

# Ascombe

Marc Los Huertos

2025-04-25

## R Markdown

Let's check out some funky data!

```
anscombe.csv="/home/mwl04747/RTricks/00_Stats/Anscombe_files/anscombe.csv"
anscombe = read.csv(anscombe.csv)

names(anscombe)
```

```
## [1] "X"      "Obs." "x1"    "x2"    "x3"    "x4"    "y1"    "y2"    "y3"    "y4"
```

## Summary of the linear model

Identical statistics for each linear model!

```
lm1 <- lm(y1 ~ x1, data = anscombe); summary(lm1)
```

```
##
## Call:
## lm(formula = y1 ~ x1, data = anscombe)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92127 -0.45577 -0.04136  0.70941  1.83882
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0001      1.1247   2.667  0.02573 *
## x1             0.5001      0.1179   4.241  0.00217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6665, Adjusted R-squared:  0.6295
## F-statistic: 17.99 on 1 and 9 DF, p-value: 0.00217
```

```
lm2 <- lm(y2 ~ x2, data = anscombe); summary(lm2)
```

```
##
## Call:
## lm(formula = y2 ~ x2, data = anscombe)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```

## -1.9009 -0.7609 0.1291 0.9491 1.2691
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.001      1.125   2.667 0.02576 *
## x2            0.500      0.118   4.239 0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6662, Adjusted R-squared:  0.6292
## F-statistic: 17.97 on 1 and 9 DF, p-value: 0.002179
lm3 <- lm(y3 ~ x3, data = anscombe); summary(lm3)

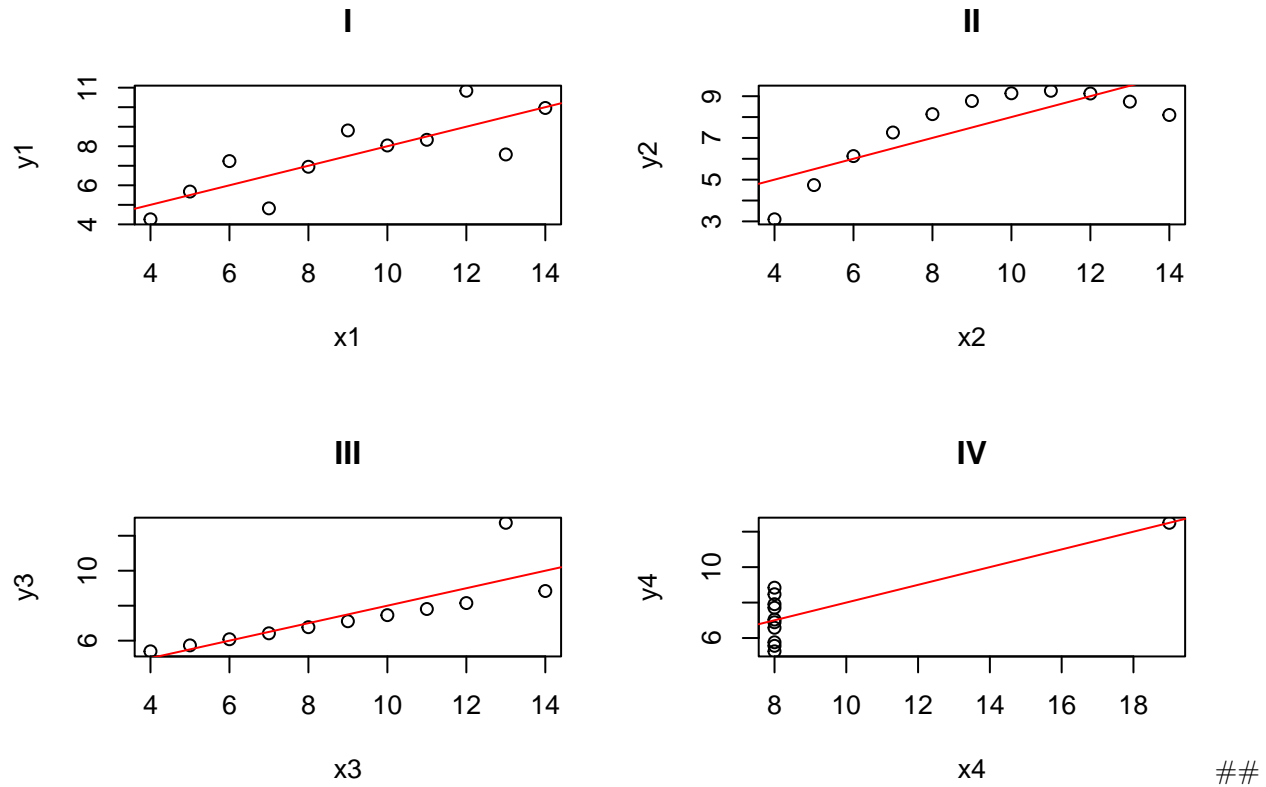
##
## Call:
## lm(formula = y3 ~ x3, data = anscombe)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1586 -0.6146 -0.2303  0.1540  3.2411
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0025      1.1245   2.670 0.02562 *
## x3            0.4997      0.1179   4.239 0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.236 on 9 degrees of freedom
## Multiple R-squared:  0.6663, Adjusted R-squared:  0.6292
## F-statistic: 17.97 on 1 and 9 DF, p-value: 0.002176
lm4 <- lm(y4 ~ x4, data = anscombe); summary(lm4)

##
## Call:
## lm(formula = y4 ~ x4, data = anscombe)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.751 -0.831  0.000  0.809  1.839
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0017      1.1239   2.671 0.02559 *
## x4            0.4999      0.1178   4.243 0.00216 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.236 on 9 degrees of freedom
## Multiple R-squared:  0.6667, Adjusted R-squared:  0.6297
## F-statistic: 18 on 1 and 9 DF, p-value: 0.002165

```

## Including Plots

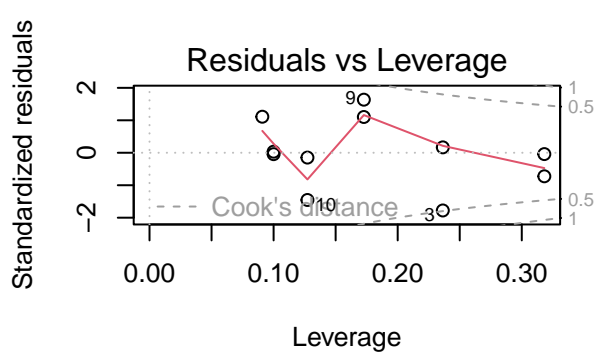
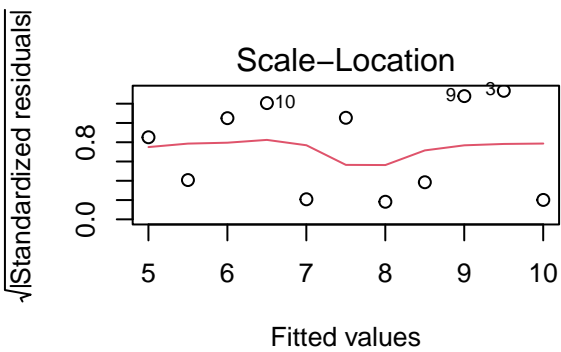
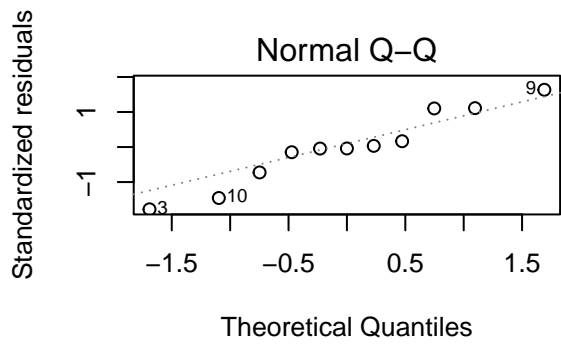
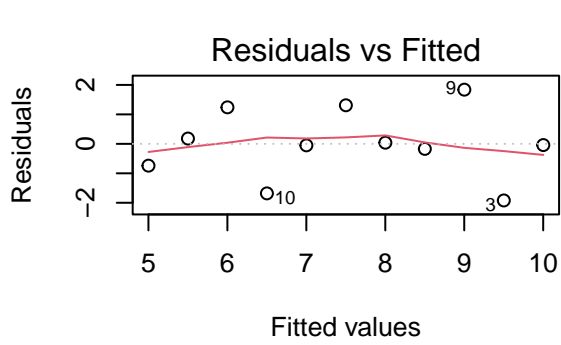
But the plots show major assumption issues.



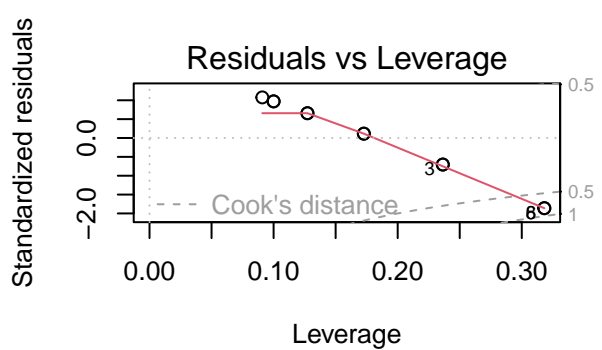
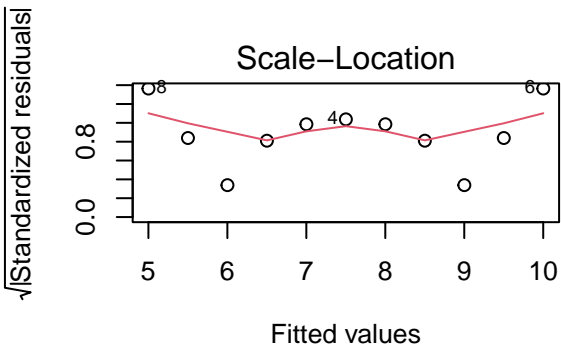
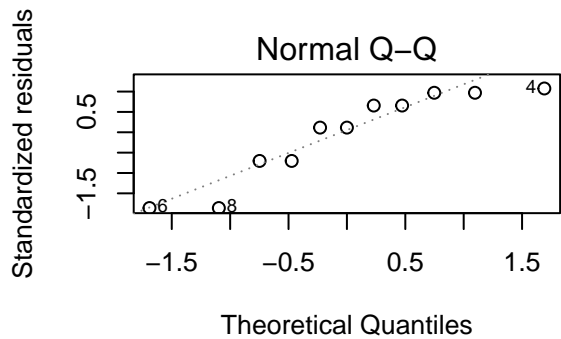
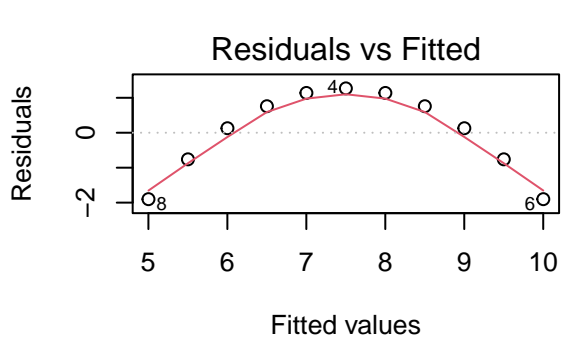
Plot Linear Model – Diagnostic Plots

Confirmed by the diagnostic plots.

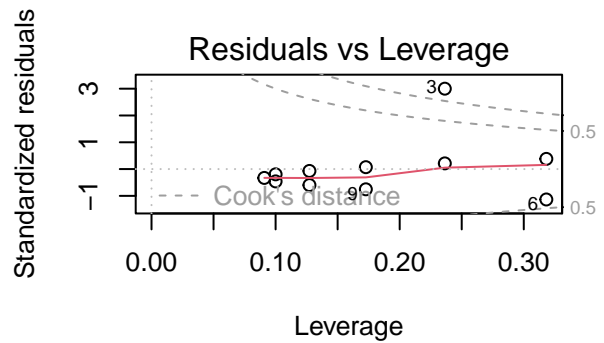
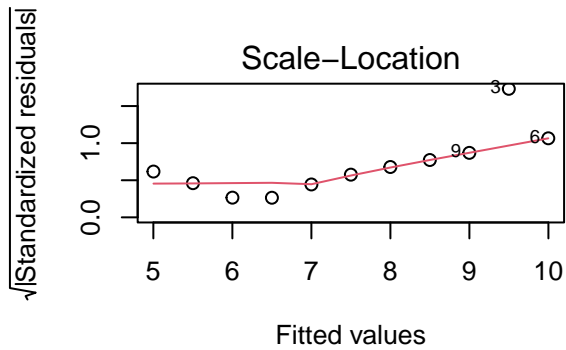
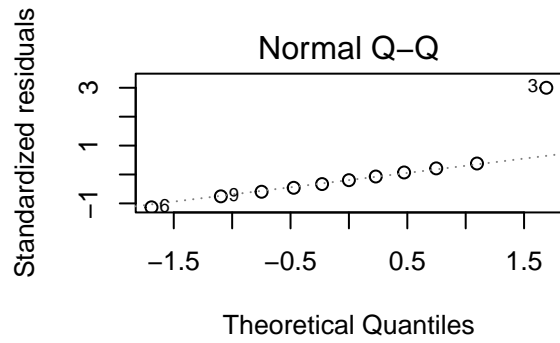
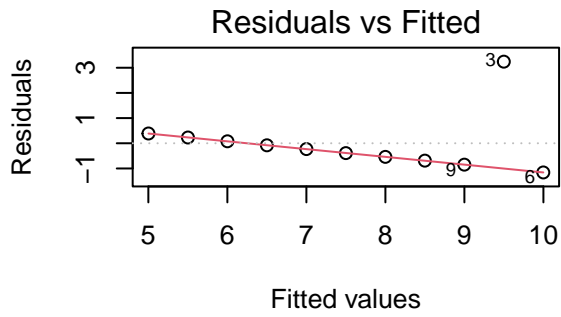
```
par(mfrow=c(2,2))  
plot(lm1)
```



```
plot(lm2)
```



```
plot(lm3)
```



```
plot(lm4)
```

```
## Warning: not plotting observations with leverage one:
##      8
```

