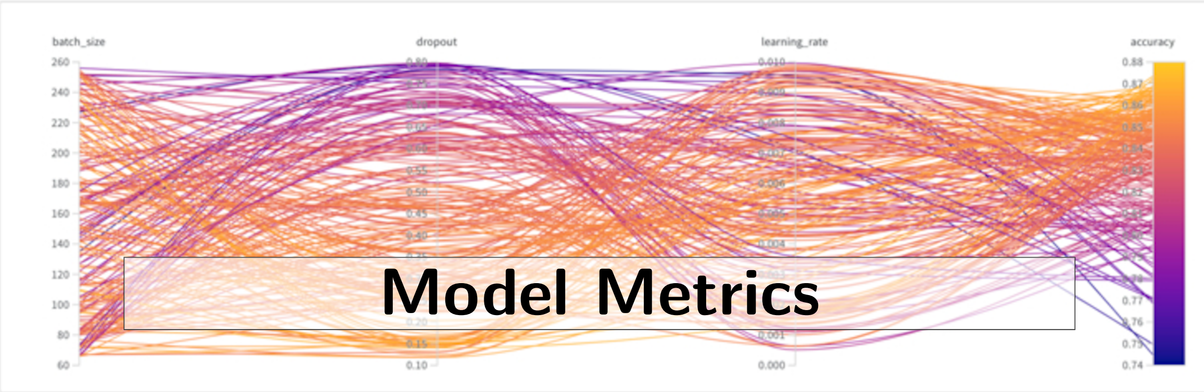
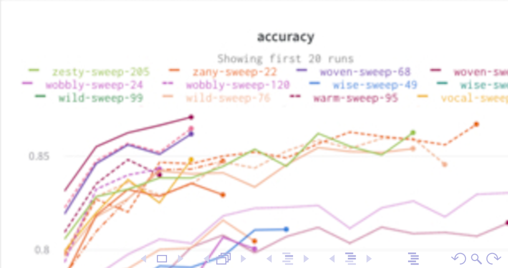
Model Metrics





Today:

- Introduce metrics for linear regression;
- Introduce metrics for binary dependent variables.

How good is the model?

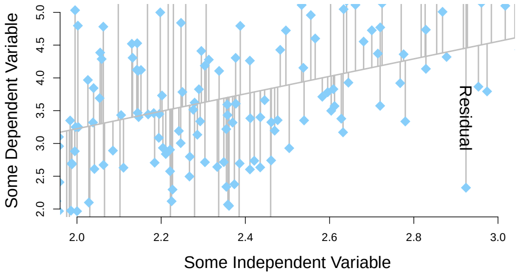
How well does the model predict?

Linear Regresssion: Root Mean Squared Error

- We could measure model fit by looking at the 'typical' error of the model when it makes predictions;

- Remember how regression works? It finds the βs by minimizing this:

$$\underbrace{\sum_i (\overbrace{y_i}^{\text{obs}} - \overbrace{(\beta_0 + \beta_1 x_i)}^{\text{prediction}})^2}_{\text{sum of squared residuals}}$$

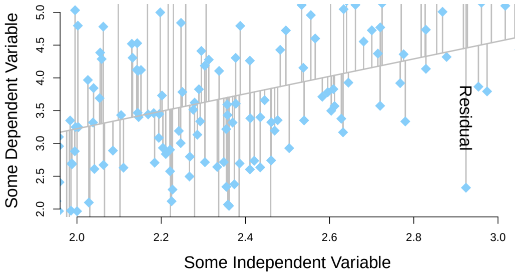


Linear Regresssion: Root Mean Squared Error

- We could measure model fit by looking at the 'typical' error of the model when it makes predictions;

- Remember how regression works? It finds the βs by minimizing this:

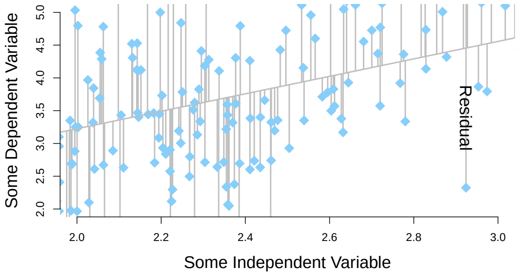
$$\underbrace{\sum_i (\overbrace{y_i}^{\text{obs}} - \overbrace{(\beta_0 + \beta_1 x_i)}^{\text{prediction}})^2}_{\text{total squared error}}$$



Linear Regresssion: Root Mean Squared Error

- We could measure model fit by looking at the 'typical' error of the model when it makes predictions;
- Remember how regression works? It finds the βs by minimizing this:

$$\underbrace{\frac{1}{n}\sum_i(\overbrace{y_i}^{\text{obs}} - \overbrace{(\beta_0 + \beta_1 x_i)}^{\text{prediction}})^2}_{\text{mean squared error}}$$

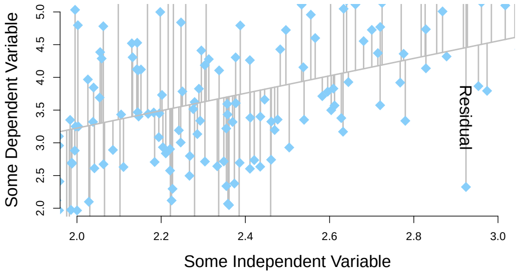


Linear Regresssion: Root Mean Squared Error

- We could measure model fit by looking at the 'typical' error of the model when it makes predictions;

- Remember how regression works? It finds the βs by minimizing this:

$$\underbrace{\sqrt{\frac{1}{n}\sum_i(\overbrace{y_i}^{\text{obs}} - \overbrace{(\beta_0 + \beta_1 x_i)}^{\text{prediction}})^2}}_{\text{root mean squared error}}$$



Linear Regresssion: Root Mean Squared Error

- We could measure model fit by looking at the 'typical' error of the model when it makes predictions;
- Remember how regression works? It finds the βs by minimizing this:

$$\underbrace{\sqrt{\frac{1}{n}\sum_i(\overbrace{y_i}^{\text{obs}} - \overbrace{(\beta_0 + \beta_1 x_i)}^{\text{prediction}})^2}}_{\text{root mean squared error}}$$

- **For HW7 root mean squared error: 3.3907 years.**

	Estimate	Pr(>|t|)
Intercept	62.3701	0.0000
health exp % GDP	0.1993	0.0567
log(GNI PC)	1.4278	0.0001
infant mortality	-0.2288	0.0000

Linear regression: R^2

- We could measure model fit by comparing:
  - 'Typical' model prediction error;
  - Error associated with just predicting the mean of the dependent variable;
- One way to do this would be:

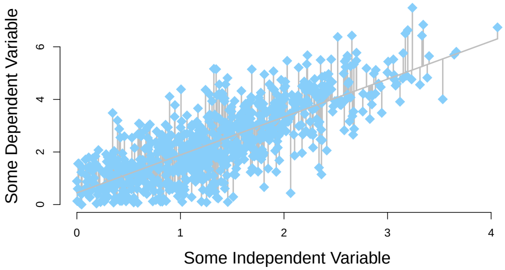
$$R^2 = 1 - \frac{\text{sum of squared residuals}}{\text{error from predicting mean}}$$

Linear regression: R^2

- We could measure model fit by comparing:
  - 'Typical' model prediction error;
  - Error associated with just predicting the mean of the dependent variable;

- One way to do this would be:

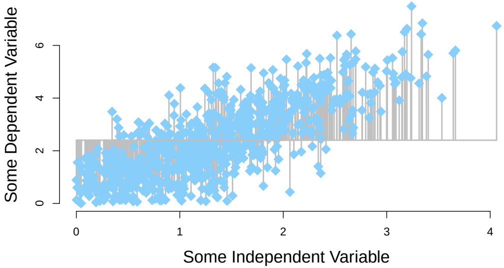
$$R^2 = 1 - \frac{\textbf{sum of squared residuals}}{\text{error from predicting mean}}$$



Linear regression: R^2

- We could measure model fit by comparing:
  - 'Typical' model prediction error;
  - Error associated with just predicting the mean of the dependent variable;
- One way to do this would be:

$$R^2 = 1 - \frac{\text{sum of squared residuals}}{\textbf{error from predicting mean}}$$



Linear regression: R^2

- We could measure model fit by comparing:
  - 'Typical' model prediction error;
  - Error associated with just predicting the mean of the dependent variable;
- One way to do this would be:

$$R^2 = 1 - \frac{\text{sum of squared residuals}}{\text{error from predicting mean}}$$

	Estimate	Pr(>|t|)
Intercept	62.3701	0.0000
health exp % GDP	0.1993	0.0567
log(GNI PC)	1.4278	0.0001
infant mortality	-0.2288	0.0000

Multiple R-squared: 0.8854

Linear regression: R^2

- We could measure model fit by comparing:
  - 'Typical' model prediction error;
  - Error associated with just predicting the mean of the dependent variable;
- One way to do this would be:

$$R^2 = 1 - \frac{\text{sum of squared residuals}}{\text{error from predicting mean}}$$

- Beware! R^2 will always increase as you add more independent variables.

	Estimate	Pr(>|t|)
Intercept	62.3701	0.0000
health exp % GDP	0.1993	0.0567
log(GNI PC)	1.4278	0.0001
infant mortality	-0.2288	0.0000

Multiple R-squared: 0.8854

Linear Regression: *F*-test

- We could measure model fit by doing hypothesis testing;

- Given a model with some independent variables ask does the model fit the data well?
  - $\mathbf{H_0}$:
  - $\mathbf{H_A}$:
  - **Test stat**:
  - **Rejection criterion**:

	Estimate	Pr(>|t|)
Intercept	62.3701	0.0000
health exp % GDP	0.1993	0.0567
log(GNI PC)	1.4278	0.0001
infant mortality	-0.2288	0.0000

Multiple R-squared: 0.8854

Linear Regression: F-test

- We could measure model fit by doing hypothesis testing;

- Given a model with some independent variables ask does the model fit the data well?
  - $\mathbf{H_0}$: The fit of the 'intercept-only' model is no different than the fit of the 'intercept-plus-variables' model;
  - $\mathbf{H_A}$: The fit of the 'intercept-only' model is significantly worse;
  - **Test stat**:

  - **Rejection criterion**:

	Estimate	Pr(>|t|)
Intercept	62.3701	0.0000
health exp % GDP	0.1993	0.0567
log(GNI PC)	1.4278	0.0001
infant mortality	-0.2288	0.0000

Multiple R-squared: 0.8854

Linear Regression: F-test

- We could measure model fit by doing hypothesis testing;

- Given a model with some independent variables ask does the model fit the data well?
  - $\mathbf{H_0}$: The fit of the 'intercept-only' model is no different than the fit of the 'intercept-plus-variables' model;
  - $\mathbf{H_A}$: The fit of the 'intercept-only' model is significantly worse;
  - **Test stat**: F-statistic – measures the variation in the dependent variable explained by the model;
  - **Rejection criterion**: p-value < 0.05.

	Estimate	Pr(>|t|)
Intercept	62.3701	0.0000
health exp % GDP	0.1993	0.0567
log(GNI PC)	1.4278	0.0001
infant mortality	-0.2288	0.0000

Multiple R-squared: 0.8854

Linear Regression: F-test

- We could measure model fit by doing hypothesis testing;

- Given a model with some independent variables ask does the model fit the data well?
  - $\mathbf{H_0}$: The fit of the 'intercept-only' model is no different than the fit of the 'intercept-plus-variables' model;
  - $\mathbf{H_A}$: The fit of the 'intercept-only' model is significantly worse;
  - **Test stat**: F-statistic – measures the variation in the dependent variable explained by the model;
  - **Rejection criterion**: p-value < 0.05.

	Estimate	Pr(>|t|)
Intercept	62.3701	0.0000
health exp % GDP	0.1993	0.0567
log(GNI PC)	1.4278	0.0001
infant mortality	-0.2288	0.0000

Multiple R-squared: 0.8854
F-statistic: p-value: $< 2.2\text{e-}16$

Logistic Regression: predictions?

- A logit actually outputs a **predicted probability** when it makes a prediction;
- To turn a predicted probability into a **predicted value**:
  - Choose a threshold t;
  - If predicted probability $> t$ then predict a 1;
  - If predicted probability $< t$ then predict a 0;
- Once we have predicted values we can start to assess model fit:
  - **True Positive**: model predicts 1, actual observation is 1;
  - **True Negative**: model predicts 0, actual observation is 0;
  - **False Positive**: model predicts 1, actual observation is 0;
  - **False Negative**: model predicts 0, actual observation is 1.

Logistic Regression: predictions?

Note: everything to follow will reference the Titanic example from the logistic regression lecture.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	5.2973	0.5574	9.50	0.0000
Pclass	-1.1777	0.1461	-8.06	0.0000
Age	-0.0435	0.0077	-5.63	0.0000
Male	-2.7573	0.2004	-13.76	0.0000
Siblings/Spouses	-0.4018	0.1107	-3.63	0.0003
Parents/Children	-0.1065	0.1186	-0.90	0.3691
Fare	0.0028	0.0024	1.17	0.2437

Logistic Regression: predictions? Choose a threshold t

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09		
Mrs. John Bradley Cumings	1	0.91		
Miss. Laina Heikkinen	1	0.66		
Mrs. Jacques Heath Futrelle	1	0.91		
Mr. William Henry Allen	0	0.08		
Mr. Timothy J McCarthy	0	0.30		
Master. Gosta Leonard Palsson	0	0.09		
Mrs. Oscar W Johnson	1	0.60		
Mrs. Nicholas Nasser	1	0.88		
Miss. Marguerite Rut Sandstrom	1	0.76		
Miss. Elizabeth Bonnell	1	0.84		
Miss. Hulda Vestrom	0	0.76		
⋮	⋮	⋮		

Logistic Regression: predictions? Choose a threshold $t = 0.0$

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09	1	
Mrs. John Bradley Cumings	1	0.91	1	
Miss. Laina Heikkinen	1	0.66	1	
Mrs. Jacques Heath Futrelle	1	0.91	1	
Mr. William Henry Allen	0	0.08	1	
Mr. Timothy J McCarthy	0	0.30	1	
Master. Gosta Leonard Palsson	0	0.09	1	
Mrs. Oscar W Johnson	1	0.60	1	
Mrs. Nicholas Nasser	1	0.88	1	
Miss. Marguerite Rut Sandstrom	1	0.76	1	
Miss. Elizabeth Bonnell	1	0.84	1	
Miss. Hulda Vestrom	0	0.76	1	
⋮	⋮	⋮	⋮	

Logistic Regression: predictions? Choose a threshold $t = 0.0$

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09	1	FP
Mrs. John Bradley Cumings	1	0.91	1	TP
Miss. Laina Heikkinen	1	0.66	1	TP
Mrs. Jacques Heath Futrelle	1	0.91	1	TP
Mr. William Henry Allen	0	0.08	1	FP
Mr. Timothy J McCarthy	0	0.30	1	FP
Master. Gosta Leonard Palsson	0	0.09	1	FP
Mrs. Oscar W Johnson	1	0.60	1	TP
Mrs. Nicholas Nasser	1	0.88	1	TP
Miss. Marguerite Rut Sandstrom	1	0.76	1	TP
Miss. Elizabeth Bonnell	1	0.84	1	TP
Miss. Hulda Vestrom	0	0.76	1	FP
⋮	⋮	⋮	⋮	⋮

Logistic Regression: predictions? Choose a threshold $t = 0.1$

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09	0	TN
Mrs. John Bradley Cumings	1	0.91	1	TP
Miss. Laina Heikkinen	1	0.66	1	TP
Mrs. Jacques Heath Futrelle	1	0.91	1	TP
Mr. William Henry Allen	0	0.08	0	TN
Mr. Timothy J McCarthy	0	0.30	1	FP
Master. Gosta Leonard Palsson	0	0.09	0	TN
Mrs. Oscar W Johnson	1	0.60	1	TP
Mrs. Nicholas Nasser	1	0.88	1	TP
Miss. Marguerite Rut Sandstrom	1	0.76	1	TP
Miss. Elizabeth Bonnell	1	0.84	1	TP
Miss. Hulda Vestrom	0	0.76	1	FP
⋮	⋮	⋮	⋮	⋮

Logistic Regression: predictions? Choose a threshold $t = 0.5$

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09	0	TN
Mrs. John Bradley Cumings	1	0.91	1	TP
Miss. Laina Heikkinen	1	0.66	1	TP
Mrs. Jacques Heath Futrelle	1	0.91	1	TP
Mr. William Henry Allen	0	0.08	0	TN
Mr. Timothy J McCarthy	0	0.30	0	TN
Master. Gosta Leonard Palsson	0	0.09	0	TN
Mrs. Oscar W Johnson	1	0.60	1	TP
Mrs. Nicholas Nasser	1	0.88	1	TP
Miss. Marguerite Rut Sandstrom	1	0.76	1	TP
Miss. Elizabeth Bonnell	1	0.84	1	TP
Miss. Hulda Vestrom	0	0.76	1	FP
$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Logistic Regression: predictions? Choose a threshold $t = 0.75$

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09	0	TN
Mrs. John Bradley Cumings	1	0.91	1	TP
Miss. Laina Heikkinen	1	0.66	0	FN
Mrs. Jacques Heath Futrelle	1	0.91	1	TP
Mr. William Henry Allen	0	0.08	0	TN
Mr. Timothy J McCarthy	0	0.30	0	TN
Master. Gosta Leonard Palsson	0	0.09	0	TN
Mrs. Oscar W Johnson	1	0.60	0	FN
Mrs. Nicholas Nasser	1	0.88	1	TP
Miss. Marguerite Rut Sandstrom	1	0.76	1	TP
Miss. Elizabeth Bonnell	1	0.84	1	TP
Miss. Hulda Vestrom	0	0.76	1	FP
⋮	⋮	⋮	⋮	⋮

Logistic Regression: predictions? Choose a threshold $t = 0.9$

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09	0	TN
Mrs. John Bradley Cumings	1	0.91	1	TP
Miss. Laina Heikkinen	1	0.66	0	FN
Mrs. Jacques Heath Futrelle	1	0.91	1	TP
Mr. William Henry Allen	0	0.08	0	TN
Mr. Timothy J McCarthy	0	0.30	0	TN
Master. Gosta Leonard Palsson	0	0.09	0	TN
Mrs. Oscar W Johnson	1	0.60	0	FN
Mrs. Nicholas Nasser	1	0.88	0	FN
Miss. Marguerite Rut Sandstrom	1	0.76	0	FN
Miss. Elizabeth Bonnell	1	0.84	0	FN
Miss. Hulda Vestrom	0	0.76	0	TN
⋮	⋮	⋮	⋮	⋮

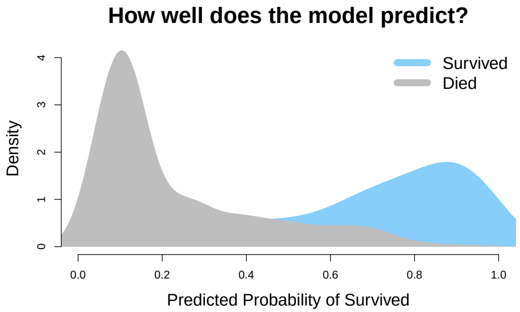
Logistic Regression: predictions? Choose a threshold $t = 1$

Name	Survived	Predicted Probability	Predicted Value	Type
Mr. Owen Harris Braund	0	0.09	0	TN
Mrs. John Bradley Cumings	1	0.91	0	FN
Miss. Laina Heikkinen	1	0.66	0	FN
Mrs. Jacques Heath Futrelle	1	0.91	0	FN
Mr. William Henry Allen	0	0.08	0	TN
Mr. Timothy J McCarthy	0	0.30	0	TN
Master. Gosta Leonard Palsson	0	0.09	0	TN
Mrs. Oscar W Johnson	1	0.60	0	FN
Mrs. Nicholas Nasser	1	0.88	0	FN
Miss. Marguerite Rut Sandstrom	1	0.76	0	FN
Miss. Elizabeth Bonnell	1	0.84	0	FN
Miss. Hulda Vestrom	0	0.76	0	TN
⋮	⋮	⋮	⋮	⋮

Logistic Regression: confusion matrix

- We could measure model fit by aggregating TP/TN/FP/FN across the entire data set;
- Add each of these up across the data:

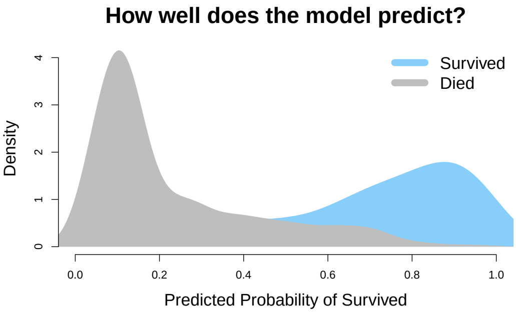
		Prediction	
		Survived	Died
Observation	Survived	#TP	#FN
	Died	#FP	#TN



Logistic Regression: confusion matrix

- We could measure model fit by aggregating TP/TN/FP/FN across the entire data set;
- Add each of these up across the data:

		Prediction	
		Survived	Died
Observation	Survived	239	103
	Died	73	472

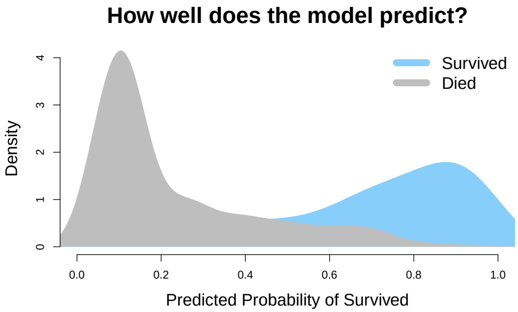


Logistic Regression: accuracy

- We could measure model fit by summarizing the confusion matrix up as a single number;

- One way to do this is by using **accuracy**:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

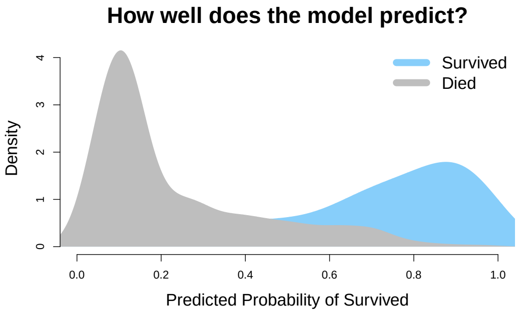


Logistic Regression: accuracy

- We could measure model fit by summarizing the confusion matrix up as a single number;

- One way to do this is by using **accuracy**:

$$\frac{239 + 472}{239 + 472 + 73 + 103} \approx 0.8016;$$

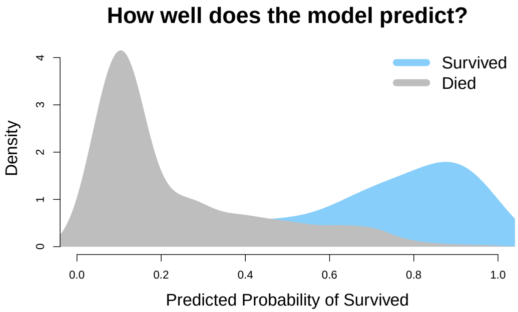


Logistic Regression: accuracy

- We could measure model fit by summarizing the confusion matrix up as a single number;
- One way to do this is by using **accuracy**:

$$\frac{239 + 472}{239 + 472 + 73 + 103} \approx 0.8016;$$

- Beware! Accuracy is only useful when each value of the dependent variable is equally important.



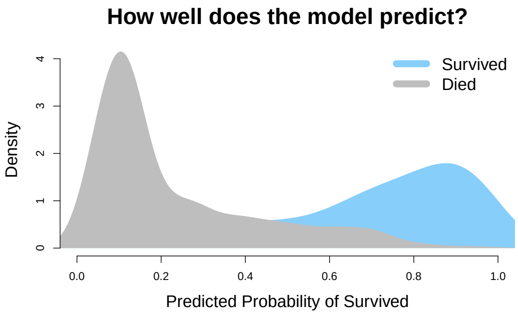
Logistic Regression: recall/precision

- We could measure model fit by trying to gauge its ability to identify the survivors and only the survivors;
- **Recall**: of all the survivors how many did the model identify?

$$R = \frac{TP}{TP + FN}$$

- **Precision**: of all the passengers predicted to survive how many actually did?

$$P = \frac{TP}{TP + FP}.$$



Logistic Regression: recall/precision

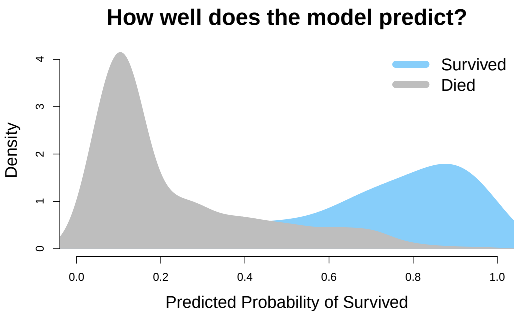
- We could measure model fit by trying to gauge its ability to identify the survivors and only the survivors;

- **Recall**: of all the survivors how many did the model identify?

$$R = \frac{239}{239 + 103} \approx 0.70$$

- **Precision**: of all the passengers predicted to survive how many actually did?

$$P = \frac{239}{239 + 73} \approx 0.77.$$



Why should we care?

Better model fit = better prediction = a more useful, impactful model.