PLSC 503 – Spring 2020
Ordinal Outcomes

April 16, 2020

Ordinal Data

Ordinal data are:

- Discrete: $Y \in \{1, 2, ...\}$
- *Grouped Continuous* Data
- *Assessed Ordered* Data

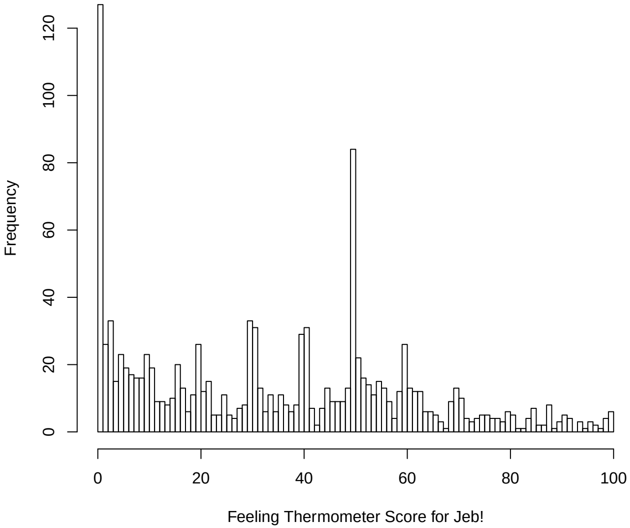
In general:

- Some things can be ordered, but shouldn't be
- Some things are ordered in some circumstances but not others
- Orderings can differ across applications

Ordinal vs. Continuous Response Models

"I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 and 100 degrees mean that you feel favorably and warm toward the person; ratings between 0 and 50 degrees mean that you don't feel favorably toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person."

Thermometer Scores for Jeb! (2016)



A Fake-Data Example

$$Y_i^* = 0 + 1.0X_i + u_i,$$

$$X_i \sim U[0, 10]$$

$$u_i \sim N(0, 1)$$

$$\begin{aligned} Y_{1i} &= 1 \text{ if } Y_i^* < 2.5 \\ &= 2 \text{ if } 2.5 \leq Y_i^* < 5 \\ &= 3 \text{ if } 5 \leq Y_i^* < 7.5 \\ &= 4 \text{ if } Y_i^* > 7.5 \end{aligned}$$

$$\begin{aligned} Y_{2i} &= 1 \text{ if } Y_i^* < 2 \\ &= 2 \text{ if } 2 \leq Y_i^* < 8 \\ &= 3 \text{ if } 8 \leq Y_i^* < 9 \\ &= 4 \text{ if } Y_i^* > 9 \end{aligned}$$

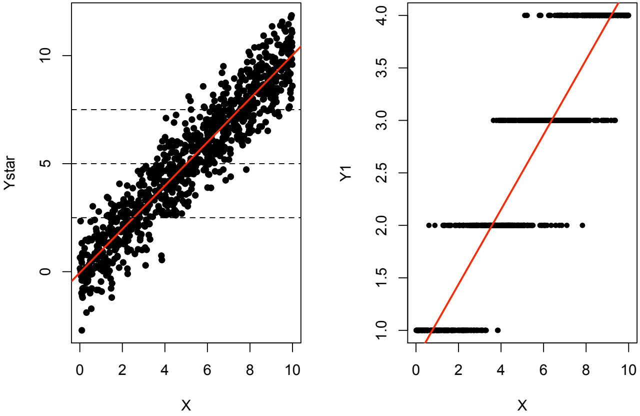
World's Best Regression

```
> summary(lm(Ystar~X))

Call:
lm(formula = Ystar ~ X)

Residuals:
   Min     1Q Median     3Q    Max
-3.006 -0.654 -0.049  0.643  3.298

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)  -0.0830     0.0609   -1.36                 0.17
X             1.0110     0.0106   95.48 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.988 on 998 degrees of freedom
Multiple R-squared:  0.901,Adjusted R-squared:  0.901
F-statistic: 9.12e+03 on 1 and 998 DF,  p-value: <0.0000000000000002
```

Also A Pretty Good Regression

```
> summary(lm(Y1~X))

Call:
lm(formula = Y1 ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2889 -0.2439  0.0158  0.2592  1.3968

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)  0.69979    0.02639    26.5 <0.0000000000000002 ***
X            0.35825    0.00459    78.0 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.428 on 998 degrees of freedom
Multiple R-squared:  0.859,Adjusted R-squared:  0.859
F-statistic: 6.09e+03 on 1 and 998 DF,  p-value: <0.0000000000000002
```

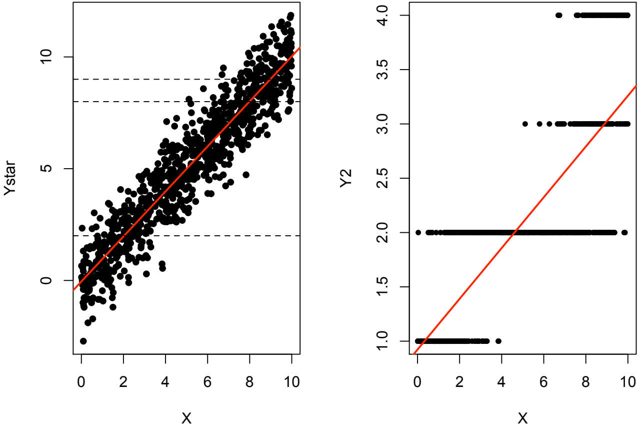
What That Looks Like



A Not-So-Good Regression

```
> summary(lm(Y2~X))

Call:
lm(formula = Y2 ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3115 -0.3205 -0.0405  0.2914  1.4876

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)  0.88919    0.03069    29.0 <0.0000000000000002 ***
X            0.24383    0.00534    45.7 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.498 on 998 degrees of freedom
Multiple R-squared:  0.676,Adjusted R-squared:  0.676
F-statistic: 2.09e+03 on 1 and 998 DF,  p-value: <0.0000000000000002
```

What That Looks Like



Models for Ordinal Responses

$$Y_i^* = \mu + u_i$$

$$Y_i = j \text{ if } \tau_{j-1} \leq Y_i^* < \tau_j,\ j \in \{1, ...J\}$$

$$\begin{aligned} Y_i &= 1 \text{ if } -\infty \leq Y_i^* < \tau_1 \\ &= 2 \text{ if } \tau_1 \leq Y_i^* < \tau_2 \\ &= 3 \text{ if } \tau_2 \leq Y_i^* < \tau_3 \\ &= 4 \text{ if } \tau_3 \leq Y_i^* < \infty \end{aligned}$$

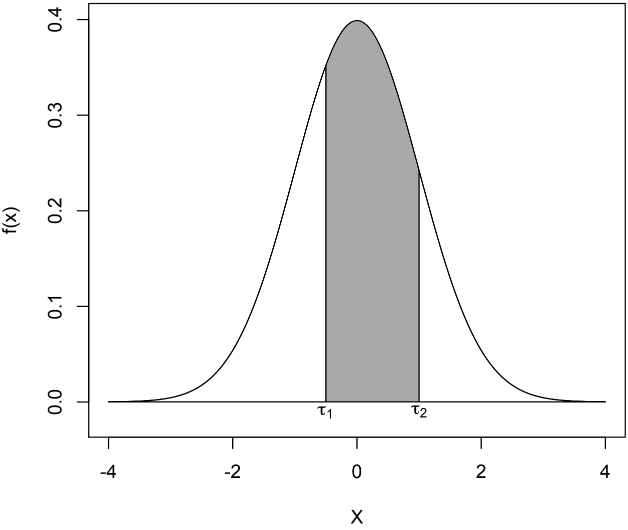
Ordinal Response Models: Probabilities

$$
\begin{aligned}
\Pr(Y_i = j) &= \Pr(\tau_{j-1} \leq Y^* < \tau_j) \\
&= \Pr(\tau_{j-1} \leq \mu_i + u_i < \tau_j) \qquad (1)
\end{aligned}
$$

$$\mu_i = \mathbf{X}_i\boldsymbol{\beta}$$

$$
\begin{aligned}
\Pr(Y_i = j|\mathbf{X}, \boldsymbol{\beta}) &= \Pr(\tau_{j-1} \leq Y_i^* < \tau_j|\mathbf{X}) \\
&= \Pr(\tau_{j-1} \leq \mathbf{X}_i\boldsymbol{\beta} + u_i < \tau_j) \\
&= \Pr(\tau_{j-1} - \mathbf{X}_i\boldsymbol{\beta} \leq u_i < \tau_j - \mathbf{X}_i\boldsymbol{\beta}) \\
&= \int_{-\infty}^{\tau_j - \mathbf{X}_i\boldsymbol{\beta}} f(u_i)du - \int_{-\infty}^{\tau_{j-1} - \mathbf{X}_i\boldsymbol{\beta}} f(u_i)du \\
&= F(\tau_j - \mathbf{X}_i\boldsymbol{\beta}) - F(\tau_{j-1} - \mathbf{X}_i\boldsymbol{\beta})
\end{aligned}
$$

What That Looks Like



Probabilities, etc.

$$
\begin{aligned}
\Pr(Y_i = 1) &= \Phi(\tau_1 - \mathbf{X}_i\beta) - 0 \\
\Pr(Y_i = 2) &= \Phi(\tau_2 - \mathbf{X}_i\beta) - \Phi(\tau_1 - \mathbf{X}_i\beta) \\
\Pr(Y_i = 3) &= \Phi(\tau_3 - \mathbf{X}_i\beta) - \Phi(\tau_2 - \mathbf{X}_i\beta) \\
\Pr(Y_i = 4) &= 1 - \Phi(\tau_3 - \mathbf{X}_i\beta)
\end{aligned}
$$

Likelihoods

Define:

$$\begin{aligned}\delta_{ij} &= 1 \text{ if } Y_i = j \\ &= 0 \text{ otherwise.}\end{aligned}$$

Likelihood:

$$L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \prod_{i=1}^{N} \prod_{j=1}^{J} [F(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - F(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]^{\delta_{ij}}$$

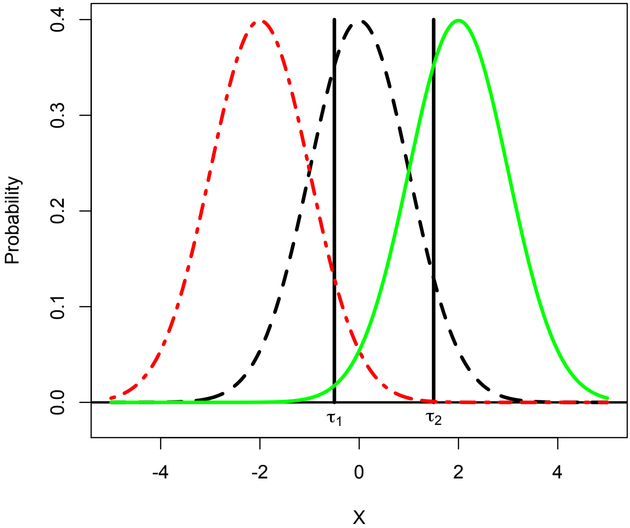
Log-Likelihood, probit:

$$\ln L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln[\Phi(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - \Phi(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]$$

Log-Likelihood, logit:

$$\ln L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln[\Lambda(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - \Lambda(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]$$

The Intuition



Identification

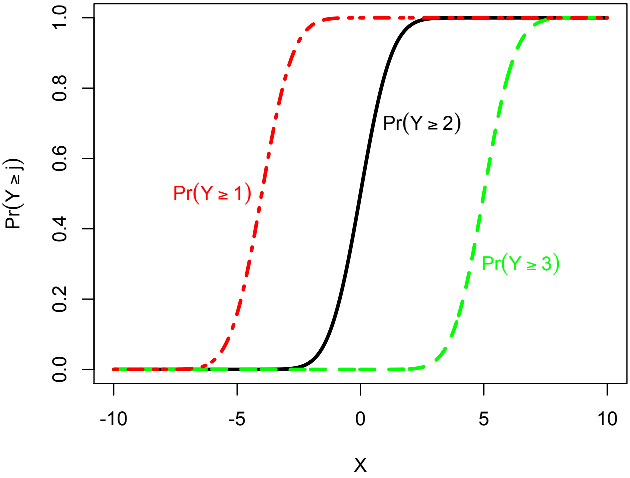
- (Usual) Assumption about $\sigma^2_{Y^*}$
- β_0 vs. the τs...
- Must either omit β_0 or drop one of the $J-1$ τs
- In practice: Stata & R omit β_0

Parallel Regressions

$$\frac{\partial \Pr(Y_i \geq j)}{\partial X} = \frac{\partial \Pr(Y_i \geq j')}{\partial X} \; \forall \; j \neq j'$$

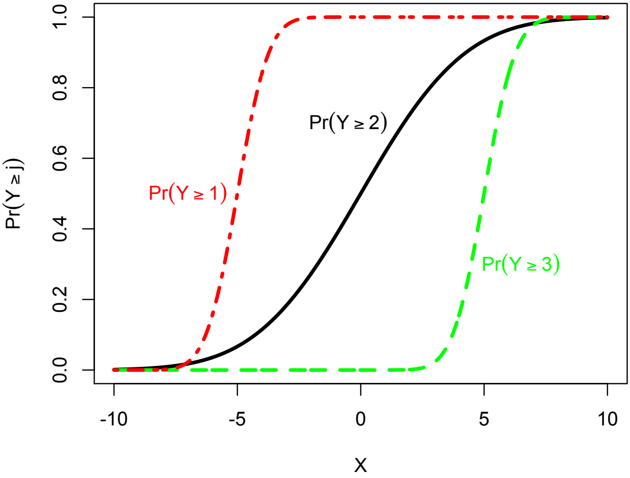
(aka "proportional odds"...)

Parallel Regressions Envisioned



$$\frac{\partial \Pr(Y_i \geq j)}{\partial X} \neq \frac{\partial \Pr(Y_i \geq j')}{\partial X} \; \forall \; j \neq j'$$

Nonparallel Regressions Envisioned



Estimation (in R)

- `polr` (in `MASS`)
- `ologit`/`oprobit` (in `Zelig`; calls `polr`)
- `vglm` (in `VGAM`)

Best Example Ever

1996 Consumer Reports Beer Survey:

```
> summary(beer)

    name              contqual        quality          price           calories
 Length:69          Min.   :24.00   Min.   :1.000   Min.   :2.360   Min.   : 58.0
 Class :character   1st Qu.:49.00   1st Qu.:2.000   1st Qu.:3.900   1st Qu.:142.0
 Mode  :character   Median :70.00   Median :3.000   Median :4.790   Median :148.0
                    Mean   :64.78   Mean   :2.536   Mean   :4.963   Mean   :142.3
                    3rd Qu.:80.00   3rd Qu.:4.000   3rd Qu.:6.240   3rd Qu.:160.0
                    Max.   :98.00   Max.   :4.000   Max.   :7.800   Max.   :201.0

    alcohol        craftbeer          bitter           malty                          class
 Min.   :0.500   Min.   :0.0000   Min.   : 8.00   Min.   : 5.00   Craft Lager          :13
 1st Qu.:4.400   1st Qu.:0.0000   1st Qu.:21.00   1st Qu.:12.00   Craft Ale            :17
 Median :4.900   Median :0.0000   Median :31.00   Median :23.00   Imported Lager       :10
 Mean   :4.471   Mean   :0.4348   Mean   :35.44   Mean   :33.13   Regular or Ice Beer:16
 3rd Qu.:5.100   3rd Qu.:1.0000   3rd Qu.:52.50   3rd Qu.:50.50   Light Beer           : 6
 Max.   :6.000   Max.   :1.0000   Max.   :80.50   Max.   :86.00   Nonalcoholic         : 7
```

Ordered Logit

```
> library(MASS)
> beer.logit<-polr(as.factor(quality)~price+calories+craftbeer+bitter
  +malty,data=beer)
> summary(beer.logit)

Call:
polr(formula = as.factor(quality) ~ price + calories + craftbeer +
    bitter + malty)

Coefficients:
           Value Std. Error t value
price     -0.451      0.293    -1.5
calories   0.047      0.012     3.8
craftbeer -1.705      0.942    -1.8
bitter    -0.030      0.042    -0.7
malty      0.051      0.025     2.1

Intercepts:
    Value  Std. Error t value
1|2  2.771  1.674      1.655
2|3  4.270  1.725      2.475
3|4  5.578  1.760      3.170
```

Ordered Probit

```
> beer.probit<-polr(as.factor(quality)~price+calories+craftbeer+bitter+malty,
+  data=beer,method="probit")
> summary(beer.probit)

Call:
polr(formula = as.factor(quality) ~ price + calories + craftbeer +
    bitter + malty, method = "probit")

Coefficients:
             Value Std. Error t value
price     -0.27914   0.172012 -1.6228
calories   0.02800   0.007184  3.8979
craftbeer -0.98427   0.559020 -1.7607
bitter    -0.01737   0.024719 -0.7025
malty      0.02855   0.014321  1.9937

Intercepts:
    Value  Std. Error t value
1|2  1.647  1.018      1.619
2|3  2.508  1.034      2.426
3|4  3.290  1.049      3.136
```

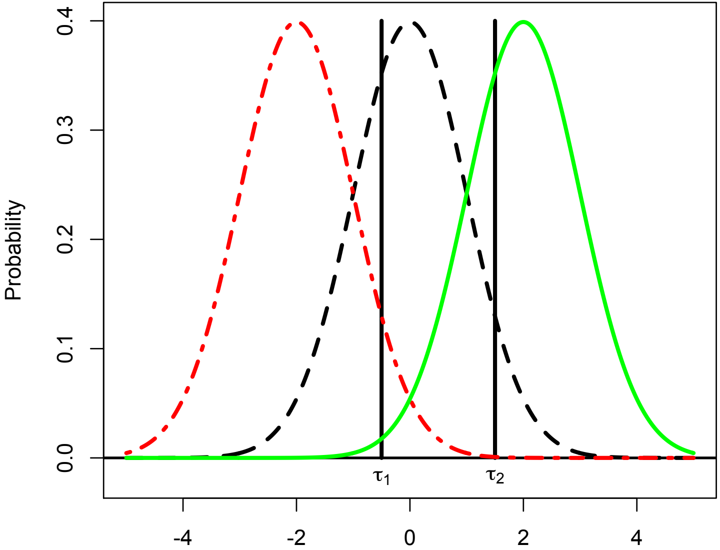
Interpretation: Marginal Effects

$$
\begin{aligned}
\frac{\partial \Pr(Y=j)}{\partial X_k} &= \frac{\partial F(\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta})}{\partial X_k} - \frac{\partial F(\hat{\tau}_{j} - \bar{\mathbf{X}}\hat{\beta})}{\partial X_k} \\
&= \hat{\beta}_k[f(\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta}) - f(\hat{\tau}_{j} - \bar{\mathbf{X}}\hat{\beta})]
\end{aligned}
$$

So:

- $\text{sign}\left(\frac{\partial \Pr(Y=1)}{\partial X_k}\right) = \text{-sign}(\hat{\beta}_k)$
- $\text{sign}\left(\frac{\partial \Pr(Y=J)}{\partial X_k}\right) = \text{sign}(\hat{\beta}_k)$
- $\frac{\partial \Pr(Y=\ell)}{\partial X_k},\ \ell \in \{2, 3, ...J-1\}$ are non-monotonic

Marginal Effects, Illustrated



For a δ-unit change in X_k:

$$\begin{aligned}\mathrm{OR}_{X_k} &= \frac{\frac{\Pr(Y>j|\mathbf{X},X_k+\delta)}{\Pr(Y\leq j|\mathbf{X},X_k+\delta)}}{\frac{\Pr(Y>j|\mathbf{X},X_k)}{\Pr(Y\leq j|\mathbf{X},X_k)}} \\ &= \exp(\delta\hat{\beta}_k)\end{aligned}$$

Calculating Odds Ratios

```
> olreg.or <- function(model)
+  {
+  coeffs <- coef(summary(model))
+  lci <- exp(coeffs[ ,1] - 1.96 * coeffs[ ,2])
+  or <- exp(coeffs[ ,1])
+  uci <- exp(coeffs[ ,1] + 1.96 * coeffs[ ,2])
+  lreg.or <- cbind(lci, or, uci)
+  lreg.or
+  }

> olreg.or(beer.logit)

             lci       or      uci
price     0.3586   0.6373    1.133
calories  1.0231   1.0479    1.073
craftbeer 0.0287   0.1818    1.152
bitter    0.8933   0.9707    1.055
malty     1.0023   1.0518    1.104
1|2       0.6003  15.9748  425.133
2|3       2.4319  71.4963 2101.961
3|4       8.4053 264.4357 8319.319
```

Odds Ratios: Explication

- `craftbeer`:
  - $\exp(-1.705) = 0.18$
  - "The odds of being rated "Good" or better (versus "Fair") are more than 80 percent lower for a craft beer than for a regular beer."
  - "The odds of being rated "Very Good" or better (versus "Fair" or "Good") are more than 80 percent lower for a craft beer than for a regular beer."
- `calories`:
  - $\exp(0.047) = 1.05$
  - "A one-calorie increase raises the odds of being in a higher set of categories (versus all lower ones) by about five percent."
  - etc.

Predicted Probabilities: Basics

$$\Pr(\widehat{Y_i = j}|\mathbf{X}) = F(\hat{\tau}_j - \bar{\mathbf{X}}_i\hat{\beta}) - F(\hat{\tau}_{j-1} - \bar{\mathbf{X}}_i\hat{\beta})$$

Means:

- `price` $= 4.96$, `calories` $= 142$, `craftbeer` $= 0$, `bitter` $= 35.4$, `malty` $= 33.1$.
- Yields:

$$\begin{aligned}
\sum_{k=1}^{K} \bar{\mathbf{X}}_k\hat{\beta}_k &= -0.45 \times 4.96 + 0.047 \times 142 - 1.70 \times 0 - \\
&\quad 0.03 \times 35.4 + 0.05 \times 33.1 \\
&= -2.23 + 6.67 - 0 - 1.06 + 1.66 \\
&= \mathbf{5.04}.
\end{aligned}$$

Predicted Probabilities: "By Hand"

$$\begin{aligned}
\Pr(Y=1) &= \Lambda(2.77-5.04)-0 \\
&= \frac{\exp(-2.27)}{1+\exp(-2.27)} \\
&= \mathbf{0.09}.
\end{aligned}$$

$$\begin{aligned}
\Pr(Y=2) &= \Lambda(4.27-5.04)-\Lambda(2.77-5.04) \\
&= \Lambda(-0.77)-\Lambda(-2.27) \\
&= 0.32-0.09 \\
&= \mathbf{0.23}.
\end{aligned}$$

$$\begin{aligned}
\Pr(Y=3) &= \Lambda(5.58-5.04)-\Lambda(4.27-5.04) \\
&= \Lambda(0.54)-\Lambda(-0.77) \\
&= 0.63-0.32 \\
&= \mathbf{0.31}.
\end{aligned}$$

$$\begin{aligned}
\Pr(Y=4) &= 1-\Lambda(5.58-5.04) \\
&= 1-\Lambda(0.54) \\
&= 1-0.63 \\
&= \mathbf{0.37}.
\end{aligned}$$

Changes in Predicted Probabilities

For `craftbeer`=1:

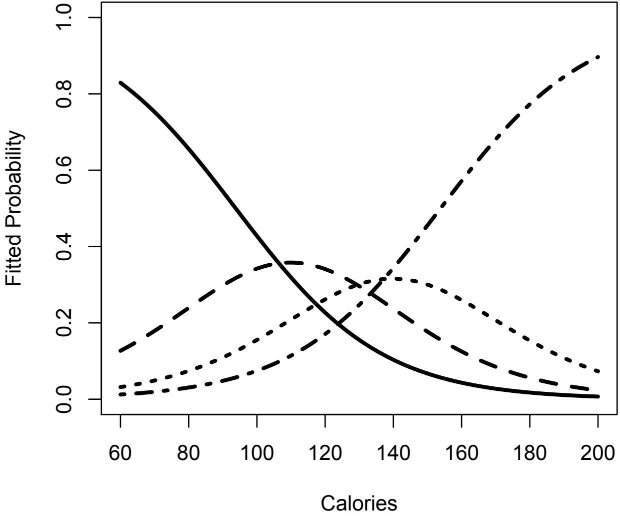
- $\Pr(Y = 1) = \Lambda(2.77 - 3.34) - 0 = \mathbf{0.36}$.
- $\Pr(Y = 2) = \Lambda(4.27 - 3.34) - \Lambda(2.77 - 3.34) = 0.72 - 0.36 = \mathbf{0.36}$.
- $\Pr(Y = 3) = \Lambda(5.58 - 3.34) - \Lambda(4.27 - 3.34) = 0.90 - 0.72 = \mathbf{0.18}$.
- $\Pr(Y = 4) = 1 - 0.90 = \mathbf{0.10}$.

Outcome	Change in Probability
ΔPr(Fair)	0.27
ΔPr(Good)	0.13
ΔPr(Very Good)	-0.13
ΔPr(Excellent)	-0.27

Predicted Probability Plots

- Can be category-specific or “cumulative”
- In-sample in `$fitted.values`
- `polr` class supports `predict`, `confint`, etc.

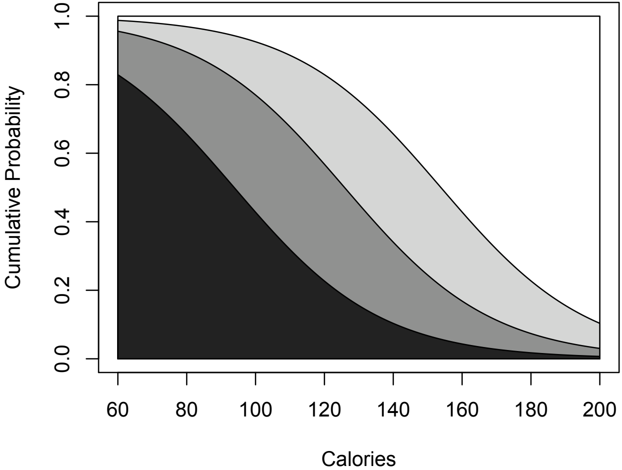
Plot by Outcome



(How'd He Do That?)

```
> calories<-seq(60,200,1)
> price<-mean(beer$price)
> craftbeer<-median(beer$craftbeer)
> bitter<-mean(beer$bitter)
> malty<-mean(beer$malty)
> beersim<-cbind(calories,price,craftbeer,bitter,malty)
> beer.hat<-predict(beer.logit,beersim,type='probs')
> plot(c(60,200), c(0,1), type='n', xlab="Calories", ylab='Fitted
  Probability')
> lines(60:200, beer.hat[1:141, 1], lty=1, lwd=3)
> lines(60:200, beer.hat[1:141, 2], lty=2, lwd=3)
> lines(60:200, beer.hat[1:141, 3], lty=3, lwd=3)
> lines(60:200, beer.hat[1:141, 4], lty=4, lwd=3)
```

Cumulative Predicted Probabilities



(code...)

```
> xaxis<-c(60,60:200,200)
> yaxis1<-c(0,beer.hat[,1],0)
> yaxis2<-c(0,beer.hat[,2]+beer.hat[,1],0)
> yaxis3<-c(0,beer.hat[,3]+beer.hat[,2]+beer.hat[,1],0)
> yaxis4<-c(0,beer.hat[,4]+beer.hat[,3]+beer.hat[,2]+beer.hat[,1],0)
>
> plot(c(60,200), c(0,1), type='n', xlab="Calories", ylab="Cumulative
  Probability")
> polygon(xaxis,yaxis4,col="white")
> polygon(xaxis,yaxis3,col="grey80")
> polygon(xaxis,yaxis2,col="grey50")
> polygon(xaxis,yaxis1,col="grey10")
```

Variants / Extensions (for PLSC 504...)

- Generalized (relaxes parallel regressions; Brant (1990))
- Heteroscedastic
- Varying τs (Maddala, Terza, Sanders)
- Models for "balanced" scales (Jones & Sobel)
- Compound Ordered Hierarchical Probit ("`chopit`") (Wand & King)
- "Zero-Inflated" Ordered Models (Hill, Bagozzi, Moore & Mukherjee)
- Latent class/mixture models (Winkelmann, etc.)