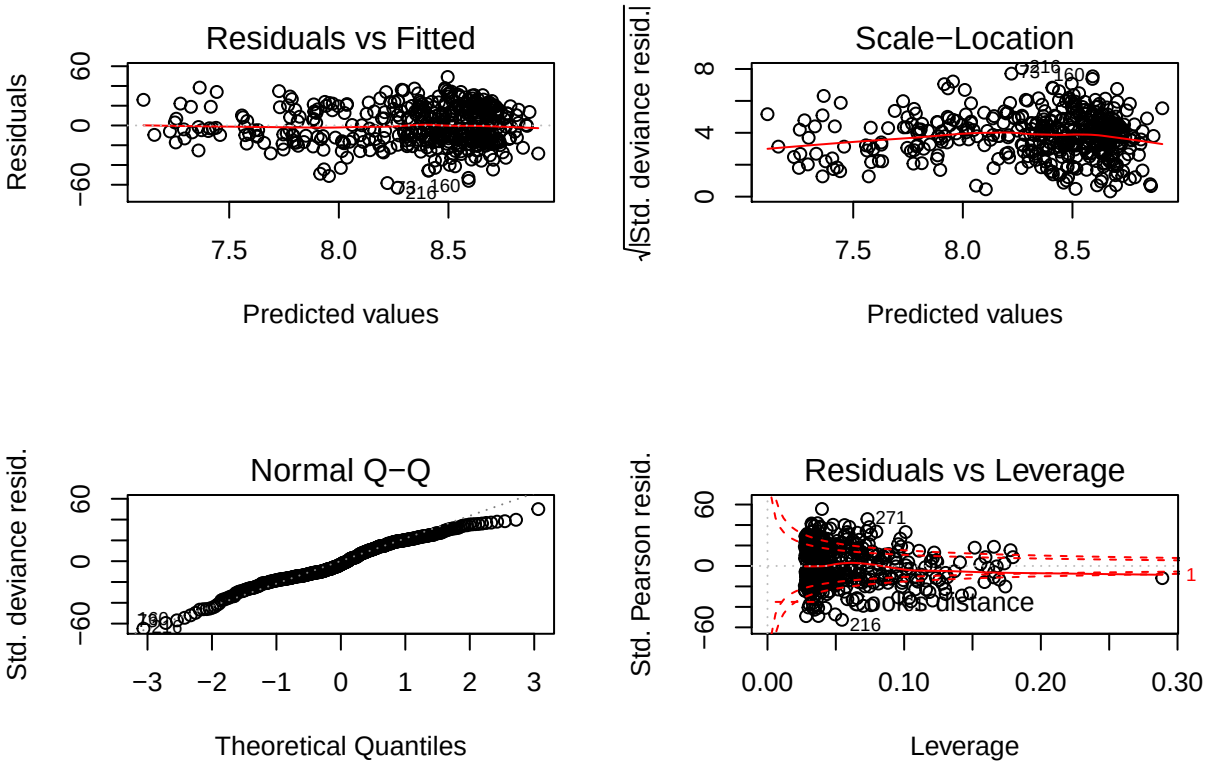
- Deviance is a measure of error; lower is better fit
- If predictor is noise deviance will decrease by 1 (on average)
- Useful predictors will decrease the deviance by more than 1.

```
## glm(formula = daily_rentals ~ month_val * log_temp_scale + workingday +
##     wind_scale, family = "poisson", data = bikes_daily)
##                              coef.est coef.se
## (Intercept)                   8.13     0.01
## month_val2                    0.05     0.01
## month_val3                    0.19     0.01
## month_val4                    0.26     0.01
## month_val5                    0.17     0.01
## month_val6                    0.84     0.01
## month_val7                    0.49     0.02
## month_val8                    0.19     0.02
## month_val9                    0.28     0.01
## month_val10                   0.40     0.01
## month_val11                   0.27     0.01
## month_val12                   0.36     0.01
## log_temp_scale                0.32     0.00
## workingday1                   0.00     0.00
## wind_scale                   -0.06     0.00
## month_val2:log_temp_scale    -0.05     0.01
## month_val3:log_temp_scale     0.17     0.01
## month_val4:log_temp_scale     0.06     0.01
## month_val5:log_temp_scale     0.26     0.01
## month_val6:log_temp_scale    -0.69     0.01
## month_val7:log_temp_scale    -0.31     0.01
## month_val8:log_temp_scale    -0.04     0.01
## month_val9:log_temp_scale    -0.03     0.01
## month_val10:log_temp_scale   -0.06     0.01
## month_val11:log_temp_scale   -0.38     0.01
## month_val12:log_temp_scale   -0.09     0.01
## ---
##   n = 456, k = 26
##   residual deviance = 166958.1, null deviance = 378784.4 (difference = 211826.2)
```

Coefficient Interpretation:

- Intercept: The intercept is the mean response for the response (on the log scale), when all of the predictors equal zero *or* the exponential of the the intercept is the mean response when all of the predictors are zero.
- Other Coefficients: The coefficients are the expected difference in the response (on the log scale) for each additional unit of the predictor, while holding all of the other predictors constant. The exponentiated coefficient is the expected multiplicative increase for each additional unit of the predictor while holding all of the other variable constant.

Residual Figures



Test for Overdispersion

```
##
##  Overdispersion test
##
## data:  pois_month_logtemp
## z = 17.291, p-value < 2.2e-16
## alternative hypothesis: true alpha is greater than 0
## sample estimates:
##    alpha
## 353.5494
```

Quasipoisson

The quasipoisson is a way include an additional term that account for extra variation, and includes appropriate standard errors.

```
## glm(formula = daily_rentals ~ month_val * log_temp_scale + workingday +
##     wind_scale, family = "quasipoisson", data = bikes_daily)
##                              coef.est coef.se
## (Intercept)                  8.13     0.13
## month_val2                   0.05     0.16
## month_val3                   0.19     0.14
## month_val4                   0.26     0.13
## month_val5                   0.17     0.15
## month_val6                   0.84     0.18
## month_val7                   0.49     0.29
## month_val8                   0.19     0.32
```

```
## month_val9                  0.28     0.16
## month_val10                 0.40     0.13
## month_val11                 0.27     0.14
## month_val12                 0.36     0.15
## log_temp_scale              0.32     0.08
## workingday1                 0.00     0.03
## wind_scale                 -0.06     0.01
## month_val2:log_temp_scale  -0.05     0.11
## month_val3:log_temp_scale   0.17     0.11
## month_val4:log_temp_scale   0.06     0.13
## month_val5:log_temp_scale   0.26     0.18
## month_val6:log_temp_scale  -0.69     0.16
## month_val7:log_temp_scale  -0.31     0.23
## month_val8:log_temp_scale  -0.04     0.27
## month_val9:log_temp_scale  -0.03     0.16
## month_val10:log_temp_scale -0.06     0.13
## month_val11:log_temp_scale -0.38     0.13
## month_val12:log_temp_scale -0.09     0.13
## ---
##   n = 456, k = 26
##   residual deviance = 166958.1, null deviance = 378784.4 (difference = 211826.2)
##   overdispersion parameter = 376.0
```

Negative Binomial

The negative binomial distribution is another alternative for modeling count data. The negative binomial distribution has two parameters, one can account for the dispersion of the data.

Ordinal Regression

```
## Parsed with column specification:
## cols(
##   age = col_double(),
##   body_type = col_character(),
##   diet = col_character(),
##   drinks = col_character(),
##   drugs = col_character(),
##   ethnicity = col_character(),
##   height = col_double(),
##   job = col_character(),
##   sex = col_character(),
##   smokes = col_character()
## )
```

```
## # A tibble: 6 x 3
##   drinks_order     n proportion
##   <ord>        <int>      <dbl>
## 1 not at all    1547    0.0704
## 2 rarely        2525    0.115
## 3 socially     15680    0.714
## 4 often         1866    0.0849
## 5 very often     227    0.0103
## 6 desperately    129    0.00587
```

```
##
## Re-fitting to get Hessian
```

```
## polr(formula = drinks_order ~ 1, data = OkCupid, method = "probit")
##                         coef.est coef.se
## not at all|rarely       -1.47     0.01
## rarely|socially         -0.90     0.01
## socially|often           1.28     0.01
## often|very often         2.14     0.02
## very often|desperately   2.52     0.03
## ---
## n = 21974, k = 5 (including 5 intercepts)
## residual deviance = 42324.4, null deviance is not computed by polr
```

Some probabilities

```
pnorm(-1.47)
```

```
## [1] 0.07078088
```

```
pnorm(-.895) - pnorm(-1.47)
```

```
## [1] 0.1146127
```

```
pnorm(1.27) - pnorm(-.895)
```

```
## [1] 0.7125641
```

Ordinal Regression

```
##
## Re-fitting to get Hessian

## polr(formula = drinks_order ~ age_scale, data = OkCupid, method = "probit")
##                         coef.est coef.se
## age_scale               -0.13     0.01
## not at all|rarely       -1.49     0.01
## rarely|socially         -0.90     0.01
## socially|often           1.29     0.01
## often|very often         2.16     0.02
## very often|desperately   2.54     0.03
## ---
## n = 21974, k = 6 (including 5 intercepts)
## residual deviance = 42041.0, null deviance is not computed by polr
```

Now the intercept thresholds represent an average age. However as age increases, the latent variables are going to shift to the left (or toward less drinking).