Data analysis for _____

Sasha D. Hafner

31 January, 2022

Summary

RAD: `dfsumm()` gives a nice summary. The `summary()` function is an alternative.

```
dfsumm(dat)
```

```
##
##  23 rows and 15 columns
##  23 unique rows
##                          id        sample       box thick.samp  temp.c  rh.tar
## Class               integer     character character    numeric integer integer
## Minimum                 124 C 2009JULY16A         A       0.15       5      70
## Maximum                 170  C2009JUNE30D         D       0.15      35      70
## Mean                    143          <NA>      <NA>       0.15    19.3      70
## Unique (excld. NA)       23            23         4          1       3       1
## Missing values            0             0         0          0       0       0
## Sorted                 TRUE         FALSE     FALSE       TRUE   FALSE    TRUE
##
##                     speed.tar headspace   speed      dm c.etoh.i   rho.d
## Class                 numeric   numeric numeric numeric  numeric numeric
## Minimum                  0.05      0.01   0.042    31.8     1970     184
## Maximum                     5       0.1    5.07    35.8     3930     306
## Mean                     1.56    0.0397    1.58    33.7     3050     272
## Unique (excld. NA)          3         3      23      23       23      23
## Missing values              0         0       0       0        0       0
## Sorted                  FALSE     FALSE   FALSE   FALSE    FALSE   FALSE
##
##                       por.g  emis.t  emis.n
## Class               numeric numeric numeric
## Minimum               0.213   0.216   0.016
## Maximum               0.54    6.53   0.501
## Mean                  0.29    1.57   0.143
## Unique (excld. NA)      23      23      23
## Missing values           0       0       0
## Sorted               FALSE   FALSE   FALSE
##
```

RAD: Check replication or balance

```
table(dat$temp.c, dat$speed.tar)
```

```
##
##      0.05 0.5 5
##   5     2   3 2
```

```
##   20    2   6 2
##   35    2   2 2
```

Plots

RAD: It is possble to include plots here, and certainly useful for data exploration prior to fitting a model.
RAD: But doing so will slow down creation of the pdf.

Regression analysis

```
mod1 <- lm(log10(emis.n) ~ temp.c + speed, data = dat)
summary(mod1)
```

```
##
## Call:
## lm(formula = log10(emis.n) ~ temp.c + speed, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.26529 -0.14522 -0.06907  0.09646  0.53691
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.634108   0.101140 -16.157 6.07e-13 ***
## temp.c       0.015997   0.004265   3.751  0.00126 **
## speed        0.162903   0.023502   6.932 9.91e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2301 on 20 degrees of freedom
## Multiple R-squared:  0.763,  Adjusted R-squared:  0.7393
## F-statistic:  32.2 on 2 and 20 DF,  p-value: 5.589e-07
```

```
dat$pred1 <- predict(mod1)
dat$resid1 <- resid(mod1)
```