Analysis of covariance

Analysis of covariance (ANCOVA) is a combination between anova and regression. It is useful when there are both continuous quantitative and categorical independent variables.

Examples of use

Performing a test of:

- Efficacy of different drugs in rats that differ in mass
- effect of time of day on zooplankton at different distances offshore.
- whether the linear relationship between two variables such as Dosage and Effect depend upon another categorical variable like sex.

Non-biological example

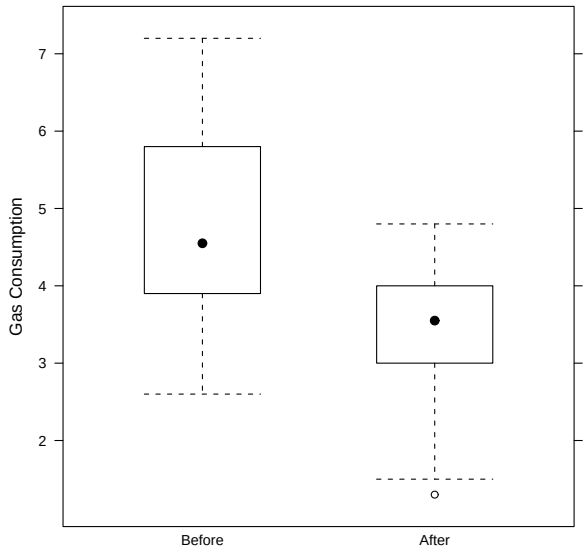
This example was derived from Venables and Ripley (1999).
The data frame 'whiteside' (in library(MASS)) contains data on energy use at a residential house over two years. In the first year the house was uninsulated. In the second the house was insulated.

```
> library(MASS)
> data(whiteside)
> summary(whiteside)
```

```
    Insul         Temp             Gas
 Before:26   Min.   :-0.800   Min.   :1.300
 After :30   1st Qu.: 3.050   1st Qu.:3.500
             Median : 4.900   Median :3.950
             Mean   : 4.875   Mean   :4.071
             3rd Qu.: 7.125   3rd Qu.:4.625
             Max.   :10.200   Max.   :7.200
```

The data

```
> library(lattice)
> print(bwplot(Gas ~ Insul, data = whiteside, ylab = "Gas Consumption"))
```

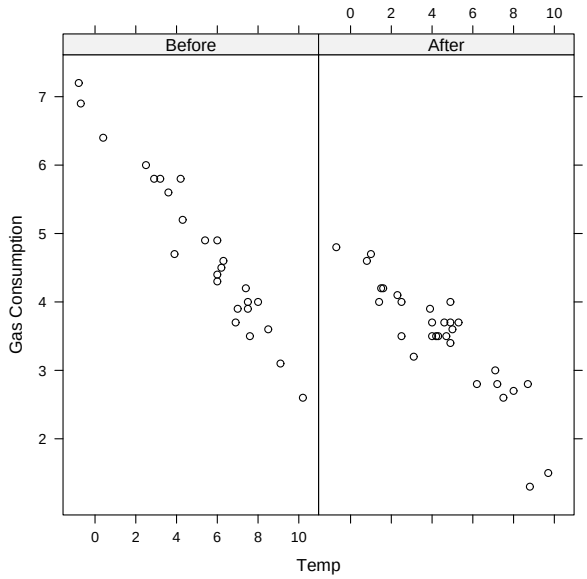


Test for time difference

```
> summary(aov(Gas ~ Insul, data = whiteside))

            Df Sum Sq Mean Sq F value    Pr(>F)
Insul        1 22.348  22.348  22.913 1.357e-05 ***
Residuals   54 52.667   0.975
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

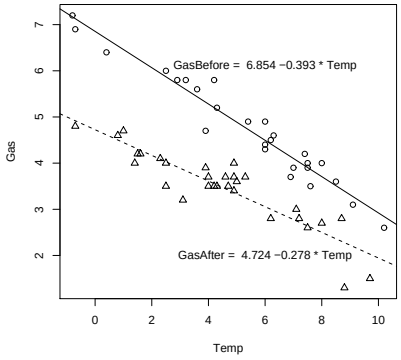
The data: view 2

```
> library(lattice)
> print(xyplot(Gas ~ Temp | Insul, data = whiteside, ylab = "Gas Consumption"))
```



Differences in slope?

```
> g1 <- lm(Gas ~ Temp, data = whiteside, subset = Insul == "Before")
> g2 <- lm(Gas ~ Temp, data = whiteside, subset = Insul != "Before")
> plot(Gas ~ Temp, data = whiteside, pch = unclass(Insul))
> abline(coef(g1), lty = 1)
> abline(coef(g2), lty = 2)
> text(6, 6.1, paste("GasBefore = ", round(coef(g1)[1], 3), round(coef(g1)[2],
+     3), "* Temp"))
> text(6, 2, paste("GasAfter = ", round(coef(g2)[1], 3), round(coef(g2)[2],
+     3), "* Temp"))
```



ANCOVA

```
> white.aov <- aov(Gas ~ Insul + Temp + Insul:Temp, data = whiteside)
> summary(white.aov)

            Df Sum Sq Mean Sq F value    Pr(>F)
Insul        1 22.348  22.348 214.198 < 2.2e-16 ***
Temp         1 45.896  45.896 439.908 < 2.2e-16 ***
Insul:Temp   1  1.345   1.345  12.893 0.0007307 ***
Residuals   52  5.425   0.104
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

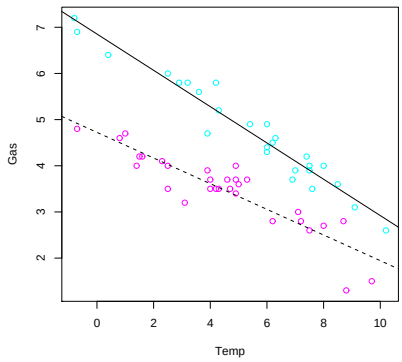
What does this mean

```
> white.lm <- lm(Gas ~ Insul + Temp + Insul:Temp, data = whiteside)
> summary(white.lm)

Call:
lm(formula = Gas ~ Insul + Temp + Insul:Temp, data = whiteside)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97802 -0.18011  0.03757  0.20930  0.63803

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     6.85383    0.13596  50.409  < 2e-16 ***
InsulAfter     -2.12998    0.18009 -11.827 2.32e-16 ***
Temp           -0.39324    0.02249 -17.487  < 2e-16 ***
InsulAfter:Temp  0.11530    0.03211   3.591  0.00073 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 0.323 on 52 degrees of freedom
Multiple R-squared: 0.9277,        Adjusted R-squared: 0.9235
F-statistic: 222.3 on 3 and 52 DF,  p-value: < 2.2e-16
```

Interpreting coefficients

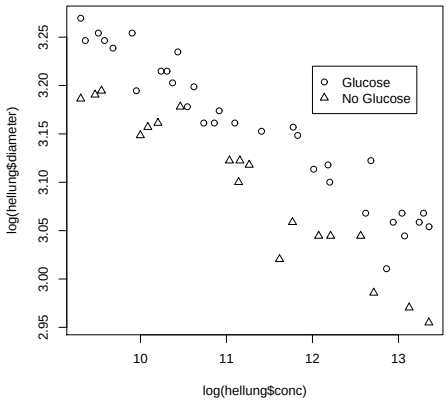
```
> coef(white.lm)

    (Intercept)      InsulAfter            Temp InsulAfter:Temp
      6.8538277      -2.1299780      -0.3932388       0.1153039
```

```
> plot(Gas ~ Temp, data = whiteside, col = (unclass(Insul) + 4))
> abline(coef(white.lm)[1], coef(white.lm)[3], lty = 1)
> abline(coef(white.lm)[1] + coef(white.lm)[2], coef(white.lm)[3] +
+     coef(white.lm)[4], lty = 2)
```



Biological example (Dalgaard; 2002)

```
> library(ISwR)
> data(hellung)
> plot(log(hellung$diameter) ~ log(hellung$conc), pch = (hellung$glucose))
> legend(list(x = 12, y = 3.22), legend = c("Glucose", "No Glucose"),
+     pch = c(1:2))
```



ANCOVA

Here 'conc' will be the covariate responsible for variation removed from the test of glucose effects.

```
> glu.lm <- lm(log(diameter) ~ log(conc) * glucose, data = hellung)
> summary(glu.lm)

Call:
lm(formula = log(diameter) ~ log(conc) * glucose, data = hellung)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0615296 -0.0112544  0.0001289  0.0086745  0.0405433

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        3.748438   0.077720  48.230  < 2e-16 ***
log(conc)         -0.046716   0.006846  -6.823 1.51e-08 ***
glucose            0.007869   0.054559   0.144    0.886
log(conc):glucose -0.006480   0.004821  -1.344    0.185
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 0.02086 on 47 degrees of freedom
Multiple R-squared: 0.9361,        Adjusted R-squared: 0.9321
F-statistic: 229.6 on 3 and 47 DF,  p-value: < 2.2e-16
```

ANCOVA 2, parallel lines

```
> glu.lm2 <- lm(log(diameter) ~ log(conc) + glucose, data = hellung)
> summary(glu.lm2)

Call:
lm(formula = log(diameter) ~ log(conc) + glucose, data = hellung)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0581233 -0.0132006 -0.0004489  0.0112696  0.0435502

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.846170   0.027700  138.85  < 2e-16 ***
log(conc)   -0.055393   0.002301  -24.07  < 2e-16 ***
glucose     -0.065020   0.006095  -10.67 2.93e-14 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

```
Residual standard error: 0.02103 on 48 degrees of freedom
Multiple R-squared: 0.9337,        Adjusted R-squared: 0.9309
F-statistic: 337.9 on 2 and 48 DF,  p-value: < 2.2e-16
```

ANOVA table

```
> glu.aov <- aov(log(diameter) ~ log(conc) * glucose, data = hellung)
> summary(glu.aov)

                  Df   Sum Sq  Mean Sq F value    Pr(>F)
log(conc)          1 0.248609 0.248609 571.436 < 2.2e-16 ***
glucose            1 0.050335 0.050335 115.698  2.89e-14 ***
log(conc):glucose  1 0.000786 0.000786   1.807    0.1853
Residuals         47 0.020448 0.000435
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

Conclusions

Analysis of covariance

- Provides a framework to determine if slopes differ in two groups
- Provides a framework to test for an effect of grouping category given some additional continuous variation

References

Dalgaard, P. (2002). *Introductory Statistics with R*, Academic Press, Inc Springer. book goes with R package ISwR.

Venables, W. N. and Ripley, B. D. (1999). *Modern Applied Statistics with S*, 4th edn, Springer, New York.