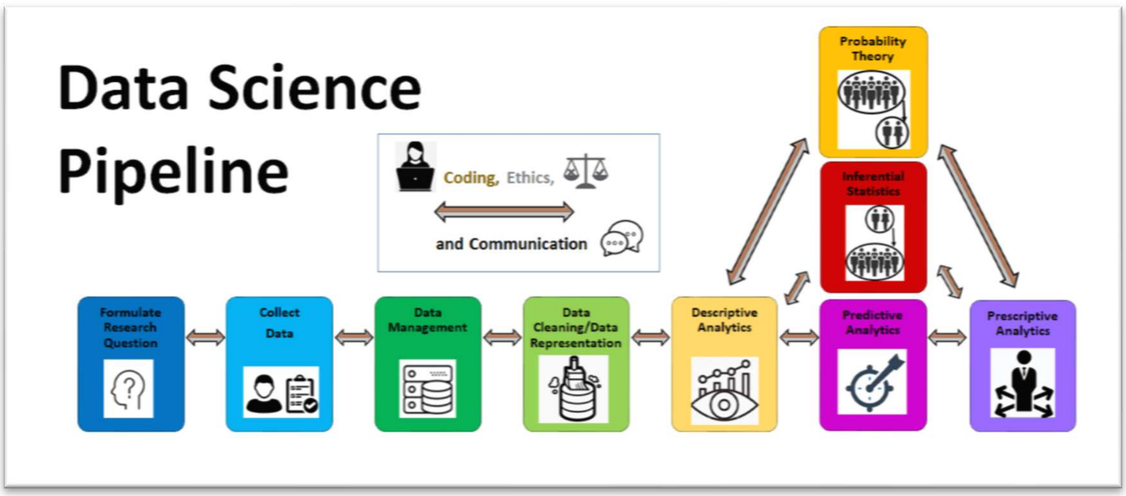
disapprove

approve

<u>Unit 17</u>: Classification Accuracy of Classifier Models

<u>Case Studies</u>:

- We will try to accurately classify whether a person in our Pew Research Survey approves or disapproves of the president's foreign policy given the following explanatory variables.
  - <u>Explanatory(s)</u>:
    - Sex
    - Age
    - Party



<u>Summary of Concepts:</u>

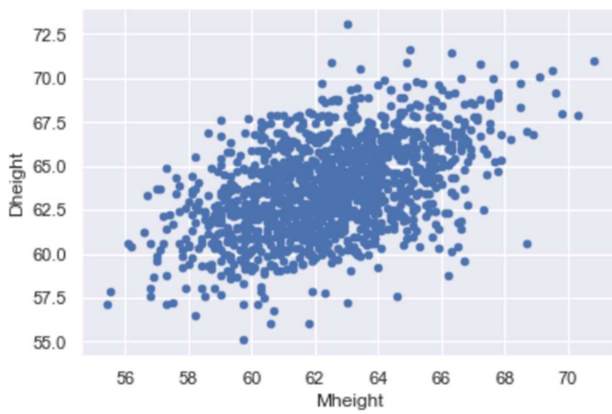
1. **How to predict the response variable value of a given observation using a logistic regression model.**
2. **How can we use a predictive probability to classify a given observation with either a y=1 or y=0?**
3. **Types of misclassifications and correct classifications.**
   - True positive rate
   - False positive rate
   - True negative rate
   - False negative rate
   - Confusion matrix
   - Sensitivity (True positive rate)
   - Specificity (True negative rate)
   - False positive rate
4. **Relationship between sensitivity and specificity**
5. **ROC and AUC** – Which classifier will give us the best combinations of (false positive rate, true positive rate) for all sets of thresholds?

1. HOW TO PREDICT *THE RESPONSE VARIABLE VALUE* OF A GIVEN OBSERVATION USING A LOGISTIC REGRESSION MODEL

Logistic regression model predictions are different from linear regression model predictions in that:

- A **linear** regression model will predict _______________________________ for a given set of explanatory variable values.

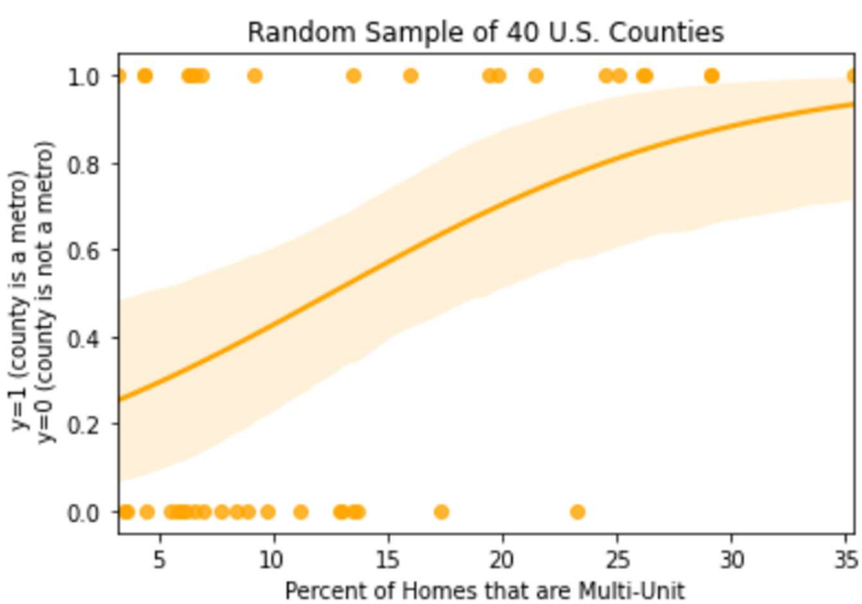
$$\widehat{DHeight} = \hat{\beta}_0 + \hat{\beta}_1(Mheight)$$



- A **logistic** regression model will predict _______________________________ for a given set of explanatory variable values.

$$log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{\beta}_0 + \hat{\beta}_1(multiunit)$$

$$\hat{p} = \frac{e^{\hat{\beta}_0+\hat{\beta}_1(multiunit)}}{e^{\hat{\beta}_0+\hat{\beta}_1(multiunit)}+1}$$



2. HOW CAN WE USE A PREDICTIVE PROBABILITY TO CLASSIFY A GIVEN OBSERVATION WITH EITHER A Y=1 OR Y=0?

A simplistic way to use a predictive probability to classify an observation with a given set of explanatory variable values in a logistic regression model as either y=1 or y=0 to is to use the following rule.

$$y = \begin{Bmatrix} 1, & if\ \hat{p} > 0.5 \\ 0, & if\ \hat{p} \leq 0.5 \end{Bmatrix}$$

Thus, we created a **predictive probability threshold of $\boldsymbol{p_0 = 0.5}$** to classify an observation with either y=1 or y=0.

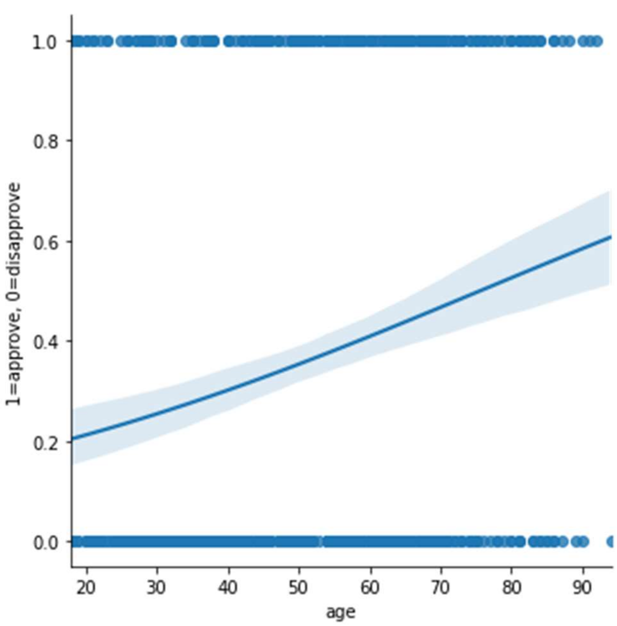
In general, we use a given predictive probability threshold p_0 to

- classify all observations with a predictive probability that is $\hat{p} > p_0$ as y=1 and
- classify all observations with a predictive probability that is $\hat{p} \leq p_0$ as y=0.

Ex: Use the logistic regression model below to classify a 70 year old as either y=1 (ie. approving of the president's foreign policy) or y=0 (ie. disapproving of the president's foreign policy.)

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -1.7872 + 0.0236(age)$$

$$\hat{p} = \frac{e^{-1.7872+0.0236(age)}}{e^{-1.7872+\ .0236(age)} + 1}$$

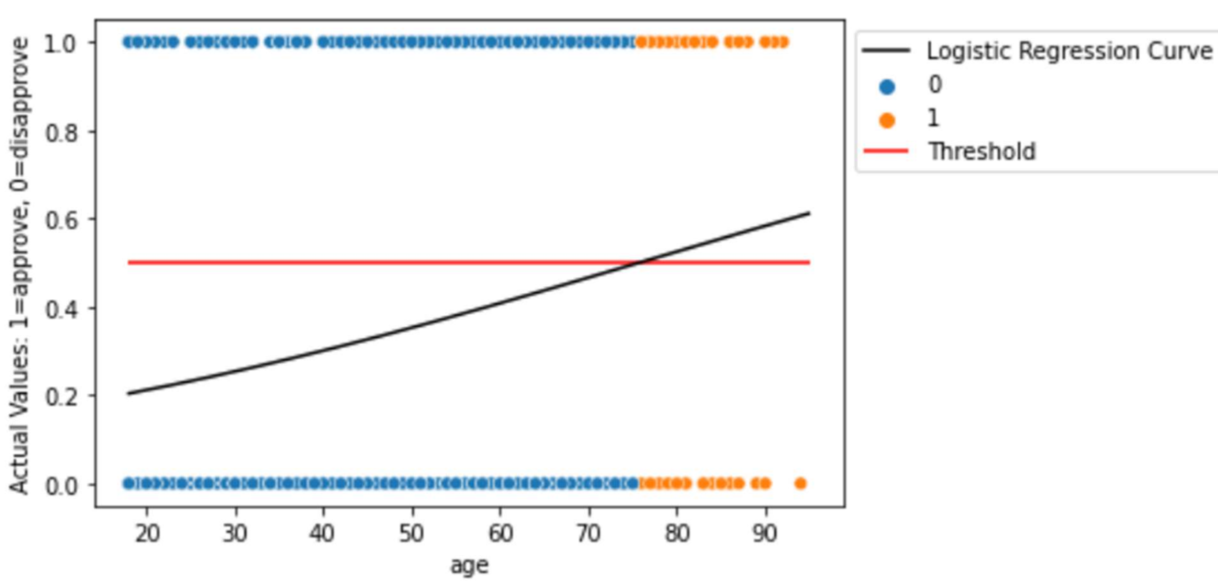


Logit Regression Results

Dep. Variable:	y	No. Observations:	691
Model:	Logit	Df Residuals:	689
Method:	MLE	Df Model:	1
Date:	Thu, 08 Apr 2021	Pseudo R-squ.:	0.03047
Time:	10:50:09	Log-Likelihood:	-437.89
converged:	True	LL-Null:	-451.65
Covariance Type:	nonrobust	LLR p-value:	1.553e-07

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.7872	0.254	-7.023	0.000	-2.286	-1.288
age	0.0236	0.005	5.128	0.000	0.015	0.033

Ex: Using this predictive probability threshold of $p_0 = 0.5$, what is the minimum age that a person needs to be in order to be classified as approving the president's foreign policy in this sample?



Many Different Thresholds

A predictive probability threshold of $p_0 = 0.5$, however, is just one threshold that we could have used to classify the observations in this dataset. In the next few units, we will assess how good a given predictive probability threshold is with respect to the particular goals that we have for using a given classifier model.

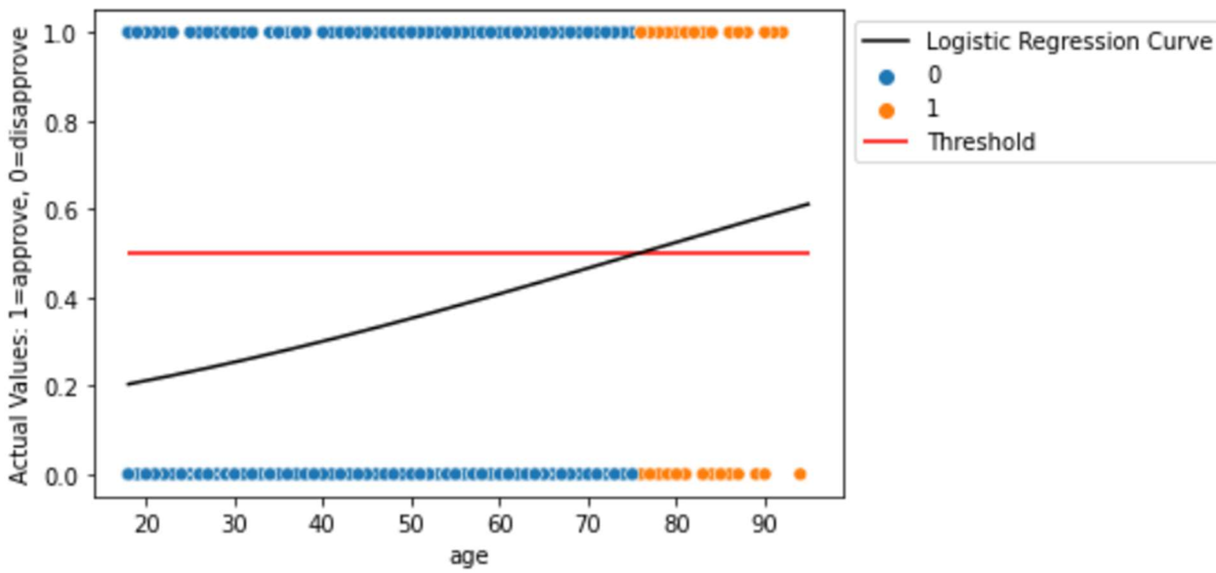
What is a classifier?

In this class, we will use a fitted logistic regression model to classify a given observation as either y=1 or y=0. But there are many types of algorithms and models that can perform the same task, which yield potentially different (better) classifications of the same set of points.

In general, we can think of a **classifier model** as a set of rules that decide which of a set of categories an observation belongs to, on the basis of a training set of data containing observations *whose category membership is known*.

3. TYPES OF MISCLASSIFICATIONS AND CORRECT CLASSIFICATIONS.

Ex: Will using a predictive probability threshold of $p_0 = 0.5$ in the logistic regression model above create misclassifications? In the plot below, label all the observations that will be misclassified and all of the observations that will be correctly classified.



Definitions about Predictions made by a Classifier (and Threshold)

We say that an observation is a **predictive positive** for a given threshold in a classifier if _________.

We say that an observation is a **predictive negative** for a given threshold in a classifier if _________.

Types of Predicted Positives

- We call an observation a **true positive** for a given threshold in a classifier if the observation *predicted* to have _________________ and it is *actually* the case that __________________. Thus the classifier has __________________ classified this observation.
- We call an observation a **false positive** for a given threshold in a classifier if the observation *predicted* to have _________________ and it is *actually* the case that __________________ for this observation. Thus the classifier has __________________ classified this observation.

Types of Predicted Negatives

- We call an observation a **true negative** for a given threshold in a classifier if the observation *predicted* to have _________________ and it is *actually* the case that __________________. Thus the classifier has __________________ classified this observation.
- We call an observation a **false negative** for a given threshold in a classifier if the observation *predicted* to have _________________ and it is *actually* the case that __________________ for this observation. Thus the classifier has __________________ classified this observation.

Confusion Matrix

We define a **confusion matrix** for a given threshold in a classifier as the following matrix below.

Classification	Actual Negative (0)	Actual Positive (1)
Predicted Negative (0)	TN = True Neg	FN = False Neg
Predicted Positive (1)	FP = False Pos	TP = True Pos

Sensitivity Rate (True Positive Rate)

We define the **sensitivity** (also known as **true positive rate**) of a given classifier model with a given threshold as the percent of observations that are *actually* a positive (ie. y=1) that are *correctly predicted* to be a positive (ie. y=1). Or in other words…

$$\boldsymbol{sensitivity\ rate} = \frac{\#\ \boldsymbol{of\ True\ Positives}}{(\#\ \boldsymbol{of\ True\ Positives}) + (\#\ \boldsymbol{of\ False\ Negatives})}$$

Ideally, we want the sensitivity rate to be ______________________.

Specificity Rate (True Negative Rate)

We define the **specificity** (also known as **true negative rate**) of a given classifier model with a given threshold as the percent of observations that are *actually* a negative (ie. y=0) that are *correctly predicted* to be a negative (ie. y=0). Or in other words…

$$\boldsymbol{specificity\ rate} = \frac{\#\ \boldsymbol{of\ True\ Negatives}}{(\#\ \boldsymbol{of\ True\ Negatives}) + (\#\ \boldsymbol{of\ False\ Positives})}$$

Ideally, we want the specificity rate to be ______________________.

False Positive Rate (1-Specificity Rate)

We define the **false positive rate** of a given classifier model with a given threshold as the percent of observations that are *actually* a negative (ie. y=0) that are *incorrectly predicted* to be a positive (ie. y=1). Or in other words…

$$\boldsymbol{specificity\ rate} = \frac{\#\ \boldsymbol{of\ False\ Positives}}{(\#\ \boldsymbol{of\ True\ Negatives}) + (\#\ \boldsymbol{of\ False\ Positives})}$$

Ideally, we want the false positive rate to be ______________________.

In the Jupyter notebook do the following.

1. Fit a logistic regression model predicting the probability that an adult in our Pew Research survey (from 2017) approves of the president's foreign policy, using age as an explanatory variable.
2. Use a predictive probability threshold of $p_0 = 0.5$ to classify all of the observations in the sample.
3. Create a confusion matrix for this threshold and classifier.
4. Find the sensitivity, specificity, and the false positive rate for this threshold and classifier.

4. Relationship between sensitivity and Specificity

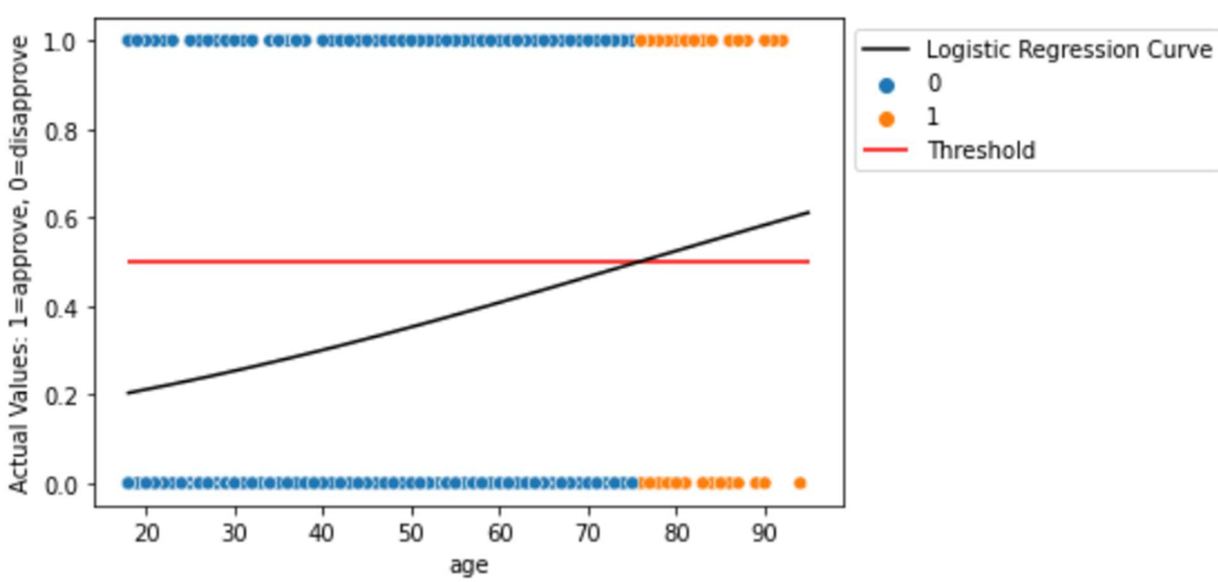
It looks like using a simplistic predictive probability threshold of $p_0 = 0.5$ provided us with:

- a low sensitivity of (0.112) (ie. the classifier was not accurate at correctly identifying people that actually approved of the president's foreign policy in the sample.), but
- a high specificity of (0.950) (ie. the classifier was accurate at correctly identifying people that actually disapproved of the president's foreign policy in the sample.), but

We can actually use any value between 0 and 1 to be our predictive probability threshold p_0.

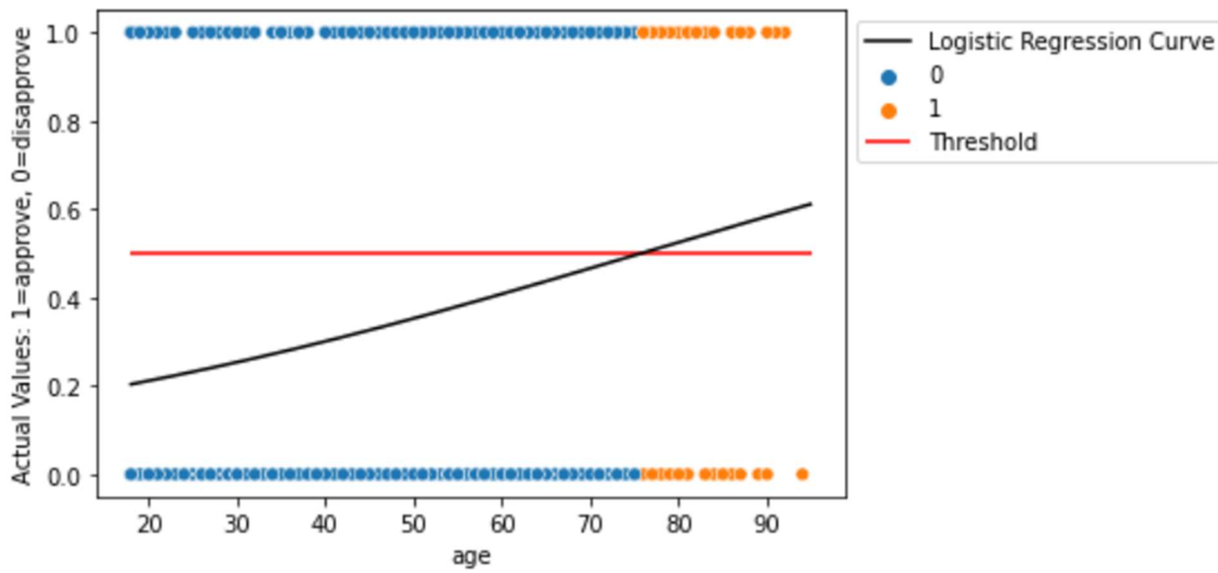
Ex: If we <u>**increased** our predictive probability threshold</u> (from it's original value of $p_0 = 0.5$), what would we expect to happen to the:

- sensitivity: ________________________________
- specificity: ________________________________



Ex: If we **decreased** our predictive probability threshold (from it's original value of $p_0 = 0.5$), what would we expect to happen to the:

- sensitivity: ________________________________
- specificity: ________________________________



In general, **increasing our predictive probability threshold** will _________________ the **sensitivity** rate and _______________________ the **specificity** rate.

In general, **decreasing our predictive probability threshold** will _________________ the **sensitivity** rate and _______________________ the **specificity** rate.

In the Jupyter notebook do the following.

1. Using the same logistic regression model as before, and now a predictive probability threshold of $p_0 = 0.3$ to classify all of the observations in the sample.
2. Find the sensitivity and specificity for this threshold and classifier.
3. Which predictive probability threshold yielded better results? $p_0 = .5$ or $p_0 = 0.3$?

5. ROC AND AUC – WHICH *CLASSIFIER* WILL GIVE US THE BEST COMBINATIONS OF (FALSE POSITIVE RATE, TRUE POSITIVE RATE) FOR ALL SETS OF THRESHOLDS?

Remember, that

- true positive rate = sensitivity
- false positive rate = (1- specificity).

So if sensitivity increases (good), then specificity will ___________________, and thus the false positive rate will _____________________ (bad).

So **an ideal classifier model** will have some predictive probability threshold p_0 that yields a **true positive rate** that is __________________ and **a false positive rate** that is ___________________.

ROC Curve

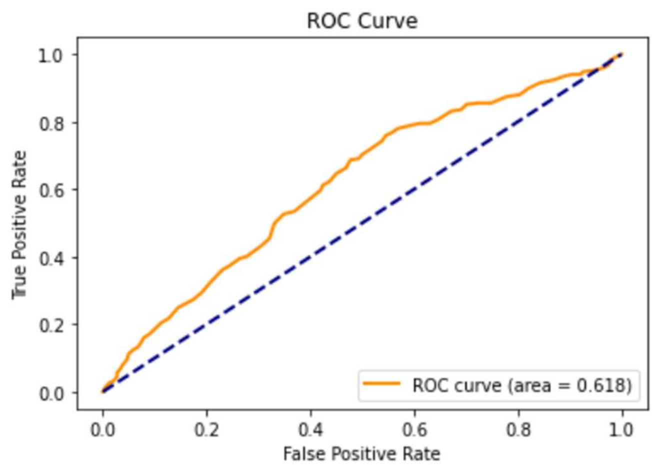
To help us select which classifier model is best (out of all possible predictive probability thresholds that they could use), we plot what we call a **Receiver Operating Characteristic curve (ROC curve)** for a given classifier. We create these plots by

1. selecting many, many values of the predictive probability threshold p_0 starting from 0 and going to 1, and then
2. plotting the resulting (false positive rate, true positive rate) as a coordinate on a line plot.

<u>Model 1</u>:

<u>**ROC:**</u>

For instance, the plot below is the ROC curve for the logistic regression model that predicts **approval of presidential foreign policy** given **<u>age</u>** for the data **in the sample.**



<u>**ROC Interpretation:**</u>

As we can see, this model ___________________ have any predictive probability threshold that will get us anywhere close to the ideal (false positive rate, true positive rate) = (0,1).

This is an indication that this classifier (in general) _______________ yield very accurate classifications of the sample data.

AUC

We can *visually* see that this is the case with this plot, but how can we quickly numerically quantify this without having to look at the plot?

Property of an Ideal ROC Curve: A ROC curve that has a predictive probability threshold that yields the ideal combination of (false positive rate, true positive rate) = (0,1) will have an area of __________________ under the ROC curve. We call this the **AUC** or the **Area Under the Curve** (of the ROC curve).

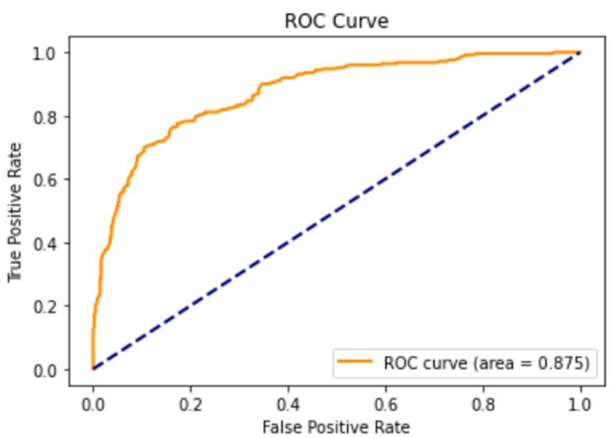
Interpretation of AUC: The closer the AUC of a given classifier model is to ___________, the closer this model will be to having a carefully selected predictive probability threshold that yields the ideal combination of (false positive rate, true positive rate) = (0,1).

Property of the ROC Curve of a Random Classifier: A ROC curve of a classifier model that randomly assigns points to either a y=0 or a y=1, is expected to have on average an AUC of _________________. You could think of a model with an AUC of ______________ as a "worst case scenario."

Model 2:

ROC:

For instance, the plot below is the ROC curve for the logistic regression model that predicts **approval of presidential foreign policy** given **age, sex, and party** for the data **in the sample.**



ROC Interpretation:

As we can see, this model does have a predictive probability threshold that will get us ____________ to the ideal (false positive rate, true positive rate) = (0,1).

This is an indication that this model 2 classifier (in general) _______________ yield more accurate classifications of the sample data than the model 1 classifier.

AUC Interpretation:

We can quantify this this higher performance of model 2 over model 1 as well because the AUC of the model 2 classifier is _______________________ than the AUC of the model 1 classifier _________________.