

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Introduction to Multinomial Probability Distribution

# Introduction

The previous three weeks provided analysis methods for when there were binary responses. The purpose of this week is to generalize some of these previous methods to allow for more than two response categories. Examples include:

- Canadian political party affiliation—Conservative, New Democratic, Liberal, Bloc Quebecois, or Green
- Chemical compounds in drug discovery experiments—positive, blocker, or neither
- Cereal shelf placement in a grocery store—bottom, middle, or top
- Beef grades—Prime, choice, select, standard, utility, and commercial
- Five-level Likert scale—strongly disagree, disagree, neutral, agree, or strongly agree

For these examples, some responses are ordinal (e.g., Likert scale) and some are not (e.g., chemical compounds).

# Multinomial Probability Distribution

The multinomial probability distribution is the extension of the binomial distribution to situations where there are more than two categories for a response.

## Notation:

- $Y$  denotes the response category with levels of  $j = 1, \dots, J$
- Each category has a probability of  $\pi_j = P(Y = j)$ .
- $n$  denotes the number of trials
- $n_1, \dots, n_J$  denote the response count for category  $j$ , where  $\sum_{j=1}^J n_j = n$

The probability mass function for observing particular values of  $n_1, \dots, n_J$  is

$$\frac{n!}{\prod_{j=1}^J n_j!} \prod_{j=1}^J \pi_j^{n_j}$$

# Example: Multinomial Simulation

As a quick way to see what a sample looks like in a multinomial setting, consider the situation (a sample) with **n = 1,000 trials**,  $\pi_1 = 0.25$ ,  $\pi_2 = 0.35$ ,  $\pi_3 = 0.2$ ,  $\pi_4 = 0.1$ , and  $\pi_5 = 0.1$ . Below is how we can simulate a sample:

```
> pi.j<-c(0.25, 0.35, 0.2, 0.1, 0.1)
> set.seed(2195) #Set a seed to be able to reproduce the sample
> n.j<-rmultinom(n = 1, size = 1000, prob = pi.j)
> data.frame(n.j, pihat.j = n.j/1000, pi.j)
   n.j pihat.j pi.j
1 242  0.242 0.25
2 333  0.333 0.35
3 188  0.188 0.20
4 122  0.122 0.10
5 115  0.115 0.10
```

Suppose there are  $m = 5$  separate sets of  $n = 1000$  trials.

```
set.seed(9182)
n.j<-rmultinom(n = 5, size = 1000, prob = pi.j)
[,1] [,2] [,3] [,4] [,5]          > n.j/1000
[1,] 259  259  237  264  247          [,1] [,2] [,3] [,4] [,5]
[2,] 341  346  374  339  341          [1,] 0.259 0.259 0.237 0.264 0.247
[3,] 200  188  198  191  210          [2,] 0.341 0.346 0.374 0.339 0.341
[4,]  92  106  89   108  107          [3,] 0.200 0.188 0.198 0.191 0.210
[5,] 108  101  102   98   95          [4,] 0.092 0.106 0.089 0.108 0.107
                                         [5,] 0.108 0.101 0.102 0.098 0.095
```

Notice the variability from one set to another.

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

Subtitle

**datascience@berkeley**

# Part 2: $I \times J$ Contingency Tables and Inference Procedures (6 minutes)

# Introduction

- In Lecture 1, we introduced a  $2 \times 2$  contingency table.
- Now we extend this concept to an  $I \times J$  contingency table.
- We will begin by focusing on two separate ways that one can think of how the counts arise in a contingency table structure through using a multinomial distribution.
- In the next lecture, we will consider another way through using a Poisson distribution.
- The text also considers a third way that uses a hypergeometric distribution, but we will not cover it in this course.

# One Multinomial Distribution

Set-up:

- $X$  denotes the row variable with levels  $i = 1, \dots, I$
- $Y$  denotes the column variable with levels  $j = 1, \dots, J$
- $P(X = i, Y = j) = \pi_{ij}$
- $\sum_{i=1}^I \sum_{j=1}^J \pi_{ij} = 1$
- $n_{ij}$  denotes the cell count for row  $i$  and column  $j$
- $\sum_{i=1}^I \sum_{j=1}^J n_{ij} = n$

Contingency tables summarizing this information are shown below:

		Y				
		1	2	...	J	
		$\pi_{11}$	$\pi_{12}$	$\cdots$	$\pi_{1J}$	$\pi_{1+}$
X	1	$\pi_{21}$	$\pi_{22}$	$\cdots$	$\pi_{2J}$	$\pi_{2+}$
	2	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	$\pi_{I1}$	$\pi_{I2}$	$\cdots$	$\pi_{IJ}$	$\pi_{I+}$
		$\pi_{+1}$	$\pi_{+2}$	$\cdots$	$\pi_{+J}$	1

		Y				
		1	2	...	J	
		$n_{11}$	$n_{12}$	$\cdots$	$n_{1J}$	$n_{1+}$
X	1	$n_{21}$	$n_{22}$	$\cdots$	$n_{2J}$	$n_{2+}$
	2	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	I	$n_{I1}$	$n_{I2}$	$\cdots$	$n_{IJ}$	$n_{I+}$
		$n_{+1}$	$n_{+2}$	$\cdots$	$n_{+J}$	n

# Multinomial Probability Model

The setup given for these contingency tables fits right into the multinomial setting of Section 3.1. We now just categorize the responses with respect to X and Y. The probability mass function for observing particular values of  $n_{11}, \dots, n_{IJ}$  is

$$\frac{n!}{\prod_{i=1}^I \prod_{j=1}^J n_{ij}!} \prod_{i=1}^I \prod_{j=1}^J \pi_{ij}^{n_{ij}}$$

The MLE of  $\pi_{ij}$  is the estimated proportion  $\hat{\pi}_{ij} = \underline{n_{ij}}/n$ .

We can also discuss marginal distributions for X and for Y as well:

- X is multinomial with counts  $\underline{n_{i+}}$  for  $i = 1, \dots, I$  and corresponding probabilities  $\pi_{i+}$ .  
The maximum likelihood estimate of  $\pi_{i+}$  is  $\hat{\pi}_{i+} = \underline{n_{i+}}/n$ .
- Y is multinomial with counts  $\underline{n_{+j}}$  for  $j = 1, \dots, J$  and corresponding probabilities  $\pi_{+j}$ .  
The MLE of  $\pi_{+j}$  is  $\hat{\pi}_{+j} = \underline{n_{+j}}/n$

# Example

As a quick way to see what a sample looks like in a  $2 \times 3$  contingency table setting, consider the situation with  $n = 1,000$  observations,  $\pi_{11} = 0.2$ ,  $\pi_{21} = 0.3$ ,  $\pi_{12} = 0.2$ ,  $\pi_{22} = 0.1$ ,  $\pi_{13} = 0.1$ , and  $\pi_{23} = 0.1$ . Below is how we can simulate a sample:

```

> pi.ij<-c(0.2, 0.3, 0.2, 0.1, 0.1, 0.1)
> pi.table<-array(data = pi.ij, dim = c(2,3), dimnames =
+   list(X = 1:2, Y = 1:3))
> pi.table
      Y
X   1   2   3
  1 0.2 0.2 0.1
  2 0.3 0.1 0.1

```

```

> set.seed(9812)
> save<-rmultinom(n = 1, size = 1000, prob = pi.ij)
> c.table1<-array(data = save, dim = c(2,3), dimnames =
+ list(X = 1:2, Y = 1:3))
> c.table1
      Y
X   1   2   3
  1 191 206 94
  2 311 95 103

```

```

> c.table1/sum(c.table1)
      Y
X   1   2   3
  1 0.191 0.206 0.094
  2 0.311 0.095 0.103

```

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Independence

# Independence

Independence between X and Y when the occurrence of  $X = i$  does not have an effect of the occurrence of  $Y = j$  for each  $i = 1, \dots, I$  and  $j = 1, \dots, J$ .

Symbolically, independence exists when  $\pi_{ij} = \pi_{i+} \pi_{+j}$

$$\pi_{ij} = \pi_{i+} \pi_{+j}$$

for  $i = 1, \dots, I$  and  $j = 1, \dots, J$ .

→ This means that the probability of an item being in cell  $(i,j)$  only involves knowing the separate marginal probabilities for X and for Y; thus, X and Y are independent of each other.

---

# Why is independence important?

Independence helps to simplify the understanding of probabilities within the contingency table!

There are only  $I - 1 + J - 1 = I + J - 2$  unknown probability parameters.

Note that the “-1” parts occur due to

the  $\sum_{i=1}^I \pi_{i+} = 1$  and  $\sum_{j=1}^J \pi_{+j} = 1$ .

$$\boxed{IJ - 1}$$

Without independence, there are  $IJ - 1$  unknown probability parameters, where the “-1” part occurs

due to the  $\sum_{i=1}^I \sum_{j=1}^J \pi_{ij} = 1$ .

# Implications of Independence

Consider the  $I$  multinomial distributions case again.

Similar to Section 1.2, it is often of interest to know if these conditional probabilities are equal across the rows of the table. Thus, we want to know if  $\pi_{j|1} = \dots = \pi_{j|I}$  for  $j = 1, \dots, J$ . Note that this is mathematically equivalent to  $\pi_{ij} = \pi_{i+}\pi_{+j}$  for  $i = 1, \dots, I$  and  $j = 1, \dots, J$ !

Recall from w203 that

$$P(Y = j | X = i) = P(X = i, Y = j) / P(X = i).$$

Thus,

$$\pi_{j|i} = \pi_{ij} / \pi_{i+} = \pi_{i+}\pi_{+j} / \pi_{i+} = \pi_{+j}$$

under independence.

Because each  $\pi_{j|I}$  is equal to  $\pi_{+j}$  for each  $i$ , we have  $\pi_{j|1} = \dots = \pi_{j|I}$ .

Because of this equivalence, we will refer to  $\pi_{j|1} = \dots = \pi_{j|I}$  for  $j = 1, \dots, J$  as “independence” as well.

# Test for Independence

The hypotheses are:

$$\begin{aligned} H_0: \pi_{ij} &= \pi_{i+} \pi_{+j} \text{ for } i = 1, \dots, I \text{ and } j = 1, \dots, J \\ H_a: \text{Not all equal} & \end{aligned}$$

Remember that a Pearson chi-square test statistic calculates

$$\frac{(\text{observed count} - \text{estimated expected count})^2}{(\text{estimated expected count})}$$

for every cell of a contingency table and sums these quantities.  
 The Pearson chi-square test for independence then uses the statistic

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - n_{i+} n_{+j} / n)^2}{n_{i+} n_{+j} / n}$$

# Remarks

- The estimated expected cell count is  $n\hat{\pi}_{i+}\hat{\pi}_{+j} = n_{i+}n_{+j} / n$
- $\chi^2$  is equivalent to the corresponding statistic used in Section 1.2 (in the text) for the Pearson chi-square test for a  $2\times 2$  contingency table.
- If the null hypothesis is true,  $\chi^2$  has a  $\chi^2_{(I-1)(J-1)}$  distribution for a large sample.
- Reject the null hypothesis if  $X^2 > \chi^2_{(I-1)(J-1), 1-\alpha}$ .

$$\chi^2_{(I-1)(J-1)}$$


$$X^2 > \chi^2_{(I-1)(J-1), 1-\alpha}$$

# Likelihood-Ratio Test (LRT) Statistic

The LRT statistic is formed the usual way with

$$\Lambda = \frac{\text{Max. lik. when parameters satisfy } H_0}{\text{Max. lik. when parameters satisfy } H_0 \text{ or } H_a}$$

The numerator of  $\Lambda$  uses  $\hat{\pi}_{i+} \hat{\pi}_{+j}$  to estimate  $\pi_{ij}$ , and the denominator of  $\Lambda$  uses  $\hat{\pi}_{jj}$  to estimate  $\pi_{ii}$ . The transformed statistic simplifies to

$$-2\log(\Lambda) = 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \left( \frac{n_{ij}}{n_{i+} n_{+j} / n} \right)$$

where we use  $0 \times \log(0) = 0$ . The large sample distribution is the same as for  $\chi^2$ .

# Degree of Freedom

A general way to find degrees of freedom for a hypothesis test is to calculate:

$$(\text{Number of free parameters under } H_a) - (\text{Number of free parameters under } H_0)$$

Under the alternative hypothesis for the test of independence, we have  $IJ \pi_{ij}$  parameters with the restriction that  $\sum_{i=1}^I \sum_{j=1}^J \pi_{ij} = 1$ . When independence is true, we need the  $I$  different  $\pi_{i+}$  and the  $J$  different  $\pi_{+j}$  parameters to find  $\pi_{ij}$  with the restriction that

$$\sum_{i=1}^I \pi_{i+} = 1 \text{ and } \sum_{j=1}^J \pi_{+j} = 1.$$

Thus, the overall degrees of freedom is

$$(IJ - 1) - (I + J - 2) = (I - 1)(J - 1)$$

# Example

- Fiber is often added to foods as a convenient way for people to consume it.
- The Data and Story Library (DASL) describes the results of a study where individuals are given a new type of fiber-enriched cracker.
- The participants ate the crackers and then a meal. Shortly afterward, the participants were instructed to describe any bloating that they experienced.
- Below is the data:

Fiber source	Bloating severity			
	None	Low	Medium	High
None	6	4	2	0
Bran	7	4	1	0
Gum	2	2	3	5
Both	2	5	3	2

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

Subtitle

**datascience@berkeley**

# An Example

# Example

The purpose of this data is to determine if the fiber source has an effect on the bloating severity.

- Notice the columns have ordinal levels. (We will take this into account later in this course.) It is instructive in a class setting to analyze the data first without taking the order into account, so that we can see the benefits of taking into account the order later.
- I would expect that each person fits in one and only one cell of the table. **Why would this be important to know?**
- Given the layout of the data, it is likely that the sample size for each row was fixed. Thus, this would correspond to the **multinomial distribution setting**.
- Fiber source could actually be analyzed as two separate explanatory variables: bran ("yes" or "no") and gum ("yes" or "no").
- We will analyze this data in a **4x4 contingency table**.  
*(Refer to the text for how this data can be analyzed as two separate explanatory variables through using regression models.)*

# Example

```
diet <- read.csv(file = "C:\\\\data\\\\Fiber.csv")  
head(diet)
```

	fiber	bloat	count
1	bran	high	0
2	gum	high	5
3	both	high	2
4	none	high	0
5	bran	medium	1
6	gum	medium	3

```
# Match order given at DASL  
diet$fiber<-factor(x = diet$fiber, levels = c("none", "bran", "gum", "both"))  
diet$bloat<-factor(x = diet$bloat, levels = c("none", "low", "medium", "high"))  
diet.table<-xtabs(formula = count ~ fiber + bloat, data = diet)  
diet.table
```

	none	low	medium	high
none	6	4	2	0
bran	7	4	1	0
gum	2	2	3	5
both	2	5	3	2

# Example

```
ind.test<-chisq.test(x = diet.table, correct = FALSE)  
ind.test
```

```
Pearson's Chi-squared test  
  
data: diet.table  
X-squared = 16.9427, df = 9, p-value = 0.04962  
  
Warning message:  
In chisq.test(diet.table, correct = FALSE) :  
Chi-squared approximation may be incorrect
```

```
library(package = vcd)  
assocstats(x = diet.table)
```

```
Likelihood Ratio    k^2   df   P(> X^2)  
Pearson            18.880   9   0.026230  
                  16.943   9   0.049621  
  
Phi-Coefficient : 0.594  
Contingency Coeff.: 0.511  
Cramer's V        : 0.343
```

# Example

```
> class(diet.table)
[1] "xtabs" "table"

> summary(diet.table)
Call: xtabs(formula = count ~ fiber + bloat, data = diet2)
Number of cases in table: 48
Number of factors: 2
Test for independence of all factors:
    Chisq = 16.943, df = 9, p-value = 0.04962
    Chi-squared approximation may be incorrect

> qchisq(p = 0.95, df = 9)
[1] 16.91898
```

# Summary

In summary,

- $\chi^2 = 16.94$
- $-2\log(\Lambda) = 18.88$
- $\chi^2_{0.95,9} = 16.92$
- P-value using  $\chi^2$  is  $P(A > 16.94) = 0.0496$  where  $A \sim \chi^2_9$
- P-value using  $-2\log(\Lambda)$  is  $P(A > 18.88) = 0.0262$  where  $A \sim \chi^2_9$
- Because the p-value is small, but not extremely so, we would say there is moderate evidence against independence (thus, moderate evidence of dependence).
- Thus, there is moderate evidence that bloating severity is dependent on the fiber source.

# Implications of Independence

Can we trust the  $\chi^2_9$  approximation?

```
> ind.test$expected
      bloat
fiber   none    low medium  high
  none 4.25 3.75    2.25 1.75
  bran 4.25 3.75    2.25 1.75
  gum  4.25 3.75    2.25 1.75
  both 4.25 3.75    2.25 1.75
```

These only partially satisfy the recommendations given earlier!

For the details of  $\chi^2_{(I-1)(J-1)}$  approximation, please refer to the text.

# Remarks

- If independence is rejected, we would like to determine why it is rejected.
- For example, perhaps only particular combinations of  $X$  and  $Y$  are causing the dependence.
- Also, we would like to determine how much dependence exists.
- There are a number of ways to examine a contingency table further to understand the dependence.
- My preference is to generally use statistical models for this purpose, while even using these models to help test for independence.

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Nominal Response Model

# Introduction

Suppose there are  $J$  **categories** for the response variable with corresponding probabilities  $\pi_1, \pi_2, \dots, \pi_J$ . Using the first category as a “**baseline**,” we can form “**baseline category logits**” as  $\log(\pi_j/\pi_1)$  for  $j = 2, \dots, J$ .

When  $J = 2$ , we have  $\log(\pi_2/\pi_1) = \log(\pi_2/(1-\pi_2))$ , which is equivalent to  $\log(\pi/(1-\pi))$  in logistic regression with  $\pi = \pi_2$ .

When there is only one explanatory variable  $x$ , we can form the multinomial regression model of

$$\log(\pi_j/\pi_1) = \beta_{j0} + \beta_{j1}x \text{ for } j = 2, \dots, J$$

One can easily compare other categories so that category 1 is not always used. For example, suppose you would like to compare category 2 to 3. Then

$$\log(\pi_2/\pi_1) - \log(\pi_3/\pi_1) = \log(\pi_2) - \log(\pi_3) = \log(\pi_2/\pi_3)$$

and

$$\beta_{20} + \beta_{21}x - \beta_{30} - \beta_{31}x = \beta_{20} - \beta_{30} + x(\beta_{21} - \beta_{31})$$

# Introduction

What is  $\pi_j$  only? Consider the case of one explanatory variable  $x$  again:

We can re-write the model as

Noting that  $\sum_{j=1}^J \pi_j = 1$ , we have

$$\pi_1 + \pi_1 e^{\beta_{j0} + \beta_{j1}x} + \dots + \pi_1 e^{\beta_{j0} + \beta_{J1}x} = 1$$

$\overbrace{\pi_1 + \pi_1 e^{\beta_{j0} + \beta_{j1}x}}^{\pi_1}$

Thus,

$$\pi_1 = \frac{1}{1 + \sum_{j=2}^J e^{\beta_{j0} + \beta_{j1}x}}$$

Also, we can now find that

$$\pi_j = \frac{e^{\beta_{j0} + \beta_{j1}x}}{1 + \sum_{j=2}^J e^{\beta_{j0} + \beta_{j1}x}}$$

for  $j = 2, \dots, J$ .

# Introduction

- Parameters are estimated using maximum likelihood estimation.
- For a sample of size  $m$ , the likelihood function is simply the product of  $m$  multinomial distributions with probability parameters as given above.
- Iterative numerical procedures are used then to find the parameter estimates.
- The **multinom() function** from the nnet package (within the default installation of R) performs the necessary computations.
- The covariance matrix for the parameter estimates follows from using standard likelihood procedures as outlined in Appendix B (of the text).
- Wald and LR-based inference methods are performed in the same ways as for likelihood procedures in earlier weeks.

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Example

# Example

- Wheat producers want to identify kernels that are in poor condition after being harvested.
- To facilitate this identification process, categorization systems have been developed to partition kernels into different categories (see Martin et al., 1998).
- For this example, we will look at the categories of "Healthy," "Sprout," or "Scab."
- In summary:
  - Healthy is the preferred condition because these kernels have not been damaged.
  - Sprout is less preferred than healthy because they have reduced weight and poorer flour quality.
  - Scab is less preferred than healthy because they come from plants that have been infected by a disease and have undesirable qualities in their appearance.

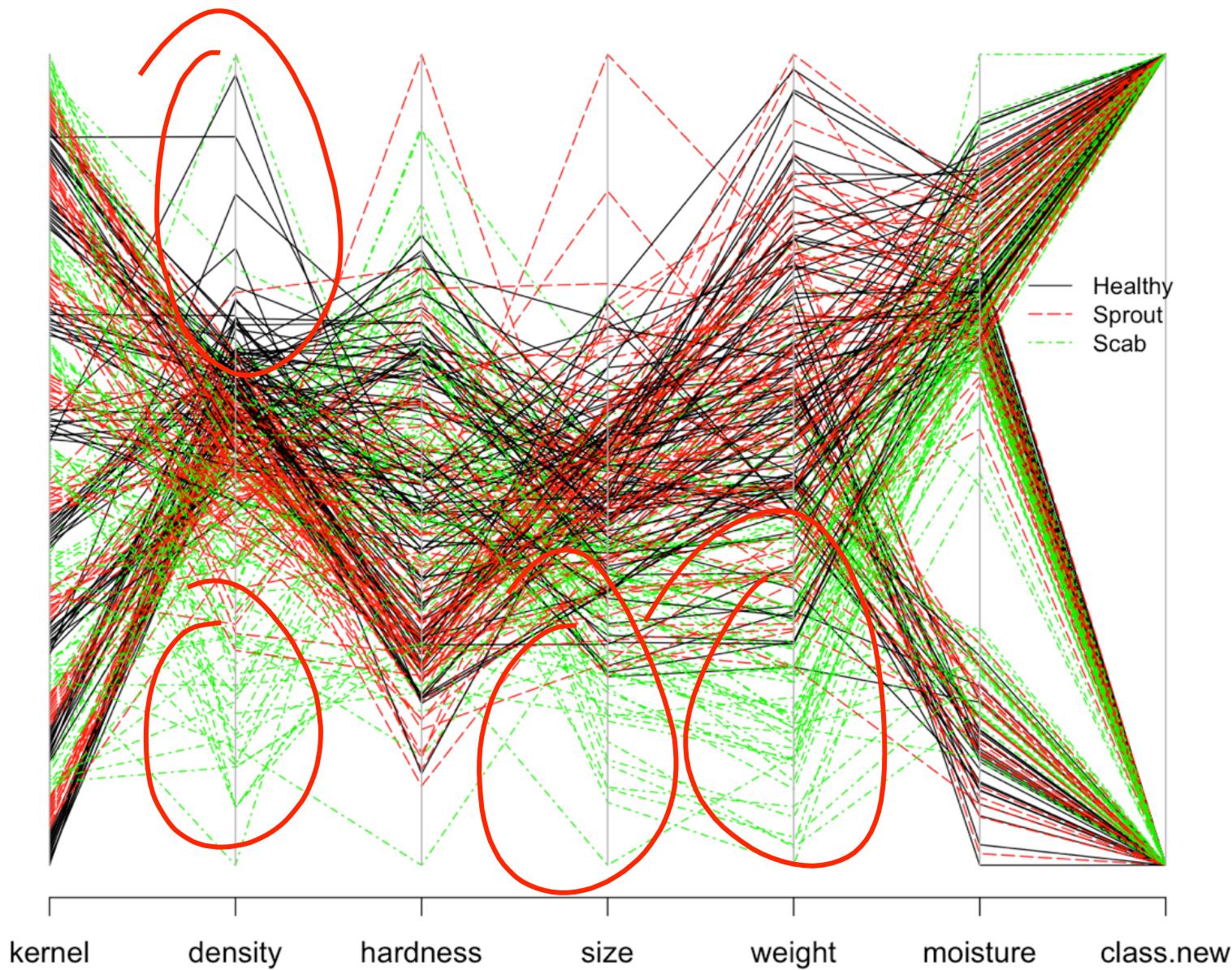
# Example

- Ideally, it would be preferred to make these categorizations for each kernel through using an automated process.
- To test a new system out, **275 wheat kernels** were classified by human examination (assumed to be perfect). The automated system uses information about the class of the wheat kernel (soft red winter or hard red winter) and measurements for density, hardness, size, weight, and moisture for the kernel.

# Example

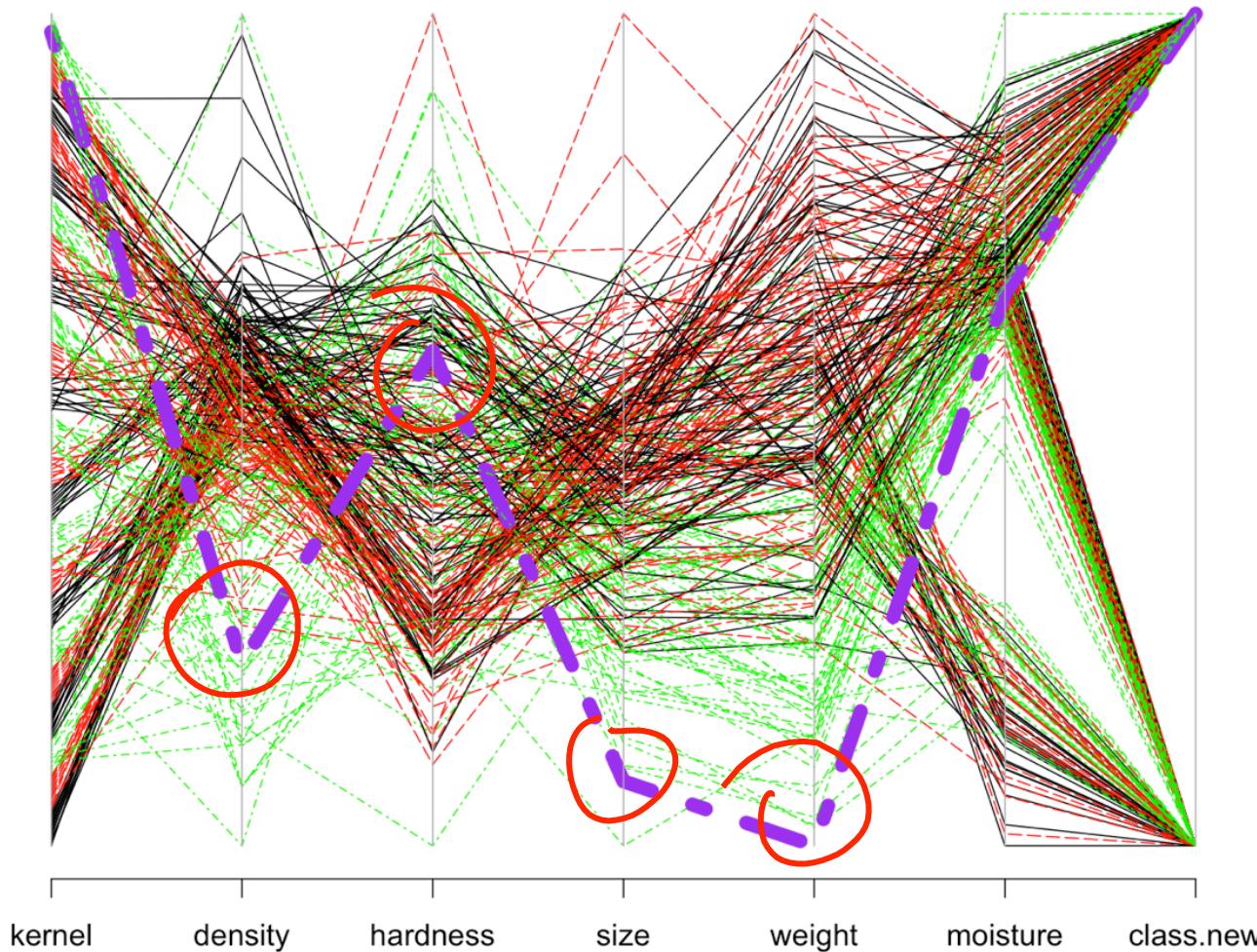
```
> head(wheat, n = 3) # n argument gives the number of rows to print
  class  density hardness   size weight moisture type
1  hrw 1.349253 60.32952 2.30274 24.6480 12.01538 Healthy
2  hrw 1.287440 56.08972 2.72573 33.2985 12.17396 Healthy
3  hrw 1.233985 43.98743 2.51246 31.7580 11.87949 Healthy
> tail(wheat, n = 3)
  class  density hardness   size weight moisture type
273  srw 0.8491887 34.06615 1.40665 12.0870 11.92744 Scab
274  srw 1.1770230 60.97838 1.05690  9.4800 12.24046 Scab
275  srw 1.0305543 -9.57063 2.05691 23.8185 12.64962 Scab
```

# Parallel Coordinate Plot



# Parallel Coordinate Plot

Parallel coordinate plot for wheat data - highlight kernel 269



# A Multinomial Logistic Regression Model

Consider the following model

$$\log(\pi_j/\pi_1) = \beta_{j0} + \beta_{j1}x_1 + \dots + \beta_{j6}x_6 \text{ for } j = 2, 3$$

```
> levels(wheat$type) #Shows the 3 categories
[1] "Healthy" "Scab"    "Sprout"
```

Thus,  $j = 1$  is healthy,  $j = 2$  is scab, and  $j = 3$  is sprout.

```
> mod.fit<-multinom(formula = type ~ class + density + hardness + size + weight + moisture, data=wheat)
# weights: 24 (14 variable)
initial value 302.118379
iter 10 value 234.991271
iter 20 value 192.127549
final value 192.112352
converged
```

# A Multinomial Logistic Regression Model

```
> summary(mod.fit)

Call:
multinom(formula = type ~ class + density + hardness + size +
  weight + moisture, data = wheat)

Coefficients:
            (Intercept) classsrw   density   hardness    size    weight   moisture
Scab        30.54650 -0.6481277 -21.59715 -0.01590741 1.0691139 -0.2896482 0.10956505
Sprout      19.16857 -0.2247384 -15.11667 -0.02102047 0.8756135 -0.0473169 -0.04299695

Std. Errors:
            (Intercept) classsrw   density   hardness    size    weight   moisture
Scab        4.289865 0.6630948 3.116174 0.010274587 0.7722862 0.06170252 0.1548407
Sprout      3.767214 0.5009199 2.764306 0.008105748 0.5409317 0.03697493 0.1127188

Residual Deviance: 384.2247
AIC: 412.2247
```

```
> class(mod.fit)
[1] "multinom" "nnet"
> methods(class = multinom)
[1] add1      anova     Anova     coef      confint   deltaMethod drop1
[8] extractAIC logLik    model.frame predict   print     summary   vcov
see '?methods' for accessing help and source code
```

# A Multinomial Logistic Regression Model

```
> summary(mod.fit)
```

Call:

```
multinom(formula = type ~ class + density + hardness + size +
  weight + moisture, data = wheat)
```

Coefficients:

	(Intercept)	classsrw	density	hardness	size	weight	moisture
Scab	30.54650	-0.6481277	-21.59715	-0.01590741	1.0691139	-0.2896482	0.10956505
Sprout	19.16857	-0.2247384	-15.11667	-0.02102047	0.8756135	-0.0473169	-0.04299695

$$\log(\hat{\pi}_{\text{scab}} / \hat{\pi}_{\text{healthy}}) = 30.55 - 0.65\text{SRW} - 21.60\text{density} \\ - 0.016\text{hardness} + 1.07\text{size} - 0.29\text{weight} \\ + 0.11\text{moisture}$$

$$\log(\hat{\pi}_{\text{sprout}} / \hat{\pi}_{\text{healthy}}) = 19.17 - 0.22\text{SRW} - 15.12\text{density} \\ - 0.021\text{hardness} + 0.88\text{size} - 0.047\text{weight} \\ - 0.043\text{moisture}$$

# The Estimated Model

$$\log(\hat{\pi}_{\text{scab}} / \hat{\pi}_{\text{healthy}}) = 30.55 - 0.65\text{SRW} - 21.60\text{density} \\ - 0.016\text{hardness} + 1.07\text{size} - 0.29\text{weight} \\ + 0.11\text{moisture}$$

$$\log(\hat{\pi}_{\text{sprout}} / \hat{\pi}_{\text{healthy}}) = 19.17 - 0.22\text{SRW} - 15.12\text{density} \\ - 0.021\text{hardness} + 0.88\text{size} - 0.047\text{weight} \\ - 0.043\text{moisture}$$

- Notice how R forms an indicator variable for the class of the wheat ("classsrw" corresponds to SRW).
- Now that we have the estimated model, many of the basic types of analyses done in the last few lectures can be performed here!
- The R code used is very similar as well. Because of the similarity, we will assign those as take-home exercise.

# Remarks

- The **moprofile package** cannot be used for likelihood-ratio-based inference methods.
- Confidence intervals for  $\pi_j$  are more complicated to calculate than what we saw in Week 2.
  - The main reason is because Brian Ripley, the author of the nnet package, does not believe that one-at-a-time intervals should be calculated. For example, my program shows how to calculate one-at-a-time 95% intervals as

$$\begin{aligned} 0.7376 &< \pi_{\text{Healthy}} &< 0.9728 \\ -0.0067 &< \pi_{\text{Scab}} &< 0.0995 \\ 0.0143 &< \pi_{\text{Sprout}} &< 0.1825 \end{aligned}$$

for the first observation.

- Of course,  $\pi_{\text{Healthy}} + \pi_{\text{Scab}} + \pi_{\text{Sprout}} = 1$  needs to occur.
- If we added the upper limits from the intervals together, we have a total greater than 1! For this reason, Ripley advocates constructing a confidence region. However, this is much more difficult to calculate, and he does not provide any code (no one else provides any code, either) to calculate it for these types of models. More discussion is included in the text.

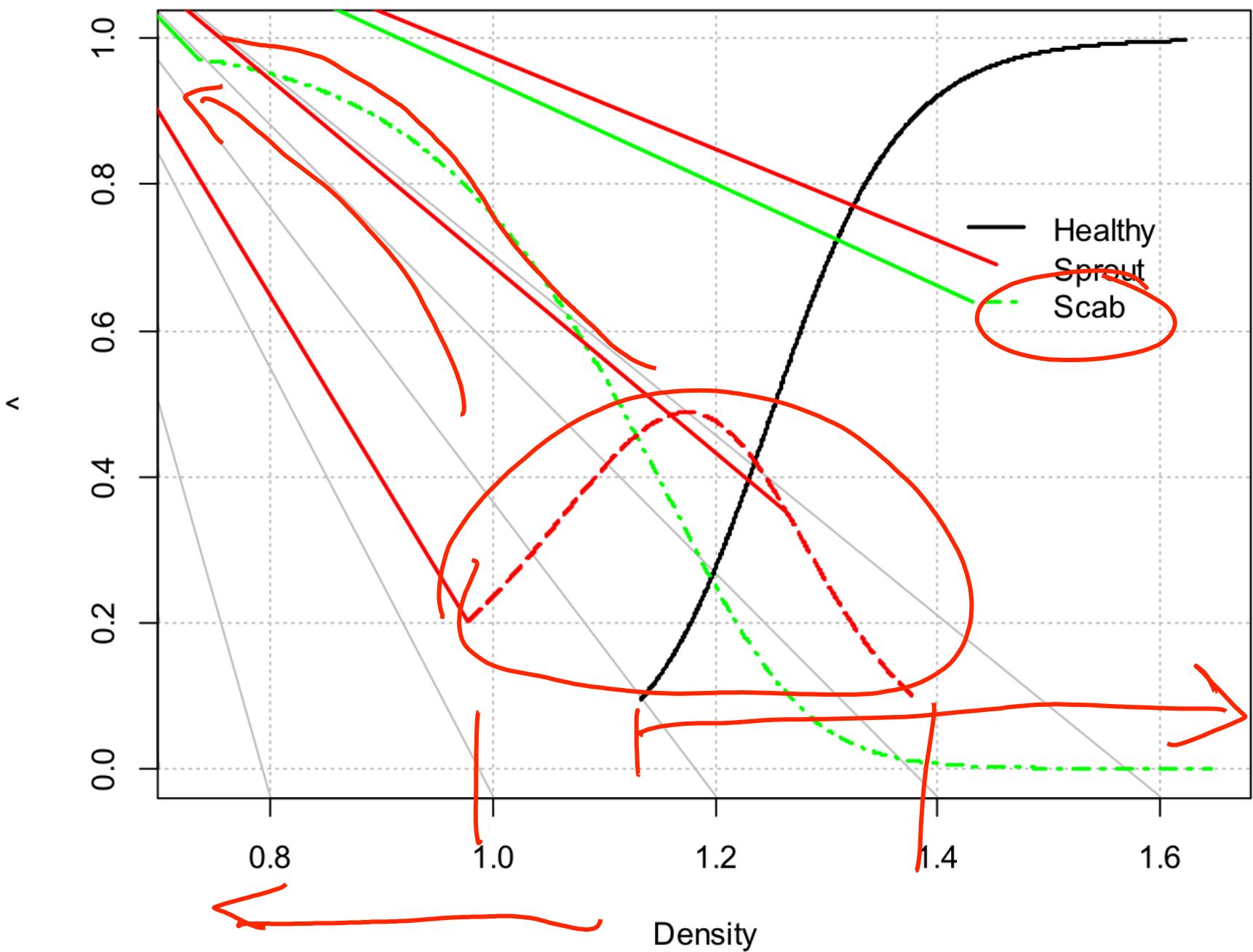
# Visualize Estimated Probability

When there is only one explanatory variable in the model, we can easily examine the estimated probabilities through a plot. The model using only density is

$$\log(\hat{\pi}_{\text{scab}} / \hat{\pi}_{\text{healthy}}) = 29.38 - 24.56 \text{density}$$

and

$$\log(\hat{\pi}_{\text{sprout}} / \hat{\pi}_{\text{healthy}}) = 19.12 - 15.48 \text{density}$$



# Estimated Probabilities

Below is a summary of the estimated probabilities for selected density values:

	<u>density.values</u>	Healthy	Scab	Sprout
1	0.8	0.00	0.95	0.05
2	0.9	0.00	0.89	0.11
3	1.0	0.01	0.76	0.24
4	1.1	0.05	0.54	0.41
5	1.2	0.27	0.25	0.48
6	1.3	0.69	0.05	0.25
7	1.4	0.92	0.01	0.07
8	1.5	0.98	0.00	0.02
9	1.6	1.00	0.00	0.00



- The lines are drawn from the smallest to the largest observed density value for a wheat kernel condition.
- We see that the estimated scab probability is the largest for the smaller density kernels. The estimated healthy probability is the largest for the high-density kernels. For density levels in the middle, sprout has the largest estimated probability. The parallel coordinates plot displays similar findings where the density levels tend to follow the scab < sprout < healthy ordering.

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Odds Ratio

# Odds Ratio

Because the log-odds are being modeled directly in a multinomial regression model, odds ratios are useful for interpreting an explanatory variable's relationship with the response.

Consider the model again of

$$\log(\pi_j/\pi_1) = \beta_{j0} + \beta_{j1}x \text{ for } j = 2, \dots, J$$

The odds of a **category j response vs. a category 1 response** are  $\exp(\beta_{j0} + \beta_{j1}x)$ . This directly leads to using odds ratios as a way to understand the explanatory variable in the model. Thus, **the odds of a category j vs. a category 1 response change by  $e^{c\beta_{j1}}$  times for every c-unit increase in x.**

In a similar manner, we could also compare category j to j' ( $j \neq j', j > 1, j' > 1$ ):

$$e^{c(\beta_{j1} - \beta_{j'1})}$$

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Odds Ratio

# Remarks

- The odds ratio interpretation specifically states “the odds of a category  $j$  vs. a category 1” comparison. In the past when  $Y$  was a binary response, we said something like “the odds of a success” only because it was assumed that a comparison was being made to the one other response category (failure).
- When there is more than one explanatory variable, we will need to include a statement like “holding the other variables in the model constant.”
- Adjustments need to be made to an odds ratio interpretation when interactions or transformations are present in the model.
- Wald and LR-based inference methods for odds ratios are performed in the same ways as for likelihood procedures discussed in earlier weeks.

# Example: Continue With the Wheat Example

Recall that we have these variables in the dataframe:

```
> summary(wheat)
   class      density      hardness      size      weight
hrw:143  Min.   :0.7352  Min.   :-44.080  Min.   :0.5973  Min.   : 8.532
srw:132  1st Qu.:1.1358  1st Qu.: 0.689  1st Qu.:1.8900  1st Qu.:21.982
          Median :1.2126  Median : 24.465  Median :2.2303  Median :27.610
          Mean   :1.1885  Mean   : 25.564  Mean   :2.2047  Mean   :27.501
          3rd Qu.:1.2687  3rd Qu.: 45.606  3rd Qu.:2.5125  3rd Qu.:32.882
          Max.   :1.6454  Max.   :111.934  Max.   :4.3100  Max.   :46.334

  moisture      type
Min.   : 6.486  Healthy:96
1st Qu.: 9.540  Scab   :83
Median :11.909  Sprout :96
Mean   :11.192
3rd Qu.:12.538
Max.   :14.514
```

# Example: Continue With the Wheat Example

- Because most of the explanatory variables are “continuous,” we use a value of **c** equal to 1 standard deviation.
- Ideally, it would be best to talk to the subject matter researcher about possible values for c.

```
> sd.wheat<-apply(X = wheat[,-c(1,7,8)], MARGIN = 2, FUN = sd)
> c.value<-c(1, sd.wheat) # class = 1 is first value
> round(c.value,2)
      density hardness      size    weight moisture
1.00     0.13     27.36     0.49     7.92     2.03
```

```
#beta.hat.jr for r = 1, ..., 6 and j = 2, 3
beta.hat2<-coefficients(mod.fit)[1,2:7]
beta.hat3<-coefficients(mod.fit)[2,2:7]
```

```
#OR for j = 2 (scab vs. healthy)
round(exp(c.value*beta.hat2), 2)
      density hardness      size    weight moisture
0.52     0.06     0.65     1.69     0.10     1.25
round(1/exp(c.value*beta.hat2), 2)
      density hardness      size    weight moisture
1.91    17.04    1.55     0.59     9.90     0.80
```

```
#OR for j = 3 (sprout vs. healthy)
round(exp(c.value*beta.hat3), 2)
      density hardness      size    weight moisture
0.80     0.14     0.56     1.54     0.69     0.92
round(1/exp(c.value*beta.hat3), 2)
      density hardness      size    weight moisture
1.25    7.28     1.78     0.65     1.45     1.09
```

# Example: Continue With the Wheat Example

```
> sd.wheat<-apply(X = wheat[,-c(1,7,8)], MARGIN = 2, FUN = sd)
> c.value<-c(1, sd.wheat) # class = 1 is first value
> round(c.value,2)
      density hardness   size  weight moisture
 1.00     0.13    27.36  0.49    7.92   2.03
 3
#OR for j = 2 (scab vs. healthy)
round(exp(c.value*beta.hat2), 2)
      density hardness   size  weight moisture
0.52     0.06    0.65   1.69   0.10   1.25
round(1/exp(c.value*beta.hat2), 2)
      density hardness   size  weight moisture
1.91    17.04   1.55   0.59   9.90   0.80
```

```
#OR for j = 3 (sprout vs. healthy)
round(exp(c.value*beta.hat3), 2)
      density hardness   size  weight moisture
0.80     0.14    0.56   1.54   0.69   0.92
round(1/exp(c.value*beta.hat3), 2)
      density hardness   size  weight moisture
1.25    7.28    1.78   0.65   1.45   1.09
```

- The estimated odds of a sprout vs. a healthy response change by **7.28 times for a 0.13 decrease in the density**, holding the other variables constant.
- The estimated odds of a scab vs. healthy response change by **9.90 times for a 7.92 decrease in the weight**, holding the other variables constant.

- The estimated odds of a scab vs. a healthy response change by **0.06 times for a 0.13 increase in the density**, holding the other variables constant.
- Likewise, the estimated odds of a scab vs. a healthy response change by **17.04 times for a 0.13 decrease in the density**, holding the other variables constant.

# Interpretation (cont.)

- The estimated odds of a sprout vs. healthy response change by 1.45 times for a 7.92 decrease in the weight, holding the other variables constant. Note that a Wald test of  $H_0: \beta_{35} = 0$  vs.  $H_a: \beta_{35} \neq 0$ , which uses the parameter needed for this sprout vs. healthy odds ratio, has a p-value of 0.2, so this odds ratio may not be interpreted in actual applications.

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Odds Ratio

# Wald Confidence Intervals

```
> conf.beta<-confint(object = mod_fit, level = 0.95)
> round(conf.beta,2) # Results are stored in a 3D array
, , Scab
```

	2.5 %	97.5 %
(Intercept)	22.14	38.95
classsrw	-1.95	0.65
density	-27.70	-15.49
hardness	-0.04	0.00
size	-0.44	2.58
weight	-0.41	-0.17
moisture	-0.19	0.41

, , Sprout

	2.5 %	97.5 %
(Intercept)	11.78	26.55
classsrw	-1.21	0.76
density	-20.53	-9.70
hardness	-0.04	-0.01
size	-0.18	1.94
weight	-0.12	0.03
moisture	-0.26	0.18

```
> conf.beta[2:7,1:2,1] # C.I.s for beta_2r
2.5 % 97.5 %
classsrw -1.94776958 0.651514098
density -27.70474380 -15.489565975
hardness -0.03604523 0.004230411
size -0.44453927 2.582767006
weight -0.41058295 -0.168713512
moisture -0.19391723 0.413047326
> conf.beta[2:7,1:2,2] # C.I.s for beta_3r
2.5 % 97.5 %
classsrw -1.20652328 0.757046542
density -20.53461137 -9.698731394
hardness -0.03690744 -0.005133494
size -0.18459306 1.935820104
weight -0.11978643 0.025152642
moisture -0.26392179 0.177927888
```

# Confidence Intervals for the Odds Ratios

```
> ci.OR2<-exp(c.value*conf.beta[2:7,1:2,1])
> ci.OR3<-exp(c.value*conf.beta[2:7,1:2,2])
> round(data.frame(low = ci.OR2[,1], up = ci.OR2[,2]), 2)
      low   up
classsrw 0.14 1.92
density   0.03 0.13
hardness  0.37 1.12
size      0.80 3.55
weight    0.04 0.26
moisture  0.67 2.32
```

```
> round(data.frame(low = 1/ci.OR2[,2], up = 1/ci.OR2[,1]), 2)[c(2,5),] # Specific rows
to cut down on output in book
      low   up
density 7.64 38.00
weight  3.80 25.79
> round(data.frame(low = ci.OR3[,1], up = ci.OR3[,2]), 2)
      low   up
classsrw 0.30 2.13
density   0.07 0.28
hardness  0.36 0.87
size      0.91 2.59
weight    0.39 1.22
moisture  0.58 1.44
> round(data.frame(low = 1/ci.OR3[,2], up = 1/ci.OR3[,1]), 2)[c(2,3),] # Specific rows
to cut down on output in book
      low   up
density 3.57 14.82
hardness 1.15  2.74
```

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Contingency Table

# Constructing a Contingency Table to Test Independence

- The multinomial regression model provides a convenient way to perform the same test for independence as in Section 3.2 (of the text).
- We can treat the row variable  $X$  as a qualitative variable by constructing **I - 1 indicator variables**.
- Using  $Y$  as the response variable with category probabilities of  $\pi_1, \dots, \pi_J$ , we have the model

$$\log(\pi_j/\pi_1) = \beta_{j0} + \beta_{j2}x_2 + \dots + \beta_{jI}x_I \text{ for } j = 2, \dots, J$$

where  $x_2, \dots, x_I$  are used as indicator variables for  $X$  (subscript matches level of  $X$ ). This is a model under dependence.



# Test for Independence

A model under independence between  $X$  and  $Y$  is simply

$$\log(\pi_j/\pi_1) = \beta_{j0} \text{ for } j = 2, \dots, J$$

Notice that each category of  $Y$  can have a different  $\pi_j$ , but they do not change as a function of  $X$ .

A test for independence involves the hypotheses of

$$\begin{aligned} H_0: \beta_{j2} &= \dots = \beta_{jI} = 0 \text{ for } j = 2, \dots, J \\ H_a: \text{Not all equal for some } j \end{aligned}$$

Equivalently, we can state these hypotheses in terms of models:

$$\begin{aligned} H_0: \log(\pi_j/\pi_1) &= \beta_{j0} \text{ for } j = 2, \dots, J \\ H_a: \log(\pi_j/\pi_1) &= \beta_{j0} + \beta_{j2}x_2 + \dots + \beta_{jI}x_I \text{ for } j = 2, \dots, J \end{aligned}$$

# Example

Using bloating severity as the response variable and fiber source as the explanatory variable, a multinomial regression is

$$\log(\pi_j / \pi_{\text{None}}) = \beta_{j0} + \beta_{j1}\text{bran} + \beta_{j2}\text{gum} + \beta_{j3}\text{both}$$

where bran, gum, and both in the model represent corresponding indicator variables and the  $j$  subscript represents categories low, medium, and high. We can estimate this model using **multinom()**:

```
library(package = nnet)
mod.fit.nom<-multinom(formula = bloat ~ fiber, weights =
count, data = diet2)

# weights: 20 (12 variable)
initial value 66.542129
iter 10 value 54.519963
iter 20 value 54.197000
final value 54.195737
converged
```

# Example

```
> summary(mod.fit.nom)
Call: multinom(formula = bloat ~ fiber, data = diet2, weights = count)
```

Coefficients:

	(Intercept)	fiberbran	fibergum	fiberboth
low	-0.4057626	-0.1538545	0.4055575	1.322135
medium	-1.0980713	-0.8481379	1.5032639	1.503764
high	-12.4401085	-4.1103893	13.3561038	12.440403

Std. Errors:

	(Intercept)	fiberbran	fibergum	fiberboth
low	0.6455526	0.8997698	1.190217	1.056797
medium	0.8163281	1.3451836	1.224593	1.224649
high	205.2385583	1497.8087307	205.240263	205.240994

Residual Deviance: 108.3915

AIC: 132.3915

# Example

The `weights = count` argument in `multinom()` is used because each row of `diet2` represents contingency table counts rather than individual observations.

To perform a LRT for independence, we can use the `Anova()` function from the `car` package:

```
> library(package = car)
> Anova(mod.fit.nom)
# weights: 8 (3 variable)
initial value 66.542129
final value 63.635876
converged

Analysis of Deviance Table (Type II tests)
Response: bloat
      LR Chisq Df Pr(>Chisq)
fiber   18.9   9     0.026 *
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Ordinal Logistical Regression Model

# Introduction

Suppose that the response categories are **ordered** in the following way:

$$\text{category 1} < \text{category 2} < \dots < \text{category}$$

- For example, a response variable may be measured using a scale with categories strongly disagree, disagree, neutral, agree, or strongly agree.
- There is a natural ordering here!
- Logit transformations of the probabilities can incorporate these orderings in a variety of ways.
- In this section, we will focus on one way where probabilities are cumulated based on these orderings.

# Introduction

The cumulative probability for  $Y$  is

$$P(Y \mid j) = \pi_1 + \dots + \pi_j$$

for  $j = 1, \dots, J$ . Note that  $P(Y \mid J) = 1$ . The logit of the cumulative probabilities can be written as

$$\text{logit}[P(Y \leq j)] = \log \left[ \frac{P(Y \leq j)}{1 - P(Y \leq j)} \right] = \log \left[ \frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \right]$$

When there is only one explanatory variable  $x$ , we can allow the log odds to vary by using a **proportional odds model**:

$$\text{logit}[P(Y \leq j)] = \beta_{j0} + \beta_1 x$$

for  $j = 1, \dots, J - 1$ . Equivalently, the model is written as

$$P(Y \leq j) = \frac{\exp(\beta_{j0} + \beta_1 x)}{1 + \exp(\beta_{j0} + \beta_1 x)}$$

# Proportional Odds

- The model assumes that **the effects of the explanatory variables are the same regardless of which cumulative probabilities are used to form the log odds**
- The proportional odds name comes from there being no  $j$  subscripts on the  $\beta$  parameter.
- This means these parameters are the same for each possible log-odds.
- This leads to each odds being a multiple of  $\exp(\beta_{j_0})$ .
- $\beta_{1_0} < \dots < \beta_{J_0}$  due to the cumulative probabilities. Thus, the odds increasingly become larger for  $j = 1, \dots, J - 1$ .
- A proportional odds model actually is a special case of a cumulative probability model, which allows the parameter coefficient on each explanatory variable to vary as a function of  $j$ .

# More Than One Explanatory Variable

For more than one explanatory variable, the model becomes:

$$\text{logit}[P(Y \leq j)] = \beta_{j0} + \beta_1 x_1 + \dots + \beta_p x_p$$

for  $j = 1, \dots, J - 1$

What is  $\pi_j$  only? Consider the case of one explanatory variable  $x$  again:

$$\begin{aligned}\pi_j &= P(Y = j) \\ &= P(Y \leq j) - P(Y \leq j - 1) \\ &= \frac{e^{\beta_{j0} + \beta_1 x}}{1 + e^{\beta_{j0} + \beta_1 x}} - \frac{e^{\beta_{j-1,0} + \beta_1 x}}{1 + e^{\beta_{j-1,0} + \beta_1 x}}\end{aligned}$$

for  $j = 2, \dots, J - 1$

# Proportional Odds

$$\text{For } j = 1, \pi_1 = P(Y = 1) = P(Y | 1) = e^{\beta_{10} + \beta_1 x} / (1 + e^{\beta_{10} + \beta_1 x})$$

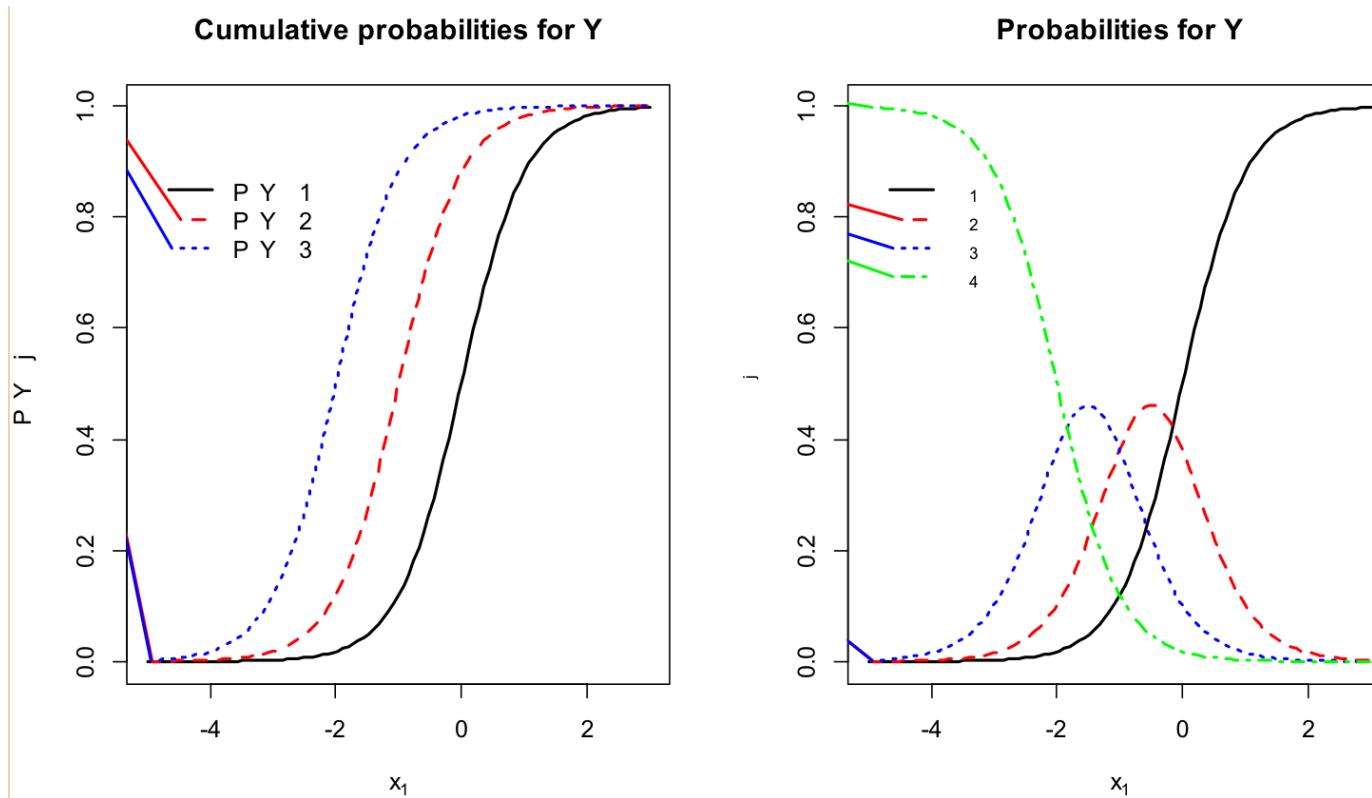
$$\begin{aligned}\text{For } j = J, \pi_J &= P(Y = J) = P(Y | J) - P(Y | J - 1) \\ &= 1 - \cancel{P(Y | J - 1)} \\ &= 1 - e^{\beta_{J-1,0} + \beta_1 x} / (1 + e^{\beta_{J-1,0} + \beta_1 x}).\end{aligned}$$

# Example

Examine the shape of the proportional odds model.

$$\text{logit}[P(Y \leq j)] = \beta_{j0} + 2x_1$$

where  $\beta_{10} = 0$ ,  $\beta_{20} = 2$ ,  $\beta_{30} = 4$ , and  $J = 4$



# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Estimation and Inference

# Estimation

- Parameters are estimated using maximum likelihood estimation.
- For a sample of size  $m$ , the likelihood function is simply the product of  $m$  multinomial distributions with probability parameters  $\pi_j$ , given the i.i.d. assumption.
- Iterative numerical procedures are used then to find the parameter estimates.
- The polr() function from the MASS package is used.

It is especially important to have the levels of the **categorical response ordered** in the desired way when using `polr()`; otherwise, the ordering of the levels of  $Y$  will not be correctly taken into account.

The covariance matrix for the parameter estimates follows from using standard likelihood procedures outlined in the appendix of the text.

# Inference

Suppose there is one explanatory variable ( $p = 1$ ), and the hypotheses of interest are

$$\begin{aligned} H_0: \beta_1 &= 0 \\ H_a: \beta_1 &\neq 0 \end{aligned}$$

- If the null hypothesis is not rejected, this says that the log-odds comparing  $P(Y \leq j)$  to  $P(Y > j)$  do not depend on this specific explanatory variable. In the case of one categorical explanatory variable  $X$ , this is equivalent to independence between  $X$  and  $Y$ .
- If the null hypothesis is rejected, the ordering of the log-odds comparing  $P(Y \leq j)$  to  $P(Y > j)$  holds; i.e., the log-odds progressively grow larger or smaller depending on the sign of  $\beta_1$ .

# Example

As an example, imagine that **the presence of disease** makes sprouted kernels more desirable than scab kernels.

**scab (Y = 1) < sprout (Y = 2) < healthy (Y = 3)**

```
> levels(wheat$type)
[1] "Healthy" "Scab"    "Sprout"
> wheat$type.order<-factor(wheat$type, levels = c("Scab",
  "Sprout", "Healthy"))
> #head(wheat)  #excluded to save space
> levels(wheat$type.order)
[1] "Scab"      "Sprout"   "Healthy"
```

We order it ourselves

Estimate the following model:

$$\text{logit}[P(Y \leq j)] = \beta_{j0} + \beta_1 x_1 + \dots + \beta_6 x_6$$

|for j = 1, 2

# Example

```

library(package = MASS) ↑
mod.fit.ord<-polr(formula = type.order ~ class + density
+ hardness + size + weight + moisture, data = wheat,
→ method = "logistic")
summary(mod.fit.ord) ↑

```

Re-fitting to get Hessian

Call: polr(formula = type.order ~ class + density + hardness +
size + weight + moisture, data = wheat, method = "logistic") )

Coefficients:

	Value	Std. Error	t value
classrw	0.17370	0.391764	0.4434
density	13.50534	1.713009	7.8840
hardness	0.01039	0.005932	1.7522
size	-0.29253	0.413095	-0.7081
weight	0.12721	0.029996	4.2411
moisture	-0.03902	0.088396	-0.4414

Intercepts:

	Value	Std. Error	t value
Scab Sprout	17.5724	2.2460	7.8237
Sprout Healthy	20.0444	2.3395	8.5677

Residual Deviance: 422.4178

AIC: 438.4178

# Example

Because the actual model estimated by polr() is

$$\text{logit}(\hat{P}(Y \leq j)) = \beta_{j0} - \eta_1 X_1 - \dots - \eta_p X_p$$

where  $-\eta_r$  is  $\beta_r$  in our notation. Thus, we will always need to change the sign of the estimated parameter given by polr(). The estimated model is:

$$\begin{aligned}\text{logit}(\hat{P}(Y \leq j)) &= \hat{\beta}_{j0} - 0.17\text{SRW} - 13.51\text{density} \\ &\quad - 0.01\text{hardness} + 0.29\text{size} \\ &\quad - 0.13\text{weight} + 0.04\text{moisture}\end{aligned}$$

where  $\hat{\beta}_{10} = 17.57$  and  $\hat{\beta}_{20} = 20.04$ .

# Example

The "t value" column in the coefficients table provides the Wald statistic for testing

$$H_0: \beta_r = 0 \text{ vs. } H_a: \beta_r \neq 0$$

for  $r = 1, \dots, 6$ , and the Anova() function provides the corresponding LRTs:

```
> library(package = car) #If not done already
> Anova(mod.fit.ord)
Analysis of Deviance Table (Type II tests)
```

	Response: type.order	LR	Chisq	Df	Pr(>Chisq)	
class			0.197	1	0.65749	
density			98.437	1	< 2.2e-16	***
hardness			3.084	1	0.07908	.
size			0.499	1	0.47982	
weight			18.965	1	1.332e-05	***
moisture			0.195	1	0.65872	
---						
Signif. codes:	0	'****'	0.001	'**'	0.01	'*'
					0.05	'.'
					0.1	' '
					1	

# Example

```
> pi.hat.ord<-predict(object = mod.fit.ord, type = "probs")
```

```
> head(pi.hat.ord)
```

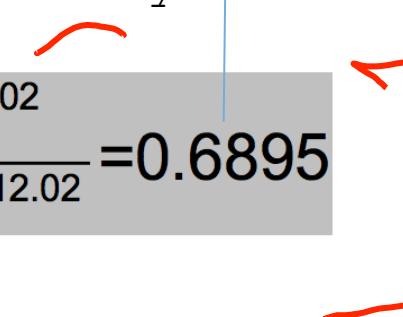
	Scab	Sprout	Healthy
1	0.03661601	0.2738502	0.6895338
2	0.03351672	0.2576769	0.7088064
3	0.08379891	0.4362428	0.4799583
4	0.01694278	0.1526100	0.8304472
5	0.11408176	0.4899557	0.3959626
6	0.02874814	0.2308637	0.7403882

```
> head(predict(object = mod.fit.ord, type = "class"))
```

[1] Healthy Healthy Healthy Healthy Sprout Healthy

Levels: Scab Sprout Healthy

The estimated probability of being healthy for the first observation is

$$\hat{\pi}_{\text{healthy}} = 1 - \frac{e^{20.04 - 0.17 \times 0 + 0.04 \times 12.02}}{1 + e^{20.04 - 0.17 \times 0 + 0.04 \times 12.02}} = 0.6895$$


# Example

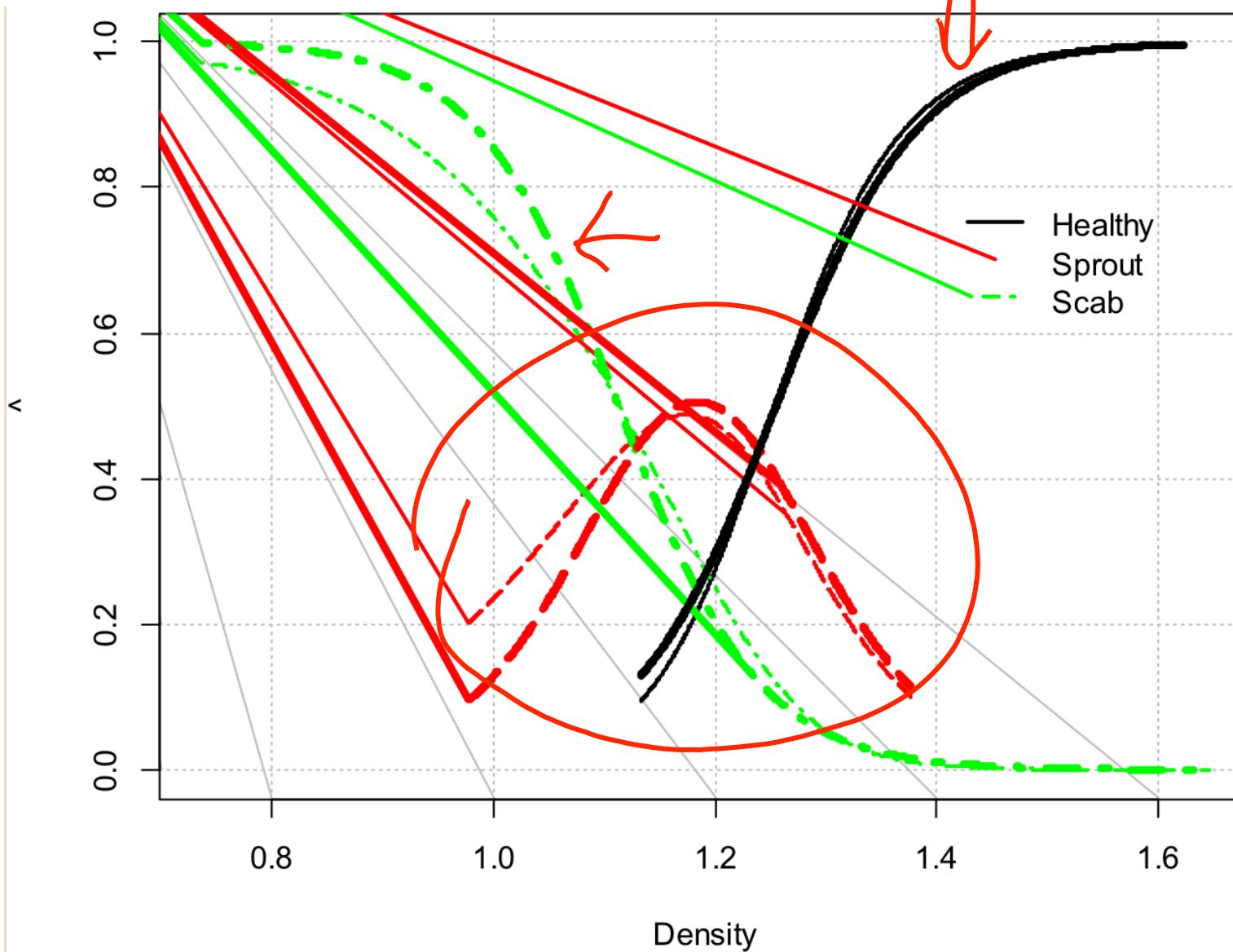
When there is only one explanatory variable in the model, we can easily examine the estimated probabilities through a plot.

The model using only density is

$$\text{logit}(\hat{P}(Y \leq j)) = \hat{\beta}_{j0} - 15.64\text{density}$$

$$\hat{\beta}_{10} = 17.41 \text{ and } \hat{\beta}_{20} = 19.63.$$

# Example



Berkeley

SCHOOL OF  
INFORMATION

# Discrete Response Model

## Lecture 4

---

**datascience@berkeley**

# Odds Ratio

# Odds Ratio

Odds ratios are easily formed because the proportional odds model equates log-odds to the linear predictor. The main difference now is the odds involve cumulative probabilities .

Consider the model again of

$$\text{logit}[P(Y \leq j)] = \beta_{j0} + \beta_1 x$$

The odds ratio is

$$\frac{\text{Odds}_{x+c}(Y \leq j)}{\text{Odds}_x(Y \leq j)} = \frac{e^{\beta_{j0} + \beta_1(x+c)}}{e^{\beta_{j0} + \beta_1 x}} = e^{c\beta_1}$$

where  $\text{Odds}_x(Y \leq j)$  denotes the odds of observing category  $j$  or smaller for  $Y$ .

# Odds Ratios: Interpretation

The formal interpretation of the odds ratio is

$\exp(c\beta_1)$

The odds of  $Y \leq j$  vs.  $Y > j$  change by  $\exp(\beta_1)$  times for a c-unit increase in x

Interestingly, this odds ratio stays the same no matter what response category is used for  $j$ .

This is again due the absence of a  $j$  subscript on  $\beta_1$  in the model.

## Notes:

1. When there is more than one explanatory variable, we will need to include a statement like "holding the other variables in the model constant."
2. Adjustments need to be made to an odds ratio interpretation when interactions or transformations are present in the model.
3. Wald and LR-based inference methods for odds ratios are performed in the same ways as for likelihood procedures discussed in earlier weeks.

# Example

The estimated odds ratios for each explanatory

variable are calculated as  $\exp(c\hat{\beta}_r)$  for  $r = 1, \dots, 6$ .

$c$  is set to be equal  
to one standard deviation for each continuous  
explanatory variable and  $c = 1$  for the SRW variable.

Below are the calculations (remember that  $-\hat{\eta}_r$  is  $\hat{\beta}_r$ ):

```
# Information about each variable to help with choosing c
summary(wheat)
sd.wheat<-apply(X = wheat[,-c(1,7,8)], MARGIN = 2, FUN = sd)
c.value<-c(1, sd.wheat)
round(c.value, 2) # class = 1 is first value
```

# Example

```
> head(wheat)
  class density hardness size weight moisture type type.order
1  hrw 1.349253 60.32952 2.30274 24.6480 12.01538 Healthy   Healthy
2  hrw 1.287440 56.08972 2.72573 33.2985 12.17396 Healthy   Healthy
3  hrw 1.233985 43.98743 2.51246 31.7580 11.87949 Healthy   Healthy
4  hrw 1.336534 53.81704 2.27164 32.7060 12.11407 Healthy   Healthy
5  hrw 1.259040 44.39327 2.35478 26.0700 12.06487 Healthy   Healthy
6  hrw 1.300258 48.12066 2.49132 33.2985 12.18577 Healthy   Healthy
> levels(wheat$type.order)
[1] "Scab"   "Sprout" "Healthy"
```

```
> summary(wheat)
  class      density      hardness      size      weight
hrw:143  Min.   :0.7352  Min.   :-44.080  Min.   :0.5973  Min.   : 8.532
srw:132  1st Qu.:1.1358  1st Qu.:  0.689  1st Qu.:1.8900  1st Qu.:21.982
          Median :1.2126  Median : 24.465  Median :2.2303  Median :27.610
          Mean   :1.1885  Mean   : 25.564  Mean   :2.2047  Mean   :27.501
          3rd Qu.:1.2687  3rd Qu.: 45.606  3rd Qu.:2.5125  3rd Qu.:32.882
          Max.   :1.6454  Max.   :111.934  Max.   :4.3100  Max.   :46.334
  moisture      type      type.order
  Min.   : 6.486  Healthy:96  Scab    :83
  1st Qu.: 9.540  Scab    :83  Sprout :96
  Median :11.909  Sprout :96  Healthy:96
  Mean   :11.192
  3rd Qu.:12.538
  Max.   :14.514
> sd.wheat<-apply(X = wheat[,-c(1,7,8)], MARGIN = 2, FUN = sd)
> c.value<-c(1, sd.wheat)
> round(c.value, 2) # class = 1 is first value
      density hardness      size      weight      moisture
1.00      0.13     27.36      0.49      7.92      2.03
```

# Example—Interpretation

```
> round(exp(c.value*(-mod.fit.ord$coefficients)),2)
      density hardness   size weight moisture
 0.84     0.17    0.75   1.15   0.37   1.08
> round(1/exp(c.value*(-mod.fit.ord$coefficients)),2)
      density hardness   size weight moisture
1.19    5.89   1.33   0.87   2.74   0.92
```

Example interpretations include:

- The **estimated odds of a scab ( $Y \leq 1$ ) vs. sprout or healthy ( $Y > 1$ ) response are 0.84 times as large for soft rather than hard red winter wheat.** Note that the corresponding 95% confidence interval for the class variable contains 1, as we will see shortly.
- The **estimated odds of a scab vs. sprout or healthy response change by 5.89 times for a 0.13 decrease in the density, holding the other variables constant.**
- The **estimated odds of a scab vs. sprout or healthy response change by 2.74 times for a 7.92 decrease in the weight, holding the other variables constant.**

# Example—Interpretation

- Because of the proportional odds, each of the previous interpretations can start with "**The estimated odds of a scab or sprout vs. healthy response are ... ,**" and the same estimated odds ratios would be used in the interpretation.
- One could put the interpretation in the following form (due to the proportional odds):

**The estimated odds of kernel quality being below a particular level change by \_\_ times for a \_\_ increase in \_\_, holding the other variables constant.**

Overall, we see that the larger the density and weight, the more likely a kernel is healthy. We can again relate these results back to parallel coordinates plot to see why these interpretations make sense.

# Profile Likelihood Ratio Confidence Intervals

```

> conf.beta<-confint(object = mod.fit.ord, level = 0.95)
Waiting for profiling to be done...
Re-fitting to get Hessian
→
> conf.beta
      2.5 %    97.5 %
classsrw -0.595305729  0.9435846
density   10.315429541 17.0363926
hardness  -0.001207582  0.0221078
size      -1.103021561  0.5245184
weight    0.069318186  0.1872189
moisture  -0.213254701  0.1339876
> c.value*(-conf.beta)
      2.5 %    97.5 %
classsrw  0.5953057 -0.9435846
density   -1.3544373 -2.2369136
hardness  0.0330348 -0.6047844
size      0.5411561 -0.2573352
weight    -0.5486839 -1.4819199
moisture  0.4335923 -0.2724253
> c.value[2]*(-conf.beta[2,])
      2.5 %    97.5 %
-1.354437 -2.236914
→
> ci<-exp(c.value*(-conf.beta))
> round(data.frame(low = ci[,2], up = ci[,1]), 2)
      low   up
classsrw 0.39 1.81
density   0.11 0.26
hardness  0.55 1.03
size      0.77 1.72
weight    0.23 0.58
moisture  0.76 1.54
> round(data.frame(low = 1/ci[,1], up = 1/ci[,2]), 2)
      low   up
classsrw 0.55 2.57
density  3.87 9.36
hardness 0.97 1.83
size     0.58 1.29
weight   1.73 4.40
moisture 0.65 1.31

```

The density odds ratio can be interpreted as: **With 95% confidence, the odds of a scab instead of a sprout or healthy response change by 3.87 to 9.36 times when density is decreased by 0.13, holding the other variables constant.**

Berkeley

SCHOOL OF  
INFORMATION