

Lab 11 – ANCOVA

October 29 & 30, 2018
FANR 6750

Richard Chandler and Bob Cooper

Scenario

- We are interested in doing a one-way ANOVA
- However, we need to account for variation associated with a continuous predictor variable

Additive model

$$y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \bar{x}) + \varepsilon_{ij}$$

ANCOVA can be thought of as a hybrid between ANOVA and regression

ANOVA, regression, and ANCOVA are linear models

REGRESSION

ONE-WAY ANOVA

ANCOVA

2 / 19

THE DIET DATA

Import the data and view the levels of the factor

```
dietData <- read.csv("dietData.csv")
levels(dietData$diet)

## [1] "Control" "High" "Low" "Med"
```

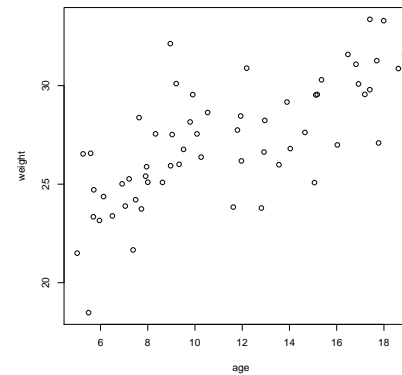
Reorder the levels of the factor, just for convenience

```
levels(dietData$diet) <- list(Control="Control", Low="Low",
                             Med="Med", High="High")
levels(dietData$diet)

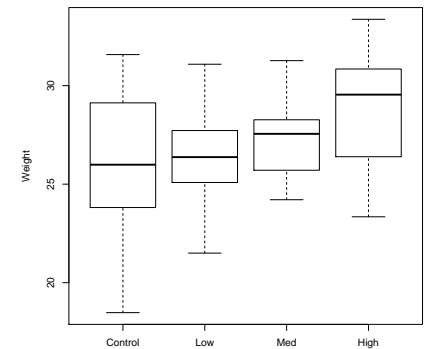
## [1] "Control" "Low" "Med" "High"
```

THE DIET DATA

```
plot(weight ~ age, dietData)
```



```
boxplot(weight ~ diet, dietData, ylab="Weight")
```



```
fm1 <- lm(weight ~ age, dietData)
summary(fm1)

##
## Call:
## lm(formula = weight ~ age, data = dietData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.6906 -1.2625  0.0522  1.0233  6.1680
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 21.32523    0.80685   26.430 < 2e-16 ***
## age         0.51807    0.06742    7.685 2.07e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.171 on 58 degrees of freedom
## Multiple R-squared:  0.5045, Adjusted R-squared:  0.496
## F-statistic: 59.05 on 1 and 58 DF, p-value: 2.072e-10
```

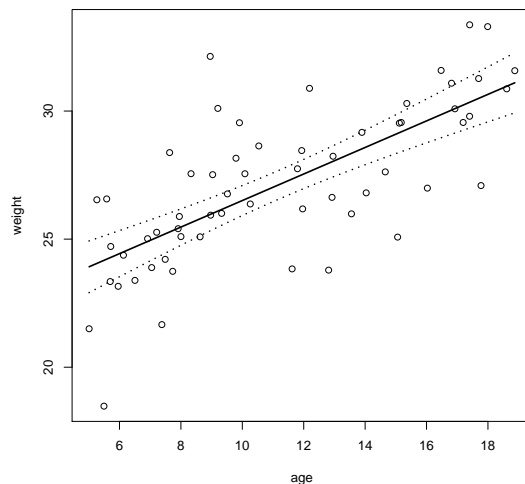
The two estimates correspond to the intercept and slope parameters

Regression lines and CIs can be created using `predict`

- (1) Create a new `data.frame` containing a sequence of values of the predictor variable `age`
- (2) Predict `weight` using these values of `age`
- (3) Put predictions and data together for easy plotting

```
age <- dietData$age
predData1 <- data.frame(age=seq(min(age), max(age), length=50))
pred1 <- predict(fm1, newdata=predData1, se.fit=TRUE,
                 interval="confidence")
predictions1 <- data.frame(pred1$fit, predData1)
```

```
plot(weight ~ age, data=dietData) # raw data
lines(fit ~ age, data=predictions1, lwd=2) # fitted line
lines(lwr ~ age, data=predictions1, lwd=2, lty=3) # lower CI
lines(upr ~ age, data=predictions1, lwd=2, lty=3) # upper CI
```



Change the `contrasts` option so that the estimates will correspond to the additive model, and then fit the ANOVA

```
options(contrasts=c("contr.sum", "contr.poly"))
fm2 <- lm(weight ~ diet, dietData)
summary.aov(fm2)

##              Df Sum Sq Mean Sq F value Pr(>F)
## diet          3   54.6  18.216   2.053  0.117
## Residuals    56  496.9   8.873
```

The `aov` function gives identical results

```
summary(aov(weight ~ diet, dietData))

##              Df Sum Sq Mean Sq F value Pr(>F)
## diet          3   54.6  18.216   2.053  0.117
## Residuals    56  496.9   8.873
```

```
summary(fm2)

##
## Call:
## lm(formula = weight ~ diet, data = dietData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.6371 -1.9253 -0.0366  1.9770  5.4576
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.13962    0.38456  70.573 <2e-16 ***
## diet1       -1.02179    0.66608  -1.534  0.131
## diet2       -0.56593    0.66608  -0.850  0.399
## diet3        0.08027    0.66608   0.121  0.905
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.979 on 56 degrees of freedom
## Multiple R-squared:  0.09908, Adjusted R-squared:  0.05082
## F-statistic: 2.053 on 3 and 56 DF,  p-value: 0.1169
```

Because we changed the `contrast` option to `contr.sum`, the intercept is the grand mean (μ) and the other estimates are the effect sizes (α_i)

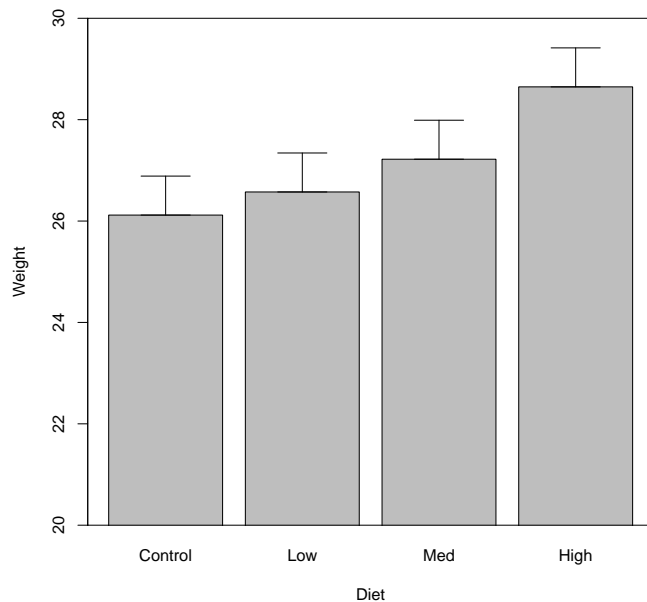
The `predict` function can also be used to obtain group means, SEs, and CIs from a one-way ANOVA

```
predData2 <- data.frame(diet=levels(dietData$diet))
pred2 <- predict(fm2, newdata=predData2,
                 se.fit=TRUE, interval="confidence")
```

```
predictions2 <- data.frame(pred2$fit, SE=pred2$se, predData2)
predictions2
```

##	fit	lwr	upr	SE	diet
## 1	26.11783	24.57710	27.65856	0.7691199	Control
## 2	26.57368	25.03295	28.11442	0.7691199	Low
## 3	27.21988	25.67915	28.76062	0.7691199	Med
## 4	28.64707	27.10634	30.18780	0.7691199	High

ONE-WAY ANOVA



ANCOVA PRELIMINARIES

Additive model

$$y_{ij} = \mu + \alpha_i + \beta(x_{ij} - \bar{x}) + \varepsilon_{ij}$$

Make sure the contrasts are set as before

```
options(contrasts=c("contr.sum", "contr.poly"))
```

Centering the covariate isn't required, but doing so allow the intercept to be interpreted as the grand mean

```
dietData$ageCentered <- dietData$age - mean(dietData$age)
```

Put the covariate before the treatment variable in the formula.

```
fm3 <- lm(weight ~ ageCentered + diet, dietData)
```

```
summary(fm3)
```

```
##
## Call:
## lm(formula = weight ~ ageCentered + diet, data = dietData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.8214 -1.2213 -0.2519  1.2161  4.9185
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.1396    0.2406  112.787 < 2e-16 ***
## ageCentered   0.5573    0.0594   9.382 5.2e-13 ***
## diet1        -1.7446    0.4238  -4.116 0.00013 ***
## diet2         -0.3758    0.4173  -0.901 0.37171
## diet3         0.7819    0.4234   1.847 0.07020 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.864 on 55 degrees of freedom
## Multiple R-squared:  0.6536, Adjusted R-squared:  0.6284
## F-statistic: 25.94 on 4 and 55 DF,  p-value: 4.147e-12
```

The null hypothesis of no diet effect is rejected, even though it was not rejected before.

```
summary.aov(fm3)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## ageCentered  1 278.25   278.25  80.095 2.54e-12 ***
## diet         3  82.22    27.41   7.889 0.000182 ***
## Residuals   55 191.07     3.47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

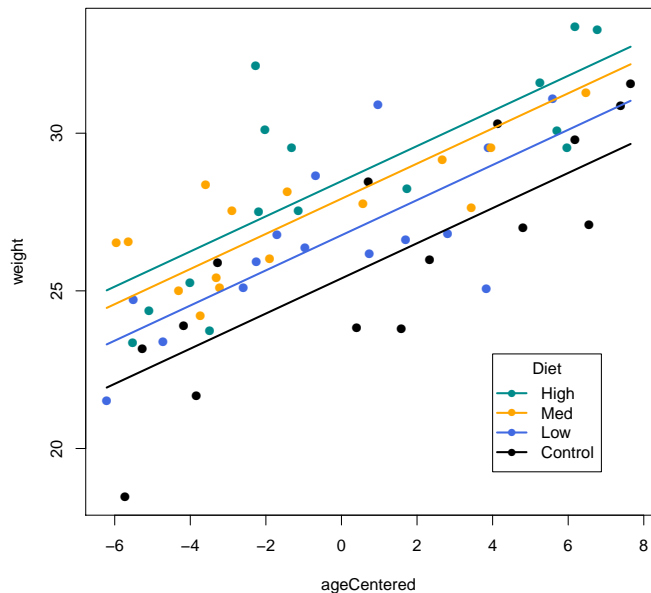
PREDICT WEIGHT

Create predictions of weight over a sequences of ages, for every level of diet

```
ageC <- dietData$ageCentered
predData3 <- data.frame(
  diet=rep(c("Control", "Low", "Med", "High"), each=20),
  ageCentered=rep(seq(min(ageC), max(ageC),
    length=20),
    times=4))
pred3 <- predict(fm3, newdata=predData3, se.fit=TRUE,
  interval="confidence")
predictions3 <- data.frame(pred3$fit, predData3)
```

PLOT THE REGRESSION LINES

```
colrs <- c("black", "royalblue", "orange", "darkcyan")
plot(weight ~ ageCentered, dietData, pch=16, cex=1.2,
  col=rep(colrs, each=15))
lines(fit ~ ageCentered, predictions3, subset=diet=="Control",
  col=colrs[1], lwd=2)
lines(fit ~ ageCentered, predictions3, subset=diet=="Low", lty=1,
  col=colrs[2], lwd=2)
lines(fit ~ ageCentered, predictions3, subset=diet=="Med", lty=1,
  col=colrs[3], lwd=2)
lines(fit ~ ageCentered, predictions3, subset=diet=="High", lty=1,
  col=colrs[4], lwd=2)
legend(4, 23, c("High", "Med", "Low", "Control"), pch=16,
  title="Diet", lwd=2, col=rev(colrs))
```



```
## install.packages("multcomp")
library(multcomp)
summary(glht(fm3, linfct=mcp(diet="Tukey")))
```

```
##
## Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = weight ~ ageCentered + diet, data = dietData)
##
## Linear Hypotheses:
##              Estimate Std. Error t value Pr(>|t|)
## Low - Control == 0    1.3688    0.6875   1.991 0.20389
## Med - Control == 0    2.5265    0.6973   3.623 0.00336 **
## High - Control == 0   3.0830    0.6832   4.513 < 0.001 ***
## Med - Low == 0        1.1577    0.6828   1.696 0.33583
## High - Low == 0       1.7143    0.6817   2.515 0.06861 .
## High - Med == 0       0.5566    0.6869   0.810 0.84931
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

ASSIGNMENT

Complete the following and upload your R script to ELC before lab next week

- (1) Fit an ANCOVA model to the data in `treeData.csv`, which represent the height of trees following a fertilizer experiment. The covariate is pH.
- (2) Use: `options(contrasts=c("contr.sum", "contr.poly"))` so that your estimates correspond to the additive model from the lecture notes
- (3) Interpret each of the estimates from `lm`. What is the null hypothesis associated with each *p*-value?
- (4) Plot the data and the regression lines. Use different colors or symbols to distinguish the treatment groups.
- (5) Which fertilizer treatments are significantly different?