ANCOVA

Klinkenberg

19 oct 2017

Inhoud

- ancova 1
  - define the model
  - group means
  - model fit no covar
  - model fit only covar
  - model fit with covar
  - total variance
  - total variance visual
  - model variance group visual
  - model variance covariate visual
  - model variance group and covariate visual
  - error variance with covariate
  - sums of squares
  - f ratio
  - adjusted means

ANCOVA

ANCOVA

Analysis of covariance (ANCOVA) is a general linear model which blends ANOVA and regression. ANCOVA evaluates whether population means of a dependent variable (DV) are equal across levels of a categorical independent variable (IV) often called a treatment, while statistically controlling for the effects of other continuous variables that are not of primary interest, known as covariates (CV).

WIKIPEDIA

ANCOVA

Determine main effect while correcting for covariate

- 1 dependent variable
- 1 or more independent variables
- 1 or more covariates

A covariate is a variable can be a confounding variable biasing your results. By adding a covariate we reduce error/residual in the model.

Assumptions

- Same as ANOVA
- Independence of the covariate and treatment effect §12.3.1.
  - No difference on ANOVA with covar and independent variable
  - Matching experimental groups on the covariate
- Homogeneity of regression slopes §12.3.2.
  - Visual: scatterplot dep var * covar per condition
  - Testing: interaction indep. var * covar

Data example

We want to test the difference in national extraversion but want to controll for openness to experience.

- Dependent variable: Extraversion
- Independent variabele: Nationality
    - Dutch
    - German
    - Belgian
- Covariate: Openness to experience

Simulate data

```
# Simulate data
n = 20
k = 3
group       = round(runif(n,1,k),0)
mu.covar    = 8
sigma.covar = 1
covar       = round(rnorm(n,mu.covar,sigma.covar),2)

# Create dummy variables
dummy.1 <- ifelse(group == 2, 1, 0)
dummy.2 <- ifelse(group == 3, 1, 0)

# Set parameters
b.0 = 15 # initial value for group 1
b.1 = 3  # difference between group 1 and 2
b.2 = 4  # difference between group 1 and 3
b.3 = 3  # Weight for covariate

# Create error
error = rnorm(n,0,1)
```

Define the model

$$outcome = model + error$$

$$model = indvar + covariate = nationality + openness$$

Formal model

$$b_0 + b_1 dummy_1 + b_2 dummy_2 + b_3 covar$$

```
# Define model
outcome = b.0 + b.1 * dummy.1 + b.2 * dummy.2 + b.3 * covar + error
```

Dummies

The data

Group means

```
aggregate(outcome ~ group, data, mean)
```

```
##   group  outcome
## 1     1 38.67000
## 2     2 42.16625
## 3     3 43.79800
```

Model fit no covar

What are the beta coëfficients based on the data without the covariate?

```
fit.group <- lm(outcome ~ factor(group), data); fit.group
```

```
##
## Call:
## lm(formula = outcome ~ factor(group), data = data)
##
## Coefficients:
##    (Intercept)  factor(group)2  factor(group)3
##         38.670           3.496           5.128
```

```
fit.group$coefficients[2:3] + fit.group$coefficients[1]
```

```
## factor(group)2 factor(group)3
##       42.16625       43.79800
```


$$Dutch = 38.67 \; German = 42.16625 \; Belgian = 43.798$$

Model fit only covar

What is the weight of only the covariate?

```
fit.covar <- lm(outcome ~ covar, data)
fit.covar
```

```
##
## Call:
## lm(formula = outcome ~ covar, data = data)
##
## Coefficients:
## (Intercept)        covar
##      15.667        3.185
```

Model fit with covar

```
fit <- lm(outcome ~ factor(group) + covar, data); fit
```

```
##
## Call:
## lm(formula = outcome ~ factor(group) + covar, data = data)
##
## Coefficients:
##    (Intercept)  factor(group)2  factor(group)3           covar
##         15.965           2.769           4.181           2.881
```

```
fit$coefficients[2:3] + fit$coefficients[1]
```

```
## factor(group)2 factor(group)3
##       18.73401       20.14609
```

$$Dutch = 15.96 \; German = 18.73 \; Belgian = 20.14$$

Total variance

What is the total variance

$$MS_{total} = s^2_{outcome} = \frac{SS_{outcome}}{df_{outcome}}$$

```
ms.t = var(data$outcome); ms.t
```

```
## [1] 11.97756
```

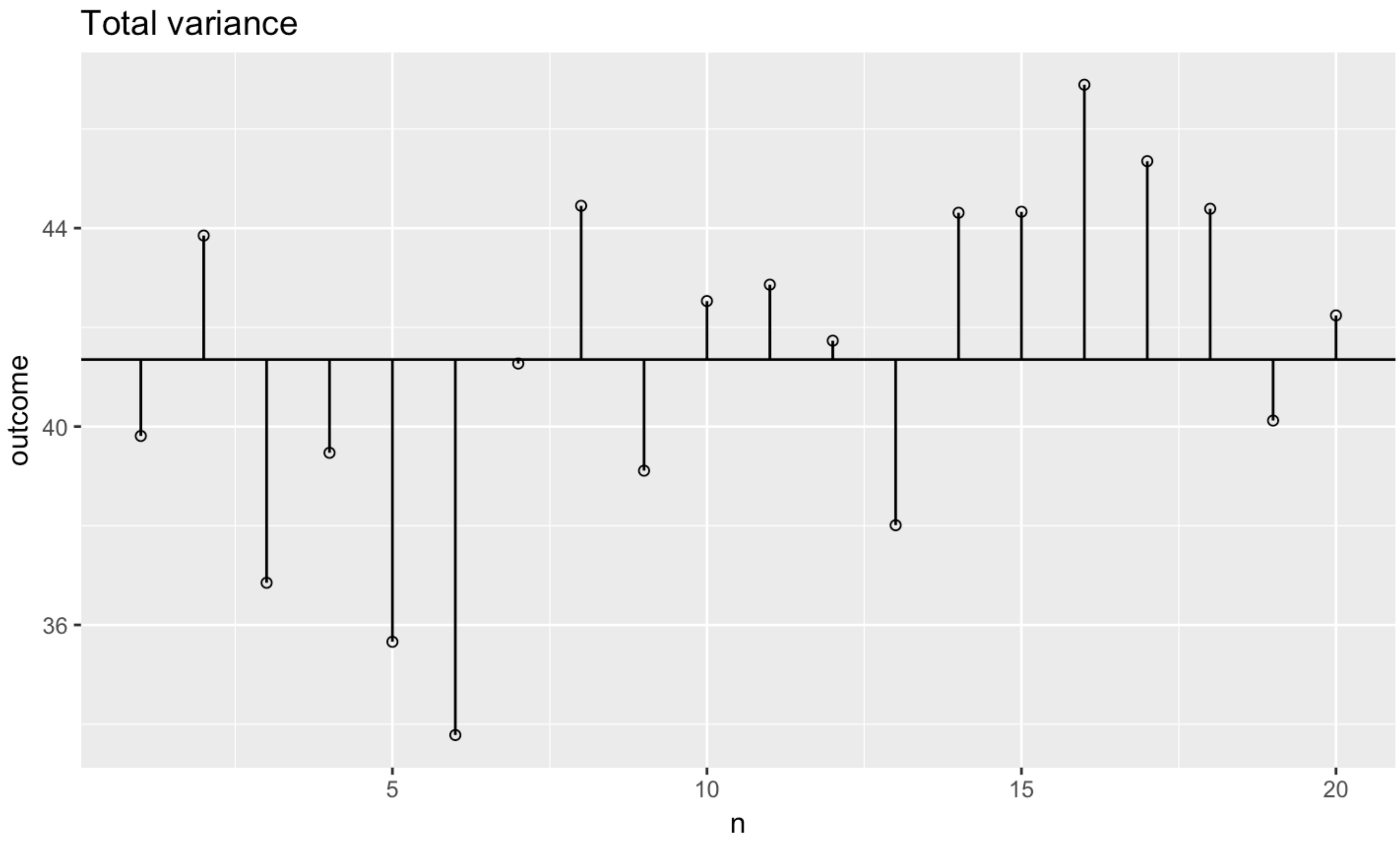
```
ss.t = var(data$outcome) * (length(data$outcome) - 1); ss.t
```

```
## [1] 227.5737
```

The data

Total variance visual

```r
plot <- ggplot(data, aes(x=n, y=outcome)) + geom_point(shape=1) +
        geom_hline(yintercept=mean(outcome)) +
        geom_segment(aes(x = n, y = grand.mean, xend = n, yend = outcome)) +
        ggtitle("Total variance")
plot
```

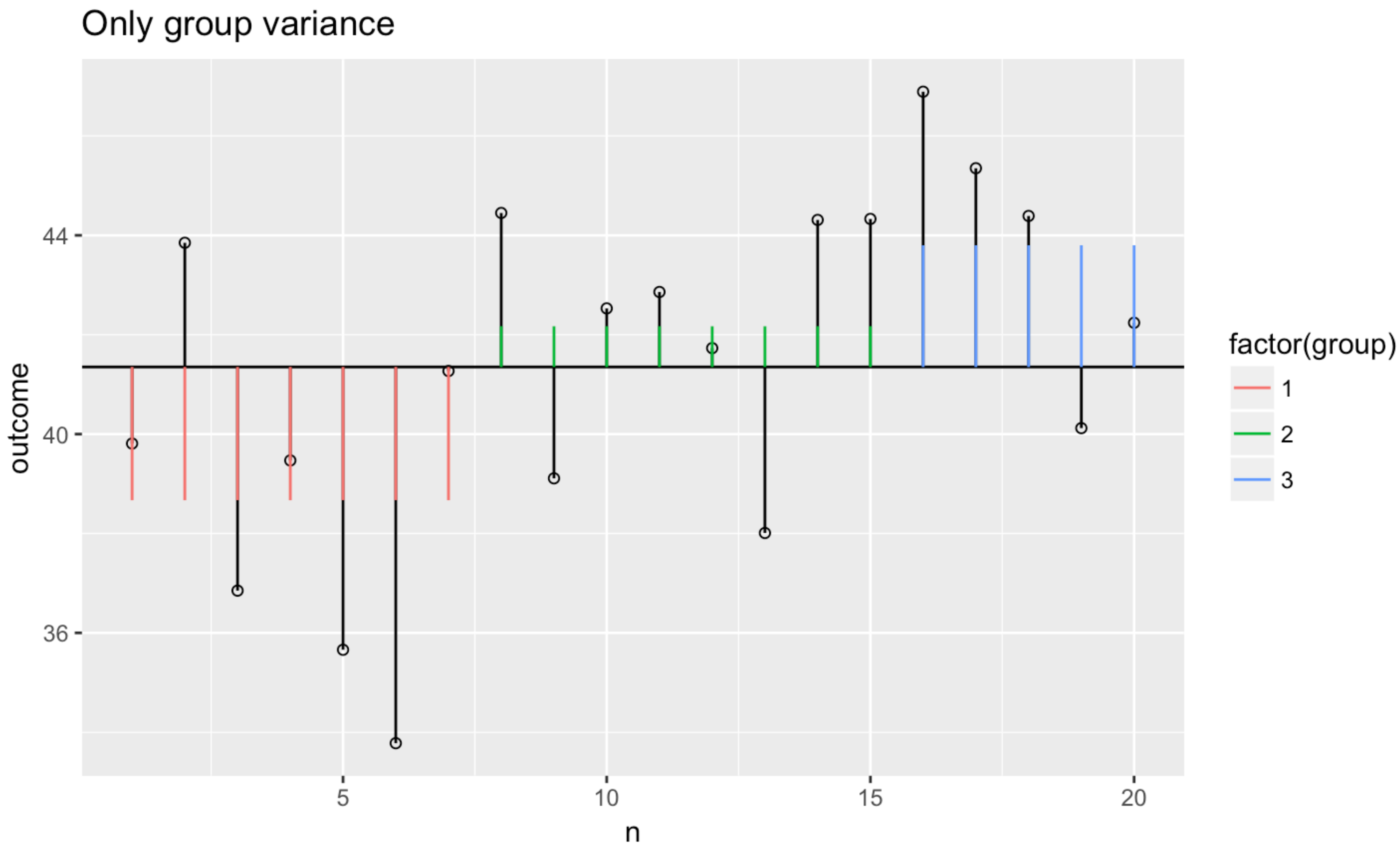

Total variance
outcome
44
40
36
5
10
15
20
n

Model variance group

The model variance consists of two parts. One for the independent variable and one for the covariate. Lets first look at the independent variable.

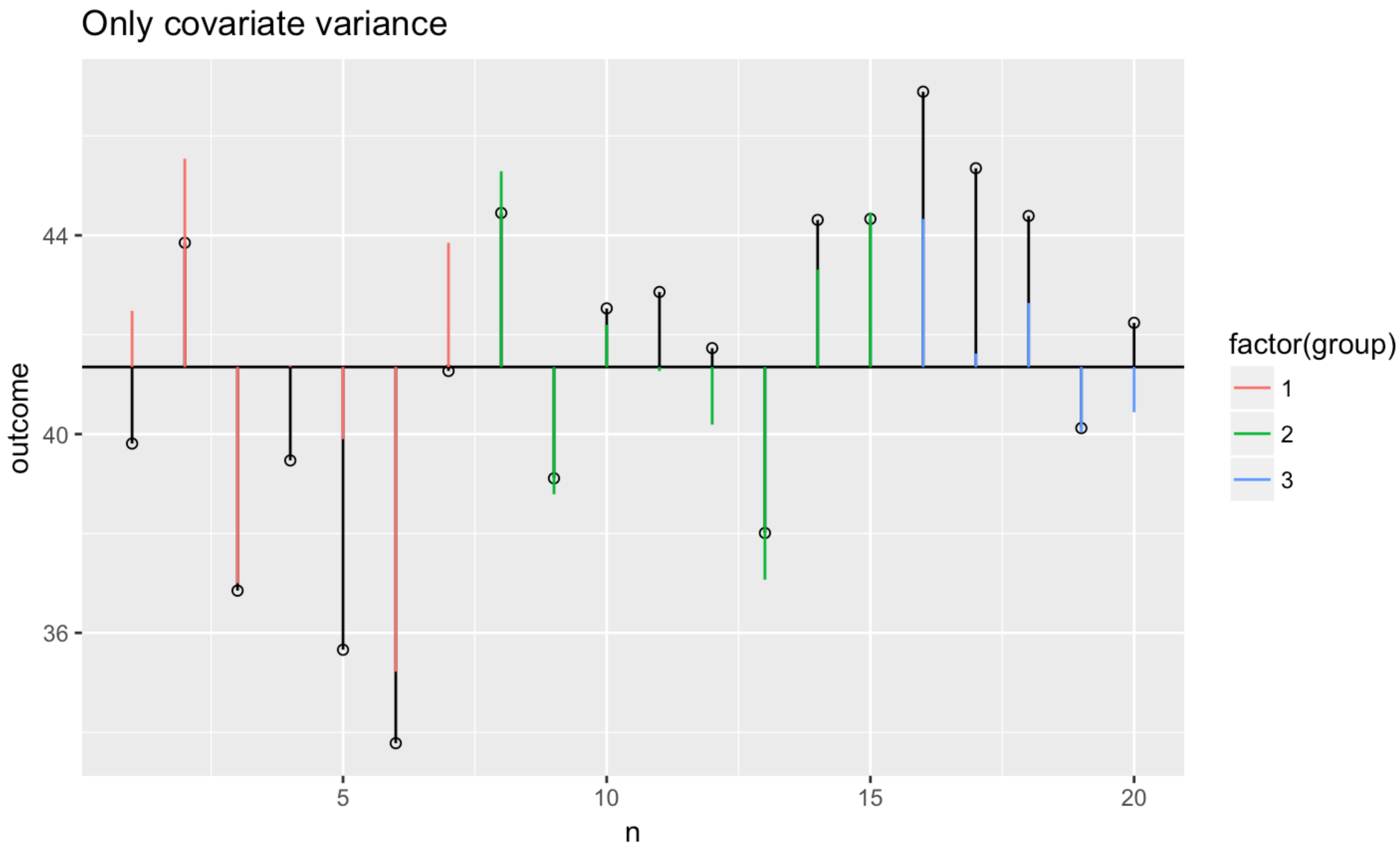
Model variance group visual

```
plot + geom_segment(aes(x = n, y = grand.mean, xend = n, yend = data$model.gro
       ggtitle("Only group variance")
```


Only group variance
outcome
44
40
36
n
5
10
15
20
factor(group)
1
2
3

Model variance covariate visual

```
data$model.covar <- round(fit.covar$fitted.values,2)
plot + geom_segment(aes(x = n, y = grand.mean, xend = n, yend = data$model.cov
       ggtitle("Only covariate variance")
```

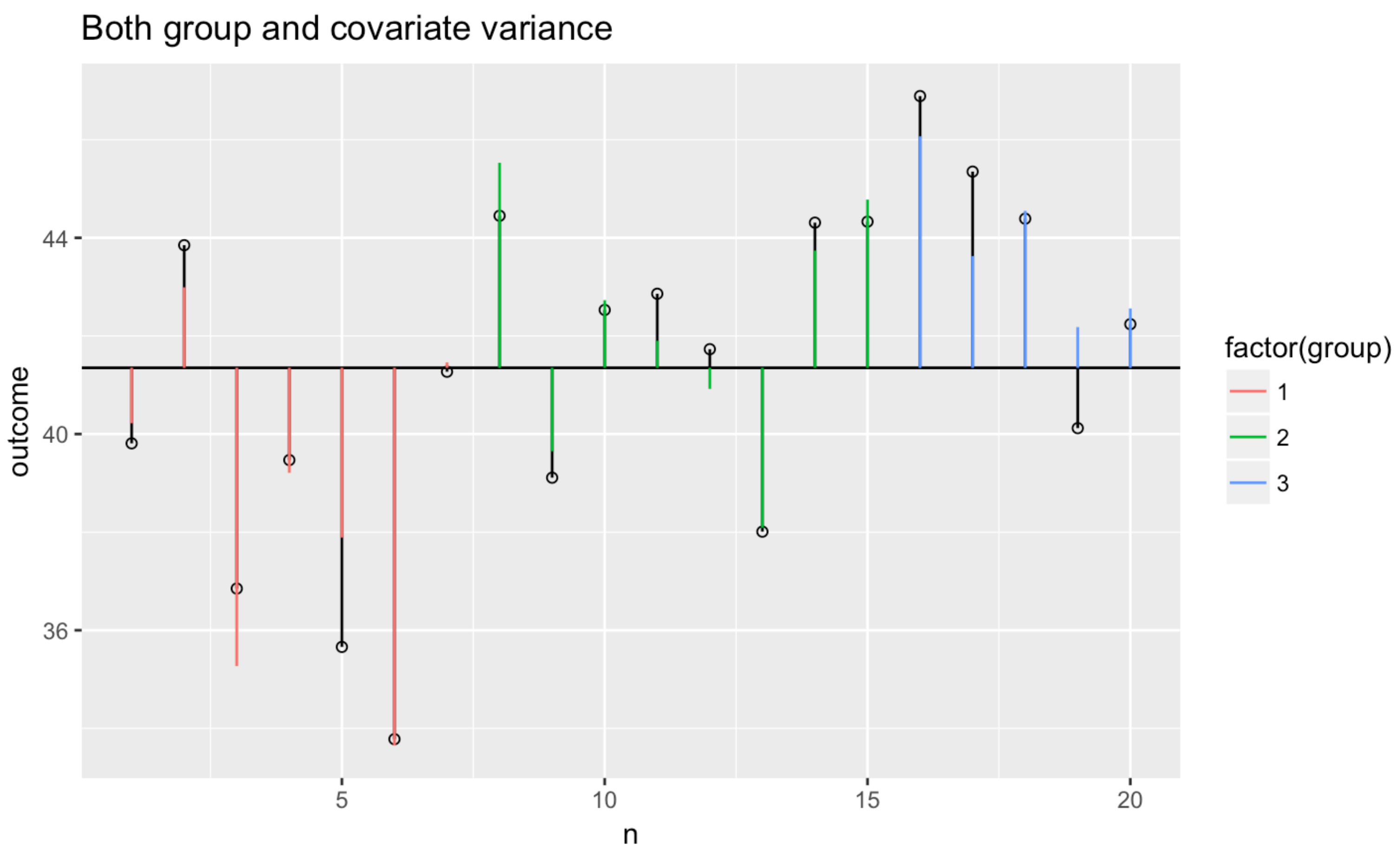


Only covariate variance
outcome
44
40
36
5
10
15
20
n
factor(group)
1
2
3

Model variance group and covariate

Model variance group and covariate visual

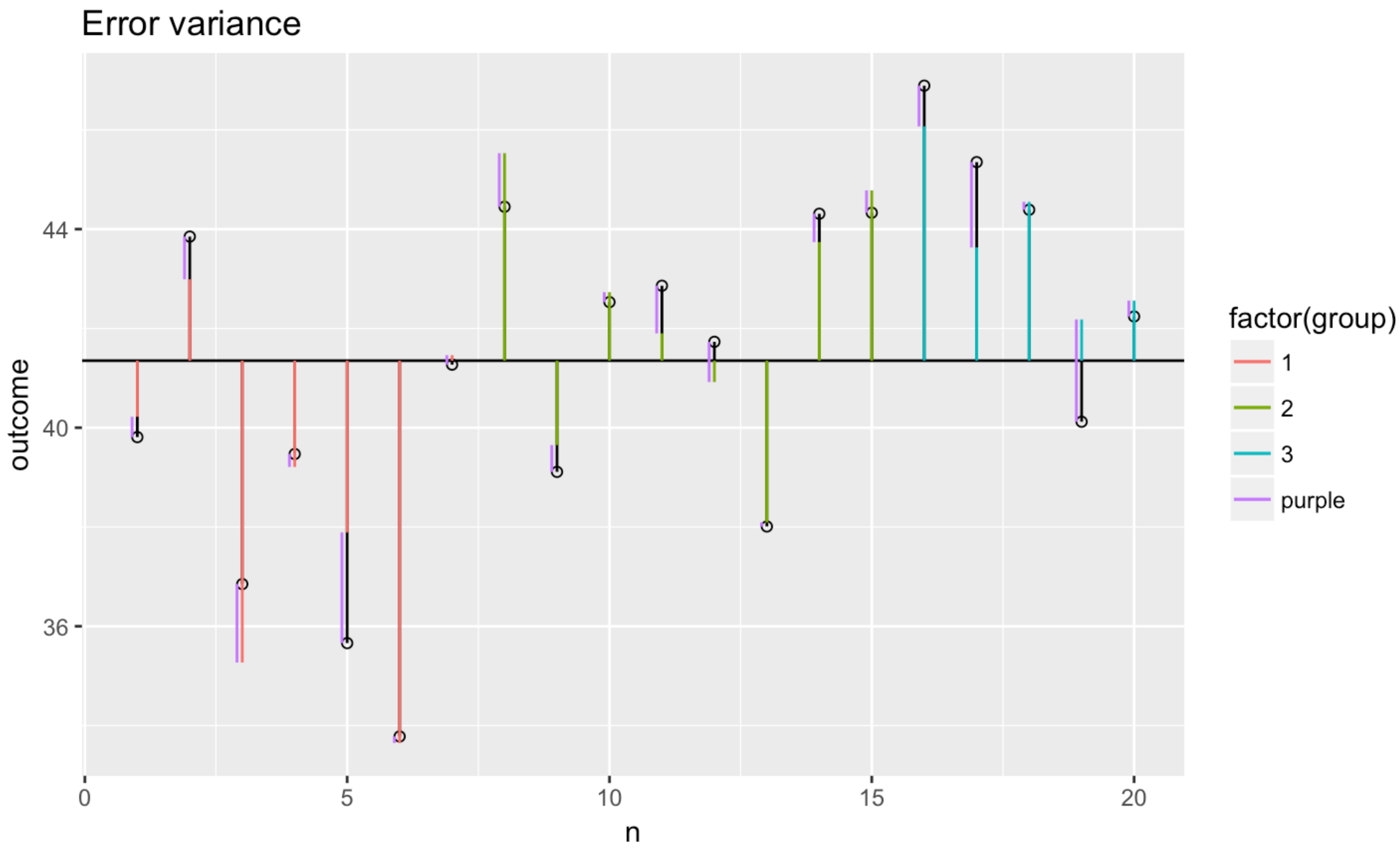
```
plot + geom_segment(aes(x = n, y = grand.mean, xend = n, yend = data$model, co

plot + ggtitle("Both group and covariate variance")
```



Both group and covariate variance
outcome
44
40
36
n
5
10
15
20
factor(group)
1
2
3

Error variance with covariate

```
plot + geom_segment(aes(x       = n-.1,
                        y       = outcome,
                        xend    = n-.1,
                        yend    = data$model,
                        colour = "purple")) +
       ggtitle("Error variance")
```



Error variance
outcome
44
40
36
0
5
10
15
20
n
factor(group)
1
2
3
purple

Sums of squares

```
SS.model = with(data, sum((model - grand.mean)^2))
SS.error = with(data, sum((outcome - model)^2))

# Sums of squares for individual effects
SS.model.group = with(data, sum((model.group - grand.mean)^2))
SS.model.covar = with(data, sum((model.covar - grand.mean)^2))

SS.covar = SS.model - SS.model.group; SS.covar ## SS.covar corrected for group
```

```
## [1] 121.8463
```

```
SS.group = SS.model - SS.model.covar; SS.group ## SS.group corrected for covar
```

```
## [1] 54.65778
```

F-ratio

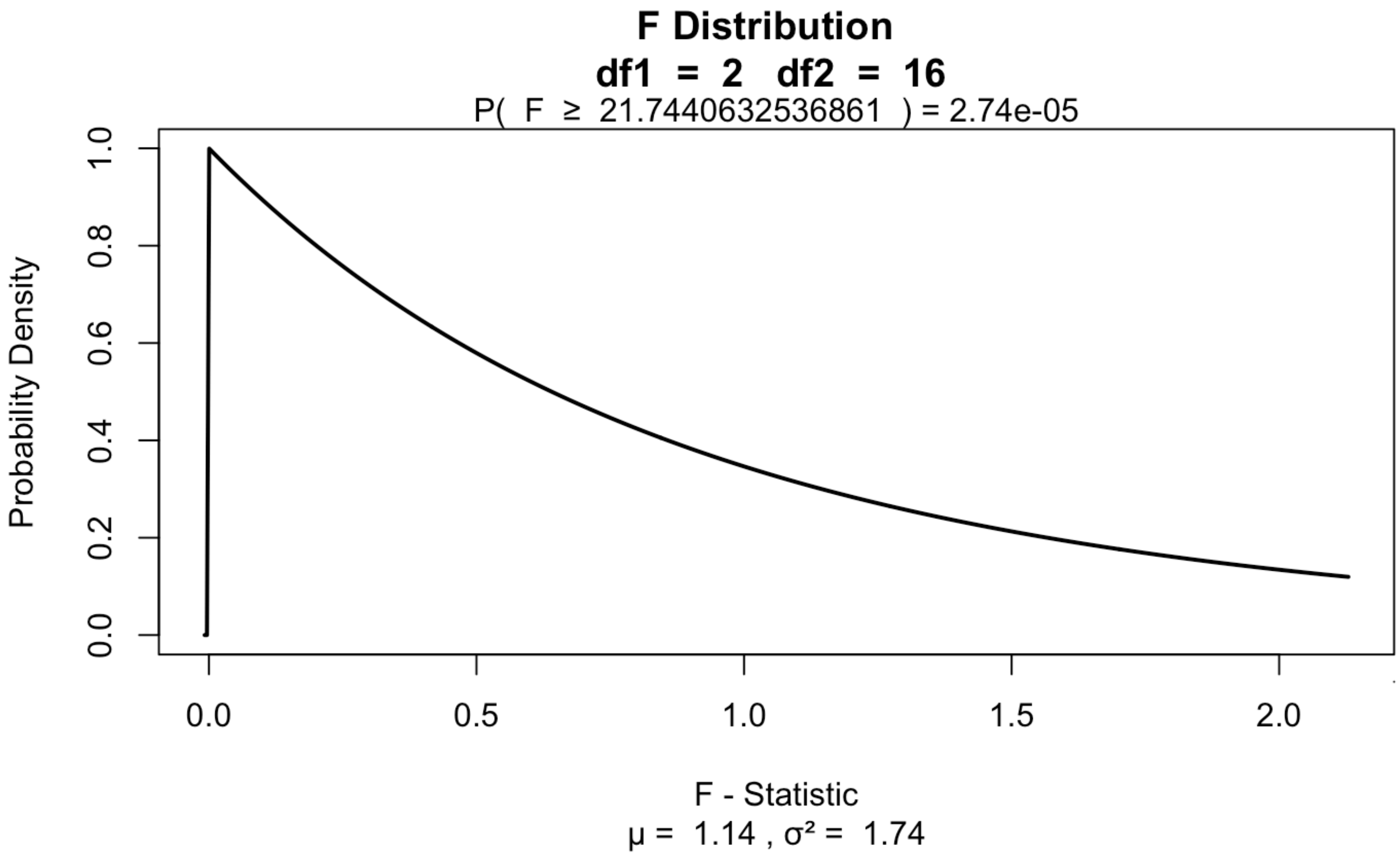
$$F = \frac{MS_{model}}{MS_{error}} = \frac{SIGNAL}{NOISE}$$

```
n = 20
k = 3
df.model = k - 1
df.error = n - k - 1

MS.model = SS.group / df.model
MS.error = SS.error / df.error

F = MS.model / MS.error
F
```

```
## [1] 21.74406
```

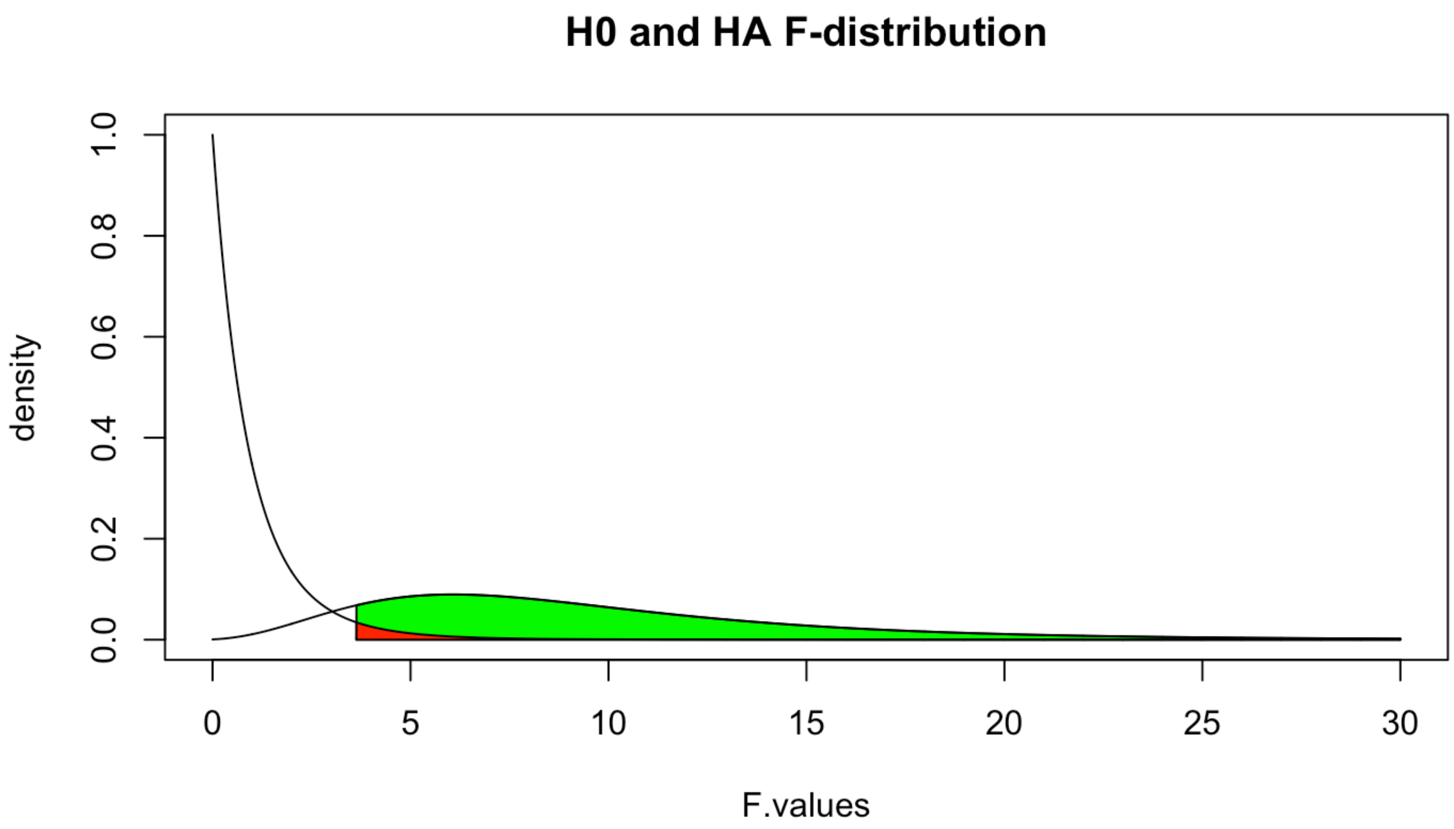
P-value

```r
library("visualize")
visualize.f(F, df.model, df.error, section = "upper")
```


F Distribution
df1 = 2 df2 = 16
P(F ≥ 21.7440632536861) = 2.74e-05
Probability Density
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
F - Statistic
μ = 1.14 , σ² = 1.74

Alpha & Power

```
F.values = seq(0, 30, .01)

plot(F.values, df(F.values, df.model, df.error), type = "l", ylab="density", m

critical.value = qf(.95, df.model, df.error)

critical.range = seq(critical.value, 30, .01)

polygon(c(critical.range,rev(critical.range)),
        c(critical.range*0, rev(df(critical.range, df.model, df.error, ncp = 1

lines(F.values, df(F.values, df.model, df.error, ncp = 15))

polygon(c(critical.range,rev(critical.range)),
        c(critical.range*0, rev(df(critical.range, df.model, df.error))), col
```


H0 and HA F-distribution
density
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
F.values

Adjusted means

```r
# Add dummy variables
data$dummy.1 <- ifelse(data$group == 2, 1, 0)
data$dummy.2 <- ifelse(data$group == 3, 1, 0)

# b coefficients
b.cov = fit$coefficients["covar"];           b.int = fit$coefficients["(Intercept)"]
b.2   = fit$coefficients["factor(group)2"]; b.3   = fit$coefficients["factor(group)3"]

# Adjustment factor for the means of the independent variable
data$mean.adj <- with(data, b.int + b.cov * mean(covar) + b.2 * dummy.1 + b.3 * dummy.2)

aggregate(mean.adj ~ group, data, mean)
```

```
##   group mean.adj
## 1     1 39.18576
## 2     2 41.95576
## 3     3 43.36576
```

Real β's

```
b.0 = 15 # initial value for group 1
b.1 = 3  # difference between group 1 and 2
b.2 = 4  # difference between group 1 and 3
b.3 = 3  # Weight for covariate

cbind(m.covar = mu.covar*3,
      BETA    = c(b.0, b.0+b.1, b.0+b.2),
      sum     = mu.covar*3 + c(b.0, b.0+b.1, b.0+b.2))
```

```
##      m.covar BETA sum
## [1,]      24   15  39
## [2,]      24   18  42
## [3,]      24   19  43
```

- @shklinkenberg
- Klinkenberg
- S.Klinkenberg@UvA.nl
- ShKlinkenberg

END