Regression and ANOVA for Professor ratings

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(readr)

theme_set(theme_bw(base_size = 16))
```

Load in professor rating data

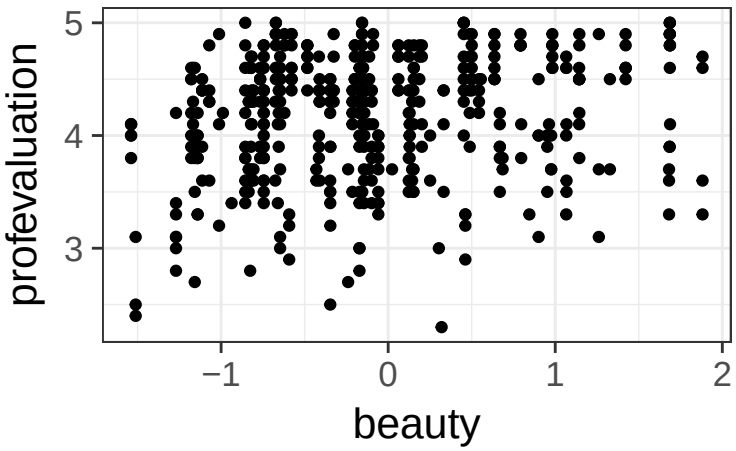
```
prof_data <- read.csv("https://dyurovsky.github.io/psyc20100/data/demos/prof_data.csv")

head(prof_data)
```

```
##   X minority age profevaluation formal native     beauty lower gender
## 1 1      yes  36            4.7     no native  0.2015666    no female
## 2 2       no  59            4.6     no native -0.8260813    no   male
## 3 3       no  51            4.1     no native -0.6603327    no   male
## 4 4       no  40            4.5     no native -0.7663125    no female
## 5 5       no  31            4.8     no native  1.4214450    no female
## 6 6       no  62            4.4    yes native  0.5002196    no   male
##   students       tenure
## 1       43 tenure track
## 2       20      tenured
## 3       55      tenured
## 4       46      tenured
## 5       48 tenure track
## 6      282      tenured
```

Scatter plot of beauty vs. professor evaluation

```
qplot(data = prof_data, x = beauty, y = profevaluation)
```



Full model

```
prof_lm <- lm(profevaluation ~ beauty + gender + age + formal + lower +
                native + minority + students + tenure, data = prof_data)

summary(prof_lm)
```

```
##
## Call:
## lm(formula = profevaluation ~ beauty + gender + age + formal +
##     lower + native + minority + students + tenure, data = prof_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.79845 -0.37270  0.09849  0.39052  0.93273
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          4.6282155  0.1720227  26.905  < 2e-16 ***
## beauty               0.1079530  0.0329357   3.278 0.001127 **
## gendermale           0.2040127  0.0527509   3.867 0.000126 ***
## age                 -0.0089405  0.0032458  -2.755 0.006115 **
## formalyes            0.1511348  0.0749453   2.017 0.044328 *
## loweryes             0.0581603  0.0553270   1.051 0.293723
## nativenonnative     -0.2157998  0.1146764  -1.882 0.060503 .
## minorityyes         -0.0706677  0.0762621  -0.927 0.354607
## students            -0.0003726  0.0003603  -1.034 0.301536
## tenuretenure track -0.1932547  0.0846549  -2.283 0.022903 *
## tenuretenured       -0.1574315  0.0655919  -2.400 0.016791 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5206 on 452 degrees of freedom
## Multiple R-squared:  0.1037, Adjusted R-squared:  0.0839
## F-statistic: 5.231 on 10 and 452 DF,  p-value: 2.748e-07
```

Stepwise regression

```
lm_null <- lm(profevaluation ~ 1, data = prof_data)

lm_all <- lm(profevaluation ~ ., data = prof_data)
```

```r
# Backwards stepwise using defaults
lm_step <- step(lm_all)
```

```
## Start:  AIC=-592.4
## profevaluation ~ X + minority + age + formal + native + beauty +
##     lower + gender + students + tenure
##
##            Df Sum of Sq    RSS     AIC
## - lower     1    0.1406 122.43 -593.86
## - X         1    0.1871 122.48 -593.69
## - minority  1    0.2223 122.52 -593.56
## - students  1    0.2623 122.56 -593.40
## <none>                  122.29 -592.40
## - native    1    1.0038 123.30 -590.61
## - formal    1    1.0989 123.39 -590.25
## - tenure    2    2.0098 124.30 -588.85
## - age       1    2.1826 124.47 -586.21
## - beauty    1    2.8862 125.18 -583.60
## - gender    1    4.0279 126.32 -579.39
##
## Step:  AIC=-593.86
## profevaluation ~ X + minority + age + formal + native + beauty +
##     gender + students + tenure
##
##            Df Sum of Sq    RSS     AIC
## - minority  1    0.1631 122.60 -595.25
## - students  1    0.1866 122.62 -595.16
## - X         1    0.3460 122.78 -594.56
## <none>                  122.43 -593.86
## - formal    1    1.0472 123.48 -591.92
## - native    1    1.1527 123.59 -591.53
## - tenure    2    2.2373 124.67 -589.48
## - age       1    2.3070 124.74 -587.22
## - beauty    1    2.8668 125.30 -585.15
## - gender    1    4.2024 126.64 -580.24
##
## Step:  AIC=-595.25
## profevaluation ~ X + age + formal + native + beauty + gender +
##     students + tenure
##
##            Df Sum of Sq    RSS     AIC
## - students  1    0.1727 122.77 -596.60
## - X         1    0.3255 122.92 -596.02
## <none>                  122.60 -595.25
## - formal    1    1.1098 123.71 -593.08
## - native    1    1.5040 124.10 -591.60
## - tenure    2    2.3212 124.92 -590.56
## - age       1    2.2926 124.89 -588.67
## - beauty    1    2.8489 125.44 -586.61
## - gender    1    4.3613 126.96 -581.06
##
## Step:  AIC=-596.6
## profevaluation ~ X + age + formal + native + beauty + gender +
```

```
##     tenure
##
##           Df Sum of Sq    RSS     AIC
## - X        1    0.3144 123.08 -597.41
## <none>                  122.77 -596.60
## - formal   1    0.9515 123.72 -595.02
## - native   1    1.3870 124.16 -593.39
## - tenure   2    2.4247 125.19 -591.54
## - age      1    2.1842 124.95 -590.43
## - beauty   1    2.7691 125.54 -588.27
## - gender   1    4.2231 126.99 -582.94
##
## Step:  AIC=-597.41
## profevaluation ~ age + formal + native + beauty + gender + tenure
##
##           Df Sum of Sq    RSS     AIC
## <none>                  123.08 -597.41
## - formal   1    0.9307 124.01 -595.92
## - native   1    1.3696 124.45 -594.29
## - tenure   2    2.4496 125.53 -592.29
## - age      1    2.0412 125.12 -591.80
## - beauty   1    2.7960 125.88 -589.01
## - gender   1    4.3401 127.42 -583.37
```

```r
# Forward stepwise starting from null model
lm_forward <- step(lm_null, scope = list(lower = lm_null, upper = lm_all), direction = "forward")
```

```
## Start:  AIC=-562.99
## profevaluation ~ 1
##
##             Df Sum of Sq    RSS     AIC
## + beauty     1    4.5938 132.06 -576.82
## + gender     1    2.2602 134.39 -568.71
## + native     1    1.6023 135.05 -566.45
## + age        1    1.5655 135.09 -566.32
## + tenure     2    1.5891 135.06 -564.40
## + lower      1    0.9575 135.70 -564.24
## + minority   1    0.7857 135.87 -563.66
## <none>                   136.65 -562.99
## + X          1    0.2576 136.40 -561.86
## + formal     1    0.1959 136.46 -561.65
## + students   1    0.0922 136.56 -561.30
##
## Step:  AIC=-576.82
## profevaluation ~ beauty
##
##             Df Sum of Sq    RSS     AIC
## + gender     1    3.1935 128.87 -586.15
## + native     1    1.6588 130.40 -580.67
## + tenure     2    1.5450 130.52 -578.27
## + minority   1    0.9173 131.14 -578.05
## + lower      1    0.8266 131.23 -577.73
## <none>                   132.06 -576.82
## + age        1    0.4120 131.65 -576.26
## + X          1    0.2933 131.77 -575.85
```

```
## + formal    1    0.0498 132.01 -574.99
## + students  1    0.0082 132.05 -574.85
##
## Step:  AIC=-586.15
## profevaluation ~ beauty + gender
##
##            Df Sum of Sq    RSS     AIC
## + native    1   1.64722 127.22 -590.11
## + age       1   1.32050 127.55 -588.92
## + tenure    2   1.86166 127.00 -588.89
## + lower     1   0.64210 128.22 -586.46
## + minority  1   0.58276 128.28 -586.25
## <none>                  128.87 -586.15
## + X         1   0.19539 128.67 -584.85
## + students  1   0.02592 128.84 -584.25
## + formal    1   0.00176 128.87 -584.16
##
## Step:  AIC=-590.11
## profevaluation ~ beauty + gender + native
##
##            Df Sum of Sq    RSS     AIC
## + age       1   1.31525 125.91 -592.92
## + tenure    2   1.60691 125.61 -591.99
## <none>                  127.22 -590.11
## + lower     1   0.38814 126.83 -589.52
## + X         1   0.22238 127.00 -588.92
## + minority  1   0.16356 127.06 -588.70
## + formal    1   0.11902 127.10 -588.54
## + students  1   0.07599 127.14 -588.39
##
## Step:  AIC=-592.92
## profevaluation ~ beauty + gender + native + age
##
##            Df Sum of Sq    RSS     AIC
## + tenure    2   1.89037 124.01 -595.92
## <none>                  125.91 -592.92
## + formal    1   0.37147 125.53 -592.29
## + X         1   0.31820 125.59 -592.09
## + lower     1   0.25385 125.65 -591.85
## + minority  1   0.24235 125.66 -591.81
## + students  1   0.08879 125.82 -591.25
##
## Step:  AIC=-595.92
## profevaluation ~ beauty + gender + native + age + tenure
##
##            Df Sum of Sq    RSS     AIC
## + formal    1   0.93072 123.08 -597.41
## <none>                  124.01 -595.92
## + X         1   0.29359 123.72 -595.02
## + minority  1   0.19473 123.82 -594.65
## + lower     1   0.10857 123.91 -594.33
## + students  1   0.01221 124.00 -593.97
##
## Step:  AIC=-597.41
```

```
## profevaluation ~ beauty + gender + native + age + tenure + formal
##
##            Df Sum of Sq    RSS     AIC
## <none>                  123.08 -597.41
## + X         1   0.31440 122.77 -596.60
## + students  1   0.16154 122.92 -596.02
## + minority  1   0.13036 122.95 -595.90
## + lower     1   0.12349 122.96 -595.88
```

```
# Backward stepwise without defaults
lm_backward <- step(lm_all, scope = list(lower = lm_null, upper = lm_all), direction = "backward")
```

```
## Start:  AIC=-592.4
## profevaluation ~ X + minority + age + formal + native + beauty +
##     lower + gender + students + tenure
##
##            Df Sum of Sq    RSS     AIC
## - lower     1    0.1406 122.43 -593.86
## - X         1    0.1871 122.48 -593.69
## - minority  1    0.2223 122.52 -593.56
## - students  1    0.2623 122.56 -593.40
## <none>                  122.29 -592.40
## - native    1    1.0038 123.30 -590.61
## - formal    1    1.0989 123.39 -590.25
## - tenure    2    2.0098 124.30 -588.85
## - age       1    2.1826 124.47 -586.21
## - beauty    1    2.8862 125.18 -583.60
## - gender    1    4.0279 126.32 -579.39
##
## Step:  AIC=-593.86
## profevaluation ~ X + minority + age + formal + native + beauty +
##     gender + students + tenure
##
##            Df Sum of Sq    RSS     AIC
## - minority  1    0.1631 122.60 -595.25
## - students  1    0.1866 122.62 -595.16
## - X         1    0.3460 122.78 -594.56
## <none>                  122.43 -593.86
## - formal    1    1.0472 123.48 -591.92
## - native    1    1.1527 123.59 -591.53
## - tenure    2    2.2373 124.67 -589.48
## - age       1    2.3070 124.74 -587.22
## - beauty    1    2.8668 125.30 -585.15
## - gender    1    4.2024 126.64 -580.24
##
## Step:  AIC=-595.25
## profevaluation ~ X + age + formal + native + beauty + gender +
##     students + tenure
##
##            Df Sum of Sq    RSS     AIC
## - students  1    0.1727 122.77 -596.60
## - X         1    0.3255 122.92 -596.02
## <none>                  122.60 -595.25
## - formal    1    1.1098 123.71 -593.08
## - native    1    1.5040 124.10 -591.60
```

```
## - tenure    2    2.3212 124.92 -590.56
## - age       1    2.2926 124.89 -588.67
## - beauty    1    2.8489 125.44 -586.61
## - gender    1    4.3613 126.96 -581.06
##
## Step:  AIC=-596.6
## profevaluation ~ X + age + formal + native + beauty + gender +
##     tenure
##
##          Df Sum of Sq    RSS     AIC
## - X       1    0.3144 123.08 -597.41
## <none>                 122.77 -596.60
## - formal  1    0.9515 123.72 -595.02
## - native  1    1.3870 124.16 -593.39
## - tenure  2    2.4247 125.19 -591.54
## - age     1    2.1842 124.95 -590.43
## - beauty  1    2.7691 125.54 -588.27
## - gender  1    4.2231 126.99 -582.94
##
## Step:  AIC=-597.41
## profevaluation ~ age + formal + native + beauty + gender + tenure
##
##          Df Sum of Sq    RSS     AIC
## <none>                 123.08 -597.41
## - formal  1    0.9307 124.01 -595.92
## - native  1    1.3696 124.45 -594.29
## - tenure  2    2.4496 125.53 -592.29
## - age     1    2.0412 125.12 -591.80
## - beauty  1    2.7960 125.88 -589.01
## - gender  1    4.3401 127.42 -583.37
```

Doing anova in R

```
tenure_anova <- aov(profevaluation ~ tenure, data = prof_data)

summary(tenure_anova)
```

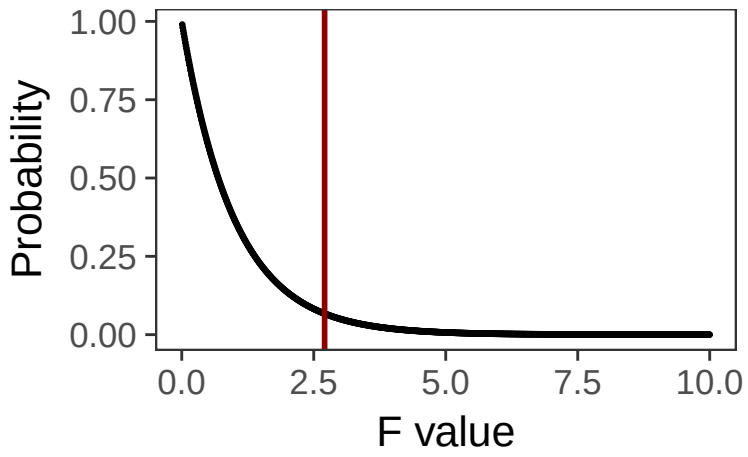
```
##              Df Sum Sq Mean Sq F value Pr(>F)
## tenure        2   1.59  0.7946   2.706 0.0679 .
## Residuals   460 135.07  0.2936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
group_means <- prof_data %>%
  select(profevaluation, tenure) %>%
  group_by(tenure) %>%
  summarise(mean = mean(profevaluation), n = n())
group_means
```

```
## # A tibble: 3 × 3
##             tenure     mean     n
##             <fctr>    <dbl> <int>
## 1 non-tenure track 4.284314   102
## 2     tenure track 4.154630   108
## 3          tenured 4.139130   253
```

```r
overall_mean <- prof_data %>%
  select(profevaluation) %>%
  summarise(mean = mean(profevaluation), n = n())

qplot(x = seq(.01,10,.01), y = df(seq(.01,10,.01), 2, 450),
      geom = "point", size = I(.5)) +
  theme(panel.grid = element_blank()) +
  xlab("F value") +
  ylab("Probability") +
  geom_vline(aes(xintercept = 2.706), color = "darkred", size = 1)
```



Comparing anova and linear regression

```r
summary(aov(profevaluation ~ tenure, data = prof_data))
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## tenure        2   1.59  0.7946   2.706 0.0679 .
## Residuals   460 135.07  0.2936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lm(profevaluation ~ tenure, data = prof_data))
```

```
##
## Call:
## lm(formula = profevaluation ~ tenure, data = prof_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8546 -0.3391  0.1157  0.4305  0.8609
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          4.28431    0.05365  79.853   <2e-16 ***
## tenuretenure track -0.12968    0.07482  -1.733   0.0837 .
## tenuretenured      -0.14518    0.06355  -2.284   0.0228 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5419 on 460 degrees of freedom
## Multiple R-squared:  0.01163,    Adjusted R-squared:  0.007332
```

```
## F-statistic: 2.706 on 2 and 460 DF,  p-value: 0.06786
```