Linear Modelling 3

Starting point:

- Clarity about what is your response variable and what are explanatory variables
- Scientific questions about the relationship between response and explanatory variables
- Data in a long-form dataframe or tibble

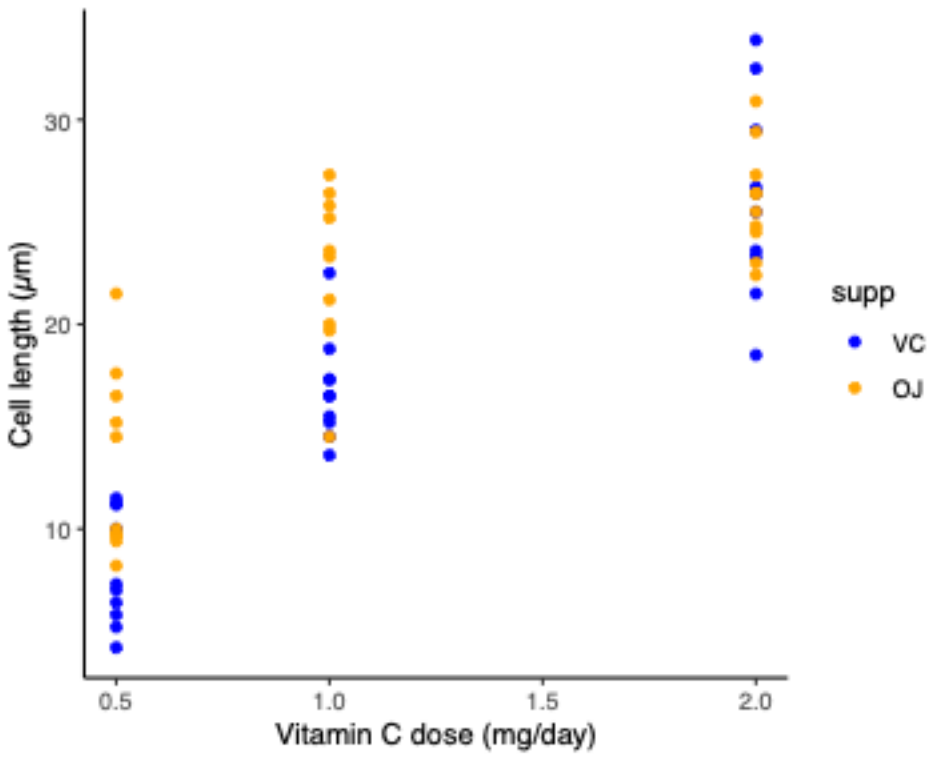
Example Experiment – variables

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)

- Roles in experiment
  - **Response variable**
    - What we care about
    - What we measured
    - Want to know if it responds to explanatory variables
    - Not set as part of experiment
    - Biological and technical '**noise**' in the measurement
  - **Explanatory variable**
    - Something we can manipulate
    - What we manipulated
    - Want to know if it affects the response variable
    - Set precisely as part of the experiment
    - Little or no '**noise**' in the value

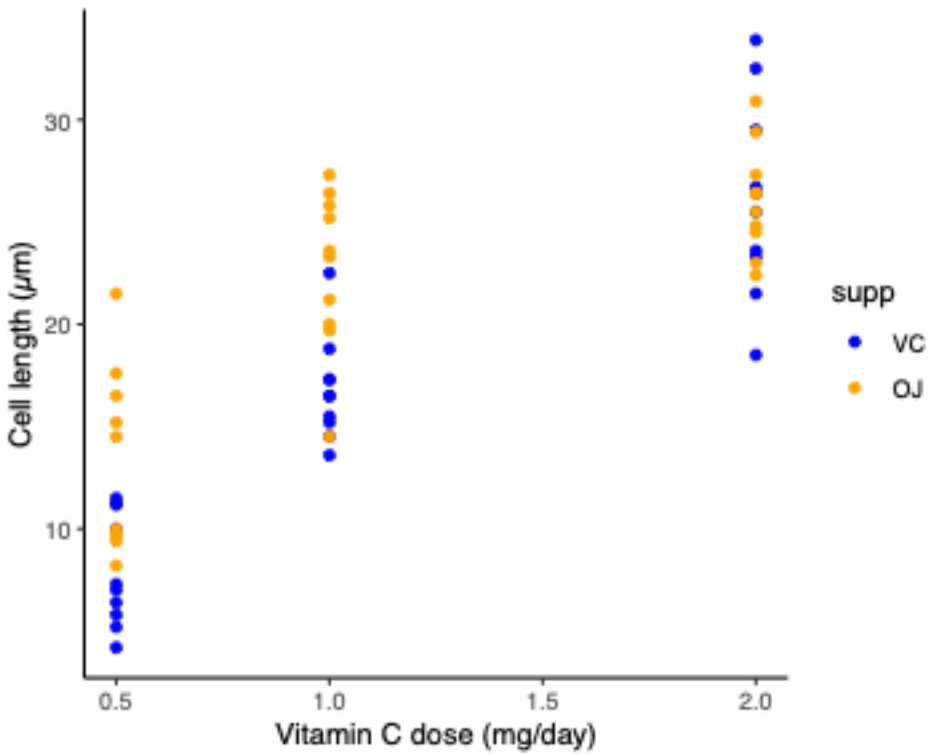
- Clarity about what is your response variable and what are explanatory variables



Crampton, E. W. (1947).

Example Experiment – questions

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs
- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)
- Questions
  - Does the cell length response to dose depend on supplement type?
  - Does cell length depend on dose?
  - Does cell length depend on supplement type?
  - How much does cell length change with dose for each supplement?
  - …

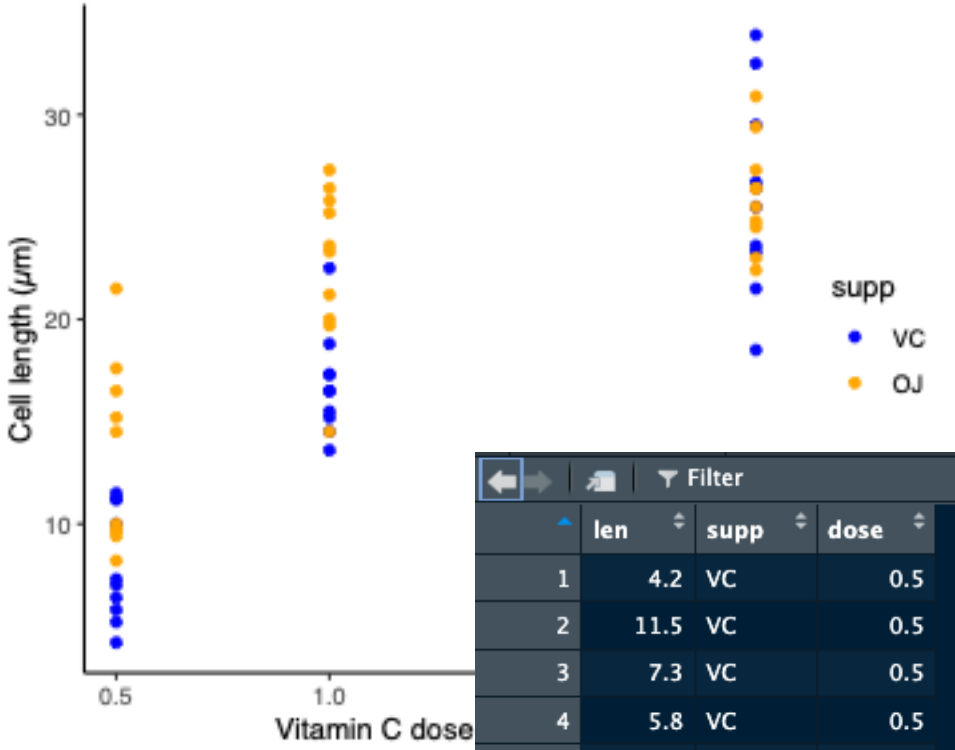


Crampton, E. W. (1947).

- Scientific questions about the relationship between response and explanatory variables
- Clarity about what is your response variable and what are explanatory variables

Example Experiment – data

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs
- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)
- Questions
  - Does the cell length response to dose depend on supplement type?
  - Does cell length depend on dose?
  - Does cell length depend on supplement type?
  - How much does cell length change with dose for each supplement?
  - …



Cra

```
> ToothGrowth
```

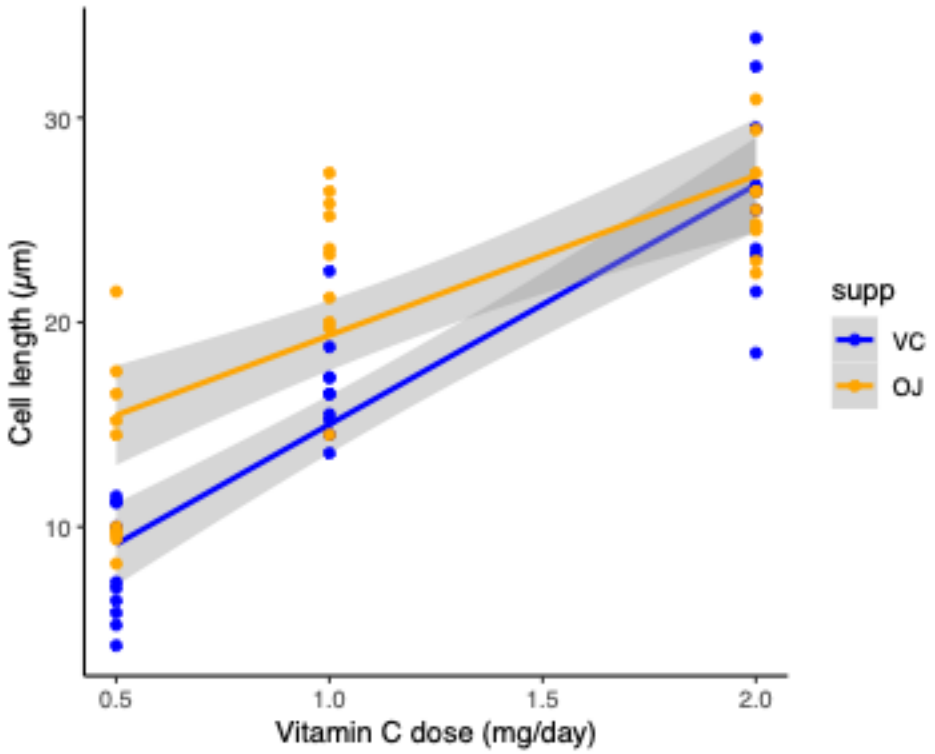
Filter

	len	supp	dose
1	4.2	VC	0.5
2	11.5	VC	0.5
3	7.3	VC	0.5
4	5.8	VC	0.5
5	6.4	VC	0.5
6	10.0	VC	0.5
7	11.2	VC	0.5
8	11.2	VC	0.5
9	5.2	VC	0.5
10	7.0	VC	0.5
11	16.5	VC	1.0
12	16.5	VC	1.0
13	15.2	VC	1.0
14	17.3	VC	1.0
15	22.5	VC	1.0

- Data in a long-form dataframe or tibble
- Scientific questions about the relationship between response and explanatory variables
- Clarity about what is your response variable and what are explanatory variables

Example Experiment – answering questions

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs
- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)
- Questions
  - Does the cell length response to dose depend on supplement type?
  - Does cell length depend on dose?
  - Does cell length depend on supplement type?
  - How much does cell length change with dose for each supplement?
  - …
- Put lines through the data – can answer the questions:
  - Are the lines **significantly** different from one another?
    - Are the slopes of each line significantly different from one another?
    - Are the intercepts of each line significantly different from one another?
  - What are the slope and intercept of each line?
  - …
- Data in a long-form dataframe or tibble
- Scientific questions about the relationship between response and explanatory variables
- Clarity about what is your response variable and what are explanatory variables



Crampton, E. W. (1947).

Example Experiment – Do the analysis

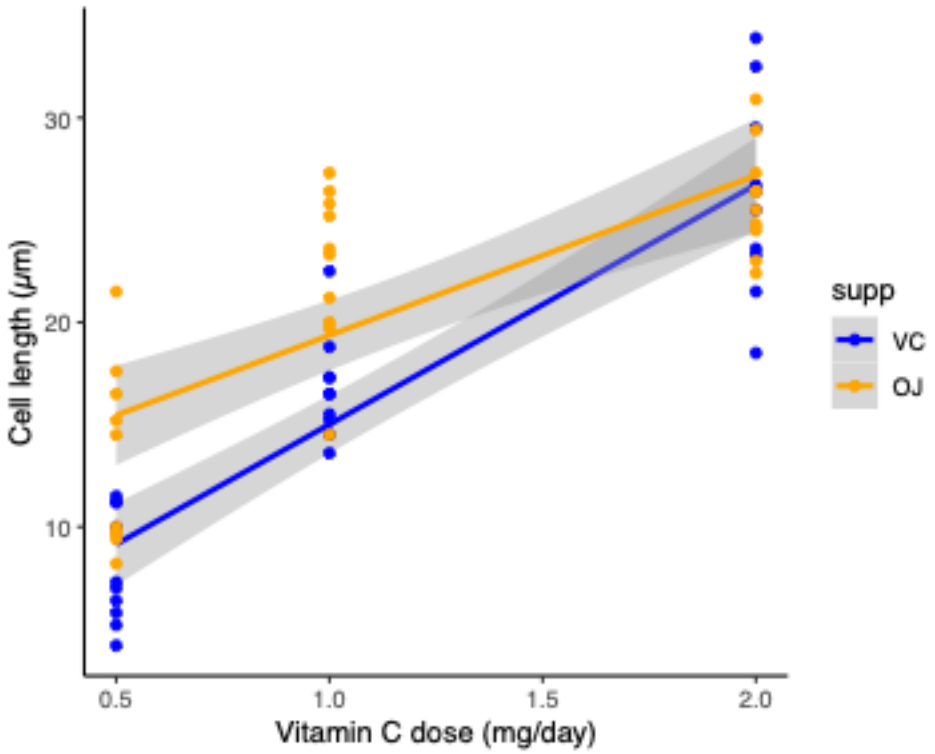
- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)

```
> model <- lm(len ~ supp * dose, data = ToothGrowth)
```

- Minimises the squared distance from points to lines
  - Only '**error**' in the response variable
  - Distance from the line the '**residual**'
  - Returns the value for one line and the differences to the other

```
> summary(model)
```



Crampton, E. W. (1947).

- Data in a long-form dataframe or tibble
- Scientific questions about the relationship between response and explanatory variables
- Clarity about what is your response variable and what are explanatory variables

Example Experiment – Do the analysis

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

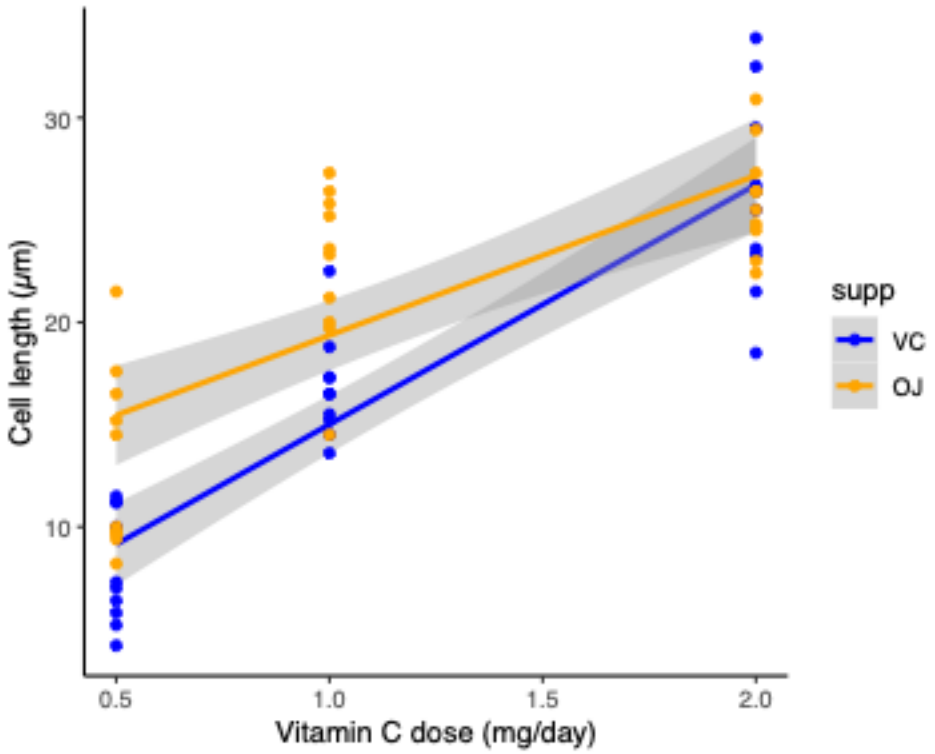
```
> model <- lm(len ~ supp*dose, data = ToothGrowth)
> summary(model)

Call:
lm(formula = len ~ supp * dose, data = ToothGrowth)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2264 -2.8462  0.0504  2.2893  7.9386

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.581   2.084 0.041772 *
suppOJ         8.255      2.236   3.691 0.000507 ***
dose          11.716      1.195   9.800 9.44e-14 ***
suppOJ:dose   -3.904      1.691  -2.309 0.024631 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.083 on 56 degrees of freedom
Multiple R-squared:  0.7296,    Adjusted R-squared:  0.7151
F-statistic: 50.36 on 3 and 56 DF,  p-value: 6.521e-16
```

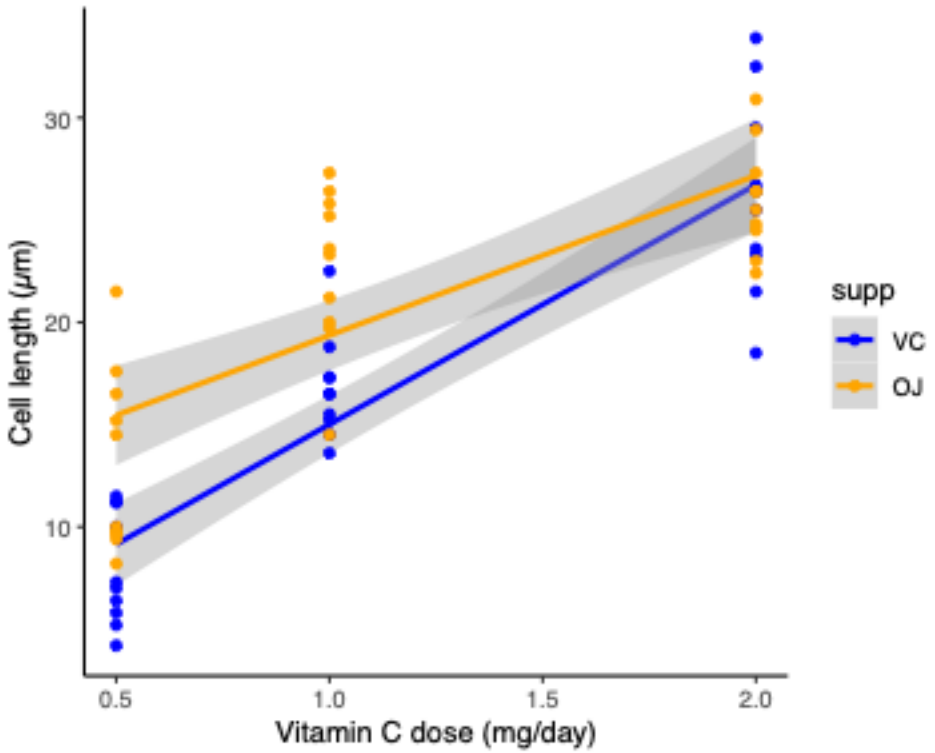


Crampton, E. W. (1947).

Example Experiment – Do the analysis

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.581   2.084 0.041772 *
supp0J         8.255      2.236   3.691 0.000507 ***
dose          11.716      1.195   9.800 9.44e-14 ***
supp0J:dose   -3.904      1.691  -2.309 0.024631 *
---
```

- Are the lines **significantly** different from one another?
  - Are the slopes of each line significantly different from one another?
  - Are the intercepts of each line significantly different from one another?
- What are the slope and intercept of each line?
- …

Crampton, E. W. (1947).

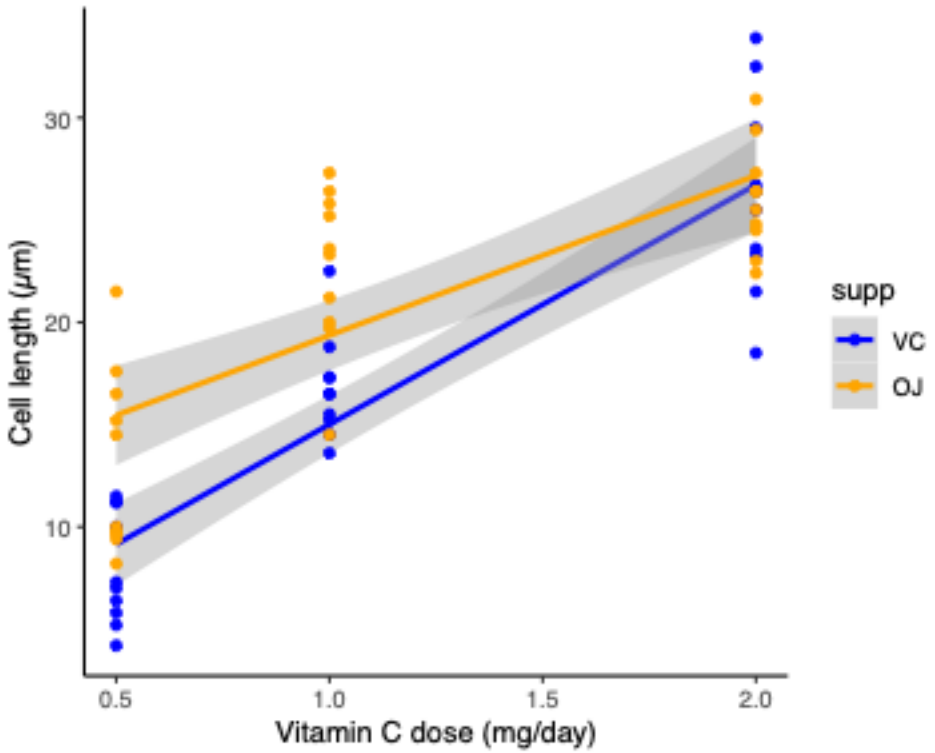
- Questions
  - Does the cell length response to dose depend on supplement type? Yes, $P = 0.025$
  - Does cell length depend on dose?
  - Does cell length depend on supplement type?
  - How much does cell length change with dose for each supplement? Need to look closely
  - …

Example Experiment – Do the analysis

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.581   2.084 0.041772 *
suppOJ         8.255      2.236   3.691 0.000507 ***
dose          11.716      1.195   9.800 9.44e-14 ***
suppOJ:dose   -3.904      1.691  -2.309 0.024631 *
---
```



- Called a linear model, because it all just adds up in a linear way (not just because we're fitting straight lines!)

$$\hat{y}_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

Crampton, E. W. (1947).

- When x is a categorical variable, works like a binary: 1 = 'it's in this category' 0 = 'it's in another category'
- So for vitamin C (supp = "VC" i.e. $x_1 = 0$) and dose = 1mg/day (i.e. $x_2 = 1$)

('**Dummy variable**')

  - $\hat{y}_i$ = Intercept + suppOJ x 0 + dose x 1 + suppOJ:dose x 0 x 1
    = 3.295 + 0 + 11.716 + 0 = **15. 011**

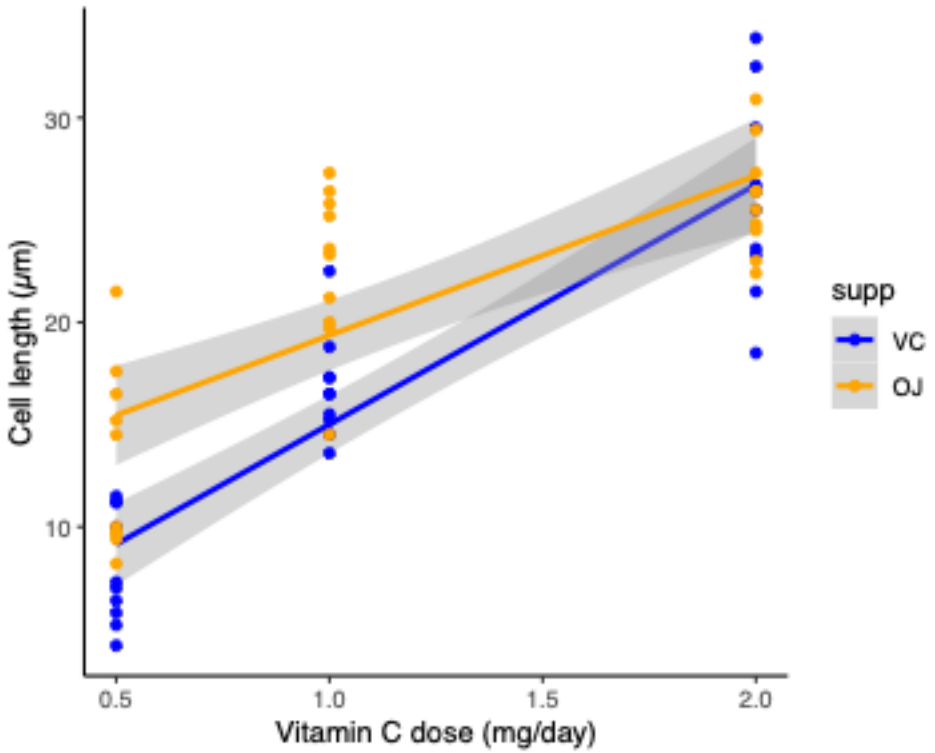
Check with plot!

- How much does cell length change with dose for each supplement? Need to look closely

Example Experiment – Do the analysis

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)




```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.581   2.084 0.041772 *
suppOJ         8.255      2.236   3.691 0.000507 ***
dose          11.716      1.195   9.800 9.44e-14 ***
suppOJ:dose   -3.904      1.691  -2.309 0.024631 *
---
```

- Called a linear model, because it all just adds up in a linear way (not just because we're fitting straight lines!)

$$\hat{y}_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

Crampton, E. W. (1947).

- When x is a categorical variable, works like a binary: 1 = 'it's in this category' 0 = 'it's in another category'
- So for OJ (supp = "OJ" i.e. $x_1 = 1$) and dose =1mg/day ($x_2 = 1$)
  - $\hat{y}_i$ = Intercept + suppOJ x 1 + dose x 1 + suppOJ:dose x 1 x 1
    = 3.295 + 8.255 + 11.716 + -3.94 = **19.326**

Check with plot!

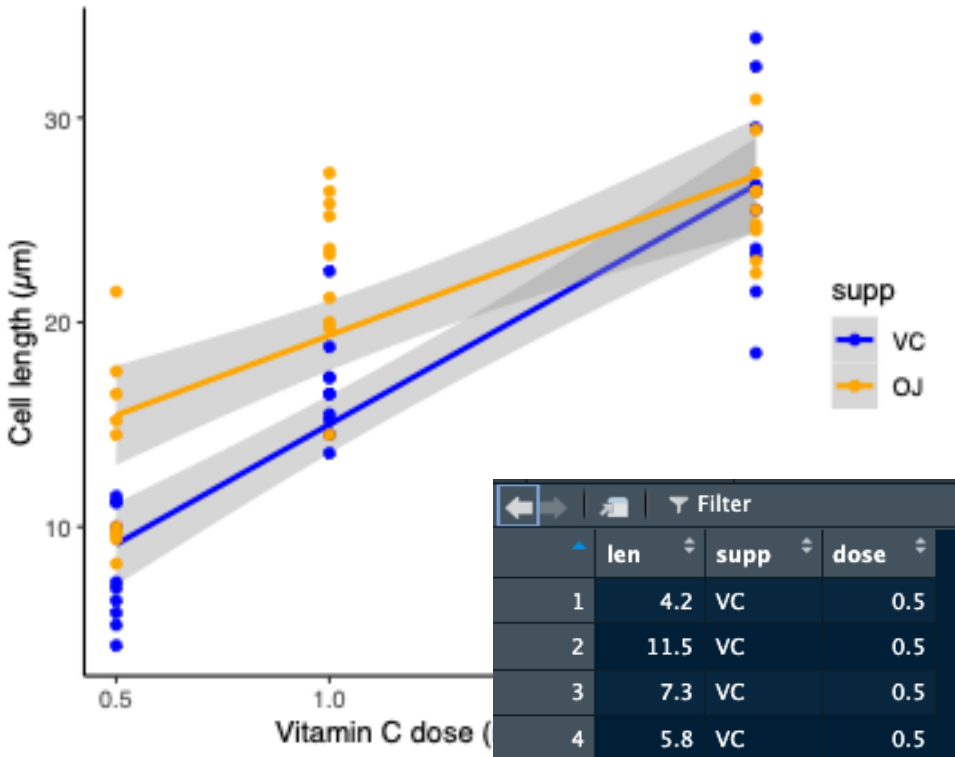
- How much does cell length change with dose for each supplement? Need to look closely

Example Experiment – Do the analysis

- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

- Three variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
  - Type of supplement: vitamin C versus orange juice (categorical)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.581   2.084 0.041772 *
supp0J         8.255      2.236   3.691 0.000507 ***
dose          11.716      1.195   9.800 9.44e-14 ***
supp0J:dose   -3.904      1.691  -2.309 0.024631 *
---
```



	len	supp	dose
1	4.2	VC	0.5
2	11.5	VC	0.5
3	7.3	VC	0.5
4	5.8	VC	0.5
5	6.4	VC	0.5
6	10.0	VC	0.5
7	11.2	VC	0.5
8	11.2	VC	0.5
9	5.2	VC	0.5
10	7.0	VC	0.5
11	16.5	VC	1.0
12	16.5	VC	1.0
13	15.2	VC	1.0
14	17.3	VC	1.0
15	22.5	VC	1.0

- How much does cell length change with dose for each supplement? Need to look closely
  - When the supplement is vitamin C, the change with dose is 12 µl /(mg/day)
  - When the supplement is OJ, the change with dose is 11.7 -3.9 = 7.8 µl /(mg/day)
- Vitamin C ('VC') is described as the '**reference level**' of the categorical variable (supp)
  - Change with `relevel()` or `factor()` functions
- Describe it as an **Analysis of Co-Variance**, ANCOVA, or just a **linear model** (many other sorts too)

Crampton, E. W. **The growth of the odontoblasts of the incisor tooth as a criterion of the vitamin C intake of the guinea pig**. *J Nutr* 33, 491-504 (1947). https://doi.org/10.1093/jn/33.5.491