

Data slicing

Jeffrey Leek Johns Hopkins Bloomberg School of Public Health

SPAM Example: Data splitting

```
[1] 3451 58
```

SPAM Example: K-fold

```
folds[[1]][1:10]
```

```
[1] 1 2 3 4 5 6 7 8 9 10
```

SPAM Example: Return test

```
Fold01 Fold02 Fold03 Fold04 Fold05 Fold06 Fold07 Fold08 Fold09 Fold10 460 461 460 459 461 459 460 460 461 460
```

```
folds[[1]][1:10]
```

```
[1] 24 27 32 40 41 43 55 58 63 68
```

SPAM Example: Resampling

```
folds[[1]][1:10]
```

```
[1] 1 2 3 3 5 5 7 8 12
```

SPAM Example: Time Slices

```
[1] "train" "test"
```

```
folds$train[[1]]
```

```
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

```
folds$test[[1]]
```

Further information

- · Caret tutorials:
 - http://www.edii.uclm.es/~useR-2013/Tutorials/kuhn/user_caret_2up.pdf
 - http://cran.r-project.org/web/packages/caret/vignettes/caret.pdf
- · A paper introducing the caret package
 - http://www.jstatsoft.org/v28/i05/paper