TRUMAN STATE UNIVERSITY

Professional Program in DATA SCIENCE

# Algorithmic Bias

DISTINCT BY DESIGN

# What is Bias?

# What do we mean by Bias?

- **In common usage:** A tendency to favor or dislike a particular person, thing or idea that often affects our actions.

# What do we mean by Bias?

- **In common usage:** A tendency to favor or dislike a particular person, thing or idea that often affects our actions.
- **In statistical analysis:** Any time the results of an analysis are not consistent with what's really true.

# What do we mean by Bias?

- **In common usage:** A tendency to favor or dislike a particular person, thing or idea that often affects our actions.
- **In statistical analysis:** Any time the results of an analysis are not consistent with what's really true.
- *These are not unrelated! Often, one leads to the other, in either order!*

# As we go forward, it’s good to consider our own biases.

- **Unconscious Bias:** Collections of unconscious associations with a group or class that affect perceptions related to that group.

A nice long list is available on Wikipedia (of course):
List of Biases in Judgment and Decision Making (wikipedia.org)

# As we go forward, it’s good to consider our own biases.

- **Unconscious Bias:** Collections of unconscious associations with a group or class that affect perceptions related to that group.
- **Confirmation Bias:** The tendency to believe/remember observations that agree with our existing beliefs.

A nice long list is available on Wikipedia (of course):
List of Biases in Judgment and Decision Making (wikipedia.org)

# As we go forward, it’s good to consider our own biases.

- **Unconscious Bias:** Collections of unconscious associations with a group or class that affect perceptions related to that group.
- **Confirmation Bias:** The tendency to believe/remember observations that agree with our existing beliefs.
- **Egocentric Bias:** The tendency to believe too highly in one’s own opinions/perception of reality.

A nice long list is available on Wikipedia (of course):
List of Biases in Judgment and Decision Making (wikipedia.org)

# As we go forward, it's good to consider our own biases.

- **Unconscious Bias:** Collections of unconscious associations with a group or class that affect perceptions related to that group.
- **Confirmation Bias:** The tendency to believe/remember observations that agree with our existing beliefs.
- **Egocentric Bias:** The tendency to believe too highly in one's own opinions/perception of reality.
- **Truthiness Bias:** Easier-to-understand statements, statements that "seem" right, or statements repeated multiple times tend to be believed more easily.

A nice long list is available on Wikipedia (of course):
List of Biases in Judgment and Decision Making (wikipedia.org)

# Statistical Bias is the difference between an estimator's expected value and a true parameter value.

| Name | Pop. Param. | Estimator | Bias |
|---|---|---|---|
| Mean | $\mu$ | $\bar{x}$ | $E(\bar{x}) - \mu$ |
| Std. Dev. | $\sigma$ | $s$ | $E(s) - \sigma$ |
| Variance | $\sigma^2$ | $s^2$ | $E(s^2) - \sigma^2$ |
| Proportion | $p$ | $\hat{p}$ | $E(\hat{p}) - p$ |
| Generically | $\theta$ | $\hat{\theta}$ | $E(\hat{\theta}) - \theta$ |

- An estimator whose bias is zero is called an *unbiased estimator*.

# Statistical Bias is the difference between an estimator's expected value and a true parameter value.

| Name | Pop. Param. | Estimator | Bias |
|---|---|---|---|
| Mean | $\mu$ | $\bar{x}$ | $E(\bar{x}) - \mu$ |
| Std. Dev. | $\sigma$ | $s$ | $E(s) - \sigma$ |
| Variance | $\sigma^2$ | $s^2$ | $E(s^2) - \sigma^2$ |
| Proportion | $p$ | $\hat{p}$ | $E(\hat{p}) - p$ |
| Generically | $\theta$ | $\hat{\theta}$ | $E(\hat{\theta}) - \theta$ |

- An estimator whose bias is zero is called an *unbiased estimator*.
- $\bar{x}$ and $s^2$ are unbiased, but $s$ is not.

# Statistical Bias is the difference between an estimator's expected value and a true parameter value.

| Name | Pop. Param. | Estimator | Bias |
|---|---|---|---|
| Mean | $\mu$ | $\bar{x}$ | $E(\bar{x}) - \mu$ |
| Std. Dev. | $\sigma$ | $s$ | $E(s) - \sigma$ |
| Variance | $\sigma^2$ | $s^2$ | $E(s^2) - \sigma^2$ |
| Proportion | $p$ | $\hat{p}$ | $E(\hat{p}) - p$ |
| Generically | $\theta$ | $\hat{\theta}$ | $E(\hat{\theta}) - \theta$ |

- An estimator whose bias is zero is called an *unbiased estimator*.
- $\bar{x}$ and $s^2$ are unbiased, but $s$ is not.
- A mathematical statistics class might explore what makes a good estimator, and that knowledge *can* be helpful in applied machine learning.

# Sources of Bias

# Data collection is part of the calculation of any statistic, and can lead to bias!

Who is in the population?

↓

How do we take a sample?

↓

How do we ask a question?

↓

Statistic(s):
As simple as $\bar{x}$,
or as complicated as ML parameters.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn't return a representative sample from the population.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn’t return a representative sample from the population.
    - Convenience sampling – No good method was used.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn't return a representative sample from the population.
  - Convenience sampling – No good method was used.
  - Frame error – List of who's in the population was skewed.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn't return a representative sample from the population.
  - Convenience sampling – No good method was used.
  - Frame error – List of who's in the population was skewed.
- **Non-response Bias:** Those *in the sample* who chose not to respond are different in some way from those who chose to respond.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn't return a representative sample from the population.
    - Convenience sampling – No good method was used.
    - Frame error – List of who's in the population was skewed.
- **Non-response Bias:** Those *in the sample* who chose not to respond are different in some way from those who chose to respond.
- **Response Bias:** Data that is collected is not accurate.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn't return a representative sample from the population.
  - Convenience sampling – No good method was used.
  - Frame error – List of who's in the population was skewed.
- **Non-response Bias:** Those *in the sample* who chose not to respond are different in some way from those who chose to respond.
- **Response Bias:** Data that is collected is not accurate.
  - Question wording, modality of survey can affect responses.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn't return a representative sample from the population.
  - Convenience sampling – No good method was used.
  - Frame error – List of who's in the population was skewed.
- **Non-response Bias:** Those *in the sample* who chose not to respond are different in some way from those who chose to respond.
- **Response Bias:** Data that is collected is not accurate.
  - Question wording, modality of survey can affect responses.
  - Interviewers/observers can have skewed perceptions.

# In traditional sampling, three types of bias are often described.

- **Sampling Bias:** The sampling method doesn't return a representative sample from the population.
  - Convenience sampling – No good method was used.
  - Frame error – List of who's in the population was skewed.
- **Non-response Bias:** Those *in the sample* who chose not to respond are different in some way from those who chose to respond.
- **Response Bias:** Data that is collected is not accurate.
  - Question wording, modality of survey can affect responses.
  - Interviewers/observers can have skewed perceptions.
  - Sometimes respondents just don't tell the truth.

# Beyond sampling and response, bias can come from our *definition* of variables.

- Humans choose *what* to measure.

# Beyond sampling and response, bias can come from our *definition* of variables.

- Humans choose *what* to measure.
- Do we understand the direction of causal relationships between predictor and response?

# Beyond sampling and response, bias can come from our *definition* of variables.

- Humans choose *what* to measure.
- Do we understand the direction of causal relationships between predictor and response?
- Have we thought through the possible lurking variables?

# Beyond sampling and response, bias can come from our *definition* of variables.

- Humans choose *what* to measure.
- Do we understand the direction of causal relationships between predictor and response?
- Have we thought through the possible lurking variables?
- Does a response variable have the real-world meaning that we think it does?

# Example: Medical professionals adjust lung function measurements according to age, sex, height and *race*.

- Readings for those categorized as "black" are adjusted up by 10–15%. Poorer lung function is assumed to be normal.

# Example: Medical professionals adjust lung function measurements according to age, sex, height and *race*.

- Readings for those categorized as “black” are adjusted up by 10–15%. Poorer lung function is assumed to be normal.
- Pro’s for using race:

# Example: Medical professionals adjust lung function measurements according to age, sex, height and *race*.

- Readings for those categorized as “black” are adjusted up by 10–15%. Poorer lung function is assumed to be normal.
- Pro’s for using race:
  - It often has predictive power.

# Example: Medical professionals adjust lung function measurements according to age, sex, height and *race*.

- Readings for those categorized as “black” are adjusted up by 10–15%. Poorer lung function is assumed to be normal.
- Pro’s for using race:
  - It often has predictive power.
  - Historically viewed as “easy” to measure.

# Example: Medical professionals adjust lung function measurements according to age, sex, height and *race*.

- Readings for those categorized as “black” are adjusted up by 10–15%. Poorer lung function is assumed to be normal.
- Pro’s for using race:
  - It often has predictive power.
  - Historically viewed as “easy” to measure.
- Con’s for using race:

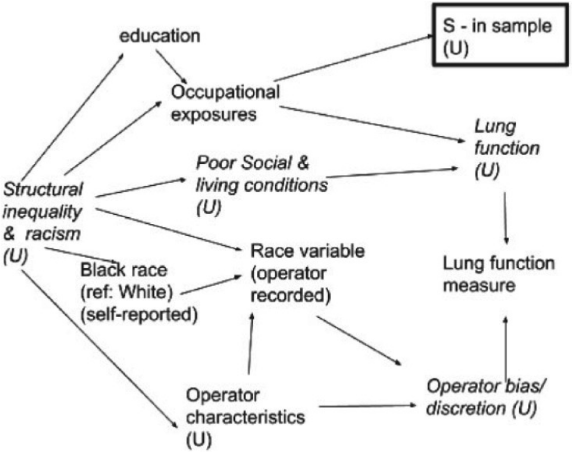
# Example: Medical professionals adjust lung function measurements according to age, sex, height and *race*.

- Readings for those categorized as “black” are adjusted up by 10–15%. Poorer lung function is assumed to be normal.
- Pro’s for using race:
  - It often has predictive power.
  - Historically viewed as “easy” to measure.
- Con’s for using race:
  - Ill-defined, and usually a stand-in for other variables.

# Example: Medical professionals adjust lung function measurements according to age, sex, height and *race*.

- Readings for those categorized as "black" are adjusted up by 10–15%. Poorer lung function is assumed to be normal.
- Pro's for using race:
  - It often has predictive power.
  - Historically viewed as "easy" to measure.
- Con's for using race:
  - Ill-defined, and usually a stand-in for other variables.
  - Promotes a view that race is *causally* related to outcomes.

# Example: What are the causal connections between race and lung function?




Source: Teaching Yourself about Structural Racism Will Improve Your Machine Learning (nih.gov)

See also: Race, Ethnicity and Lung Function: A Brief History (nih.gov)

# Example: What happens when patient health care costs are used as a proxy for health care needs?

- An algorithm was used to recommend extra support for patients with complex health-care needs.

Source: A Health Care Algorithm Offered Less Care to Black Patients (arstechnica.com)

# Example: What happens when patient health care costs are used as a proxy for health care needs?

- An algorithm was used to recommend extra support for patients with complex health-care needs.
- Patients likely to have higher health-care costs were recommended for these programs. (Cost was presumably easier to measure than “need for care.”)

Source: A Health Care Algorithm Offered Less Care to Black Patients (arstechnica.com)

# Example: What happens when patient health care costs are used as a proxy for health care needs?

- An algorithm was used to recommend extra support for patients with complex health-care needs.
- Patients likely to have higher health-care costs were recommended for these programs. (Cost was presumably easier to measure than “need for care.”)
- Context: Patients who are not white have historically generated lower costs because they have more barriers to accessing health care.

Source: A Health Care Algorithm Offered Less Care to Black Patients (arstechnica.com)

# Example: What happens when patient health care costs are used as a proxy for health care needs?

- An algorithm was used to recommend extra support for patients with complex health-care needs.
- Patients likely to have higher health-care costs were recommended for these programs. (Cost was presumably easier to measure than "need for care.")
- Context: Patients who are not white have historically generated lower costs because they have more barriers to accessing health care.
- Result: Non-white patients had to be sicker before they were referred to support programs.

Source: A Health Care Algorithm Offered Less Care to Black Patients (arstechnica.com)

# Example: Bias in recidivism risk assessments illustrates the importance of algorithm evaluation on subgroups.

- Pro Publica published an evaluation of the proprietary COMPAS risk assessment algorithm.

| Race | Actual | Low Risk | High Risk | True % | Accuracy |
|---|---|---|---|---|---|
| All | Didn’t Recidivate | 2681 | 1281 | 68% | 65% |
| All | Recidivated | 1216 | 2035 | 63% | |
| Black | Didn’t Recidivate | 990 | 805 | 55% | 64% |
| Black | Recidivated | 532 | 1369 | 72% | |
| White | Didn’t Recidivate | 1139 | 349 | 77% | 67% |
| White | Recidivated | 461 | 505 | 52% | |

Source: Machine Bias (propublica.org)

See Also: Flaws Plague a Tool Meant to Help Low-Risk Federal Prisoners Win Early Release (npr.org)

# Example: Bias in recidivism risk assessments illustrates the importance of algorithm evaluation on subgroups.

- Pro Publica published an evaluation of the proprietary COMPAS risk assessment algorithm.
- Contingency tables compare risk assessment to actual recidivism within two years of arrest.

| Race | Actual | Low Risk | High Risk | True % | Accuracy |
|---|---|---|---|---|---|
| All | Didn't Recidivate | 2681 | 1281 | 68% | 65% |
| All | Recidivated | 1216 | 2035 | 63% | |
| Black | Didn't Recidivate | 990 | 805 | 55% | 64% |
| Black | Recidivated | 532 | 1369 | 72% | |
| White | Didn't Recidivate | 1139 | 349 | 77% | 67% |
| White | Recidivated | 461 | 505 | 52% | |

Source: Machine Bias (propublica.org)

See Also: Flaws Plague a Tool Meant to Help Low-Risk Federal Prisoners Win Early Release (npr.org)

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.
- Algorithms need *balanced* training data. Otherwise…

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.
- Algorithms need *balanced* training data. Otherwise…
  - *groups* with less data see less accurate predictions, and

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.
- Algorithms need *balanced* training data. Otherwise…
  - *groups* with less data see less accurate predictions, and
  - more prevalent *outcomes* see more accurate predictions.

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.
- Algorithms need *balanced* training data. Otherwise…
  - *groups* with less data see less accurate predictions, and
  - more prevalent *outcomes* see more accurate predictions.
- Training data can reflect existing human bias—regardless of balance, ML will replicate those patterns.

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.
- Algorithms need *balanced* training data. Otherwise…
  - *groups* with less data see less accurate predictions, and
  - more prevalent *outcomes* see more accurate predictions.
- Training data can reflect existing human bias—regardless of balance, ML will replicate those patterns.
- Humans tend to invest too much authority in algorithms.

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.
- Algorithms need *balanced* training data. Otherwise…
  - *groups* with less data see less accurate predictions, and
  - more prevalent *outcomes* see more accurate predictions.
- Training data can reflect existing human bias—regardless of balance, ML will replicate those patterns.
- Humans tend to invest too much authority in algorithms.
- Algorithms get out of date, losing accuracy over time.

# “Machine Learning” presents particular algorithmic bias issues.

- Overall accuracy/lack of bias is not good enough. We need to look at subgroups.
- Algorithms need *balanced* training data. Otherwise…
  - *groups* with less data see less accurate predictions, and
  - more prevalent *outcomes* see more accurate predictions.
- Training data can reflect existing human bias—regardless of balance, ML will replicate those patterns.
- Humans tend to invest too much authority in algorithms.
- Algorithms get out of date, losing accuracy over time.
- Algorithms can create feedback loops: negative decisions can create negative outcomes which reinforce negative decisions.

## How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.
- Research toward “explainable” algorithmic decisions—eliminate the black box.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.
- Research toward “explainable” algorithmic decisions—eliminate the black box.
    - Open source code.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.
- Research toward “explainable” algorithmic decisions—eliminate the black box.
  - Open source code.
  - Algorithms report individual decisions with a rationale.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.
- Research toward “explainable” algorithmic decisions—eliminate the black box.
    - Open source code.
    - Algorithms report individual decisions with a rationale.
- Legal remedies.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.
- Research toward “explainable” algorithmic decisions—eliminate the black box.
    - Open source code.
    - Algorithms report individual decisions with a rationale.
- Legal remedies.
    - Restrictions on algorithm use.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.
- Research toward “explainable” algorithmic decisions—eliminate the black box.
    - Open source code.
    - Algorithms report individual decisions with a rationale.
- Legal remedies.
    - Restrictions on algorithm use.
    - Right to challenge decisions.

# How do we fix problems of algorithmic bias?

- Audit training data for obvious human bias.
- Work to create balanced training data sets.
- Research toward “explainable” algorithmic decisions—eliminate the black box.
  - Open source code.
  - Algorithms report individual decisions with a rationale.
- Legal remedies.
  - Restrictions on algorithm use.
  - Right to challenge decisions.
- Diversify data science teams to help eliminate blind spots.

# Some Technical Approaches: Balancing Data

# Example: Balancing Data in Predicting Insurance Expense for Smokers and Non-Smokers

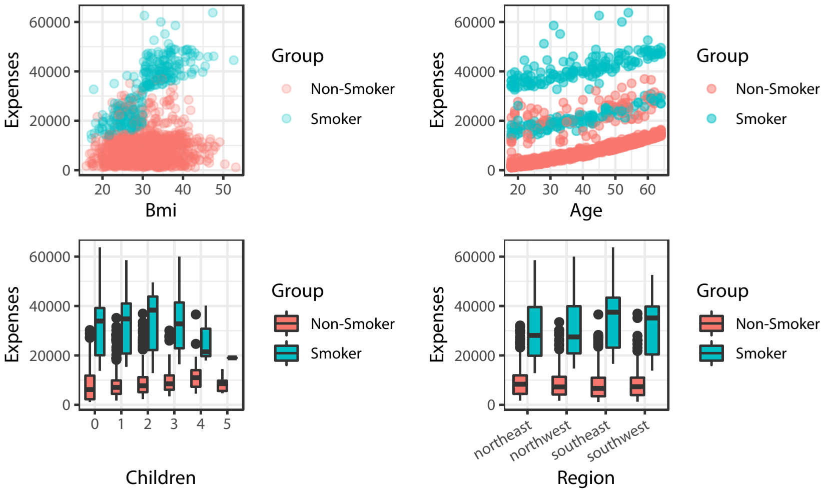
```r
# Load the data.
ins <- read.csv("data/Insurance_A.csv", stringsAsFactors=TRUE) %>%
  mutate(smoker = factor(if_else(smoker=="yes", "Smoker", "Non-Smoker"))) %>
  rename(Group = smoker)
names(ins) = str_to_title(names(ins))             # Just nice graph labels.
head(ins, 2)
```

```
##   Age    Sex  Bmi Children      Group    Region Expenses
## 1  19 female 27.9        0     Smoker southwest 16884.92
## 2  18   male 33.8        1 Non-Smoker southeast  1725.55
```

```r
# Make a training and test set.
set.seed(892)                                     # For repeatability
ind <- sample(1:nrow(ins), round(0.7*nrow(ins)))
ins.train <- ins[ind, ]
ins.test <- ins[-ind, ]
```

Data set in the public domain, from *Machine Learning with R* by Brett Lantz. See here for deatils (github.com)

# Expenses depend most strongly on BMI and Age.



# A Helpful Visualization Function for Bias and RMSE

```
library(randomForest)
library(caret) # For the downSample and upSample commands.
fit.stats <- function(fit, test){
  test$res <- test$Expenses - predict(fit, newdata=test)
  rbind(test %>% group_by(Group) %>%
          summarize(Bias = -mean(res),
                    RMSE=sqrt(mean(res^2))),
        test %>%
          summarize(Group="ALL", Bias = -mean(res),
                    RMSE=sqrt(mean(res^2))))
}
```

# Initial Random Forest Fit with Original Data Set

```r
fit.orig <- randomForest(Expenses ~ ., data=ins.train)
```

```r
table(ins.train$Group) %>%
  kable()
```

| Var1 | Freq |
|---|---|
| Non-Smoker | 750 |
| Smoker | 187 |

```r
fit.stats(fit.orig, ins.test) %>
  kable()
```

| Group | Bias | RMSE |
|---|---|---|
| Non-Smoker | -411.5110 | 5438.676 |
| Smoker | -772.9750 | 3384.287 |
| ALL | -489.9334 | 5064.254 |

# Initial Random Forest Fit with Non-Smokers Downsampled

```
ins.train.ds <- downSample(x=select(ins.train, -Group),
                           y=ins.train$Group, yname="Group")
fit.ds <- randomForest(Expenses ~ ., data=ins.train.ds)
```

```
table(ins.train.ds$Group) %>%
  kable()
```

| Var1 | Freq |
|---|---|
| Non-Smoker | 187 |
| Smoker | 187 |

```
fit.stats(fit.ds, ins.test) %>%
  kable()
```

| Group | Bias | RMSE |
|---|---|---|
| Non-Smoker | 255.7121 | 5395.480 |
| Smoker | -321.5732 | 2957.973 |
| ALL | 130.4657 | 4969.266 |

# Initial Random Forest Fit with Smokers Upsampled

```r
ins.train.us <- upSample(x=select(ins.train, -Group),
                         y=ins.train$Group, yname="Group")
fit.us <- randomForest(Expenses ~ ., data=ins.train.us)
```

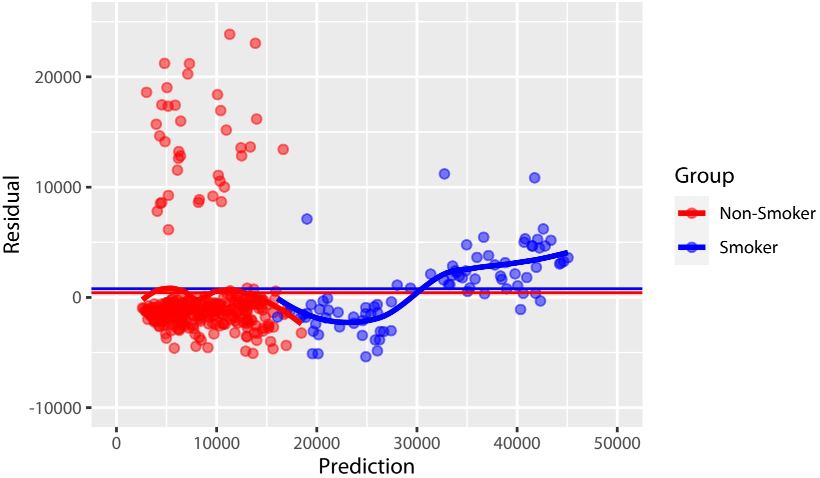
```r
table(ins.train.us$Group) %>%
  kable()
```

| Var1 | Freq |
|---|---|
| Non-Smoker | 750 |
| Smoker | 750 |

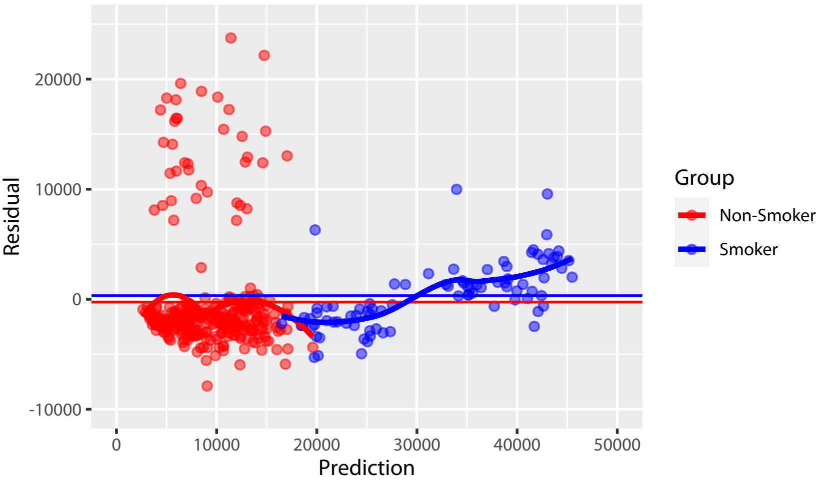
```r
fit.stats(fit.us, ins.test) %>%
  kable()
```

| Group | Bias | RMSE |
|---|---|---|
| Non-Smoker | -355.9894 | 5436.022 |
| Smoker | -379.3473 | 2760.703 |
| ALL | -361.0571 | 4979.227 |

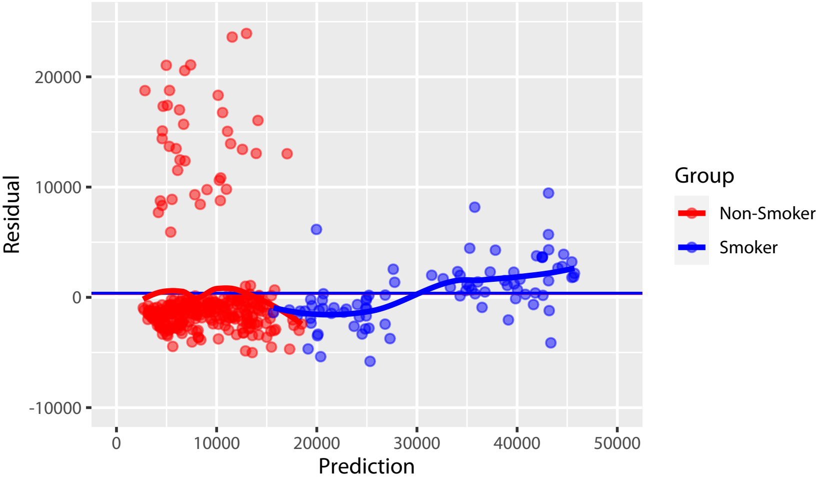
# Residuals vs. Fit for Original Data



# Residuals vs. Fit for Downsampled Data



# Residuals vs. Fit for Upsampled Data



# Example: Weighting a Categorical Response

- Our data set has measurements of several varieties of dry beans.

```r
beans <- read.csv("data/dry_bean_dataset.csv")
names(beans)
```

```
##  [1] "Area"            "Perimeter"       "MajorAxisLength" "MinorAxisLength"
##  [5] "AspectRation"    "Eccentricity"    "ConvexArea"      "EquivDiameter"
##  [9] "Extent"          "Solidity"        "roundness"       "Compactness"
## [13] "ShapeFactor1"    "ShapeFactor2"    "ShapeFactor3"    "ShapeFactor4"
## [17] "Class"
```

```r
table(beans$Class)/sum(table(beans$Class))
```

```
## 
##   BARBUNYA     BOMBAY       CALI   DERMASON      HOROZ      SEKER       SIRA 
## 0.09712732 0.03835133 0.11975608 0.26052458 0.14165014 0.14892366 0.19366689
```

KOKLU, M. and OZKAN, I.A., (2020), “Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques.” Computers and Electronics in Agriculture, 174, 105507. DOI: https://doi.org/10.1016/j.compag.2020.105507

# Example: Weighting a Categorical Response

- Our data set has measurements of several varieties of dry beans.
- We'd like to predict whether a bean is of the "Sira" variety.

```
beans <- read.csv("data/dry_bean_dataset.csv")
names(beans)
```

```
##  [1] "Area"            "Perimeter"       "MajorAxisLength" "MinorAxisLength"
##  [5] "AspectRation"    "Eccentricity"    "ConvexArea"      "EquivDiameter"
##  [9] "Extent"          "Solidity"        "roundness"       "Compactness"
## [13] "ShapeFactor1"    "ShapeFactor2"    "ShapeFactor3"    "ShapeFactor4"
## [17] "Class"
```

```
table(beans$Class)/sum(table(beans$Class))
```

```
##
##   BARBUNYA     BOMBAY       CALI   DERMASON      HOROZ      SEKER       SIRA
## 0.09712732 0.03835133 0.11975608 0.26052458 0.14165014 0.14892366 0.19366689
```

KOKLU, M. and OZKAN, I.A., (2020), "Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques." Computers and Electronics in Agriculture, 174, 105507. DOI: https://doi.org/10.1016/j.compag.2020.105507

# Example: Weighting a Categorical Response

- Our data set has measurements of several varieties of dry beans.
- We'd like to predict whether a bean is of the "Sira" variety.
- Note that Sira makes up only about 19% of the data.

```
beans <- read.csv("data/dry_bean_dataset.csv")
names(beans)
```

```
##  [1] "Area"            "Perimeter"       "MajorAxisLength" "MinorAxisLength"
##  [5] "AspectRation"    "Eccentricity"    "ConvexArea"      "EquivDiameter"
##  [9] "Extent"          "Solidity"        "roundness"       "Compactness"
## [13] "ShapeFactor1"    "ShapeFactor2"    "ShapeFactor3"    "ShapeFactor4"
## [17] "Class"
```

```
table(beans$Class)/sum(table(beans$Class))
```

```
##
##   BARBUNYA     BOMBAY       CALI   DERMASON      HOROZ      SEKER       SIRA
## 0.09712732 0.03835133 0.11975608 0.26052458 0.14165014 0.14892366 0.19366689
```

KOKLU, M. and OZKAN, I.A., (2020), "Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques." Computers and Electronics in Agriculture, 174, 105507. DOI: https://doi.org/10.1016/j.compag.2020.105507

# An unweighted model has good accuracy, but

```r
library(e1071)
# Focus only on Sira.
beans$Class <- fct_other(beans$Class, keep="SIRA")
summary(beans$Class)
```

```
##  SIRA Other
##  2636 10975
```

```r
# Create training and test set.
set.seed(9084)
ind <- sample(1:nrow(beans), round(0.70*nrow(beans)))
beans.train <- beans[ind, ]
beans.test <- beans[-ind, ]
unweighted.svm <- svm(Class ~ ., data=beans.train)
unweighted.pred <- predict(unweighted.svm, newdata=beans.test)
```

# An unweighted model has good accuracy, but is better at predicting beans that aren't Sira.

```r
cm <- confusionMatrix(unweighted.pred, beans.test$Class, positive="SIRA")
cm$table
```

```
##            Reference
## Prediction SIRA Other
##      SIRA   660   102
##      Other  101  3220
```

```r
c(cm$overall["Accuracy"], cm$byClass[c("Sensitivity", "Specificity")])
```

```
##    Accuracy Sensitivity Specificity
##   0.9502817   0.8672799   0.9692956
```

# We can balance the model's predictions by weighting the levels of the categorical response.

- Most predictive model commands allow this with a command option.

```
class.pcts <- table(beans.train$Class)/sum(table(beans.train$Class))
class.pcts
```

```
## 
##      SIRA     Other 
## 0.1967884 0.8032116
```

```
weighted.svm <- svm(Class ~ ., data=beans.train,
                    class.weights=c("SIRA"=1/.197, "Other"=1/.803))
weighted.pred <- predict(weighted.svm, newdata=beans.test)
```

# We can balance the model's predictions by weighting the levels of the categorical response.

- Most predictive model commands allow this with a command option.
- Weights determine how much each element counts towards the cost function that determines the model's fit.

```
class.pcts <- table(beans.train$Class)/sum(table(beans.train$Class))
class.pcts
```

```
##
##      SIRA     Other
## 0.1967884 0.8032116
```

```
weighted.svm <- svm(Class ~ ., data=beans.train,
                    class.weights=c("SIRA"=1/.197, "Other"=1/.803))
weighted.pred <- predict(weighted.svm, newdata=beans.test)
```

# We can balance the model's predictions by weighting the levels of the categorical response.

- Most predictive model commands allow this with a command option.
- Weights determine how much each element counts towards the cost function that determines the model's fit.
- You might try weights inverse to each category's proportion.

```
class.pcts <- table(beans.train$Class)/sum(table(beans.train$Class))
class.pcts
```

```
##
##      SIRA     Other
## 0.1967884 0.8032116
```

```
weighted.svm <- svm(Class ~ ., data=beans.train,
                    class.weights=c("SIRA"=1/.197, "Other"=1/.803))
weighted.pred <- predict(weighted.svm, newdata=beans.test)
```

# The weighted model improved sensitivity at the cost of sensitivity and overall accuracy.

```r
cm <- confusionMatrix(weighted.pred, beans.test$Class, positive="SIRA")
cm$table
```

```
##             Reference
## Prediction SIRA Other
##      SIRA   721   251
##      Other   40  3071
```

```r
c(cm$overall["Accuracy"], cm$byClass[c("Sensitivity", "Specificity")])
```

```
##    Accuracy Sensitivity Specificity
##   0.9287289   0.9474376   0.9244431
```

# Further Information on Weighting.

- Algorithms like SMOTE and ROSE combine up and down-sampling, while *synthesizing* new data points in the minority class, as opposed to simply duplicating existing data points. See the `caret` documentation for examples (github.io).

# Further Information on Weighting.

- Algorithms like SMOTE and ROSE combine up and down-sampling, while *synthesizing* new data points in the minority class, as opposed to simply duplicating existing data points. See the `caret` documentation for examples (github.io).
- Sampling techniques should be applied *only to the training set*.

# Further Information on Weighting.

- Algorithms like SMOTE and ROSE combine up and down-sampling, while *synthesizing* new data points in the minority class, as opposed to simply duplicating existing data points. See the `caret` documentation for examples (github.io).
- Sampling techniques should be applied *only to the training set*.
- In cross-validation, apply inside the cross-validation step.

# Further Information on Weighting.

- Algorithms like SMOTE and ROSE combine up and down-sampling, while *synthesizing* new data points in the minority class, as opposed to simply duplicating existing data points. See the `caret` documentation for examples (github.io).
- Sampling techniques should be applied *only to the training set*.
- In cross-validation, apply inside the cross-validation step.
- More complicated weighting by case (row), rather than category, is possible. (e.g., `case.weights` option in the `ranger` package.)

# Further Information on Weighting.

- Algorithms like SMOTE and ROSE combine up and down-sampling, while *synthesizing* new data points in the minority class, as opposed to simply duplicating existing data points. See the `caret` documentation for examples (github.io).
- Sampling techniques should be applied *only to the training set*.
- In cross-validation, apply inside the cross-validation step.
- More complicated weighting by case (row), rather than category, is possible. (e.g., `case.weights` option in the `ranger` package.)
- Weighting may reduce overall performance to predict one group better. You can't really create something out of nothing.

# Further Information on Weighting.

- Algorithms like SMOTE and ROSE combine up and down-sampling, while *synthesizing* new data points in the minority class, as opposed to simply duplicating existing data points. See the `caret` documentation for examples (github.io).
- Sampling techniques should be applied *only to the training set*.
- In cross-validation, apply inside the cross-validation step.
- More complicated weighting by case (row), rather than category, is possible. (e.g., `case.weights` option in the `ranger` package.)
- Weighting may reduce overall performance to predict one group better. You can't really create something out of nothing.
- Weighting can accentuate any problems with subgroup data.

# Further Information on Weighting.

- Algorithms like SMOTE and ROSE combine up and down-sampling, while *synthesizing* new data points in the minority class, as opposed to simply duplicating existing data points. See the `caret` documentation for examples (github.io).
- Sampling techniques should be applied *only to the training set*.
- In cross-validation, apply inside the cross-validation step.
- More complicated weighting by case (row), rather than category, is possible. (e.g., `case.weights` option in the `ranger` package.)
- Weighting may reduce overall performance to predict one group better. You can't really create something out of nothing.
- Weighting can accentuate any problems with subgroup data.
- **Remember:** bias doesn't always have a technical solution!

# The Final Case Weighting Example from the Lecture

An unweighted data set might contain more beans grown in wet conditions that are not Sira variety. To counteract that imbalance, we might weight each case, or row, perhaps with the inverse of its frequency.

| Moisture | Variety | Weight |
|---|---|---|
| Dry | Other | 1/1=1.0 |
| Dry | Sira | 1/1=1.0 |
| Wet | Other | 1/5=0.2 |
| Wet | Other | 1/5=0.2 |
| Wet | Other | 1/5=0.2 |
| Wet | Sira | 1/1=1.0 |
| Wet | Other | 1/5=0.2 |
| Wet | Other | 1/5=0.2 |