

Contents

Contents	1
4 Logistic Regression	2
4.1 The Logistic Regression Model	2
4.2 Statistical Inference for Logistic Regression	7
4.3 Logistic Regression with Categorical Predictors	11
4.4 Multiple Logistic Regression	14
4.5 Summarizing Predictive Power: Classification Tables, ROC Curves, and Multiple Correlation	22

Chapter 4

Logistic Regression

This chapter:

- Focuses on the statistical modeling of binary response variables, for which the response outcome for each subject is a success or a failure.
- Presents **logistic regression** which is the most popular model for binary data with a single and multiple explanatory variables.
- Presents ways to interpret the model, to conduct statistical inference for its parameters, and to summarize effects and predictive power.

4.1 The Logistic Regression Model

Consider:

Y : binary response variable

X : explanatory variable

$P(Y = 1) = \pi(x)$ (shows the value of $P(Y = 1)$ depends on the value x of the explanatory variable).

Assumption: The observations are independent binomial variates with parameter $\pi(x)$, which itself varies according to the value of x .

The Logistic Regression Model

The logistic regression model has a linear form for the logit of the success probability, that is, the logarithm of the odds,

$$\text{logit}[\pi(x)] = \log\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \alpha + \beta x$$

$$\pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

- The logit increases by β for every 1-unit increase in x .
- For a numerical explanatory variable, $\pi(x)$ changes as an S -shaped function of x .
- The effect parameter β determines the rate of increase or decrease of the S -shaped curve for $\pi(x)$.
- The sign of β indicates whether the curve ascends ($\beta > 0$) or descends ($\beta < 0$).

- The rate of change increases as $|\beta|$ increases.
- $\beta = 0 \longrightarrow$ the curve flattens to a horizontal straight line. The binary response variable is then independent of the explanatory variable.

Odds Ratio and Linear Approximation Interpretations

The odds of a success are:

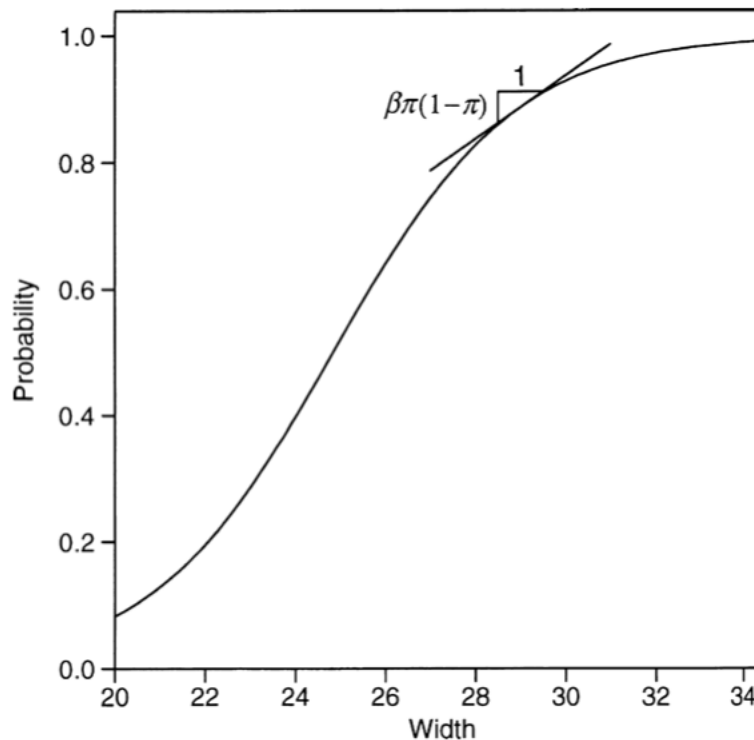
$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\alpha + \beta x) = e^\alpha (e^\beta)^x$$

By substituting $x + 1$ into formula above, we can conclude that:

For every 1-unit increase in x the odds multiply by e^β .

When $\beta = 0$, the odds do not change as x changes. ($e^\beta = e^0 = 1$)

The following graph is the linear approximation to the logistic regression curve which shows S -shaped appearance of the model for $\pi(x)$.



- Since it is curved rather than a straight line, the rate of change in $\pi(x)$ per 1-unit increase in x depends on the value of x .
- A straight line drawn tangent to the curve at a particular x value, describes the rate of change at that point.
- The slope of the line is: $\beta\pi(x)[1 - \pi(x)]$

- The slope of the line approaches 0 as $\pi(x)$ approaches 1.0 or 0.
- The steepest slope occurs when $\pi(x) = 0.50 \rightarrow \log\left(\frac{0.5}{1-0.5}\right) = \log(1) = 0 = \alpha + \beta x \rightarrow x = -\frac{\alpha}{\beta}$.

Median Effective Level

The median effective level represents the point at which each outcome has a 50% chance.

Example 4.1.1 Example: Whether a Female Horseshoe Crab Has Satellites

Consider the horseshoe crab data introduced in chapter 3. Dataset: Crabs

Here, we let y indicate whether a female crab has any satellites (other males who could mate with her).

If a female crab has at least one satellite $\rightarrow y = 1$

If she has no satellite $\rightarrow y = 0$ x : the female crab's shell width, in centimeters (cm); $21 < x < 31.5$

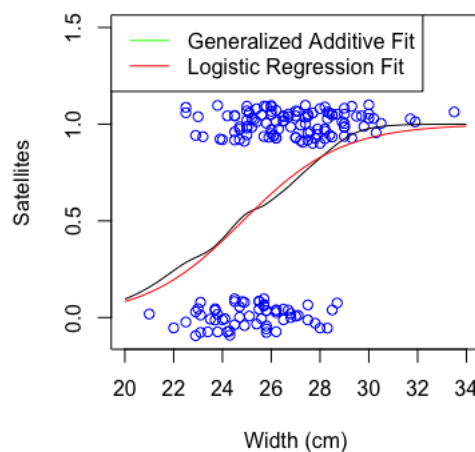
The following jitter plot helps visualize the relationship between the two variables. The data points are jittered in y -axis direction so we can see coordinates at which multiple observations occur.

It appears that $y = 1$ occurs relatively more often at higher x values. Since y takes only values 0 and 1, however, it is difficult to determine from a scatterplot whether a logistic regression model is appropriate.

The black curve is a curve that is the fit of a generalized additive model.

This smoothing curve shows an increasing trend, so we proceed with fitting a model that implies such a trend.

The red curve is the logistic regression fit which is added to the plot.



The following is logistic regression fit output:

From the output above, for probability $\pi(x)$ that a female horseshoe crab of width x has a satellite, the logistic regression model fit is:

$$\text{logit}[\hat{\pi}(x)] = -12.3508 + 0.4972x$$

```

> summary(fit)

Call:
glm(formula = y ~ width, family = binomial, data = Crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0281  -1.0458   0.5480   0.9066   1.6942

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.3508     2.6287  -4.698 2.62e-06 ***
width         0.4972     0.1017   4.887 1.02e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 194.45  on 171  degrees of freedom
AIC: 198.45

Number of Fisher Scoring iterations: 4

> predict(fit, data.frame(width=c(min(width), max(width))), type="response")
           1           2
0.1290960 0.9866974
> predict(fit, data.frame(width = mean(Crabs$width)), type="response")
           1
0.6738768

```

The estimated probability of a satellite is:

$$\hat{\pi}(x) = \frac{e^{-12.3508+0.4972x}}{1 + e^{-12.3508+0.4972x}}$$

Since $\hat{\beta} > 0$, the estimated probability is higher at larger width values.

At the minimum width in this sample of 21.0 cm, the estimated probability that the crab has at least one satellite is 0.129096.

At the maximum sample width of 33.5 cm, the estimated probability equals 0.9866974.

The effect of width seems relatively strong, in the sense that $\hat{\pi}(x)$ changes substantially over the range of x values.

Question: Calculate the median effective level.

$$\hat{\pi}(x) = 0.50 \rightarrow \log\left(\frac{0.5}{1-0.5}\right) = \log(1) = 0 = \hat{\alpha} + \hat{\beta}x \rightarrow x = -\frac{\hat{\alpha}}{\hat{\beta}} = -\frac{-12.3508}{0.4972} = 24.83922.$$

Question: Calculate the incremental rate of change in the fitted probability at the sample mean and interpret the value.

$$\text{mean}(\text{width}) = 26.29884 \rightarrow \hat{\pi}(x) = 0.6738768$$

$$\text{The slope of the line is: } 0.4972 \times 0.6738768[1 - 0.6738768] = 1.02744$$

For female crabs near the mean width, the estimated probability of having at least one satellite increases at the rate of 1.02744 per 1-cm increase in width.

Question: Interpret the value of $\hat{\beta}$.

The value of $\hat{\beta} = 0.4972$ indicates that the estimated odds of a satellite multiply by $\exp(\hat{\beta}) = \exp(0.4972) = 1.644162$ for each 1-cm increase in width; that is, there is a 64% increase.

Logistic Regression with Retrospective Studies

In retrospective sampling designs (such as case-control studies), the explanatory variable X rather than the response variable Y is random.

Consider Example 2.3.3, the study of the association between lung cancer and smoking.

For samples of subjects, cases $\rightarrow y = 1$

For samples of subjects, controls $\rightarrow y = 0$

For binary X , we then treat the observations on X , given y , as binomial samples, instead of the observations on Y , given x .

The odds ratio for X given y equals that for Y given x , so we can estimate odds ratios in such studies. Logistic regression effect parameters refer to odds and odds ratios, so we can estimate them.

Recall calculating the odds ratio of 3.0 using the conditional distribution of X given Y .

Using logistic regression:

Response variable, y is the lung cancer indicator.

Explanatory variable, x :

Smokers $\rightarrow x = 1$

Nonsmokers $\rightarrow x = 0$

Describe the following outputs:

```
> Smlung
  Smoker Cases Controls
1      1   688      650
2      0    21       59
> oddsratio(c(688,650,21,59), method="wald", conf=0.95, correct=FALSE)
$data
      Outcome
Predictor Disease1 Disease2 Total
Exposed1    688     650   1338
Exposed2     21      59     80
Total       709     709   1418

$measure
      odds ratio with 95% C.I.
Predictor estimate lower upper
Exposed1 1.000000      NA      NA
Exposed2 2.973773 1.786737 4.949427

--
> fit<- glm(Cases/n~Smoker, family=binomial, weights=n, data=Smlung)
> summary(fit)

Call:
glm(formula = Cases/n ~ Smoker, family = binomial, data = Smlung,
    weights = n)

Deviance Residuals:
[1]  0  0

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0330      0.2541  -4.065 4.80e-05 ***
Smoker        1.0898      0.2599   4.193 2.75e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1.9878e+01 on 1 degrees of freedom
Residual deviance: 2.6201e-14 on 0 degrees of freedom
AIC: 16.237

Number of Fisher Scoring iterations: 3

> exp(fit$coefficients[2])
  Smoker
2.973773
```

Question: Can we estimate the intercept term α in the model?

Note:

With case-control studies, it is not possible to estimate effects in binary models with link functions other than the **logit**. The effect measure is not then an odds ratio, so the effect for the conditional distribution of X given y does not then equal that for Y given x .

Normally Distributed X Implies Logistic Regression for Y

The logistic regression model necessarily holds if:

$$X \sim N(\mu_1, \sigma^2) \quad \text{for } y = 1$$

and

$$X \sim N(\mu_0, \sigma^2) \quad \text{for } y = 0$$

That is, both normal, with possibly different means but with the same variance.

Then for the logistic regression curve, the effect of x is:

$$\beta = \frac{(\mu_1 - \mu_0)}{\sigma^2}$$

In particular, β has the same sign as $\mu_1 - \mu_0$. If those with $y = 1$ tend to have higher values of x , then $\beta > 0$.

Note:

If the distributions of X are bell-shaped but with highly different spreads when $y = 1$ and when $y = 0$, then a logistic model containing also a quadratic term (i.e., both x and x^2) often fits well. In that case, the relationship is not monotone. Instead, $P(Y = 1)$ increases and then decreases, or the reverse.

4.2 Statistical Inference for Logistic Regression

Statistical inference for the logistic regression model parameters helps us judge the significance and size of the effects of explanatory variables.

As the total sample size n increases, the SE values tend to decrease. If the sample size is fixed, the SE values are relatively large when the estimated probabilities $\{\hat{\pi}_i\}$ are mainly close to 0 or close to 1.

In other words, it is more difficult to estimate effects of explanatory variables well when nearly all the observations are successes (e.g., as in modeling the occurrence of a rare disease) compared to when a similar number of successes and failures occurs.

Confidence Intervals for Effects

Wald confidence interval:

A Wald confidence interval for the effect in the logistic regression model:

$$\hat{\beta} \pm z_{\frac{\alpha}{2}} SE$$

Exponentiating the endpoints yields an interval for e^{β} , the multiplicative effect on the odds of a 1-unit increase in x .

Note: When n is small or when fitted probabilities are mainly near 0 or 1, it is preferable to construct a **profile likelihood confidence** interval.

Example 4.2.1 Consider example Example 4.1.1, the logistic regression modeling of the horseshoe crab data in terms of the probability that a female crab has at least one male satellite, with shell width as the explanatory variable.

a) Confidence Intervals for β :

```
> fit <- glm(y ~ width, family=binomial, data=Crabs) # link=logit is default
> summary(fit)
```

```
Call:
glm(formula = y ~ width, family = binomial, data = Crabs)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0281  -1.0458   0.5480   0.9066   1.6942
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.3508      2.6287  -4.698 2.62e-06 ***
width         0.4972      0.1017   4.887 1.02e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 225.76 on 172 degrees of freedom
Residual deviance: 194.45 on 171 degrees of freedom
AIC: 198.45
```

```
Number of Fisher Scoring iterations: 4
```

```
> confint(fit)
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept) -17.8100090 -7.4572470
width         0.3083806  0.7090167
```

```
> library(car)
> Anova(fit)
Analysis of Deviance Table (Type II tests)

Response: y
      LR Chisq Df Pr(>Chisq)
width  31.306  1  2.204e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> confint.default(fit)
              2.5 %      97.5 %
(Intercept) -17.5030100 -7.1986254
width         0.2978326  0.6966286
```

```
> ##Profile CI
> exp(P[2,])
      2.5 %      97.5 %
1.361219  2.031992
> #Wald CI
> exp(W[2,])
      2.5 %      97.5 %
1.346936  2.006975
```


We infer that a 1-cm increase in width has at least a 36% increase and at most a doubling in the odds that a female crab has a satellite.

b) Significance Testing

$H_0 : \beta = 0$ (the probability of success is independent of x)

$H_a : \beta \neq 0$

Wald test:

$$z = \frac{0.4972}{.1017} = 4.887 \rightarrow P\text{-value} = 1.02e^{-06} < 0.$$

Likelihood-ratio test:

$$G = 31.306 \rightarrow P\text{-value} = 2.204e^{-08} < 0.$$

Both tests show strong evidence of a positive effect of width on the presence of satellites.

Fitted Values and Confidence Intervals for Probabilities

A **fitted value** each observation using the for logistic regression is:

$$P(\widehat{Y} = 1) = \frac{\exp(\hat{\alpha} + \hat{\beta}x)}{[1 + \exp(\hat{\alpha} + \hat{\beta}x)]}$$

Example 4.2.2 For the horseshoe crab data, calculate the estimated probability of a satellite for female crabs of width $x = 26.5$ and construct a confidence interval for $P(Y = 1)$ at this value.

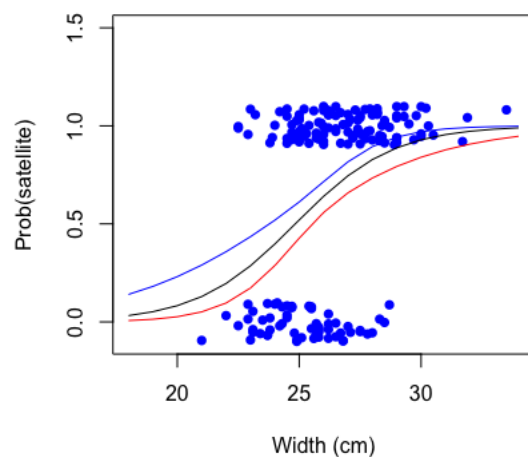
Note: To obtain the confidence interval, first find a confidence interval for the linear predictor value, $\alpha + \beta x$ and then apply the $\frac{\exp(\cdot)}{[1 + \exp(\cdot)]}$ transform to the endpoints.

```
> New=data.frame(width=c(26.5))
> Fitted=predict(fit, New, type="response")
> Fitted
      1
0.6954646
> lp=predict(fit,New, se.fit=TRUE)
> alfa=0.05
> a=1-(alfa/2)
> z = qnorm(a)
> CI=exp(lp$fit + c(-1, 1) * z * lp$se.fit)/(1+exp(lp$fit + c(-1, 1) * z * lp$se.fit))
> # fitted value and confidence bounds for P(Y=1)
> cbind(Fitted,"CI:LB"=CI[1], "CI:UB"=CI[2])
      Fitted      CI:LB      CI:UB
1 0.6954646 0.6120544 0.7677464
```

For female crabs of width $x = 26.5$, which is near the mean width, the estimated probability of a satellite is $P(\widehat{Y} = 1) = 0.695$ and a 95% confidence interval for $P(Y = 1)$ is $(0.61, 0.77)$.

The following graph shows the estimated probabilities and the lower and upper 95% confidence bands.

Why Use a Model to Estimate Probabilities?



We could use the sample proportion to estimate $P(Y = 1)$ instead of using the logistic model fit. Six crabs in the sample had width 26.5, and four of them had satellites.

```
n=length(which(Crabs$width ==26.5))
x=length(which(Crabs$width ==26.5&Crabs$y==1))
phat=x/n
phat
] 0.6666667
```

The sample proportion estimate at $x = 26.5$ is $\hat{\pi} = \frac{4}{6} = 0.67$, similar to the model-based estimate.

When the logistic regression model holds, the model-based estimator is much better than the sample proportion. It uses all the data rather than only the data at the particular x value. The result is a more precise estimate.

Compare the SE for the model-based estimate with SE for the sample proportion.

```
> SE=sqrt(phat*(1-phat)/n)
> SE
[1] 0.1924501
> predict(fit, New, type="response",se.fit=TRUE)
$fit
      1
0.6954646

$se.fit
      1
0.03996454

$residual.scale
[1] 1
```

The model-based estimator tends to be much closer than the sample proportion to the true value, unless the sample size on which that sample proportion is based is extremely large. The model smooths the sample data, somewhat dampening the observed variability.

4.3 Logistic Regression with Categorical Predictors

In this section, logistic regressions with multiple explanatory variables including some categorical explanatory variables, often called factors.

Indicator Variables Represent Categories of Predictors

Suppose that a binary response has two binary explanatory variables, which we denote by x and z . We can then display the data in a $2 \times 2 \times 2$ contingency table, such as analyzed in the next example.

We denote the two categories of x and z by the values 0 and 1. Then x and z are called indicator or dummy variables, because each indicates the category of an explanatory variable.

$$\text{logit}(P(Y = 1)) = \alpha + \beta_1 x + \beta_2 z$$

Question: Show the logit values at the four combinations of values of the two explanatory variables.

x	z	Logit
0	0	α
1	0	$\alpha + \beta_1$
0	1	$\alpha + \beta_2$
1	1	$\alpha + \beta_1 + \beta_2$

Note: This model assumes an absence of interaction. The effect of one factor is the same at each category of the other factor.

At a fixed category z , the effect on the logit of changing from $x = 0$ to $x = 1$ is:

$$\alpha + \beta_1(1) + \beta_2 z - (\alpha + \beta_1(0) + \beta_2 z) = \beta_1$$

This difference between two logits equals the difference of log odds. Equivalently, that difference equals the log odds ratio between x and y , at that category z . Thus, $\exp(\beta_1)$ equals the conditional odds ratio between x and y . For each category z , the odds of success at $x = 1$ equal $\exp(\beta_1)$ times the odds of success at $x = 0$. The lack of an interaction term implies a common value of the odds ratio for the partial tables at the two categories of z . The model satisfies homogeneous association.

Homogeneous Association

Two binary variables X and Y satisfy homogeneous association when

$$\theta_{XY(1)} = \theta_{XY(2)} = \dots,$$

that is, when all $\theta_{XY(k)}$ are identical. Conditional independence of X and Y is the special case in which each conditional odds ratio equals 1.0.

In a three-way table, homogeneous XY association means that any conditional odds ratio formed using two categories of X and two categories of Y is the same at each category of Z . Inference about associations in multi-way contingency tables is best handled in the context of models.

Example 4.3.1 *Survey about Marijuana Use*

The following table shows the result of a survey that asked students in their final year of a high school near Dayton, Ohio, whether they had ever used marijuana. The explanatory variables are gender and race.

Race	Gender	Marijuana Use	
		Yes	No
White	Female	420	620
	Male	483	579
Other	Female	25	55
	Male	32	62

Following are some results of fitting the logistic model with main effects of gender and race for predicting marijuana use ($1 = \text{yes}$, $0 = \text{no}$), using the grouped data file:

```
> fit <- glm(yes/(yes+no) ~ gender + race, weights = n, family=binomial, data=Marijuana)
> summary(fit)
```

Call:
glm(formula = yes/(yes + no) ~ gender + race, family = binomial,
data = Marijuana, weights = n)

Deviance Residuals:

1	2	3	4
-0.04513	0.04402	0.17321	-0.15493

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.83035	0.16854	-4.927	8.37e-07 ***
gendermale	0.20261	0.08519	2.378	0.01739 *
racewhite	0.44374	0.16766	2.647	0.00813 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 12.752784 on 3 degrees of freedom
Residual deviance: 0.057982 on 1 degrees of freedom
AIC: 30.414

Number of Fisher Scoring iterations: 3

RStudio has set up indicator variables for $x = \text{gender}$ ($1 = \text{male}$, $0 = \text{female}$) and for $z = \text{race}$ ($1 = \text{white}$, $0 = \text{other}$).

The ML estimated effects are:

For gender: $\hat{\beta}_1 = 0.20261$

For race: $\hat{\beta}_2 = 0.44374$

The estimated conditional odds ratio between marijuana use and gender and race respectively:

Gender:

$$\hat{\theta} = \exp(0.20261) = 1.2246$$

Interpretation

For each race, the estimated odds that a male ($x = 1$) had used marijuana were 1.22 times the estimated odds that a female had used marijuana.

Race:

$$\hat{\theta} = \exp(0.44374) = 1.558519$$

Interpretation

For each gender, the estimated odds that a person of white race ($z = 1$) had used marijuana were 1.56 times the estimated odds that a person of other races had used marijuana.

Question: Conduct a test to measure overall goodness-of-fit of the model.

H_0 : The model provides an adequate fit to the data

H_a : The model does not adequately fit the data

```
> p_value=pchisq(D,df, lower.tail = FALSE)
> p_value
[1] 0.8097152
> D=deviance(fit)
> D
[1] 0.05798151
> df=df.residual(fit)
> df
[1] 1
> p_value=pchisq(D,df, lower.tail = FALSE)
> p_value
[1] 0.8097152
```

For testing the null hypothesis that the model holds, the P - value is 0.8097152. The model seems to be adequate.

Model Comparison Using the Deviance

H_0 : $\beta_1 = \beta_2 = 0$; The simpler model with no effect of gender and race

H_a : At least one $\beta_i \neq 0$; A model with the effect of gender or race

```
- -
> G=fit$null.deviance-fit$deviance
> G
[1] 12.6948
> df_G=fit$df.null-fit$df.residual
> df_G
[1] 2
> p_value2=pchisq(G,df_G, lower.tail = FALSE)
> p_value2
[1] 0.001751292
```

This has chi- squared P -value = 0.001751292. We can conclude that at least one of the explanatory variables has an effect.

The likelihood-ratio test

```
> lrtest(fit)
Likelihood ratio test

Model 1: yes/(yes + no) ~ gender + race
Model 2: yes/(yes + no) ~ 1
  #Df LogLik Df  Chisq Pr(>Chisq)
1   3 -12.207
2   1 -18.555 -2  12.695  0.001751 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The comparison of the null and residual deviances, $12.75 - 0.06 = 12.69$, is the likelihood-ratio statistic with $df = 3 - 1 = 2$ for testing $H_0 : \beta_1 = \beta_2 = 0$. This has chi- squared P -value = 0.001751. We can conclude that at least one of the explanatory variables has an effect.

Tests for individual explanatory variables

Question:

Conduct the Wald and the likelihood-ratio test to determine whether gender has any effect on marijuana use, adjusting for race.

```
> Anova(fit)
Analysis of Deviance Table (Type II tests)

Response: yes/(yes + no)
      LR Chisq Df Pr(>Chisq)
gender  5.6662  1  0.017295 *
race    7.2770  1  0.006984 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.4 Multiple Logistic Regression

The general logistic regression model has multiple explanatory variables that can be quantitative, categorical, or both. For p explanatory variables, the model for the log odds is:

$$\text{logit}(P(Y = 1)) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

β_j : refers to the effect of x_j on the log odds that $Y = 1$, adjusting for the other x 's.

$\exp(\beta_j)$: multiplicative effect on the odds of a 1-unit increase in x_j , at a fixed value for $\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \dots + \beta_p x_p$, such as when we can hold constant $x_1, \dots, x_{j-1} + x_{j+1} + \dots + x_p$.

Example 4.4.1 Consider example Horseshoe Crabs with Color and Width Predictors; dataset: Crabs.

We continue the analysis of the horseshoe crab data by using both the female crab's shell width (quantitative) and color (categorical) as explanatory variables. Color has categories (1 = medium light, 2 = medium, 3 = medium dark, 4 = dark). Color is a surrogate for age, with older crabs tending to have a darker shell color. It is an ordinal variable, but we first treat it as a nominal-scale factor by using three indicator variables for the four color categories.

The model is:

$$\text{logit}(P(Y = 1)) = \alpha + \beta_1 x + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4$$

when:

x : shell width

$$c_2 = \begin{cases} 1 & \text{color} = \text{medium} \\ 0 & \text{otherwise} \end{cases} \quad c_3 = \begin{cases} 1 & \text{color} = \text{medium dark} \\ 0 & \text{otherwise} \end{cases} \quad c_4 = \begin{cases} 1 & \text{color} = \text{dark} \\ 0 & \text{otherwise} \end{cases} \quad \text{Base Level: color} = \text{medium light}$$

Here is edited R output for the model fit:

Question: Find the prediction equation for medium-light colored and for dark crabs.

```

> fit2 <- glm(y ~ width+factor(color), family=binomial, data=Crabs) # link=logit is default
> summary(fit2)

Call:
glm(formula = y ~ width + factor(color), family = binomial, data = Crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1124  -0.9848   0.5243   0.8513   2.1413

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -11.38519    2.87346  -3.962 7.43e-05 ***
width           0.46796    0.10554   4.434 9.26e-06 ***
factor(color)2  0.07242    0.73989   0.098   0.922
factor(color)3 -0.22380    0.77708  -0.288   0.773
factor(color)4 -1.32992    0.85252  -1.560   0.119
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 187.46  on 168  degrees of freedom
AIC: 197.46

Number of Fisher Scoring iterations: 4

```

$$\text{logit}(P(\widehat{Y} = 1)) = -11.38519 + 0.46796x + 0.07242c_2 - 0.22380c_3 - 1.32992c_4$$

For medium-light colored crabs:

$$c_2 = c_3 = c_4 = 0 \rightarrow$$

$$\text{logit}(P(\widehat{Y} = 1)) = -11.38519 + 0.46796x$$

For dark crabs:

$$c_2 = c_3 = 0, c_4 = 1 \rightarrow$$

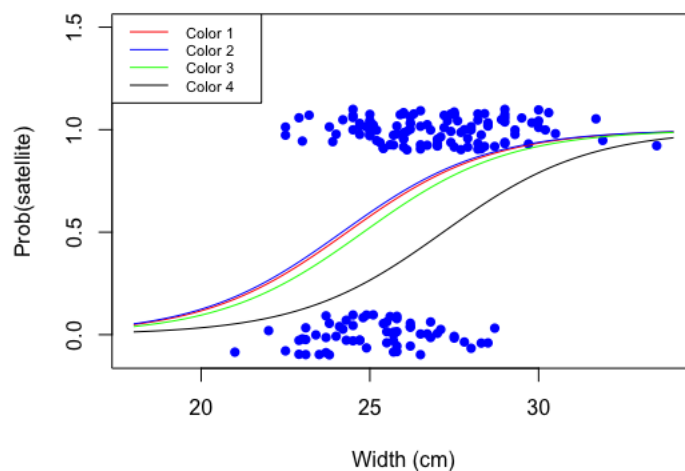
$$\text{logit}(P(\widehat{Y} = 1)) = -11.38519 + 0.46796x - 1.32992(1) = -12.71511 + 0.46796x$$

Question: Interpret the parameter estimate for the variable width.

For each color, a 1-cm increase in width has a multiplicative effect of $\exp(0.46796) = 1.596727$ on the odds that $Y = 1$.

Note: The model assumes a lack of interaction between width and color. Width has the same effect, with coefficient 0.468, for all colors. This implies that the shapes of the four curves relating width to $P(Y = 1)$ for the four colors are identical.

The following graph displays the fitted model:



Any one curve is any other curve shifted to the right or to the left.

The parallelism of curves in the horizontal dimension implies that two curves never cross. At all width values, for example, color 4 (black) has a lower estimated probability of a satellite than the other colors.

Compare the estimated probabilities of various colors of crabs with an average width.

```
> X=cbind(round(Predicted_P_Color_1,4),round(Predicted_P_Color_2,4),
+         round(Predicted_P_Color_3,4),round(Predicted_P_Color_4,4))
> colnames(X)=c("Color 1", "Color 2", "Color 3", "Color 4")
> rownames(X)=c("Predicted Probability")
> X
```

	Color 1	Color 2	Color 3	Color 4
Predicted Probability	0.7153	0.7299	0.6677	0.3993

Question: Show the exponentiated difference between two color parameter estimates is an odds ratio comparing those colors.

1- The difference in color parameter estimates between medium-light crabs and dark crabs equals 1.32992.

$$\text{Color 1: } \log(\hat{odds}_1) = \hat{\alpha} + \hat{\beta}_1 x + \hat{\beta}_2(0) + \hat{\beta}_3(0) + \hat{\beta}_4(0) = \hat{\alpha} + \hat{\beta}_1 x$$

$$\text{Color 4: } \log(\hat{odds}_4) = \hat{\alpha} + \hat{\beta}_1 x + \hat{\beta}_2(0) + \hat{\beta}_3(0) + \hat{\beta}_4(1) = \hat{\alpha} + \hat{\beta}_1 x + \hat{\beta}_4$$

$$\log(\hat{odds}_1) - \log(\hat{odds}_4) = \log(\hat{\theta}) = \hat{\alpha} + \hat{\beta}_1 x - [\hat{\alpha} + \hat{\beta}_1 x + \hat{\beta}_4] = -\hat{\beta}_4 \rightarrow$$

$$\hat{\theta} = \exp(-\hat{\beta}_4) = \exp(1.32992) = 3.780738$$

Interpretation

Therefore, at any fixed value of width, the estimated odds that a medium-light crab has a satellite are $\exp(1.32992) = 3.780738$ times the estimated odds for a dark crab.

Note:

Using the probabilities just calculated at width 26.3, the odds equal $\frac{0.3993}{1-0.3993} = 0.664706$ for a dark crab and $\frac{2.513079}{0.664706} = 3.780738$ for a medium-light crab, for which $\frac{2.513079}{0.664706} = 3.780738$.

Model Comparison to Check Whether a Term is Needed

To check certain terms are needed in a model, we can compare the deviance values for that model and for the simpler model without those terms.

Question: Does color contribute significantly to the final model, fit1, in example Example 4.4.1?

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

H_a : The presence of at least one satellite is dependent on color

Reduced Model under H_0 :

$$\text{logit}(P(Y = 1)) = \alpha + \beta_1 x$$

Full model is:

$$\text{logit}(P(Y = 1)) = \alpha + \beta_1 x + \beta_2 c_2 + \beta_3 c_3 + \beta_4 c_4$$

The likelihood-ratio test compares the maximized log-likelihood L_1 for the full model to the maximized log-likelihood L_0 for the simpler model in which those parameters equal 0, the fit of which is shown again in the following output. The test statistic $2(L_1 - L_0)$ is identical to the difference between the deviances for the two models, $194.45 - 187.46 = 6.99$. Under H_0 , this test statistic has an approximate chi-squared distribution with $df = 3$, the difference between the numbers of parameters in the two models.

```
> fit3 <- glm(y ~ width, family=binomial, data=Crabs) # link=logit is default
> summary(fit3)
Call:
glm(formula = y ~ width, family = binomial, data = Crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0281  -1.0458   0.5480   0.9066   1.6942

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.3508      2.6287  -4.698 2.62e-06 ***
width         0.4972      0.1017   4.887 1.02e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 194.45  on 171  degrees of freedom
AIC: 198.45

Number of Fisher Scoring iterations: 4

> fit1 <- glm(y ~ width+c2+c3+c4, family=binomial, data=Crabs) # link=logit is default
> summary(fit1)
Call:
glm(formula = y ~ width + c2 + c3 + c4, family = binomial, data = Crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1124  -0.9848   0.5243   0.8513   2.1413

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.38519      2.87346  -3.962 7.43e-05 ***
width         0.46796      0.10554   4.434 9.26e-06 ***
c2           0.07242      0.73989   0.098  0.922
c3          -0.22380      0.77708  -0.288   0.773
c4          -1.32992      0.85252  -1.560   0.119
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 187.46  on 168  degrees of freedom
AIC: 197.46

Number of Fisher Scoring iterations: 4
```

$$G = 2(L_1 - L_0) = 2(L_S - L_0) - 2(L_S - L_1) = \text{Deviance}(e_0) - \text{Deviance}(e_1) = 194.45 - 187.46 = 6.995631$$

$$df = 171 - 168 = 3$$

$$P\text{-value} = 0.07203714$$

```
> G = fit3$deviance - fit1$deviance
> G
[1] 6.995631
> df=fit3$df.residual-fit1$df.residual
> P_Value=pchisq(G,df,lower.tail = FALSE)
> P_Value
[1] 0.07203714
```

The P -value of 0.07 provides slight evidence of a color effect.

We can obtain the test in R either by comparing the model deviances or by using the Anova function in the car package:

```
> Anova(fit2)
Analysis of Deviance Table (Type II tests)

Response: y
      LR Chisq Df Pr(>Chisq)
width      24.6038 1  7.041e-07 *** LR test of width effect
factor(color)  6.9956 3  0.07204 .   LR test of color effect
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: Since the analysis in the previous subsection noted that estimated probabilities are quite different for dark-colored crabs, it seems safest to leave color in the model.

Example 4.4.2 Consider the data set *Crabs* in Example 4.4.1. Treat color in a quantitative manner and fit a model.

We use scores $c = (1, 2, 3, 4)$ for the color categories and fit the model:

$$\text{logit}[P(Y = 1)] = \alpha + \beta_1 x + \beta_2 c$$

```
> fit6 <- glm(y ~ width+color, family=binomial, data=Crabs)
> summary(fit6)

Call:
glm(formula = y ~ width + color, family = binomial, data = Crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1692  -0.9889   0.5429   0.8700   1.9742

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.0708     2.8068  -3.588 0.000333 ***
width         0.4583     0.1040   4.406 1.05e-05 ***
color        -0.5090     0.2237  -2.276 0.022860 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 189.12  on 170  degrees of freedom
AIC: 195.12

Number of Fisher Scoring iterations: 4
```

$$\text{logit}(\widehat{P}(Y = 1)) = -10.0708 + 0.4583x - 0.5090c$$

It supposes a linear effect, on the logit scale, for a set of scores assigned to its categories. It is advantageous to treat **ordinal explanatory variables** in a **quantitative** manner, when such models fit well. The model is simpler and easier to interpret, and tests of the effect of the variable are generally more powerful when it has a single parameter rather than several parameters.

Note: The results show **strong evidence** of an effect for each predictor.

Interpretation:

At a given width, for every one-category increase in color darkness, the estimated odds of a satellite multiply by $\exp(-0.509) = 0.60$.

For example, the estimated odds of a satellite for dark- colored crabs are 60% of those for medium-dark crabs.

Since $\exp[3(-0.509)] = 0.22$, the estimated odds of a satellite for dark-colored crabs are 22% of those for medium-light crabs.

Question:

Conduct a likelihood-ratio test to compare the previous model with the full model that has a separate parameter for each color.

H_0 : Model M (Simpler Model)

H_a : Model S (Saturated Model)

```
> anova(fit6, fit1, test="LRT")
Analysis of Deviance Table

Model 1: y ~ width + color
Model 2: y ~ width + c2 + c3 + c4
      Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1          170      189.12
2          168      187.46  2    1.6641    0.4351
```

The difference between the deviances equals $189.12 - 187.46 = 1.6641$, based on $df = 2$ (P -value = 0.4351). It tests that the color parameters in the full model, when plotted against the color scores, follow a linear trend. The simpler model seems to be adequate.

Potential Color Scoring

The estimates of the color parameters in the full model that treats color as a nominal- scale factor are $(0, 0.07, -0.22, -1.33)$; see model fit1. The 0 value for the first category reflects the lack of an indicator variable for that category. Although these values do not depart significantly from a linear trend, the first three are similar compared to the last one. This suggests that another potential color scoring for model (4.4) is $(0, 0, 0, 1)$; that is, $c = 1$ for dark-colored crabs, and $c = 0$ otherwise.

```
> fit5 <- glm(y ~ width+c4, family=binomial, data=Crabs) # link=logit is default
> summary(fit5)

Call:
glm(formula = y ~ width + c4, family = binomial, data = Crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0821  -0.9932   0.5274   0.8606   2.1553

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -11.6790     2.6925  -4.338 1.44e-05 ***
width         0.4782     0.1041   4.592 4.39e-06 ***
c4          -1.3005     0.5259  -2.473  0.0134 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 187.96  on 170  degrees of freedom
AIC: 193.96

Number of Fisher Scoring iterations: 4
```

```

> anova(fit5, fit1, test="LRT")
Analysis of Deviance Table

Model 1: y ~ width + c4
Model 2: y ~ width + c2 + c3 + c4
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      170      187.96
2      168      187.46  2   0.50085   0.7785

```

At a given width, the estimated odds that a dark crab has a satellite are $\exp(-1.3005) = 0.27$ times the estimated odds for a lighter-colored crab. The likelihood-ratio statistic comparing this model with binary scoring to full model with color factor suggests that this simpler model also seems adequate.

In summary, the model with color factor, the model with quantitative color scores (1, 2, 3, 4), and the model with binary color scores (0, 0, 0, 1) all suggest that dark crabs are the least likely to have satellites. When the sample size is not very large, typically several models can fit adequately.

Allowing Interaction between Explanatory Variables

In Example 4.4.1, we assumed a lack of interaction between width and color for the models fitted. Let us check now whether this is sensible. We can allow interaction by adding cross-products of terms for width and color. Each color then has a different-shaped curve relating width to the probability of a satellite, so a comparison of two colors varies according to the value of width.

Example 4.4.3 Consider Example 4.4.1 with the data set *Crabs*. Fit a model with an indicator variable for dark-colored crabs, width, and their interaction.

```

> fit4 <- glm(y ~ width+c4+width:c4, family=binomial, data=Crabs)
> summary(fit4)

Call:
glm(formula = y ~ width + c4 + width:c4, family = binomial, data = Crabs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1366  -0.9344   0.4996   0.8554   1.7753

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.8117     2.9576  -4.332 1.48e-05 ***
width         0.5222     0.1146   4.556 5.21e-06 ***
c4           6.9578     7.3182   0.951  0.342
width:c4     -0.3217     0.2857  -1.126  0.260
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 186.79  on 169  degrees of freedom
AIC: 194.79

Number of Fisher Scoring iterations: 4

```

The prediction equation is:

$$\text{logit}(P(\widehat{Y} = 1)) = -12.8117 + 0.5222x + 6.9578c4 - 0.3217(x \times c4)$$

For dark-colored crabs:

$$\text{logit}(P(\widehat{Y} = 1)) = -12.8117 + 0.5222x + 6.9578 - 0.3217(x \times 1) = -5.85383 + 0.2004296x$$

For lighter-colored crabs:

$$\text{logit}(P(\widehat{Y} = 1)) = -12.8117 + 0.5222x$$

The curve for lighter-colored crabs has a stronger effect for x .

The curves cross at x such that:

$$-5.85383 + 0.2004296x = -12.8117 + 0.5222x \rightarrow$$

$$12.61122 = 0.3217458x \rightarrow x = 21.62521$$

That is, at $x = 21.62521$ cm.

The sample widths range between 21.0 and 33.5 cm, so the lighter-colored crabs have a higher estimated probability of a satellite over nearly the entire range.

Question: Compare this model to the simpler model without interaction to analyze whether the fit is significantly better.

$H_0 : \beta_3 = 0$ (no-interaction model)

$H_a : \beta_3 \neq 0$ (model with interaction)

Reduced Model under H_0 :

$$\text{logit}(P(\widehat{Y} = 1)) = -11.6790 + 0.4782x - 1.3005c4$$

Full model is:

$$\text{logit}(P(\widehat{Y} = 1)) = -12.8117 + 0.5222x + 6.9578c4 - 0.3217(x \times c4)$$

Since $P\text{-value} = 0.2790881 > \alpha = 0.05 \rightarrow$ we fail to reject $H_0 \rightarrow$ The evidence of interaction is not strong.

```
> G = fit5$deviance - fit4$deviance
> G
[1] 1.171525
> df=fit5$df.residual-fit4$df.residual
> P_Value=pchisq(G,df,lower.tail = FALSE)
> P_Value
[1] 0.2790881
> |
```

Note: Although the sample slopes for the width effect are quite different for the two colors, the sample had only 22 crabs of dark color. Therefore, effects involving it have relatively large standard errors. The no-interaction model has the advantage of simpler interpretation.

4.5 Summarizing Predictive Power: Classification Tables, ROC Curves, and Multiple Correlation

For comparing models, it is often useful to summarize their predictive power, that is, how well we can predict the response variable outcome using the model fit. This section shows three ways to do this.

Summarizing Predictive Power: Classification Tables

A classification table cross-classifies the binary outcome y with a prediction of whether $y = 0$ or 1 .

The prediction for observation i for some cutoff $\hat{\pi}_0$ is:

- $\hat{y} = 1$ if $\hat{\pi}_i > \hat{\pi}_0$
- $\hat{y} = 0$ if $\hat{\pi}_i \leq \hat{\pi}_0$

Note: One possibility is to take $\hat{\pi}_0 = 0.5$. However, if a low (high) proportion of observations have $y = 1$, the model fit may never (always) have $\hat{\pi}_i > 0.5$, in which case one never (always) predicts $\hat{y} = 1$. Another possibility takes $\hat{\pi}_0$ as the sample proportion of 1 outcomes, which is $\hat{\pi}_i$ for the model containing only an intercept term.

Example 4.5.1 *We illustrate for the model for the presence of horseshoe crab satellites using width and the color factor as explanatory variables. (Model: fit 1)*

Sample proportion: $\hat{\pi} = \frac{111}{173} = 0.6416$

Of the 173 crabs, 111 had a satellite.

```
> n=length(width)
> yes=sum(y)
> phat=yes/n
> phat
[1] 0.6416185
> |
```

```
> predicted <- as.numeric(fitted(fit1) > phat) # predict y=1 when est.> 0.6416
> xtabs(~y + predicted) # Classification table with sample proportion cutoff
      predicted
y      0      1
0     43     19
1     36     75
```

Of 62 cases with $y = 0$, the model predicts $\hat{y} = 0$ for 43.

Of 111 cases with $y = 1$, the model predicts $\hat{y} = 1$ for 75.

Let us repeat the calculations with $\hat{\pi}_0 = 0.5$

```
> predicted2 <- as.numeric(fitted(fit1) > 0.5) # predict y=1 when est.> 0.5
> A=xtabs(~y + predicted2) # Classification table with sample proportion cutoff
> addmargins(A,2)
      predicted2
y      0      1 Sum
0     31     31  62
1     15     96 111
```

Of 62 cases with $y = 0$, the model predicts $\hat{y} = 0$ for 31.

Of 111 cases with $y = 1$, the model predicts $\hat{y} = 1$ for 96.

Two useful summaries of predictive power are:

Sensitivity = $P(\hat{y} = 1|y = 1)$ and Specificity = $P(\hat{y} = 0|y = 0)$

When $\pi_0 = \hat{\pi} = 0.6416$:

$$\widehat{Sensitivity} = \frac{75}{111} = 0.676$$

$$\widehat{Specificity} = \frac{43}{62} = 0.694.$$

$$\text{The overall proportion of correct classifications} = \frac{75+43}{111+62} = 0.682$$

$$\begin{aligned} P(\text{correct classif.}) &= P(y = 1 \cap \hat{y} = 1) + P(y = 0 \cap \hat{y} = 0) \\ &= P(\hat{y} = 1|y = 1)P(y = 1) + P(\hat{y} = 0|y = 0)P(y = 0) \\ &= \widehat{Sensitivity} \times P(y = 1) + \widehat{Specificity} \times (1 - P(y = 1)) \end{aligned}$$

which is a weighted average of sensitivity and specificity.

These sample summaries of predictive power are overly optimistic because they use $\hat{\pi}$ from the model fitted to the data set of which y_i was one element. It is better to make the prediction with the leave-one-out cross-validation approach by which $\hat{\pi}$ is based on the model fitted to the other $n - 1$ observations.

Limitations of a classification table:

- It collapses continuous predictive values $\hat{\pi}$ into binary ones.
- π_0 is arbitrary.
- Results are sensitive to the relative numbers of times that $y = 1$ and $y = 0$.

Summarizing Predictive Power: ROC Curves

A receiver operating characteristic (ROC) curve is a plot that shows the sensitivity and the specificity of the predictions for all the possible cutoffs π_0 .

This curve is more informative than a classification table, because it summarizes predictive power for all possible π_0 .

The ROC curve plots *sensitivity* on the vertical axis versus $(1 - \textit{specificity})$ on the horizontal axis.

When π_0 gets near 0, almost all predictions are $\hat{y} = 1$; then, sensitivity is near 1, specificity is near 0, and the point for $(1 - \textit{specificity}, \textit{sensitivity})$ has coordinates near $(1, 1)$.

When π_0 gets near 1, almost all predictions are $\hat{y} = 0$; then, sensitivity is near 0, specificity is near 1, and the point for $(1 - \textit{specificity}, \textit{sensitivity})$ has coordinates near $(0, 0)$.

The ROC curve usually has a concave or nearly concave shape connecting the points $(0, 0)$ and $(1, 1)$.

The following graph is the ROC curve for the model for the horseshoe crabs using width and the color factor as predictors.

When $\pi_0 = 0.6416$, for example, the *specificity* = 0.69, the *sensitivity* = 0.68, and the point plotted for the ROC curve has coordinates $(0.31, 0.68)$.

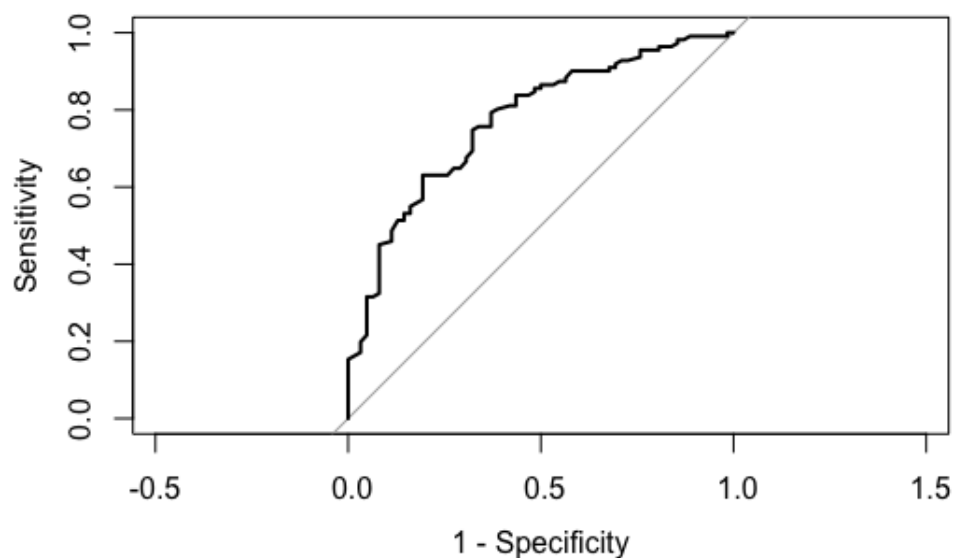
Note:

For a particular value of specificity, better predictive power corresponds to higher sensitivity. Therefore, the **better** the predictive power, the **higher** is the ROC curve. Because of this, the area under the curve provides a single value that summarizes predictive power. The greater the area, the better the predictive power. This measure of predictive power is called the **concordance index**.

Consider all pairs of observations (i, j) such that $y_i = 1$ and $y_j = 0$. The concordance index estimates the probability that the predictions and the out-comes are concordant, which means that the observation with the larger y also has the larger $\hat{\pi}$.

A concordance value of 0.50 means predictions were no better than random guessing.

This corresponds to a model having only an intercept term. Its ROC curve is a straight line connecting the points $(0, 0)$ and $(1, 1)$.



```
> auc(rocplot) # auc = area under ROC curve = concordance index
Area under the curve: 0.7714
```


Summarizing Predictive Power: Multiple Correlation

For a GLM, one summary of prediction power is the correlation R between the observed responses y_i and the model's fitted values $\hat{\mu}_i$.

For a binary regression model, R is the correlation between the n binary y_i observations (1 or 0 for each) and the fitted probabilities $\hat{\pi}_i$. The highly discrete nature of y can suppress the range of possible R values, more so when the imbalance between the frequencies of 0 and 1 values is greater. Also, like any correlation measure, the value of R depends on the range of values observed for the explanatory variables. Nevertheless, R is useful for comparing fits of different models for the same data.

```
[1] 0.4522131
> cor(y, fitted(fit1))
[1] 0.4522131
```

The following table shows the values of R for different models:

```
> R1=cor(y, fitted(fit1))
> R2=cor(y, fitted(fit3))
> R3=cor(y, fitted(fit4))
> R4=cor(y, fitted(fit5))
> X=cbind(round(R1,4),round(R2,4),
+         round(R3,4),round(R4,4))
> colnames(X)=c("width&Color", "width", "interaction", "width&color 4")
> rownames(X)=c("R")
> X
  width&Color width interaction width&color 4
R      0.4522 0.402      0.4519      0.447
```

Note:

The square of this measure does not have the proportional reduction in variation interpretation that it has for ordinary (least squares) regression.

Acknowledgement

The core content of the slides are from the textbook of this course;

AN INTRODUCTION TO CATEGORICAL DATA ANALYSIS (Third Edition)

by

Alan Agresti; WILEY SERIES IN PROBABILITY AND STATISTICS.