In-Class Exercise 01/23/2020

Team 7

1/27/2020

A regression with all variables to predict balance

```
fullBalanceModel <- lm(
  balance ~ age + job + marital + education + default + housing + loan + contact + day + month + duratio
  data=bank
)
```

Using general intuition to subset variables to predict balance

Using some general intuition about variables that may be related to predicting balance, we attempt to subset the full model to fewer predictors by eliminating variables which have no effect on predicting a customer's account balance:

```
intuitiveBalanceModel <- lm(
  balance ~ age + default + marital,
  data=bank
)
```

With the null hypothesis being intuitiveBalanceModel

```
anova(intuitiveBalanceModel, fullBalanceModel)
```

yields an F value of 6.1, thus there are more factors that meaningfully contribute to predicting balance in the fullBalanceModel that are not represented in intuitiveBalanceModel.

Updating our intuitive model

Using some more intuitive thinking, we update the intuitiveBalanceModel to attempt to better predict balance:

```
updatedIntuitiveBalanceModel <- lm(
  balance ~ age + default + marital + job + education,
  data=bank
)
```

With the null hypothesis being updatedIntuitiveBalanceModel

```
anova(updatedIntuitiveBalanceModel, fullBalanceModel)
```

yields an F value of 7.8, thus we got farther from accurately predicting balance by adding job and education, so at least one of these factors is irrelevant in predicting balance

Statistically relevant variables

Now that intuition is failing us, we can attempt to see what variables could be meaningful in predicting balance by looking at the t value column of the summary of the fullBalanceModel, and choose all categories with a t value > 2:

```
statisticallyImportantModel <- lm(
  balance ~ age + default + marital + job + loan + month,
  data=bank
)
```

With the null hypothesis being staisticallyImportantModel

```
anova(statisticallyImportantModel, fullBalanceModel)
```

yields an F value of 1.2, thus we find that this model a good predictor of balance after removing meaningless variables

Discussion

The resulting final model demonstrates that the age of the customer, whether the customer has any credit in default, the marital status of the customer, what job the customer holds, whether the customer holds a loan or not, and what month the customer joined the bank in, all hold meaningful correlations when attempting to predict the balance a customer has in an account with the bank. Which further leadsto a model which can also be useful in predicting any of the aforementioned variables, as correlations work in both directions.