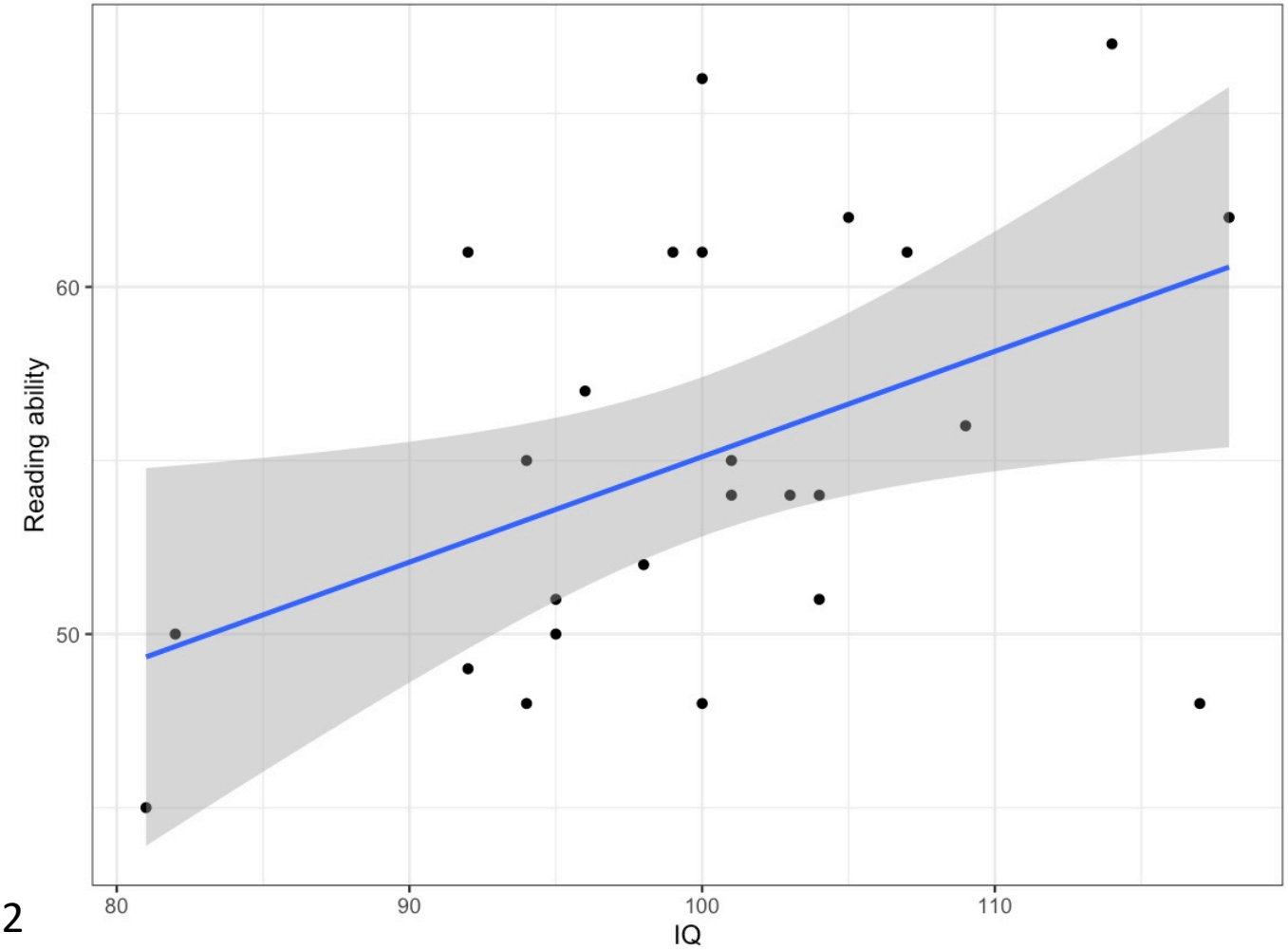
Lancaster University

How to build a linear model in R

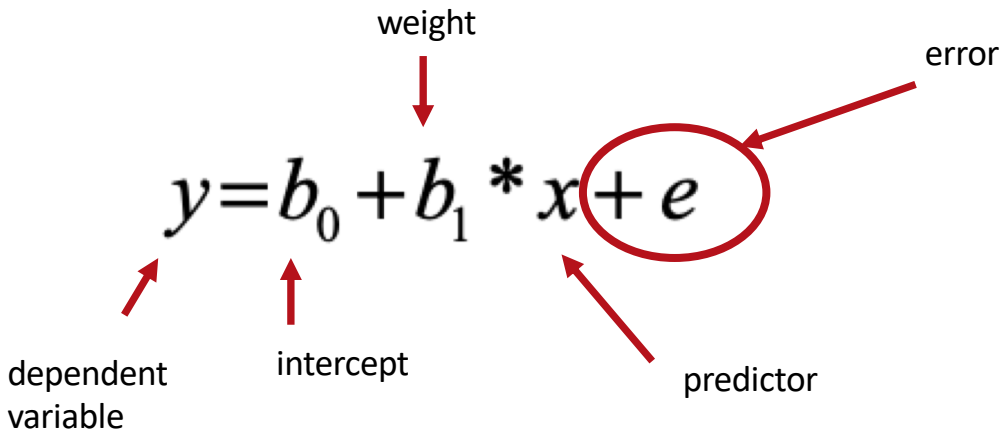
Dr. Margriet A. Groen

The data – Miller Haden

```
> head(mh)
# A tibble: 6 × 5
  Participant  Abil    IQ  Home    TV
        <dbl> <dbl> <dbl> <dbl> <dbl>
1           1    61   107   144   487
2           2    56   109   123   608
3           3    45    81   108   640
4           4    66   100   155   493
5           5    49    92   103   636
6           6    62   105   161   407
> 
```


Reading ability
60
50
80
90
100
110
IQ

Regression line



Building the model in R

$$y = b_0 + b_1 * x + e$$

```
# Run the code to build the regression model
mod <- lm(Abil ~ IQ, data = mh)
mod_summary <- summary(mod)
mod_summary
```

Interpreting the output

```
> mod_summary

Call:
lm(formula = Abil ~ IQ, data = mh)

Residuals:
    Min      1Q  Median      3Q     Max
-12.268  -3.590  -1.411   3.767  10.892

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.7517    12.5745   1.968   0.0612 .
IQ            0.3036     0.1252   2.425   0.0236 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.547 on 23 degrees of freedom
Multiple R-squared:  0.2036,    Adjusted R-squared:  0.1689
F-statistic: 5.878 on 1 and 23 DF,  p-value: 0.02359
```

Interpreting the output

```
> mod_summary

Call:
lm(formula = Abil ~ IQ, data = mh)

Residuals:
    Min      1Q  Median      3Q     Max
-12.268  -3.590  -1.411   3.767  10.892

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.7517    12.5745   1.968   0.0612 .
IQ            0.3036     0.1252   2.425   0.0236 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.547 on 23 degrees of freedom
Multiple R-squared:  0.2036,    Adjusted R-squared:  0.1689
F-statistic: 5.878 on 1 and 23 DF,  p-value: 0.02359
```

Interpreting the output

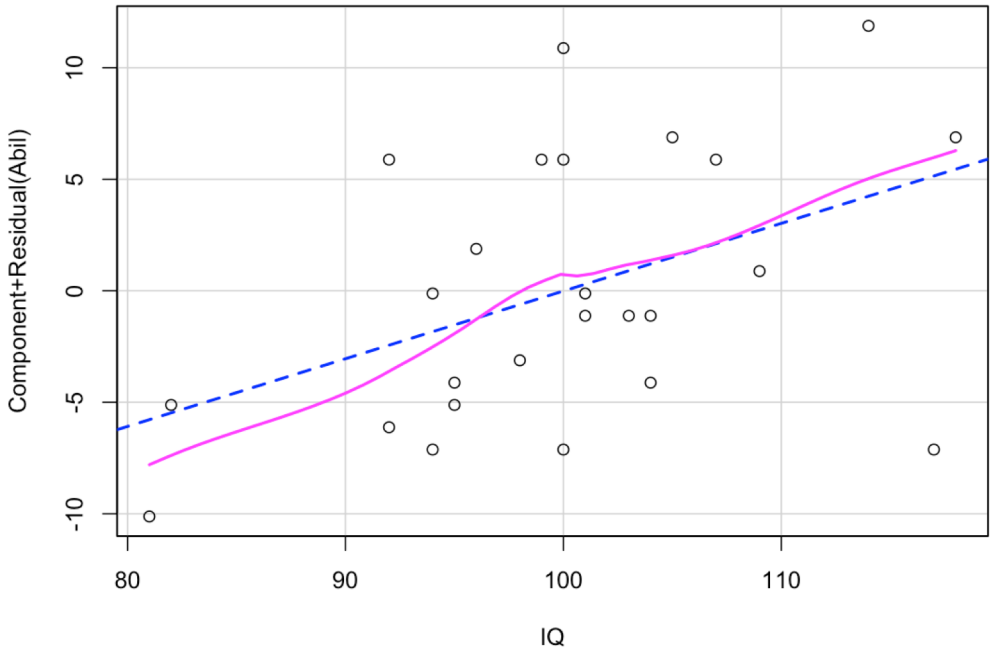
```
> mod_summary

Call:
lm(formula = Abil ~ IQ, data = mh)

Residuals:
    Min      1Q  Median      3Q     Max
-12.268  -3.590  -1.411   3.767  10.892

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.7517    12.5745   1.968   0.0612 .
IQ            0.3036     0.1252   2.425   0.0236 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.547 on 23 degrees of freedom
Multiple R-squared:  0.2036,    Adjusted R-squared:  0.1689
F-statistic: 5.878 on 1 and 23 DF,  p-value: 0.02359
```

Interpreting the output

```
> mod_summary

Call:
lm(formula = Abil ~ IQ, data = mh)

Residuals:
    Min      1Q  Median      3Q     Max
-12.268  -3.590  -1.411   3.767  10.892

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.7517    12.5745   1.968   0.0612 .
IQ            0.3036     0.1252   2.425   0.0236 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.547 on 23 degrees of freedom
Multiple R-squared:  0.2036,    Adjusted R-squared:  0.1689
F-statistic: 5.878 on 1 and 23 DF,  p-value: 0.02359
```

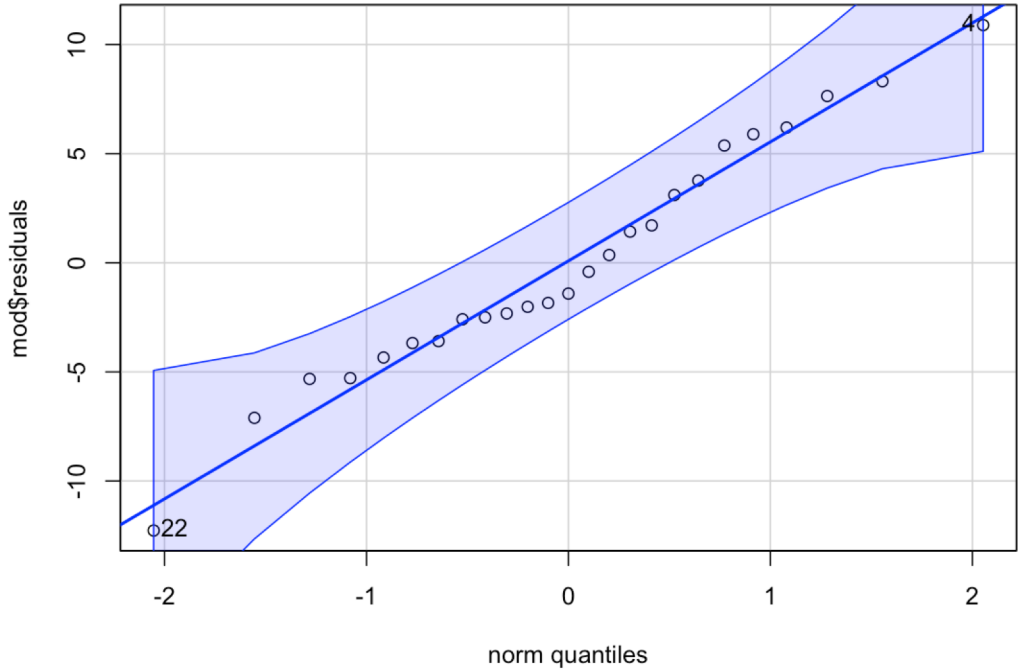
Checking assumptions - linearity

```
crPlots(mod)                    # Plot linear line and line that best fits the data to check the relationship between outcome and predic
```



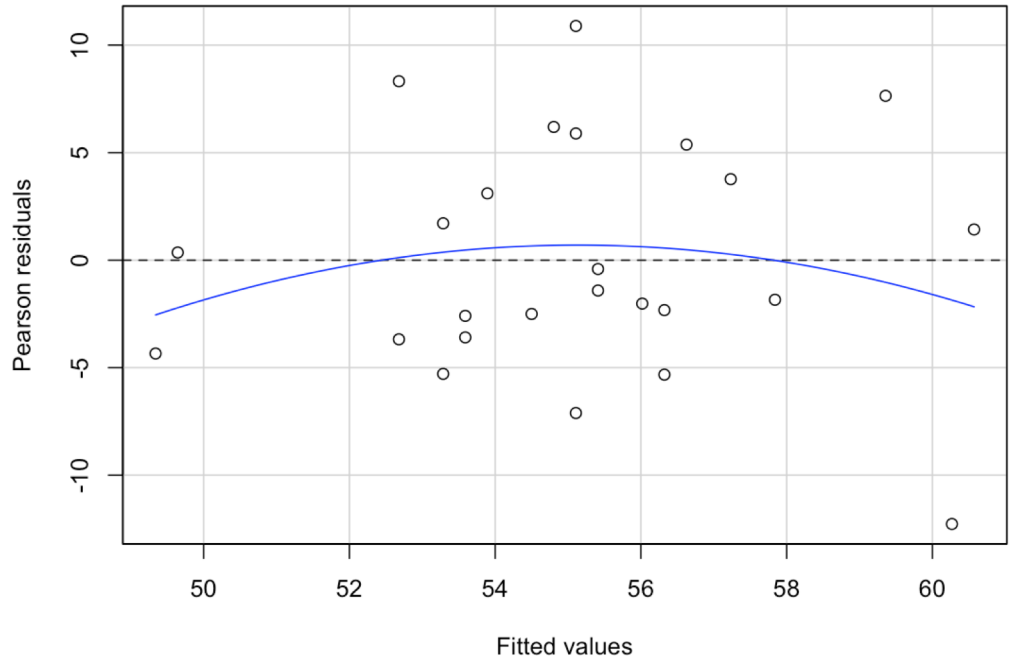
Checking assumptions – residuals normally distributed?

```
qqPlot(mod$residuals)          # Create qq-plot to check residuals are normally distributed
```



Checking assumptions – Do the residuals show constant variance?

```
residualPlot(mod)          # Create residual plot to check residual show homoscedasticity
```



Write up

A simple linear regression was performed with reading ability (*M* = 55.12, *SD* = 6.08) as the outcome variable and IQ (*M* = 100.04, *SD* = 9.04) as the predictor variable. The results of the regression indicated that the model significantly predicted reading ability ($F(1, 23) = 5.88$, $p = .024$, $R^2 = 0.20$), accounting for 20% of the variance. IQ was a significant positive predictor ($\beta = 0.30$, $p = .024$).