Lesson 06

Christopher A. Swenson (chris@cswenson.com)

11/27/2021

Heart attacks in rabbits

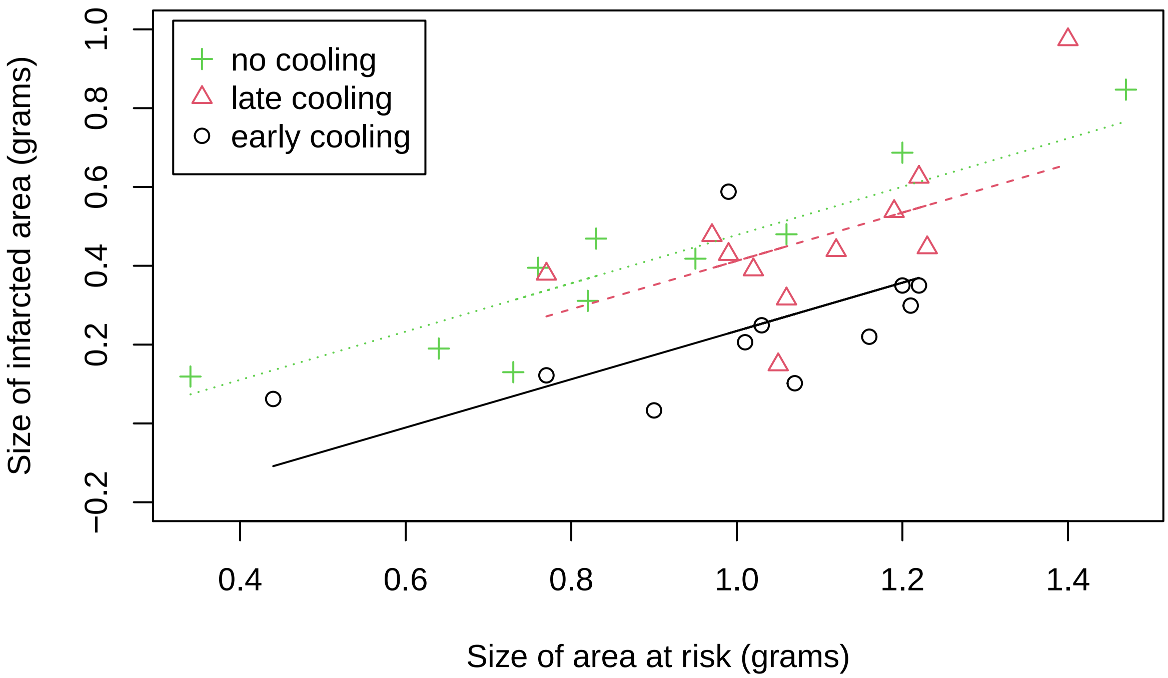
Load the coolhearts data. Fit a multiple linear regression model of Infarc on Area, X2 (early cooling), and X3 (late cooling). Display model results. Create a scatterplot of the data with points marked by group and three lines representing the fitted regression equation for each group.

```
coolhearts <- read.table("./Data/coolhearts.txt", header=T)
attach(coolhearts)

model.1 <- lm(Infarc ~ Area + X2 + X3)
summary(model.1)
```

```
##
## Call:
## lm(formula = Infarc ~ Area + X2 + X3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29410 -0.06511 -0.01329  0.07855  0.35949
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.13454    0.10402  -1.293 0.206459
## Area         0.61265    0.10705   5.723 3.87e-06 ***
## X2          -0.24348    0.06229  -3.909 0.000536 ***
## X3          -0.06566    0.06507  -1.009 0.321602
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1395 on 28 degrees of freedom
## Multiple R-squared:  0.6377, Adjusted R-squared:  0.5989
## F-statistic: 16.43 on 3 and 28 DF,  p-value: 2.363e-06
```

```
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) -0.13454    0.10402  -1.293 0.206459
# Area         0.61265    0.10705   5.723 3.87e-06 ***
# X2          -0.24348    0.06229  -3.909 0.000536 ***
# X3          -0.06566    0.06507  -1.009 0.321602

plot(Area, Infarc, type="n", ylim=c(-0.2, 1),
     xlab="Size of area at risk (grams)",
     ylab="Size of infarcted area (grams)")
for (i in 1:32) points(Area[i], Infarc[i], pch=Group[i], col=Group[i])
for (i in 1:3) lines(Area[Group==i], fitted(model.1)[Group==i], lty=i, col=i)
```

```r
legend("topleft", legend=c("no cooling",
                           "late cooling",
                           "early cooling"),
       col=3:1, pch=3:1, inset=0.02)
```

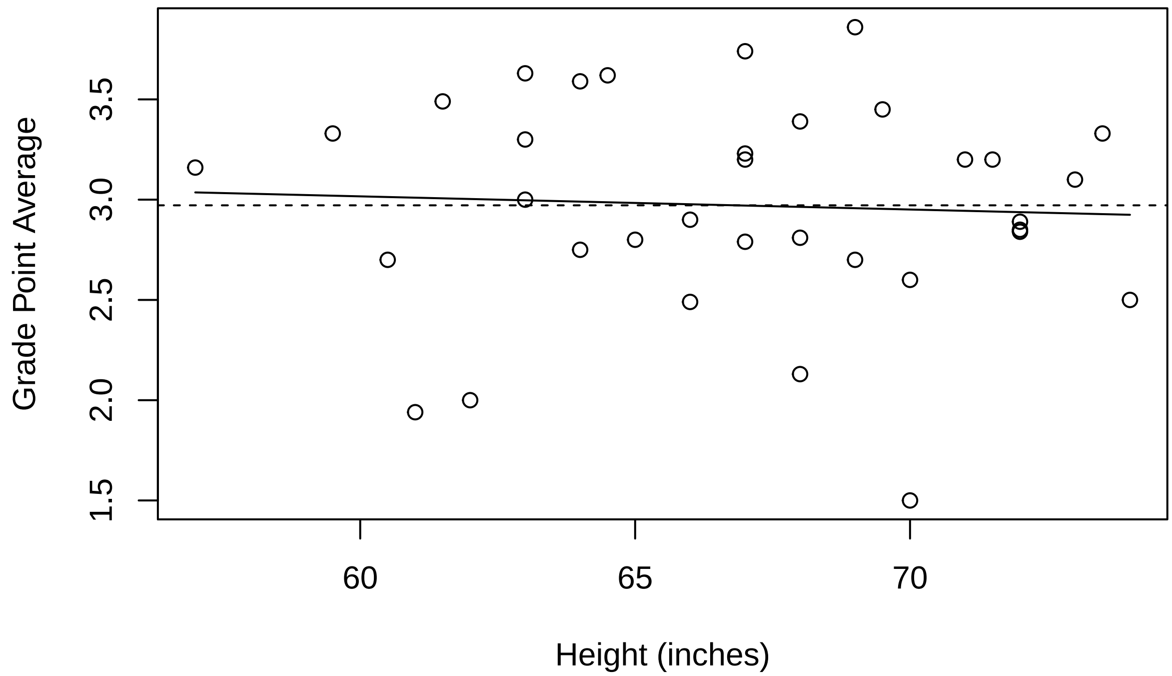



```r
detach(coolhearts)
```

Student heights and GPAs

Load the heightgpa data. Fit a simple linear regression model of gpa on height. Create a scatterplot of the data with a fitted simple linear regression line and a horizontal line at the mean of gpa. Calculate SSE for the full and reduced models. Calculate the overall F statistic by hand. Find the p-value for the overall F statistic. Display the overall F statistic and p-value in the model results.

```r
heightgpa <- read.table("./Data/heightgpa.txt", header=T)
attach(heightgpa)

model <- lm(gpa ~ height)

plot(height,gpa,xlab="Height (inches)",ylab="Grade Point Average",
     panel.last = c(lines(sort(height), fitted(model)[order(height)]),
                    abline(h=mean(gpa), lty=2)))
```




```
sum(residuals(model)^2) # SSE_F = 9.705507
```

```
## [1] 9.705507
```

```
mean(gpa) # 2.971714
```

```
## [1] 2.971714
```

```
sum((gpa-mean(gpa))^2) # SSE_R = 9.733097
```

```
## [1] 9.733097
```

```
((9.733097-9.705507)/1) / (9.705507/33) # overall F-statistic = 0.09380963
```

```
## [1] 0.09380963
```

```
pf(0.09380963, 1, 33, lower.tail=F) # p-value = 0.7613129
```

```
## [1] 0.7613129
```

```
summary(model) # F-statistic: 0.09381 on 1 and 33 DF,  p-value: 0.7613
```

```
##
## Call:
## lm(formula = gpa ~ height)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45081 -0.24878  0.00325  0.35622  0.90263
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.410214   1.434616   2.377   0.0234 *
## height      -0.006563   0.021428  -0.306   0.7613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

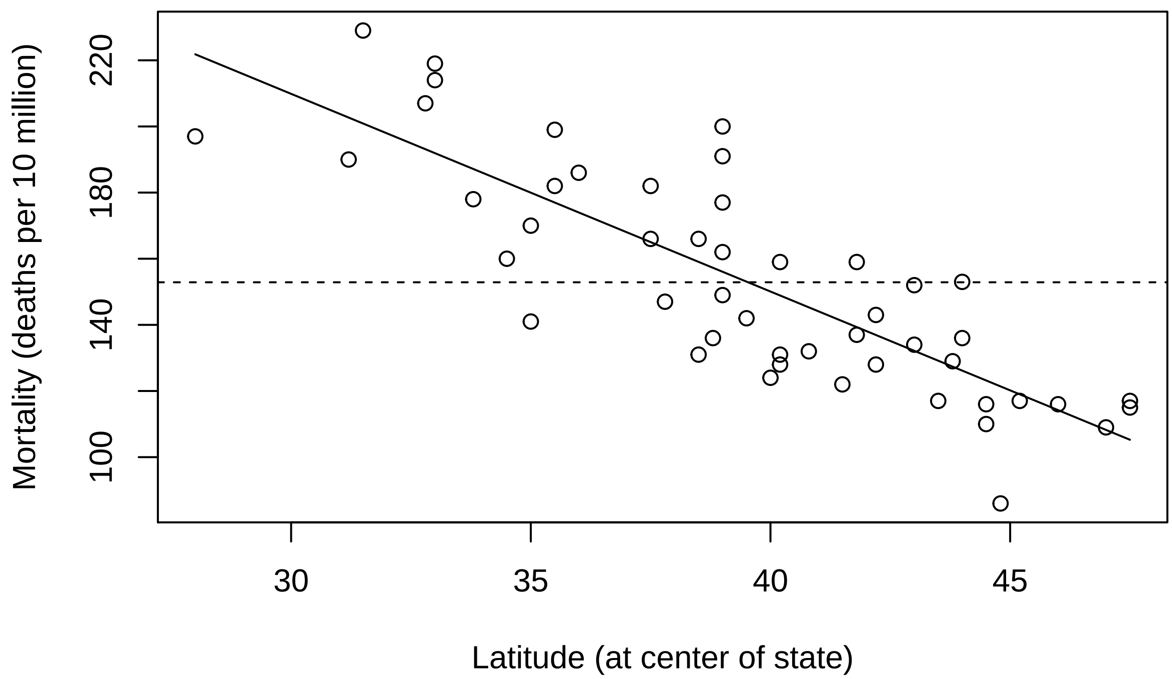
```
##
## Residual standard error: 0.5423 on 33 degrees of freedom
## Multiple R-squared:  0.002835,	Adjusted R-squared:  -0.02738
## F-statistic: 0.09381 on 1 and 33 DF,  p-value: 0.7613
```

```
detach(heightgpa)
```

Skin cancer mortality

Load the skincancer data. Fit a simple linear regression model of Mort on Lat. Create a scatterplot of the data with a fitted simple linear regression line and a horizontal line at the mean of Mort. Calculate SSE for the full and reduced models. Calculate the overall F statistic by hand. Find the p-value for the overall F statistic. Display the overall F statistic and p-value in the model results.

```
skincancer <- read.table("./Data/skincancer.txt", header=T)
attach(skincancer)

model <- lm(Mort ~ Lat)

plot(Lat,Mort,xlab="Latitude (at center of state)",
     ylab="Mortality (deaths per 10 million)",
     panel.last = c(lines(sort(Lat), fitted(model)[order(Lat)]),
                    abline(h=mean(Mort), lty=2)))
```




```
sum(residuals(model)^2) # SSE_F = 17173.07
```

```
## [1] 17173.07
```

```
mean(Mort) # 152.8776
```

```
## [1] 152.8776
```

```
sum((Mort-mean(Mort))^2) # SSE_R = 53637.27
```

```
## [1] 53637.27
```

```
((53637.27-17173.07)/1) / (17173.07/47) # overall F-statistic = 99.7968
```

```
## [1] 99.7968
```

```
pf(99.7968, 1, 47, lower.tail=F) # p-value = 3.309471e-13
```

```
## [1] 3.309471e-13
```

```
summary(model) # F-statistic:  99.8 on 1 and 47 DF,  p-value: 3.309e-13
```

```
##
## Call:
## lm(formula = Mort ~ Lat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.972 -13.185   0.972  12.006  43.938
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 389.1894    23.8123   16.34  < 2e-16 ***
## Lat          -5.9776     0.5984   -9.99 3.31e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.12 on 47 degrees of freedom
## Multiple R-squared:  0.6798, Adjusted R-squared:  0.673
## F-statistic:  99.8 on 1 and 47 DF,  p-value: 3.309e-13
```

```
detach(skincancer)
```

Alcoholism and muscle strength

Load the alcoholarm data. Fit a simple linear regression model of strength on alcohol. Calculate SSE for the full and reduced models. Calculate the overall F statistic by hand. Display the overall F statistic and p-value in the model results and in the anova table.

```
alcoholarm <- read.table("./Data/alcoholarm.txt", header=T)
attach(alcoholarm)

model <- lm(strength ~ alcohol)
sum((strength-mean(strength))^2) # SSE_R = 1224.315
```

```
## [1] 1224.315
```

```
sum(residuals(model)^2) # SSE_F = 720.2749
```

```
## [1] 720.2749
```

```
((1224.315-720.2749)/1) / (720.2749/48) # F = 33.58985
```

```
## [1] 33.58985
```

```
summary(model) # F-statistic: 33.59 on 1 and 48 DF,  p-value: 5.136e-07
```

```
##
## Call:
## lm(formula = strength ~ alcohol)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7847 -2.5450 -0.1477  2.6359  7.4815
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.36954    1.20273  21.925  < 2e-16 ***
## alcohol     -0.29587    0.05105  -5.796 5.14e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.874 on 48 degrees of freedom
## Multiple R-squared:  0.4117, Adjusted R-squared:  0.3994
## F-statistic: 33.59 on 1 and 48 DF,  p-value: 5.136e-07
```

```
anova(model)
```

```
## Analysis of Variance Table
##
## Response: strength
##           Df Sum Sq Mean Sq F value    Pr(>F)
## alcohol    1 504.04  504.04   33.59 5.136e-07 ***
## Residuals 48 720.27   15.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Analysis of Variance Table
# Response: strength
#           Df Sum Sq Mean Sq F value    Pr(>F)
# alcohol    1 504.04  504.04   33.59 5.136e-07 ***
# Residuals 48 720.27   15.01
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

detach(alcoholarm)
```

Allen cognitive level study

Load the allentest data. Calculate SSTO. Fit SLR model of ACL on Vocab and display anova table (with sequential sums of squares). Fit MLR model of ACL on Vocab and SDMT and display anova table (with sequential sums of squares). Calculate SSR(Vocab, SDMT) by hand using sequential sums of squares. Fit SLR model of ACL on SDMT and display anova table (with sequential sums of squares). Fit MLR model of ACL on SDMT and Vocab and display anova table (with sequential sums of squares). Calculate SSR(Vocab, SDMT) by hand using sequential sums of squares. Fit MLR model of ACL on SDMT, Vocab, and Abstract and display anova table (with sequential sums of squares). Calculate SSR(Vocab, Abstract | SDMT) by hand using sequential sums of squares.

```
allentest <- read.table("./Data/allentest.txt", header=T)
attach(allentest)

sum((ACL-mean(ACL))^2) # SSTO = 43.04957
```

```
## [1] 43.04957
```

```
model.1 <- lm(ACL ~ Vocab)
anova(model.1)
```

```
## Analysis of Variance Table
##
## Response: ACL
##           Df Sum Sq Mean Sq F value  Pr(>F)
## Vocab      1  2.691 2.69060  4.4667 0.03829 *
## Residuals 67 40.359 0.60237
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Analysis of Variance Table
# Response: ACL
#           Df Sum Sq Mean Sq F value  Pr(>F)
# Vocab      1  2.691 2.69060  4.4667 0.03829 *
# Residuals 67 40.359 0.60237

model.13 <- lm(ACL ~ Vocab + SDMT)
anova(model.13) # Sequential (type I) SS
```

```
## Analysis of Variance Table
##
## Response: ACL
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## Vocab      1  2.6906  2.6906  5.6786  0.02006 *
## SDMT       1  9.0872  9.0872 19.1789 4.35e-05 ***
## Residuals 66 31.2717  0.4738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df  Sum Sq Mean Sq F value   Pr(>F)
# Vocab      1  2.6906  2.6906  5.6786  0.02006 *
# SDMT       1  9.0872  9.0872 19.1789 4.35e-05 ***
# Residuals 66 31.2717  0.4738
# Calculate by hand: SSR(Vocab, SDMT) = 2.6906 + 9.0872 = 11.7778

model.3 <- lm(ACL ~ SDMT)
anova(model.3)
```

```
## Analysis of Variance Table
##
## Response: ACL
##           Df Sum Sq Mean Sq F value    Pr(>F)
## SDMT       1  11.68 11.6799  24.946 4.468e-06 ***
## Residuals 67  31.37  0.4682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#           Df Sum Sq Mean Sq F value    Pr(>F)
# SDMT       1  11.68 11.6799  24.946 4.468e-06 ***
# Residuals 67  31.37  0.4682

model.31 <- lm(ACL ~ SDMT + Vocab)
anova(model.31) # Sequential (type I) SS
```

```
## Analysis of Variance Table
##
## Response: ACL
```

```
##            Df  Sum Sq Mean Sq F value   Pr(>F)
## SDMT        1 11.6799 11.6799 24.6508 5.12e-06 ***
## Vocab       1  0.0979  0.0979  0.2067   0.6508
## Residuals 66 31.2717  0.4738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#            Df  Sum Sq Mean Sq F value   Pr(>F)
# SDMT        1 11.6799 11.6799 24.6508 5.12e-06 ***
# Vocab       1  0.0979  0.0979  0.2067   0.6508
# Residuals 66 31.2717  0.4738
# Calculate by hand: SSR(Vocab, SDMT) = 11.6799 + 0.0979 = 11.7778

model.312 <- lm(ACL ~ SDMT + Vocab + Abstract)
anova(model.312) # Sequential (type I) SS
```

```
## Analysis of Variance Table
##
## Response: ACL
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## SDMT        1 11.6799 11.6799 24.6902 5.173e-06 ***
## Vocab       1  0.0979  0.0979  0.2070    0.6506
## Abstract    1  0.5230  0.5230  1.1056    0.2969
## Residuals 65 30.7487  0.4731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#            Df  Sum Sq Mean Sq F value    Pr(>F)
# SDMT        1 11.6799 11.6799 24.6902 5.173e-06 ***
# Vocab       1  0.0979  0.0979  0.2070    0.6506
# Abstract    1  0.5230  0.5230  1.1056    0.2969
# Residuals 65 30.7487  0.4731
# Calculate by hand: SSR(Vocab, Abstract | SDMT) = 0.0979 + 0.5230 = 0.6209

detach(allentest)
```

Heart attacks in rabbits (revisited)

Load the coolhearts data. Fit a multiple linear regression model of Infarc on Area, X2 (early cooling), and X3 (late cooling). Test all slope parameters equal 0. Display the overall F statistic and p-value in the model results. Test one slope parameter is 0. Use the Anova function from the car package to display F-statistic in anova table using adjusted (type III) sums of squares. Or (easier), use t-test from model results. Test a subset of slope parameters is 0. Fit full model (with Area, X2 , and X3) and reduced model (Area only), and calculate general linear F-statistic. Or, use the anova function with full model to display anova table with sequential (type I) sums of squares, and calculate partial F-statistic. Or (easier), use the anova function with full and reduced models to display F-statistic and p-value directly. Calculate partial R-squared for (X2, X3 | Area).

```
library(car)
```

```
## Loading required package: carData
```

```
coolhearts <- read.table("./Data/coolhearts.txt", header=T)
attach(coolhearts)

model.1 <- lm(Infarc ~ Area + X2 + X3)
```

```r
summary(model.1)
```

```
##
## Call:
## lm(formula = Infarc ~ Area + X2 + X3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29410 -0.06511 -0.01329  0.07855  0.35949
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.13454    0.10402  -1.293 0.206459
## Area         0.61265    0.10705   5.723 3.87e-06 ***
## X2          -0.24348    0.06229  -3.909 0.000536 ***
## X3          -0.06566    0.06507  -1.009 0.321602
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1395 on 28 degrees of freedom
## Multiple R-squared:  0.6377, Adjusted R-squared:  0.5989
## F-statistic: 16.43 on 3 and 28 DF,  p-value: 2.363e-06
```

```r
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) -0.13454    0.10402  -1.293 0.206459
# Area         0.61265    0.10705   5.723 3.87e-06 ***
# X2          -0.24348    0.06229  -3.909 0.000536 ***
# X3          -0.06566    0.06507  -1.009 0.321602
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 0.1395 on 28 degrees of freedom
# Multiple R-squared:  0.6377,  Adjusted R-squared:  0.5989
# F-statistic: 16.43 on 3 and 28 DF,  p-value: 2.363e-06

anova(model.1) # Sequential (type I) SS
```

```
## Analysis of Variance Table
##
## Response: Infarc
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Area       1 0.62492 0.62492 32.1115 4.504e-06 ***
## X2         1 0.31453 0.31453 16.1622  0.000398 ***
## X3         1 0.01981 0.01981  1.0181  0.321602
## Residuals 28 0.54491 0.01946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Analysis of Variance Table
# Response: Infarc
#           Df  Sum Sq Mean Sq F value    Pr(>F)
# Area       1 0.62492 0.62492 32.1115 4.504e-06 ***
# X2         1 0.31453 0.31453 16.1622  0.000398 ***
# X3         1 0.01981 0.01981  1.0181  0.321602
# Residuals 28 0.54491 0.01946
```

```r
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova(model.1, type="III") # Adjusted (type III) SS
```

```
## Anova Table (Type III tests)
##
## Response: Infarc
##              Sum Sq Df F value    Pr(>F)
## (Intercept) 0.03255  1  1.6728 0.2064588
## Area        0.63742  1 32.7536 3.865e-06 ***
## X2          0.29733  1 15.2781 0.0005365 ***
## X3          0.01981  1  1.0181 0.3216018
## Residuals   0.54491 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Anova Table (Type III tests)
# Response: Infarc
#                  Sum Sq Df F value    Pr(>F)
# (Intercept) 0.03255  1  1.6728 0.2064588
# Area        0.63742  1 32.7536 3.865e-06 ***
# X2          0.29733  1 15.2781 0.0005365 ***
# X3          0.01981  1  1.0181 0.3216018
# Residuals   0.54491 28
# ---
# Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
# Hand calculation: F for Area = (0.63742/1) / (0.54491/28) = 32.75359

pf(32.75359, 1, 28, lower.tail=F) # 3.865451e-06
```

```
## [1] 3.865451e-06
```

```r
5.723^2 # 32.75273 (t-statistic squared = F-statistic)
```

```
## [1] 32.75273
```

```r
model.2 <- lm(Infarc ~ Area)
anova(model.2)
```

```
## Analysis of Variance Table
##
## Response: Infarc
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## Area       1 0.62492 0.62492  21.322 6.844e-05 ***
## Residuals 30 0.87926 0.02931
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#           Df  Sum Sq Mean Sq F value    Pr(>F)
# Area       1 0.62492 0.62492  21.322 6.844e-05 ***
# Residuals 30 0.87926 0.02931

((0.87926-0.54491)/(30-28)) / (0.54491/28) # General linear F-stat = 8.590226
```

```
## [1] 8.590226
```

```r
((0.31453+0.01981)/2) / (0.54491/28) # Partial F-stat = 8.589969
```

```
## [1] 8.589969
```

```r
pf(8.59, 2, 28, lower.tail=F) # 0.001233006
```

```
## [1] 0.001233006
```

```r
anova(model.2, model.1)
```

```
## Analysis of Variance Table
##
## Model 1: Infarc ~ Area
## Model 2: Infarc ~ Area + X2 + X3
##   Res.Df     RSS Df Sum of Sq      F   Pr(>F)
## 1     30 0.87926
## 2     28 0.54491  2   0.33435 8.5902 0.001233 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Analysis of Variance Table
# Model 1: Infarc ~ Area
# Model 2: Infarc ~ Area + X2 + X3
#   Res.Df     RSS Df Sum of Sq      F   Pr(>F)
# 1     30 0.87926
# 2     28 0.54491  2   0.33435 8.5902 0.001233 **
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

0.31453+0.01981 # SSR(X2, X3 | Area) = 0.33434
```

```
## [1] 0.33434
```

```r
# SSE(Area) = 0.87926
0.33434 / 0.87926 # Partial R-squared (X2, X3 | Area) = 0.3802516
```

```
## [1] 0.3802516
```

```r
detach(coolhearts)
```