Multiple linear regression

The ASTA team

Contents

1 Example from last lecture

1.1 Crime data set

- The data are measurements from the 67 counties of Florida.
- Our focus is on
  - The response y: `Crime` which is the crime rate
  - The predictor x_1: `Education` which is proportion of the population with high school exam
  - The predictor x_2: `Urbanisation` which is proportion of the population living in urban areas

```
FL <- read.delim("https://asta.math.aau.dk/datasets?file=fl-crime.txt")
head(FL, n = 3)
```

```
##   Crime Education Urbanisation
## 1   104      82.7         73.2
## 2    20      64.1         21.5
## 3    64      74.7         85.0
```

1.2 Multiple regression model for crime data

- Both `Education` (x_1) and `Urbanisation` (x_2) were correlated with `Crime` (y).
- We consider the model

$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

- The errors ϵ are random terms with a $\texttt{norm}(0, \sigma)$ distribution.
- The graph for the mean response is a 2-dimensional plane in a 3-dimensional space.

```
model <- lm(Crime ~ Education + Urbanisation, data = FL)
summary(model)
```

```
##
## Call:
## lm(formula = Crime ~ Education + Urbanisation, data = FL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.693 -15.742  -6.226  15.812  50.678
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   59.1181    28.3653   2.084   0.0411 *
## Education     -0.5834     0.4725  -1.235   0.2214
## Urbanisation   0.6825     0.1232   5.539 6.11e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.82 on 64 degrees of freedom
## Multiple R-squared:  0.4714, Adjusted R-squared:  0.4549
## F-statistic: 28.54 on 2 and 64 DF,  p-value: 1.379e-09
```

- From the output we find the prediction equation

$$\widehat{y} = 59 - 0.58x_1 + 0.68x_2.$$

2 The general model

2.1 Regression model

- In a multiple regression we have
  - a response variable Y.
  - k predictor variables $x_1, x_2, \ldots, x_k$.
- Multiple regression model:
$$Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon.$$
- The **systematic** part of the model says that **the mean response** is a linear function of the predictors:
$$E(Y|x_1, x_2, \ldots, x_k) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$
  - Here $E(Y|x_1, x_2, \ldots, x_k)$ is used to denote the mean value of the response when we know the value of the predictors $x_1, \ldots, x_k$.
- The **error** ϵ is a random variable having a distribution with
  - a normal distribution
  - mean 0
  - standard deviation σ

2.2 Interpretation of parameters

- In the multiple linear regression model
$$E(y|x_1, x_2, \ldots, x_k) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$
  - The parameter α is the `Intercept`, corresponding to the mean response, when all predictors are equal to zero.
  - The parameters $(\beta_1, \beta_2, \cdots, \beta_k)$ are called **partial regression coefficients**.
- Imagine that all predictors but x_1 are held fixed. Then
$$E(y|x_1, x_2, \ldots, x_k) = \tilde{\alpha} + \beta_1 x_1,$$
where
$$\tilde{\alpha} = \alpha + \beta_2 x_2 + \cdots + \beta_k x_k.$$
- So the mean response depends linearly on x_1 when all other variables are kept fixed.
  - The line has slope β_1, which describes the change in the mean response, when x_1 is changed one unit.
  - The rate of change β_1 does not depend on the value of the remaining predictors. In this case we say that there is **no interaction** between the effects of the predictors on the response.
- The above holds similarly for the other partial regression coefficients.

3 Estimation

3.1 Estimation of model parameters

- Suppose we have a sample of size n.
- Based on the sample, we estimate $(\alpha, \beta_1, \beta_2, \ldots, \beta_k)$ by the values $(a, b_1, b_2, \ldots, b_k)$.
- Based on this estimate we obtain the **prediction equation** (or **regression equation**) as
$$\widehat{y} = a + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$
- The difference between our observations and the predictions made by the prediction equation are called the **residuals** $e_i = y_i - \widehat{y}_i$.
- The estimates $(a, b_1, b_2, \ldots, b_k)$ are chosen by the **least squares method**, which seeks to minimize the sum of squared residuals
$$\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2.$$

3.2 Estimation of error variance

- Recall that σ^2 is the variance of the error terms in the model. Using the residuals as approximations of the errors in our sample, we estimate σ^2 by

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-k-1}.$$

- Rather than n we divide by **the degrees of freedom** $df = n - k - 1$. Theory shows, that this is reasonable.
  - The degrees of freedom df are determined by the sample size n minus the number of parameters in the regression equation.
  - Currently we have $k + 1$ parameters: 1 intercept and k slopes.

4 Multiple R-squared

4.1 Multiple R^2

- We can compare two models to predict the response y.
- Model 1: We do not use the predictors, and use $\bar{y}$ to predict any y-measurement. The corresponding sum of squared prediction errors is
  - $TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$ and is called the **Total Sum of Squares**.
- Model 2: We use the multiple prediction equation with predictors $x_1, \ldots, x_k$ to predict y. The sum of squared prediction errors is now
  - $SSE = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ and is called **Sum of Squared Errors**.
- We then define **the multiple coefficient of determination**

$$R^2 = \frac{TSS - SSE}{TSS}.$$

- Thus, R^2 is the relative reduction in squared prediction errors, when we use $x_1, x_2, \ldots, x_k$ as explanatory variables.
  - We say that R^2 is the proportion of the total variation in the data that can be explained by $x_1, \ldots, x_k$.
- Properties:
  - $0 \leq R^2 \leq 1$
  - $R^2 = 0$ means $TSS = SSE$, so the model does not improve when we include $x_1, \ldots, x_k$ in the model.
  - $R^2 = 1$ means $SSE = 0$, so the observations are predicted perfectly by $x_1, \ldots, x_k$.

4.2 Example

```
summary(model)
```

```
##
## Call:
## lm(formula = Crime ~ Education + Urbanisation, data = FL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.693 -15.742  -6.226  15.812  50.678
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   59.1181    28.3653   2.084   0.0411 *
## Education     -0.5834     0.4725  -1.235   0.2214
```

```
## Urbanisation   0.6825     0.1232   5.539 6.11e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.82 on 64 degrees of freedom
## Multiple R-squared:  0.4714, Adjusted R-squared:  0.4549
## F-statistic: 28.54 on 2 and 64 DF,  p-value: 1.379e-09
```

- The prediction equation is $\widehat{y} = 59 - 0.58x_1 + 0.68x_2$
- The estimate for σ is $s = 20.82$ (`Residual standard error` in **R**) with $df = 67 - 3 = 64$ degrees of freedom.
- Multiple $R^2 = 0.4714$, i.e. 47% of the variation in the response is explained by including the predictors in the model.
- The output provides a test of the hypothesis $H_0 : \beta_1 = 0$ corresponding to no effect of the predictor x_1.
  - The estimate $b_1 = -0.5834$ has standard error (`Std. Error`) $se = 0.4725$ with corresponding observed t-value (`t value`) $t_{\text{obs}} = \frac{-0.5834}{0.4725} = -1.235$.
  - This means that b_1 isn't significantly different from zero, since the p-value (`Pr(>|t|)`) is 22%. That means that we can exclude `Education` as a predictor.

4.3 Example

- Our final model is then a simple linear regression:

```
model2 <- lm(Crime ~ Urbanisation, data = FL)
summary(model2)
```

```
##
## Call:
## lm(formula = Crime ~ Urbanisation, data = FL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.766 -16.541  -4.741  16.521  49.632
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  24.54125    4.53930   5.406 9.85e-07 ***
## Urbanisation  0.56220    0.07573   7.424 3.08e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.9 on 65 degrees of freedom
## Multiple R-squared:  0.4588, Adjusted R-squared:  0.4505
## F-statistic: 55.11 on 1 and 65 DF,  p-value: 3.084e-10
```

- The coefficient of determination always decreases, when the model is simpler. Now we have $R^2 = 46\%$, where before we had 47%. But the decrease is not significant as we will see next.

4.4 F-test for comparing two models

- We can compare two models, where one is obtained from the other by setting m parameters to zero, by an F-test.
- We can compare R^2 for the two models via the following F-statistic:

$$F_{obs} = \frac{(R_2^2 - R_1^2)/\text{df}_1}{(1 - R_2^2)/\text{df}_2}$$

  - R_2^2 corresponds to the larger model and R_1^2 corresponds to the smaller model.
  - Large values of F_{obs} means that R_2^2 is large compared to R_1^2, pointing towards the alternative hypothesis.
  - $\text{df}_1 = m$ is the number of parameters set to zero in the null-hypothesis.
  - $\text{df}_2 = n - k - 1$ where n is sample size and $k + 1$ is the number of unknown parameters in the larger model.

- In R the calculations are done using `anova`:

```
anova(model2, model)
```

```
## Analysis of Variance Table
##
## Model 1: Crime ~ Urbanisation
## Model 2: Crime ~ Education + Urbanisation
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1     65 28391
## 2     64 27730  1    660.61 1.5247 0.2214
```

5 Overall F-test for effect of predictors

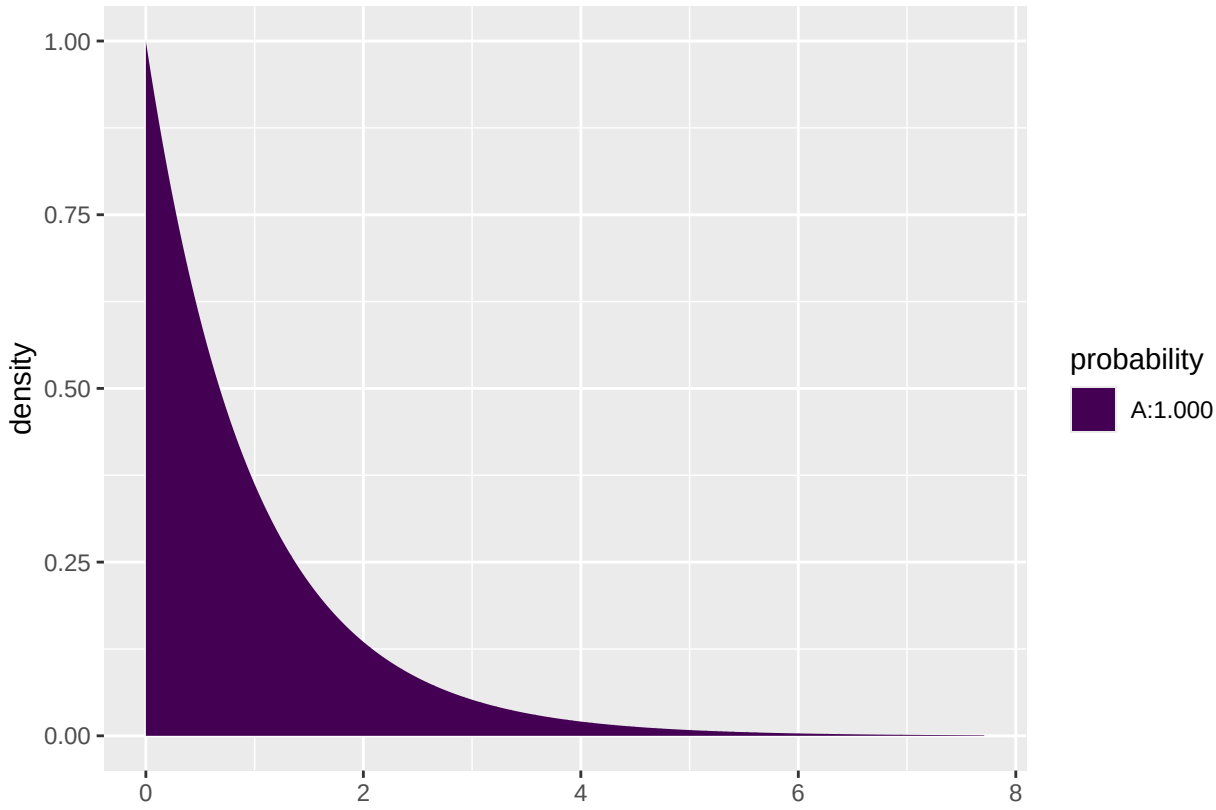
5.1 F-test

- We consider the hypothesis
$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
against the alternative that at least one β_i are non-zero.
  - This is the hypothesis that there is no effect of any of the predictors.
- As test statistic we use
$$F_{obs} = \frac{R^2/k}{(1 - R^2)/(n - k - 1)}$$
- Large values of R^2 implies large values of F_{obs}, which points to the alternative hypothesis.
- Thus, the p-value is the probability of observing something larger than the computed F_{obs}.
- The distribution of F_{obs} under the null-hypothesis is an F-distribution with degrees of freedom
  - $df_1 = k$ (the number of parameters set to zero in the null-hypothesis).
  - $df_2 = n - k - 1$ (number of observations minus number of unknown parameters in the model).

5.2 Example

- We return to `Crime` and the prediction equation $\widehat{y} = 59 - 0.58x_1 + 0.68x_2$, where $n = 67$ and $R^2 = 0.4714$.
- We test the hypothesis $H_0 : \beta_1 = \beta_2 = 0$. We have
  - $df_1 = k = 2$ since 2 parameters are set to zero under H_0.
  - $df_2 = n - k - 1 = 67 - 2 - 1 = 64$.
  - Then we can calculate $F_{obs} = \frac{R^2/k}{(1-R^2)/(n-k-1)} = 28.54$
- To evaluate the value 28.54 in the relevant F-distribution:

```
1 - pdist("f", 28.54, df1=2, df2=64)
```




```
## [1] 1.378612e-09
```

- So p-value$=1.38 \times 10^{-9}$ (notice we don't multiply by 2 since this is a one-sided test; only large values point more towards the alternative than the null hypothesis).
- All this can be found in the summary output we already have:

```
summary(model)
```

```
##
## Call:
## lm(formula = Crime ~ Education + Urbanisation, data = FL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.693 -15.742  -6.226  15.812  50.678
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   59.1181    28.3653   2.084   0.0411 *
## Education     -0.5834     0.4725  -1.235   0.2214
## Urbanisation   0.6825     0.1232   5.539 6.11e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.82 on 64 degrees of freedom
## Multiple R-squared:  0.4714, Adjusted R-squared:  0.4549
## F-statistic: 28.54 on 2 and 64 DF,  p-value: 1.379e-09
```

6 Interaction model

6.1 Interaction between effects of predictors

- Could it be possible that a combination of `Education` and `Urbanisation` is good for prediction? We investigate this using the **interaction model**

$$E(y|x_1, x_2) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2,$$

  where we have extended with a possible effect of the product $x_1 x_2$.
- If we fix x_2 in this model, the mean response is linear in x_1 with intercept $\alpha + \beta_2 x_2$ and slope $\beta_1 + \beta_3 x_2$, since

$$E(y|x_1, x_2) = (\alpha + \beta_2 x_2) + (\beta_1 + \beta_3 x_2) x_1.$$

  - The slope for x_1 now depends on the value of x_2!
  - Interaction means that the effect of x_1 on the response depends on the value of x_2.
  - Interaction **does not** mean that x_1 and x_2 affect each other.

6.2 Example - interaction model

- We fit the model for the `Crime` data set:

```
model3 <- lm(Crime ~ Education * Urbanisation, data = FL)
summary(model3)
```

```
##
## Call:
## lm(formula = Crime ~ Education * Urbanisation, data = FL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.181 -15.207  -6.457  14.559  49.889
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             19.31754   49.95871   0.387    0.700
## Education                0.03396    0.79381   0.043    0.966
## Urbanisation             1.51431    0.86809   1.744    0.086 .
## Education:Urbanisation -0.01205    0.01245  -0.968    0.337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.83 on 63 degrees of freedom
## Multiple R-squared:  0.4792, Adjusted R-squared:  0.4544
## F-statistic: 19.32 on 3 and 63 DF,  p-value: 5.371e-09
```

- When we look at the p-values in the table nothing is significant at the 5% level!
- But the F-statistic tells us that the predictors collectively have a significant prediction ability.
- Why has the highly significant effect of x_2 disappeared? Because the predictors x_1 and $x_1 x_2$ are able to explain the same as x_2.
- Previously we only had x_1 as alternative explanation to x_2 - and that wasn't enough.
- The phenomenon is called **multicollinearity**. It happens because the predictors are highly correlated.
- It also illustrates that we can have different models with equally good predictive properties.
  - In the case of an interaction model we always choose the model without interaction because it is simpler.
- However, in general it can be difficult to choose between models. For example, if both height and weight are good predictors of some response, but one of them can be left out, which one do we choose?

7 Multiple linear regression with categorical predictors

7.1 Dummy variables

- Suppose we want to predict the response variable using a categorical predictor variable x with k categories.
- We choose one group, say group k, as the reference category.
- For the remaining groups $1, \ldots, k-1$, we define dummy variables

$$z_i = \begin{cases} 0, & \text{if } x \neq i, \\ 1, & \text{if } x = i. \end{cases}$$

  for $i = 1, \ldots, k-1$.
  - The dummy variable z_i is 1 if an observation is in group i and 0 otherwise.
  - When all dummy variables $z_i = 0$, $i = 1, \ldots, k-1$, it means that the observation belongs to the reference group k.
- We can use the variables $z_1, \ldots, z_{k-1}$ in a multiple regression along with other predictor variables.

7.2 Example

- Consider the dataset `mtcars`. We are interested in how engine type `vs` (categorical) and weight of the car `wt` (quantitative, x_1) are associated with fuel consumption `mpg`.
- The variable `vs` is already coded as a dummy variable z in R, taking the value 1 if the engine is v-shaped and 0 otherwise.
- The multiple regression model becomes

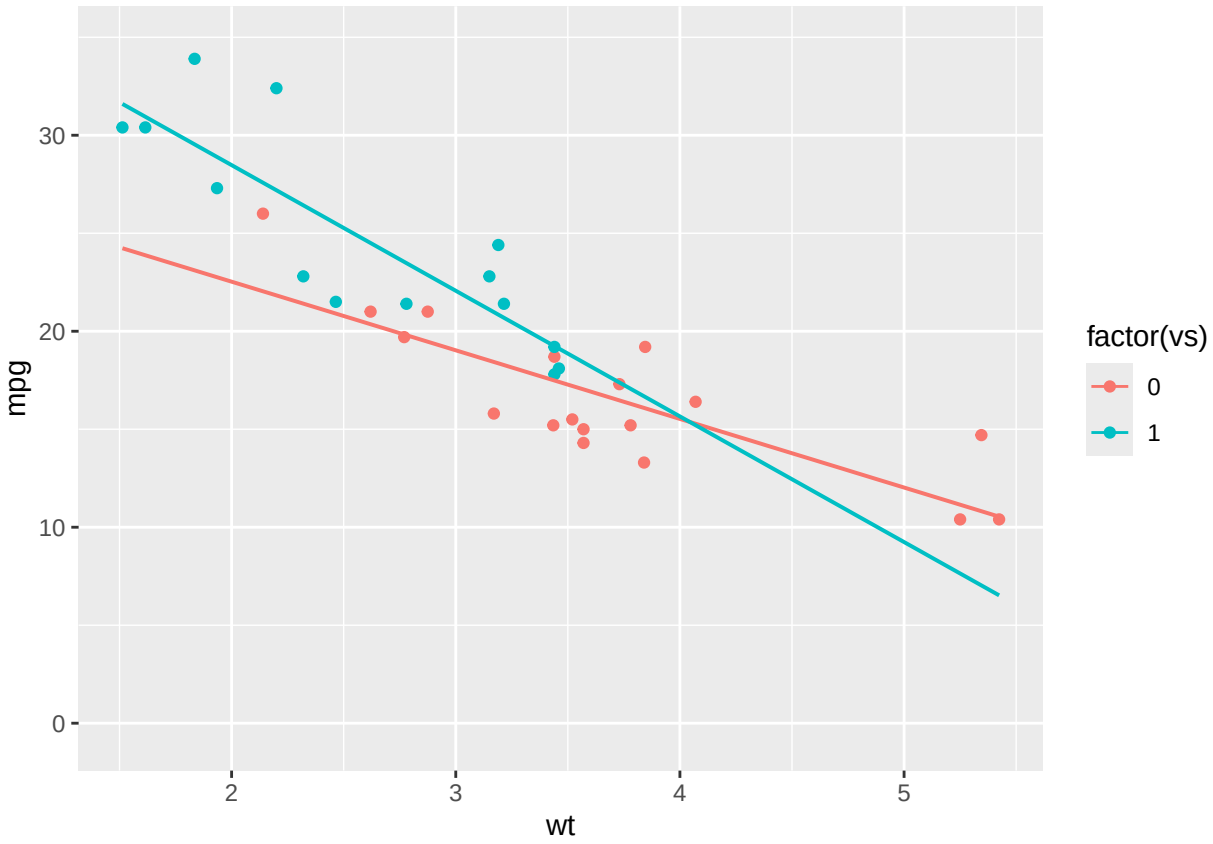
$$E(Y|x_1, z) = \alpha + \beta_1 x_1 + \beta_2 z.$$

  - For $z = 0$:
$$E(Y|x_1, z) = \alpha + \beta_1 x_1.$$
  - For $z = 1$:
$$E(Y|x_1, z) = \alpha + \beta_2 + \beta_1 x_1.$$
- So we get two different regression lines for the two groups.
  - The lines have a common slope β_1 (parallel lines).
  - The lines have different intercepts. The difference in intercepts is β_2.

7.3 Example

- We always start with some graphics (remember the function `gf_point` for plotting points and `gf_lm` for adding a regression line).

```
library(mosaic)
gf_point(mpg ~ wt, color = ~factor(vs), group=~factor(vs), data = mtcars) %>% gf_lm()
```



- An unclear picture, but a tendency to decreasing number of miles per gallon with increasing weight for both engine types.
- The slope of the lines for the two engine types look different. But is the difference significant? Or can the difference be explained by sampling variation?

7.4 Example

- We fit a multiple regression model without interaction in R:

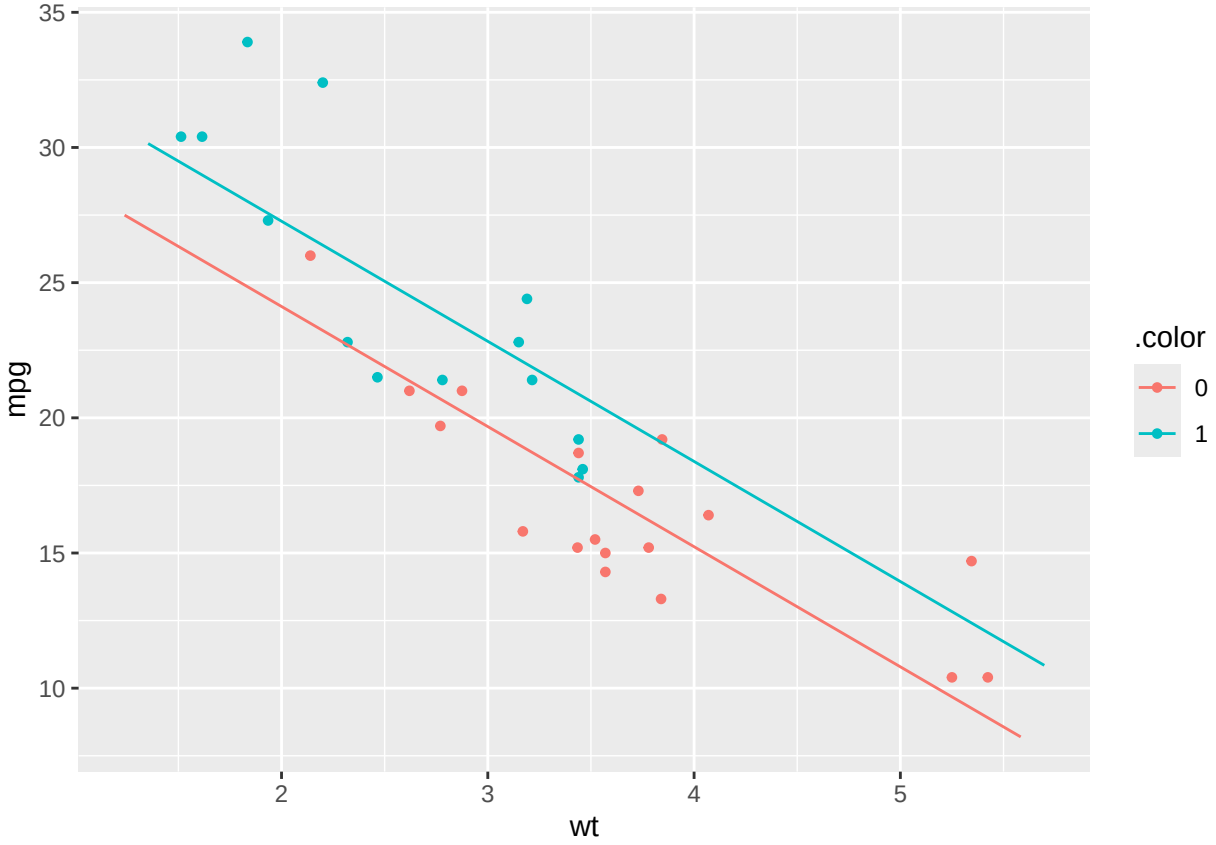
```
model1 <- lm(mpg ~ wt + factor(vs) , data = mtcars)
summary(model1)
```

```
##
## Call:
## lm(formula = mpg ~ wt + factor(vs), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7071 -2.4415 -0.3129  1.4319  6.0156
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.0042     2.3554  14.012 1.92e-14 ***
## wt           -4.4428     0.6134  -7.243 5.63e-08 ***
## factor(vs)1   3.1544     1.1907   2.649   0.0129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.78 on 29 degrees of freedom
## Multiple R-squared:  0.801,  Adjusted R-squared:  0.7873
## F-statistic: 58.36 on 2 and 29 DF,  p-value: 6.818e-11
```

- The common slope to `wt` is estimated to be $\widehat{\beta}_1 = -4.44$, with corresponding p-value $5.63 \cdot 10^{-8}$, so the effect of `wt` is significantly different from zero.
  - The estimate is negative, so increasing weight decreases the number of miles per gallon.
- The estimate for intercept in the reference group ("not v-shaped") is $\widehat{\alpha} = 33.0$, which is significantly different from zero if we test at level 5% (this test is not really of interest).
- The difference between intercepts for the two engine types is $\widehat{\beta}_1 = 3.15$, which is significant with p-value=1%.
  - This suggests that the regression lines are not the same for the two engine types.
  - The value 3.15 is the vertical distance between the two regression lines.

```
plotModel(model1)
```



7.5 Example: Prediction equations

```
summary(model1)
```

```
##
## Call:
## lm(formula = mpg ~ wt + factor(vs), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7071 -2.4415 -0.3129  1.4319  6.0156
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.0042     2.3554  14.012 1.92e-14 ***
## wt           -4.4428     0.6134  -7.243 5.63e-08 ***
## factor(vs)1   3.1544     1.1907   2.649   0.0129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.78 on 29 degrees of freedom
## Multiple R-squared:  0.801,  Adjusted R-squared:  0.7873
## F-statistic: 58.36 on 2 and 29 DF,  p-value: 6.818e-11
```

- Reference/baseline group (not v-shaped):

$$\widehat{y} = 33.0 - 4.44x$$

- V-shaped:

$$\widehat{y} = 33.0 + 3.15 - 4.44x = 36.15 - 4.44x.$$

7.6 Interaction model

- We can expand the regression model by including an interaction between x and z:

$$E(y|x,z) = \alpha + \beta_1 x + \beta_2 z + \beta_3 z \cdot x.$$

- This yields a regression line for engine type:
- Not v-shaped ($z = 0$): $E(y|x,z) = \alpha + \beta_1 x$
- V-shaped ($z = 1$): $E(y|x,z) = \alpha + \beta_2 + (\beta_1 + \beta_3)x$.
- β_2 is *the difference* in `Intercept` between the two groups, while β_3 is *the difference* in `slope` between the two groups.

7.7 Example: Prediction equations

- When we use `*` in the model formula we include interaction between `vs` and `wt`:

```
model2 <- lm(mpg ~ wt * factor(vs), data = mtcars)
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ wt * factor(vs), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9950 -1.7881 -0.3423  1.2935  5.2061
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     29.5314     2.6221  11.263 6.55e-12 ***
## wt              -3.5013     0.6915  -5.063 2.33e-05 ***
## factor(vs)1     11.7667     3.7638   3.126   0.0041 **
## wt:factor(vs)1  -2.9097     1.2157  -2.393   0.0236 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.578 on 28 degrees of freedom
## Multiple R-squared:  0.8348, Adjusted R-squared:  0.8171
## F-statistic: 47.16 on 3 and 28 DF,  p-value: 4.497e-11
```

- We use the output to write the prediction equations:
  - Reference/baseline group (not v-shaped):

$$\widehat{y} = 29.5 - 3.5x$$

  - V-shaped:

$$\widehat{y} = (29.5 + 11.8) + (-3.5 - 2.9)x = 41.3 - 6.4x.$$

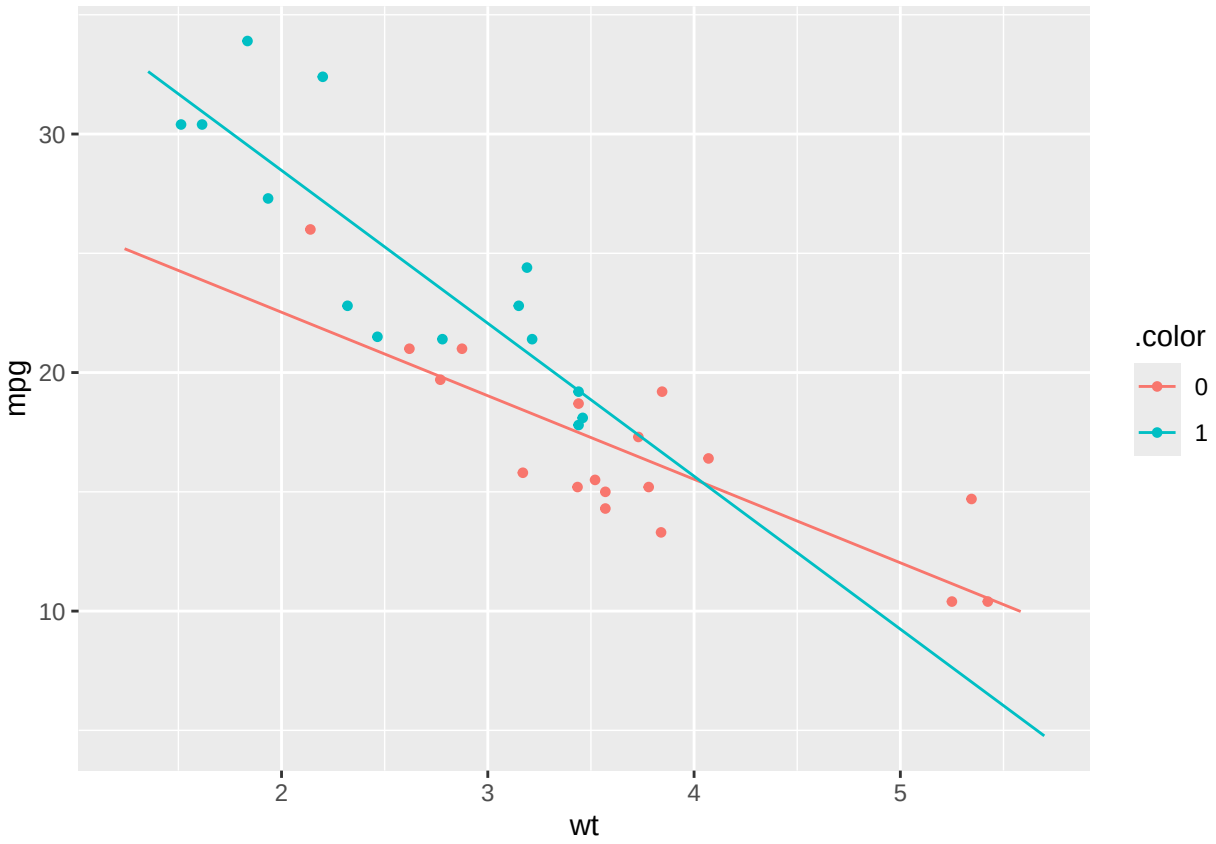
7.8 Example: Individual tests

```
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ wt * factor(vs), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9950 -1.7881 -0.3423  1.2935  5.2061
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     29.5314     2.6221  11.263 6.55e-12 ***
## wt              -3.5013     0.6915  -5.063 2.33e-05 ***
## factor(vs)1     11.7667     3.7638   3.126   0.0041 **
## wt:factor(vs)1  -2.9097     1.2157  -2.393   0.0236 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.578 on 28 degrees of freedom
## Multiple R-squared:  0.8348, Adjusted R-squared:  0.8171
## F-statistic: 47.16 on 3 and 28 DF,  p-value: 4.497e-11
```

- *The difference* in slope between the two engine types is estimated to $\widehat{\beta}_3 = -2.9$ which is significant with p-value=0.0236, so the slopes are significantly different.

```
plotModel(model2)
```



7.9 Hierarchy of models

- Always test for no interaction ($\beta_3 = 0$) before making tests for main effects ($\beta_1 = 0$ or $\beta_2 = 0$).

$$Model: E(Y|x,z) = \alpha + \beta_1 x + \beta_2 z + \beta_3 xz$$

Interaction

↓ H_0: $\beta_3 = 0$

Main effects only

H_0: $\beta_2 = 0$ ↙ ↘ H_0: $\beta_1 = 0$

Only effect of x Only effect of z

H_0: $\beta_1 = 0$ ↘ ↙ H_0: $\beta_2 = 0$

No effect of x or z

7.10 F-test

- We can compare the two models in the `mtcars` example, namely the model with and without interaction via

```
anova(model1, model2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + factor(vs)
## Model 2: mpg ~ wt * factor(vs)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     29 224.09
## 2     28 186.03  1    38.062 5.7287 0.02363 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```