Logistic Regression and the ROC curve

Schwartz

September 7, 2016

Odd, even at best

In the 2015/16 season, Leicester City was given longest odds ever seen to win the English Premier League at 5000 to 1. In fact, these were the longest odds ever recorded for any professional sporting league. To put this in perspective, the current odds out of Vegas for the most unlikely team to win the 2016/2017 NFL season – the woeful Cleveland Browns – are 200 to 1.

Since the clubs inception in 1890, Leicester City had only managed to appear in the Premier league 10 seasons. They had only been promoted for the 9th time the previous season and only escaped relegation in their final match that season. In the "modern era" of Premier league only five teams – Arsenal, Chelsea, Liverpool, Man. City, and Man. U. – have held the trophy.

Only a few stout souls put money down on Leicester City in 2015. And when Leicester City (literally against all odds) won the premiership later that season, those stout souls got paid.

For the $3,000 bet on Leicester City, $15,000,000 was paid out.

Odds

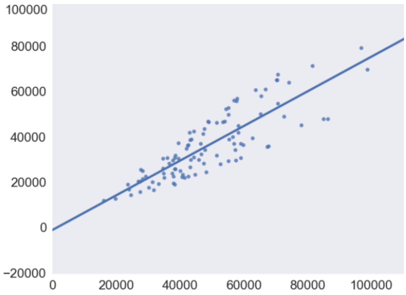
$$\mathrm{Odds} = \frac{p}{1-p} \implies \mathrm{p} = \frac{Odds}{1+Odds} = \frac{1}{1+Odds^{-1}}$$

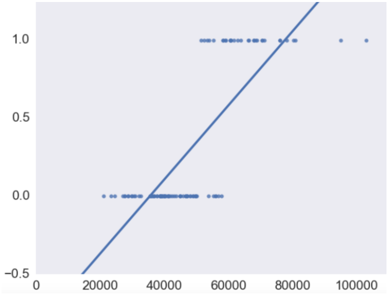
$$1-p = \frac{1}{1+Odds}$$

Objectives

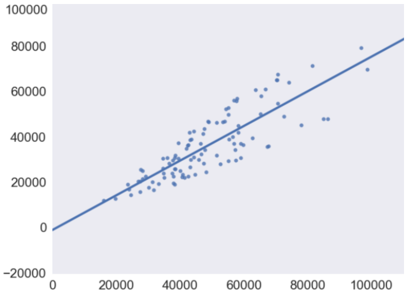
1. Know why logistic regression is a thing
2. Know how to
    - execute a logistic regression, and
    - explain what it all means
3. know ROC curves and such

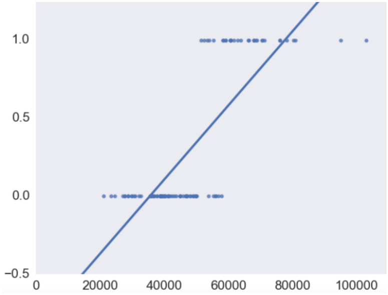
Linear Regression





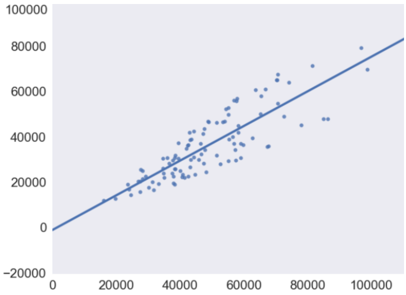
Linear Regression

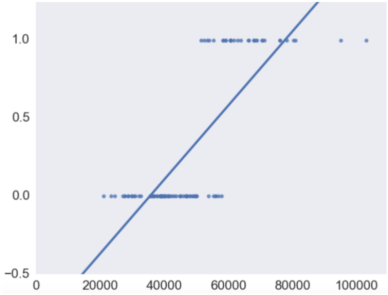




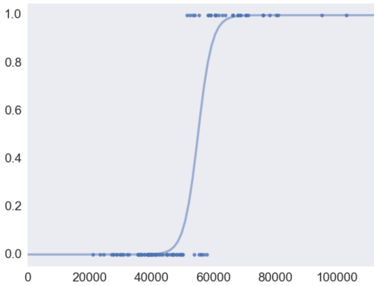
Is this satisfactory?

Linear Regression





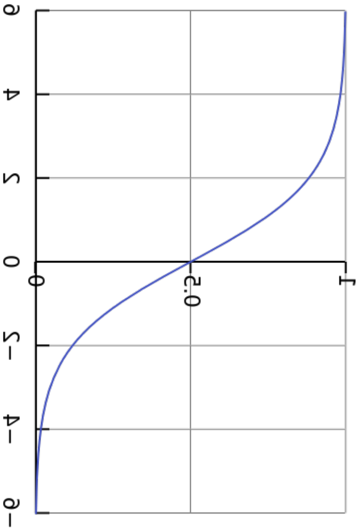
Is this satisfactory?



Link functions

- The "logit"

$$g(p) = \log\left(\frac{p}{1-p}\right)$$



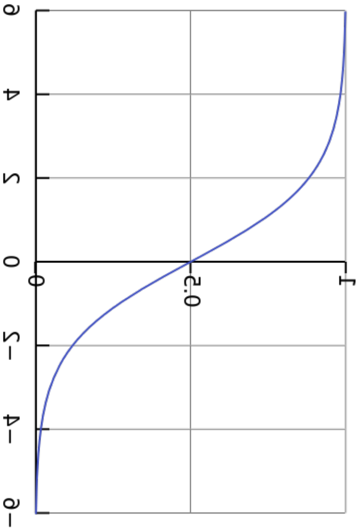
Link functions

- The "logit"

$$g(p) = \log\left(\frac{p}{1-p}\right)$$

- maps

$$p \in [0,1] \mapsto Z \in \mathbb{R}$$



Link functions

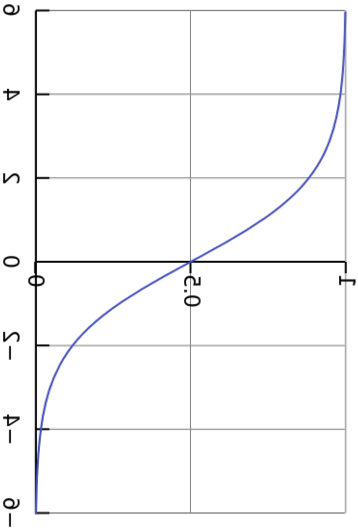
- The "logit"

$$g(p) = \log\left(\frac{p}{1-p}\right)$$

- maps

$$p \in [0,1] \mapsto Z \in \mathbb{R}$$

- Are we at odds about odds?

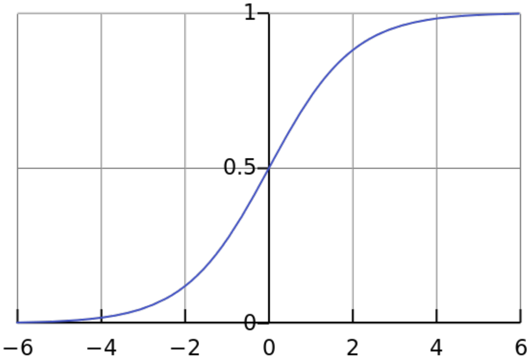


Logistic "regression"

- For a binary outcome Y, if we define

$$\mathrm{E}[Y] = \Pr(Y = 1) = g^{-1}(Z) = \frac{\exp(Z)}{1 + \exp(Z)} = \frac{1}{1 + \exp(-Z)}$$

and let $Z = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m \in \mathbb{R}$



Standard logistic (sinusoid) function

Logarithmic Scale

- So
$$\Pr(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_m)}}$$

Logarithmic Scale

- So

$$\Pr(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_m)}}$$

- And we can quickly see then that

$$\frac{Pr(Y = 1|x)}{Pr(Y = 0|x)} \quad =$$

Logarithmic Scale

- So
$$\Pr(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_m)}}$$
- And we can quickly see then that
$$\frac{Pr(Y = 1|x)}{Pr(Y = 0|x)} \quad = \quad \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_m x_m)$$

Logarithmic Scale

- So
$$\Pr(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_m)}}$$
- And we can quickly see then that
$$\frac{Pr(Y = 1|x)}{Pr(Y = 0|x)} = \exp(\beta_0) \exp(\beta_1 x_1) \cdots \exp(\beta_m x_m)$$
- So for a 1-unit increase in x_j

Logarithmic Scale

- So
$$\Pr(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_m)}}$$
- And we can quickly see then that
$$\frac{Pr(Y = 1|x)}{Pr(Y = 0|x)} = \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_m x_m)$$
- So for a 1-unit increase in x_j

  there is a $\exp(\beta_j)$ multiplicative increase in *the odds*

Logarithmic Scale

- So
$$\Pr(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_m)}}$$
- And we can quickly see then that
$$\frac{Pr(Y = 1|x)}{Pr(Y = 0|x)} = \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_m x_m)$$
- So for a 1-unit increase in x_j

  there is a $\exp(\beta_j)$ multiplicative increase in *the odds*
- I.e., *the odds* are linear in x on a multiplicative, i.e., odds increase with x on a *logorithmic* scale with base $\exp(\beta_j)$

Logarithmic Scale

- So
$$\Pr(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_m)}}$$
- And we can quickly see then that
$$\frac{Pr(Y = 1|x)}{Pr(Y = 0|x)} \quad = \quad \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_m x_m)$$
- So for a 1-unit increase in x_j

  there is a $\exp(\beta_j)$ multiplicative increase in *the odds*
- I.e., *the odds* are linear in x on a multiplicative, i.e., odds increase with x on a *logorithmic* scale with base $\exp(\beta_j)$
- The *log* odds $log\left(\frac{Pr(Y=1|x)}{Pr(Y=0|x)}\right)$ are on a linear scale

  $(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_m)$

The Odds Ratio (OR)

- Equivalently, $\exp(\beta_j)$ is the *odds ratio (OR)* between 1-unit differences in x_j (e.g., 0 versus 1) when other x's are constant

$$\exp(\beta_j) = \frac{Pr(Y = 1|x_j + 1, x_{-j})/Pr(Y = 0|x_j + 1, x_{-j})}{Pr(Y = 1|x)/Pr(Y = 0|x)}$$

since

$$\begin{aligned}
&\frac{Pr(Y = 1|x_j + 1, x_{-j})}{Pr(Y = 0|x_j + 1, x_{-j})} \\
&= \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_j(x_j + 1))\cdots\exp(\beta_m x_m) \\
&= \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_j x_j) exp(\beta_j)\cdots\exp(\beta_m x_m)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{Pr(Y = 1|x)}{Pr(Y = 0|x)} \\
&= \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_j x_j)\cdots\exp(\beta_m x_m)
\end{aligned}$$

The Odds Ratio (OR)

- Equivalently, $\exp(\beta_j)$ is the *odds ratio (OR)* between 1-unit differences in x_j (e.g., 0 versus 1) when other x's are constant

$$\exp(\beta_j) = \frac{Pr(Y=1|x_j+1, x_{-j})/Pr(Y=0|x_j+1, x_{-j})}{Pr(Y=1|x)/Pr(Y=0|x)}$$

  since

$$\frac{Pr(Y=1|x_j+1, x_{-j})}{Pr(Y=0|x_j+1, x_{-j})}$$
$$= \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_j(x_j+1))\cdots\exp(\beta_m x_m)$$
$$= \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_j x_j) exp(\beta_j)\cdots\exp(\beta_m x_m)$$

  and

$$\frac{Pr(Y=1|x)}{Pr(Y=0|x)}$$
$$= \exp(\beta_0)\exp(\beta_1 x_1)\cdots\exp(\beta_j x_j)\cdots\exp(\beta_m x_m)$$

- So β_j is the change in $log(OR)$ for one unit changes in x_j...

Logistic Regression *Likelihood* and *Deviance*

- Likelihood

$$\prod \left(\frac{1}{1+e^{-\mathbf{x}_i^T\beta}}\right)^{Y_i}\left(\frac{1}{1+e^{\mathbf{x}_i^T\beta}}\right)^{1-Y_i}$$

Logistic Regression *Likelihood* and *Deviance*

- Likelihood

$$\prod \left(\frac{1}{1+e^{-\mathbf{x}_i^T\beta}}\right)^{Y_i} \left(\frac{1}{1+e^{\mathbf{x}_i^T\beta}}\right)^{1-Y_i}$$

- Deviance

$$\begin{aligned} D_M &= -2\left(\log f(\mathbf{Y}|\hat{\theta}^M) - \log f(\mathbf{Y}|\mathbf{Y})\right) \\ &\sim \chi^2_{n-p-1} \end{aligned}$$

$$\begin{aligned} n &= \text{sample size} \\ k &= \text{number of parameters in model } M \\ f(\mathbf{Y}|\mathbf{Y}) &= \text{saturated model } (\mathbf{Y} \text{ perfectly predicted}) \end{aligned}$$

More Deviance

- In logistic regression
  $D_M = -2\left(\log f(\mathbf{Y}|\hat{\theta}^M) - \log f(\mathbf{Y}|\mathbf{Y})\right) =$

More Deviance

- In logistic regression $D_M = -2\left(\log f(\mathbf{Y}|\hat{\theta}^M) - \log f(\mathbf{Y}|\mathbf{Y})\right) =$

$$-2\sum Y_i log(p_i) + (1 - Y_i) log(1 - p_i) - Y_i log(Y_i) - (1 - Y_i) log(1 - Y_i)$$

More Deviance

- In logistic regression
  $D_M = -2\left(\log f(\mathbf{Y}|\hat{\theta}^M) - \log f(\mathbf{Y}|\mathbf{Y})\right) =$

$$-2\sum Y_i log(p_i) + (1-Y_i) log(1-p_i) - Y_i log(Y_i) - (1-Y_i) log(1-Y_i)$$

[show this]

More Deviance

- In logistic regression
  $D_M = -2\left(\log f(\mathbf{Y}|\hat{\theta}^M) - \log f(\mathbf{Y}|\mathbf{Y})\right) =$

$$-2\sum Y_i log(p_i) + (1 - Y_i) log(1 - p_i) - Y_i log(Y_i) - (1 - Y_i) log(1 - Y_i)$$

  [show this]

- In linear regression

$$\begin{aligned} D_M &= \frac{RSS}{\sigma^2} \qquad \text{[show this]} \\ &= \frac{\sum (Y_i - \hat{Y})^2}{\sigma^2} = (n - p - 1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-p-1} \end{aligned}$$

More Deviance

- In logistic regression
  $D_M = -2\left(\log f(\mathbf{Y}|\hat{\theta}^M) - \log f(\mathbf{Y}|\mathbf{Y})\right) =$

$$-2\sum Y_i log(p_i) + (1 - Y_i) log(1 - p_i) - Y_i log(Y_i) - (1 - Y_i) log(1 - Y_i)$$

  [show this]

- In linear regression

$$D_M = \frac{RSS}{\sigma^2} \qquad \text{[show this]}$$
$$= \frac{\sum (Y_i - \hat{Y})^2}{\sigma^2} = (n - p - 1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

[what are *residuals*?]

More Deviance

- In logistic regression
  $D_M = -2\left(\log f(\mathbf{Y}|\hat{\theta}^M) - \log f(\mathbf{Y}|\mathbf{Y})\right) =$

$$-2 \sum Y_i log(p_i) + (1 - Y_i) log(1 - p_i) - Y_i log(Y_i) - (1 - Y_i) log(1 - Y_i)$$

[show this]

- In linear regression

$$D_M = \frac{RSS}{\sigma^2} \qquad \text{[show this]}$$

$$= \frac{\sum(Y_i - \hat{Y})^2}{\sigma^2} = (n-p-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-p-1}$$

[what are *residuals*?] [what are *residuals* in logistic regression?]

Fitting Logistic Regression

- MLE

$$\max_{\beta} \prod \left(\frac{1}{1+e^{-\beta x}}\right)^{Y_i} \left(\frac{1}{1+e^{\beta x}}\right)^{1-Y_i}$$

$$\Longleftrightarrow$$

$$\min_{\beta} D_{\beta} = \max_{\beta} \left(\log f(\mathbf{Y}|\beta) - \log f(\mathbf{Y}|\mathbf{Y})\right)$$

Fitting Logistic Regression

- MLE

$$\max_{\beta} \prod \left(\frac{1}{1 + e^{-\beta x}} \right)^{Y_i} \left(\frac{1}{1 + e^{\beta x}} \right)^{1-Y_i}$$

$$\Longleftrightarrow$$

$$\min_{\beta} D_{\beta} = \max_{\beta} (\log f(\mathbf{Y}|\beta) - \log f(\mathbf{Y}|\mathbf{Y}))$$

- Closed form solution not available (like with linear regression)
  - Optimization done via Newton-Rhapson or Gradient Decent

Fitting Logistic Regression

- MLE

$$\max_{\beta} \prod \left(\frac{1}{1+e^{-\beta x}} \right)^{Y_i} \left(\frac{1}{1+e^{\beta x}} \right)^{1-Y_i}$$

$$\iff$$

$$\min_{\beta} D_{\beta} = \max_{\beta} \left(\log f(\mathbf{Y}|\beta) - \log f(\mathbf{Y}|\mathbf{Y}) \right)$$

- Closed form solution not available (like with linear regression)
  - Optimization done via Newton-Rhapson or Gradient Decent
- Convergence difficulties will be encountered if
  - too many features ($n/p < 10$) or data is sparse/imbalanced

Fitting Logistic Regression

- MLE

$$\max_{\beta} \prod \left(\frac{1}{1+e^{-\beta x}}\right)^{Y_i} \left(\frac{1}{1+e^{\beta x}}\right)^{1-Y_i}$$
$$\iff$$
$$\min_{\beta} D_{\beta} = \max_{\beta} (\log f(\mathbf{Y}|\beta) - \log f(\mathbf{Y}|\mathbf{Y}))$$

- Closed form solution not available (like with linear regression)
  - Optimization done via Newton-Rhapson or Gradient Decent
- Convergence difficulties will be encountered if
  - too many features ($n/p < 10$) or data is sparse/imbalanced
- Coefficient standard errors will be compromised when
  - predicted probabilities approach 1 or 0 (separated classes)
  - There is covariate multicollinearity (like linear regression)

Fitting Logistic Regression

- MLE

$$\max_{\beta} \prod \left(\frac{1}{1+e^{-\beta x}}\right)^{Y_i} \left(\frac{1}{1+e^{\beta x}}\right)^{1-Y_i}$$

$$\Longleftrightarrow$$

$$\min_{\beta} D_{\beta} = \max_{\beta} \left(\log f(\mathbf{Y}|\beta) - \log f(\mathbf{Y}|\mathbf{Y})\right)$$

- Closed form solution not available (like with linear regression)
  - Optimization done via Newton-Rhapson or Gradient Decent
- Convergence difficulties will be encountered if
  - too many features ($n/p < 10$) or data is sparse/imbalanced
- Coefficient standard errors will be compromised when
  - predicted probabilities approach 1 or 0 (separated classes)
  - There is covariate multicollinearity (like linear regression)
- What if, for some λ, we choose β to minimize

$$-\prod \left(\frac{1}{1+e^{-\mathbf{x}_i^T\beta}}\right)^{Y_i} \left(\frac{1}{1+e^{\mathbf{x}_i^T\beta}}\right) + \lambda||\beta||^2?$$

Fitting Logistic Regression

- MLE

$$\max_\beta \prod \left(\frac{1}{1+e^{-\beta x}}\right)^{Y_i} \left(\frac{1}{1+e^{\beta x}}\right)^{1-Y_i}$$

$$\Longleftrightarrow$$

$$\min_\beta D_\beta = \max_\beta \left(\log f(\mathbf{Y}|\beta) - \log f(\mathbf{Y}|\mathbf{Y})\right)$$

- Closed form solution not available (like with linear regression)
  - Optimization done via Newton-Rhapson or Gradient Decent
- Convergence difficulties will be encountered if
  - too many features ($n/p < 10$) or data is sparse/imbalanced
- Coefficient standard errors will be compromised when
  - predicted probabilities approach 1 or 0 (separated classes)
  - There is covariate multicollinearity (like linear regression)
- What if, for some λ, we choose β to minimize

$$-\prod \left(\frac{1}{1+e^{-\mathbf{x}_i^T\beta}}\right)^{Y_i} \left(\frac{1}{1+e^{\mathbf{x}_i^T\beta}}\right) + \lambda|\beta|_1?$$

Pseudo R^2

- McFadden's *pseudo* $R^2 = 1 - D_M/D_0$

Pseudo R^2

- McFadden's *pseudo* $R^2 = 1 - D_M/D_0$
  Proportion of deviance explained

Pseudo R^2

- McFadden's *pseudo* $R^2 = 1 - D_M/D_0$
  Proportion of deviance explained

- Compare to linear model $R^2 = 1 - RSS/TSS$

Pseudo R^2

- McFadden's *pseudo* $R^2 = 1 - D_M/D_0$
  Proportion of deviance explained

- Compare to linear model $R^2 = 1 - RSS/TSS$
  Proportion of variance explained

Pseudo R^2

- McFadden's *pseudo* $R^2 = 1 - D_M/D_0$
  Proportion of deviance explained
- Compare to linear model $R^2 = 1 - RSS/TSS$
  Proportion of variance explained
- http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm

Model Comparison

Remember, $D_M = -2(logf(\mathbf{Y}|\hat{\theta}_M) - logf(\mathbf{Y}|\mathbf{Y}))$

Model Comparison

Remember, $D_M = -2(logf(\mathbf{Y}|\hat{\theta}_M) - logf(\mathbf{Y}|\mathbf{Y}))$

- Model comparison can be done using

$$D_R - D_F = -2\left(\log f(Y|\hat{\theta}^R) - \log f(Y|\hat{\theta}^F)\right) \overset{approx.}{\sim} \chi^2_k$$

where model R is nested in model F with k fewer parameters

Model Comparison

Remember, $D_M = -2(logf(\mathbf{Y}|\hat{\theta}_M) - logf(\mathbf{Y}|\mathbf{Y}))$

- Model comparison can be done using

$$D_R - D_F = -2\left(\log f(Y|\hat{\theta}^R) - \log f(Y|\hat{\theta}^F)\right) \overset{approx.}{\sim} \chi^2_k$$

where model R is nested in model F with k fewer parameters

- For *non-nested* models, compare

$$AIC: \; -2\log f(Y|\hat{\theta}) + 2k$$
$$BIC: \; -2\log f(Y|\hat{\theta}) + k\log(n)$$

Uses for logistic regression

Uses for logistic regression

- Predict probabilities

Uses for logistic regression

- Predict probabilities
- Identify variables associated with outcomes

Uses for logistic regression

- Predict probabilities
- Identify variables associated with outcomes
- Classify outcomes (based on probabilities)

Uses for logistic regression

- Predict probabilities
- Identify variables associated with outcomes
- Classify outcomes (based on probabilities)

- Balancing comparison groups on *propensity scores* $\Pr(T|x)$ controls bias from group covariate composition differences

Confusion Matrix

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

Confusion Matrix

		Predicted Class	
		Yes	No
Actual Class	Yes	TP	FN
	No	FP	TN

What is Type I and Type II error?

Sensitivity & Specificity

- Sensitivity: % of "true H_a" tests we correctly call $\left(\frac{TP}{TP+FN}\right)$

  "Are we **sensitive** to variations from H_0?"

Sensitivity & Specificity

- Sensitivity: % of "true H_a" tests <u>we correctly call</u> $\left(\frac{TP}{TP+FN}\right)$

  "Are we **sensitive** to variations from H_0?"

  - Also called *True Positive Rate*

Sensitivity & Specificity

- Sensitivity: % of "true H_a" tests <u>we correctly call</u> $\left(\frac{TP}{TP+FN}\right)$

  "Are we **sensitive** to variations from H_0?"

  - Also called *True Positive Rate*

- Specificity: % of "true H_0" tests <u>we correctly call</u> $\left(\frac{TN}{TN+FP}\right)$

  "Are we **specific** about what H_0 means?"

Sensitivity & Specificity

- Sensitivity: % of "true H_a" tests <u>we correctly call</u> $\left(\frac{TP}{TP+FN}\right)$

  "Are we **sensitive** to variations from H_0?"

  - Also called *True Positive Rate*

- Specificity: % of "true H_0" tests <u>we correctly call</u> $\left(\frac{TN}{TN+FP}\right)$

  "Are we **specific** about what H_0 means?"

  - Also called *True Negative Rate*
  - The related *False Positive Rate* is $\left(\frac{FP}{TN+FP}\right)$

Sensitivity & Specificity

- Sensitivity: % of "true H_a" tests <u>we correctly call</u> $\left(\frac{TP}{TP+FN}\right)$

  "Are we **sensitive** to variations from H_0?"

  - Also called *True Positive Rate*

- Specificity: % of "true H_0" tests <u>we correctly call</u> $\left(\frac{TN}{TN+FP}\right)$

  "Are we **specific** about what H_0 means?"

  - Also called *True Negative Rate*
  - The related *False Positive Rate* is $\left(\frac{FP}{TN+FP}\right)$

- the power of a test is $\Pr(\text{Reject } H_0 \mid H_a \text{ True})$
- α-significance level is $\Pr(\text{Reject } H_0 \mid H_0 \text{ True})$

Sensitivity & Specificity

- Sensitivity: % of "true H_a" tests <u>we correctly call</u> $\left(\frac{TP}{TP+FN}\right)$

  "Are we **sensitive** to variations from H_0?"

  - Also called *True Positive Rate*

- Specificity: % of "true H_0" tests <u>we correctly call</u> $\left(\frac{TN}{TN+FP}\right)$

  "Are we **specific** about what H_0 means?"

  - Also called *True Negative Rate*
  - The related *False Positive Rate* is $\left(\frac{FP}{TN+FP}\right)$

- the power of a test is $\Pr(\text{Reject } H_0 \mid H_a \text{ True})$
- α-significance level is $\Pr(\text{Reject } H_0 \mid H_0 \text{ True})$
- **Are** power **and** α **Sensitivity and Specificity?**

Sensitivity & Specificity

- Sensitivity: % of "true H_a" tests <u>we correctly call</u> $\left(\frac{TP}{TP+FN}\right)$

  "Are we **sensitive** to variations from H_0?"

  - Also called *True Positive Rate*

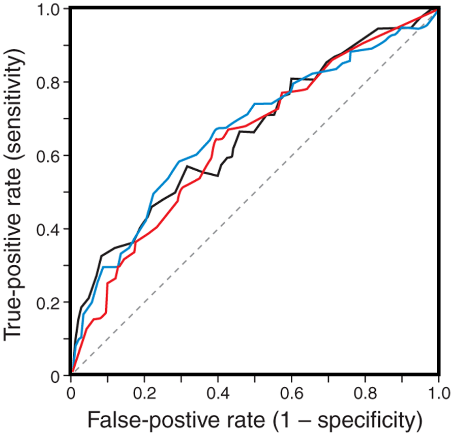
- Specificity: % of "true H_0" tests <u>we correctly call</u> $\left(\frac{TN}{TN+FP}\right)$

  "Are we **specific** about what H_0 means?"

  - Also called *True Negative Rate*
  - The related *False Positive Rate* is $\left(\frac{FP}{TN+FP}\right)$

- the power of a test is $\Pr(\text{Reject } H_0 \mid H_a \text{ True})$
- α-significance level is $\Pr(\text{Reject } H_0 \mid H_0 \text{ True})$
- **Are** power **and** α **Sensitivity and Specificity?**
- **Are** Type I **and** Type II **Sensitivity and Specificity?**

ROC/AUC



https://www.youtube.com/watch?v=JAQC59ArFJw

https://www.youtube.com/watch?v=bhvvxNUbIpo

**Notice how this is dependent upon the "+" and "-" *populations*

Precision, False Discovery Rate, and Accuracy

- Precision: % of **positives** we correctly call $\left(\frac{TP}{TP+FP}\right)$
  - "How **precise** are we when we reject H_0?"
  - Also called *Positive Predicted Value*

Precision, False Discovery Rate, and Accuracy

- Precision: % of **positives** we correctly call $\left(\frac{TP}{TP+FP}\right)$
  - "How **precise** are we when we reject H_0?"
  - Also called *Positive Predicted Value*
- FDR: % of **positives** we **incorrectly** call $\left(\frac{FP}{TP+FP}\right)$
  - False Discovery Rate: "What's our significant tests error rate?"

Precision, False Discovery Rate, and Accuracy

- Precision: % of **positives** we correctly call $\left(\frac{TP}{TP+FP}\right)$
  - "How **precise** are we when we reject H_0?"
  - Also called *Positive Predicted Value*
- FDR: % of **positives** we **incorrectly** call $\left(\frac{FP}{TP+FP}\right)$
  - False Discovery Rate: "What's our significant tests error rate?"
    VERY useful concept of multiple testing contexts
    a.k.a. for a *multiplicity adjustment* (where it's called a q-value)

Precision, False Discovery Rate, and Accuracy

- Precision: % of **positives** <u>we correctly call</u> $\left(\frac{TP}{TP+FP}\right)$
  - "How **precise** are we when we reject H_0?"
  - Also called *Positive Predicted Value*
- FDR: % of **positives** <u>we **incorrectly** call</u> $\left(\frac{FP}{TP+FP}\right)$
  - False Discovery Rate: "What's our significant tests error rate?" *VERY* useful concept of multiple testing contexts a.k.a. for a *multiplicity adjustment* (where it's called a q-value)
- Accuracy: % of tests we <u>we correctly call</u> $\left(\frac{TP+TN}{Total}\right)$
  - "Do we call hypotheses accurately?"

Precision, False Discovery Rate, and Accuracy

- Precision: % of **positives** we correctly call $\left(\frac{TP}{TP+FP}\right)$
  - "How **precise** are we when we reject H_0?"
  - Also called *Positive Predicted Value*
- FDR: % of **positives** we **incorrectly** call $\left(\frac{FP}{TP+FP}\right)$
  - False Discovery Rate: "What's our significant tests error rate?" *VERY* useful concept of multiple testing contexts a.k.a. for a *multiplicity adjustment* (where it's called a q-value)
- Accuracy: % of tests we we correctly call $\left(\frac{TP+TN}{Total}\right)$
  - "Do we call hypotheses accurately?"

		Predicted condition			
	Total population	Predicted Condition positive	Predicted Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	
True condition	condition positive	**True positive**	**False Negative** (Type II error)	True positive rate (TPR), Sensitivity, Recall $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$
	condition negative	**False Positive** (Type I error)	**True negative**	False positive rate (FPR), Fall-out $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	True negative rate (TNR), Specificity (SPC) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$
	Accuracy (ACC) $= \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	Positive predictive value (PPV), Precision $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Test outcome positive}}$	False omission rate (FOR) $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Test outcome negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$
		False discovery rate (FDR) $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Test outcome positive}}$	Negative predictive value (NPV) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Test outcome negative}}$	Negative likelihood ratio (LR–) $= \frac{\text{FNR}}{\text{TNR}}$	

Thanks, Wiki!