Linear Modelling 1

Starting point:

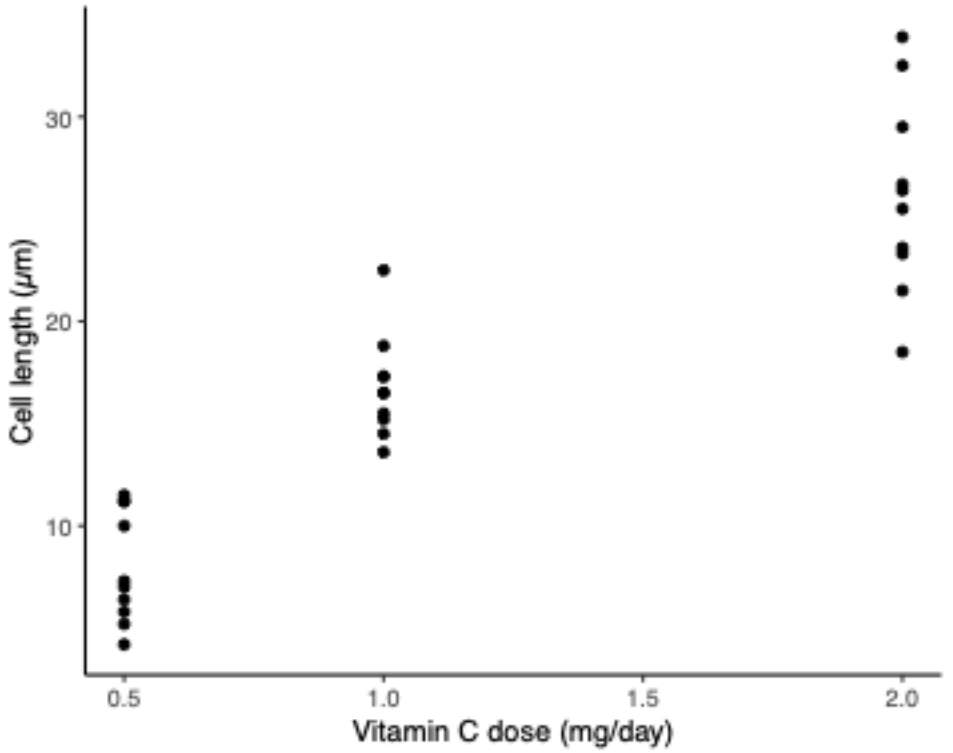
- Clarity about what is your response variable and what are explanatory variables
- Scientific questions about the relationship between response and explanatory variables
- Data in a long-form dataframe or tibble

Example Experiment – variables

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

- Two variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)

- 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 control
    - 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



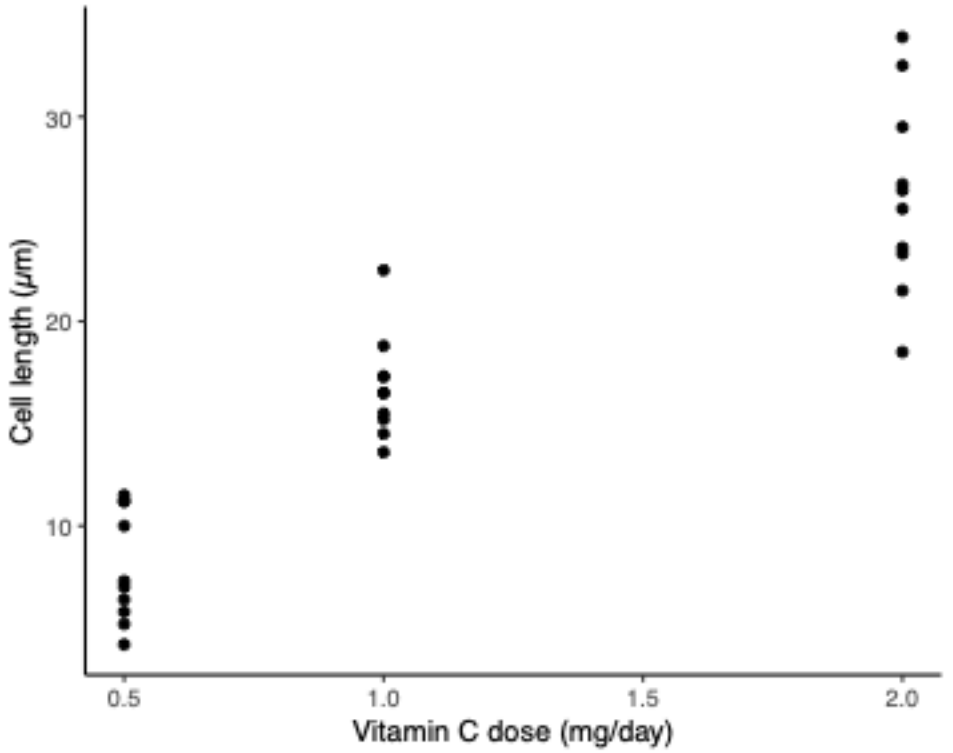
Crampton, E. W. (1947).

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

Example Experiment – questions

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

- Two variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
- Questions
  - Does cell length depend on vitamin C dose?
  - How much does cell length change with dose?
  - It would be cruel to deprive them of vitamin C entirely – Predict how much growth without any?
  - How sure are we of the answers?



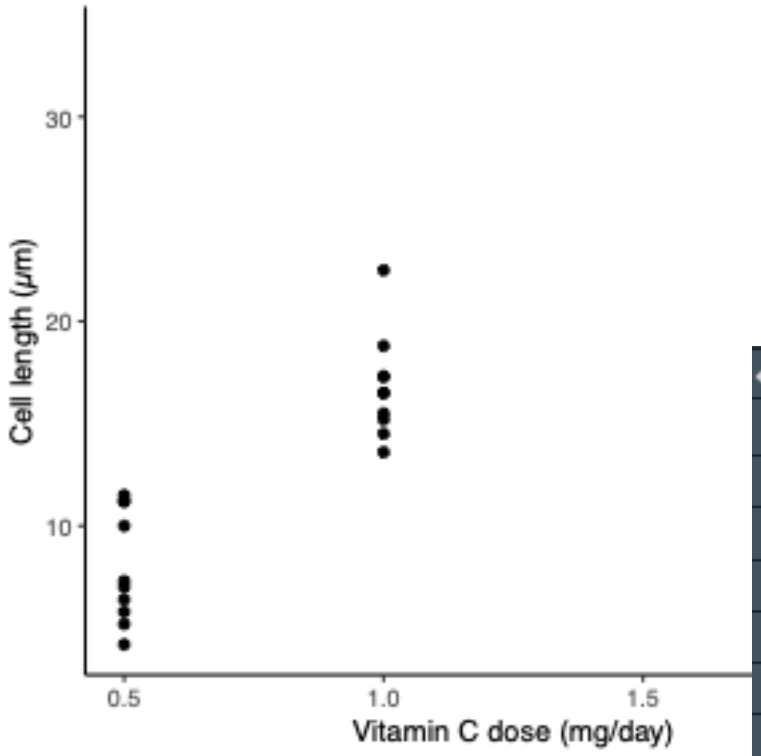
Crampton, E. W. (1947).

- 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

- Two variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
- Questions
  - Does cell length depend on vitamin C dose?
  - How much does cell length change with dose?
  - It would be cruel to deprive them of vitamin C entirely – Predict how much growth without any?
  - How sure are we of the answers?



Crampt

```
> ToothGrowth1
```

Filter

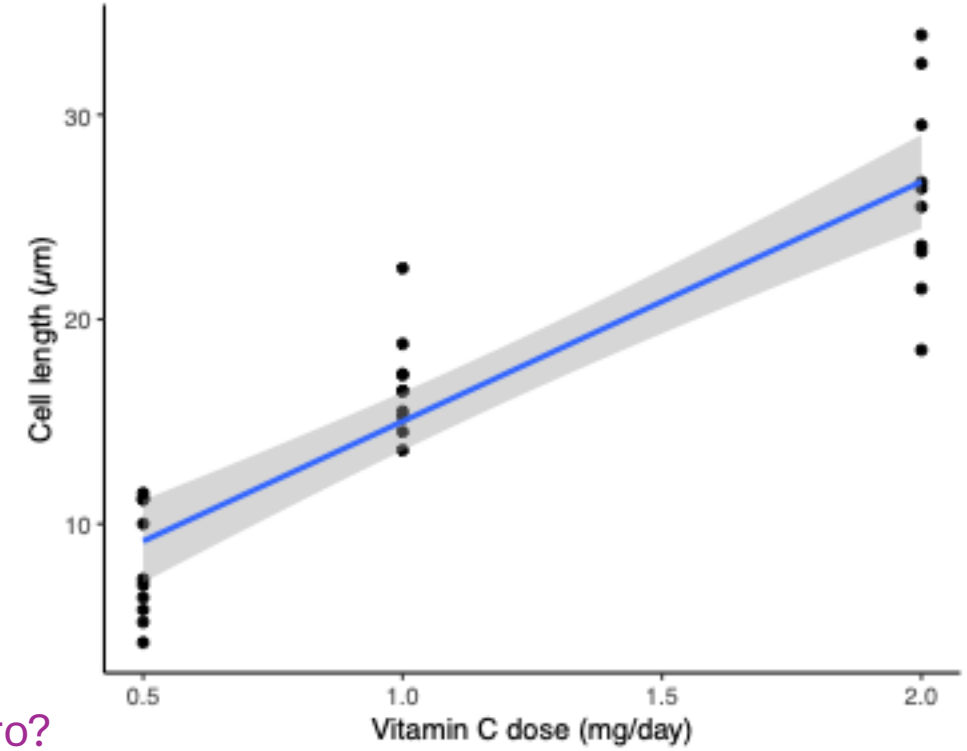
	len	dose
1	4.2	0.5
2	11.5	0.5
3	7.3	0.5
4	5.8	0.5
5	6.4	0.5
6	10.0	0.5
7	11.2	0.5
8	11.2	0.5
9	5.2	0.5
10	7.0	0.5
11	16.5	1.0
12	16.5	1.0
13	15.2	1.0
14	17.3	1.0
15	22.5	1.0

- 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

- Two variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)
- Questions
  - Does cell length depend on vitamin C dose?
  - How much does cell length change with dose?
  - It would be cruel to deprive them of vitamin C entirely – Predict how much growth without any?
  - How sure are we of the answers?
- Put a line through the data – can answer the questions:
  - Does the line have a slope **significantly** different from zero?
  - What's the slope of the line?
  - What's the intercept of the line?
  - What's the uncertainty estimate on any of the above?
- 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

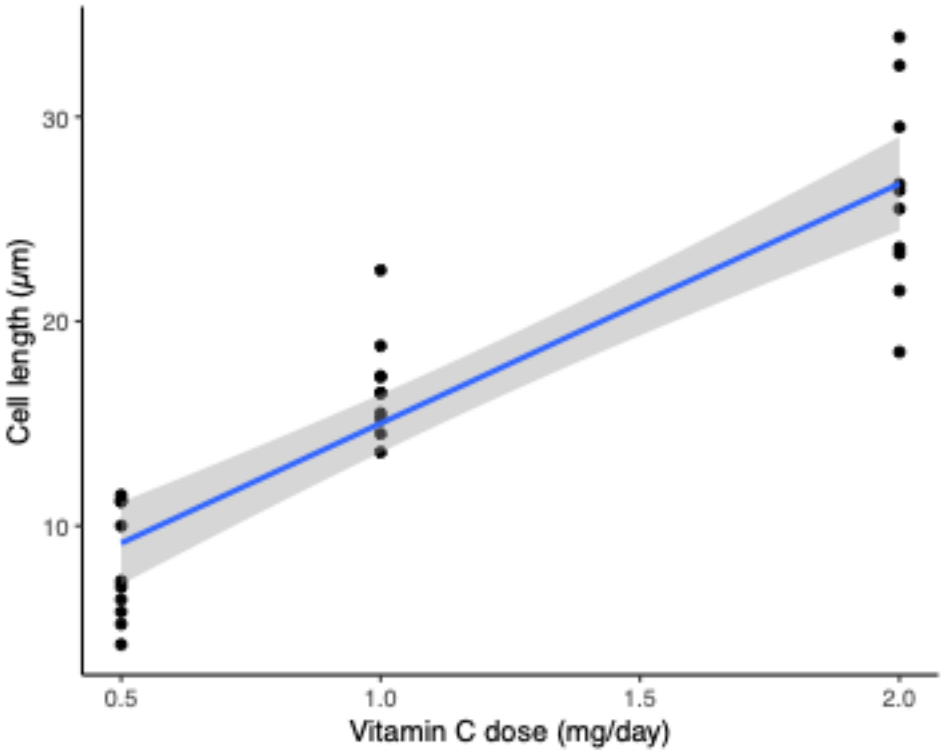
- Classic experiment giving vitamin C to look at tooth growth in guinea pigs

- Two variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)

```
> model <- lm(len ~ dose, data = ToothGrowth1)
```

- Minimises the squared distance from points to line
  - Only '**error**' in the response variable
  - Distance from the line the '**residual**'
  - Returns the slope and intercept of the line

```
> 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

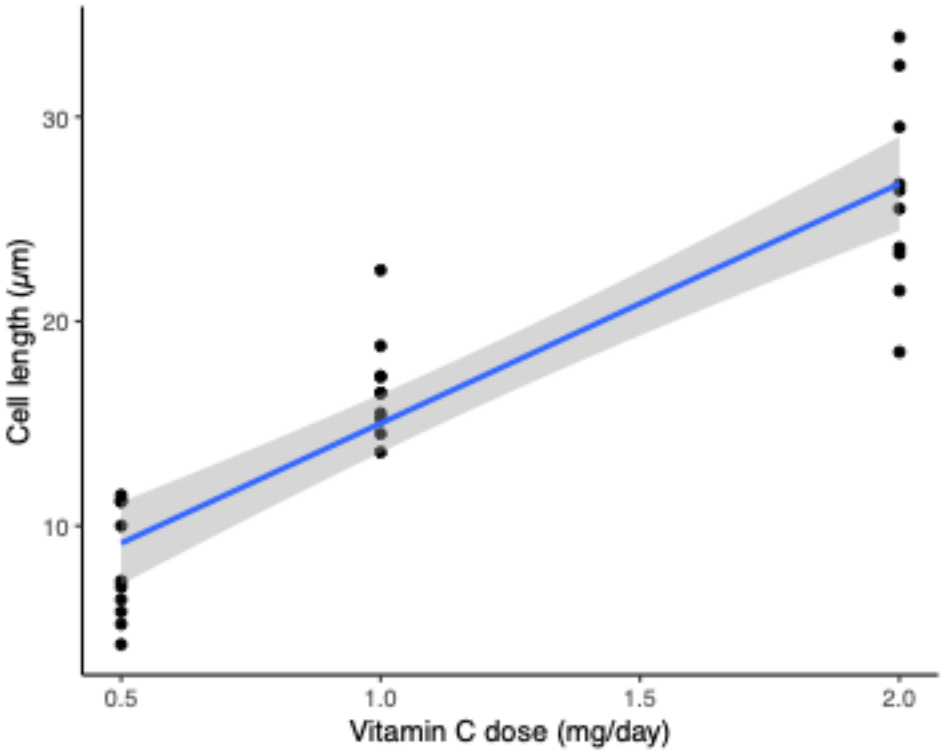
```
> model <- lm(len ~ dose, data = ToothGrowth1)
> summary(model)

Call:
lm(formula = len ~ dose, data = ToothGrowth1)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2264 -2.6029  0.0814  2.2288  7.4893

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.427   2.309   0.0285 *
dose          11.716      1.079  10.860 1.51e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.685 on 28 degrees of freedom
Multiple R-squared:  0.8082,    Adjusted R-squared:  0.8013
F-statistic: 117.9 on 1 and 28 DF,  p-value: 1.509e-11
```



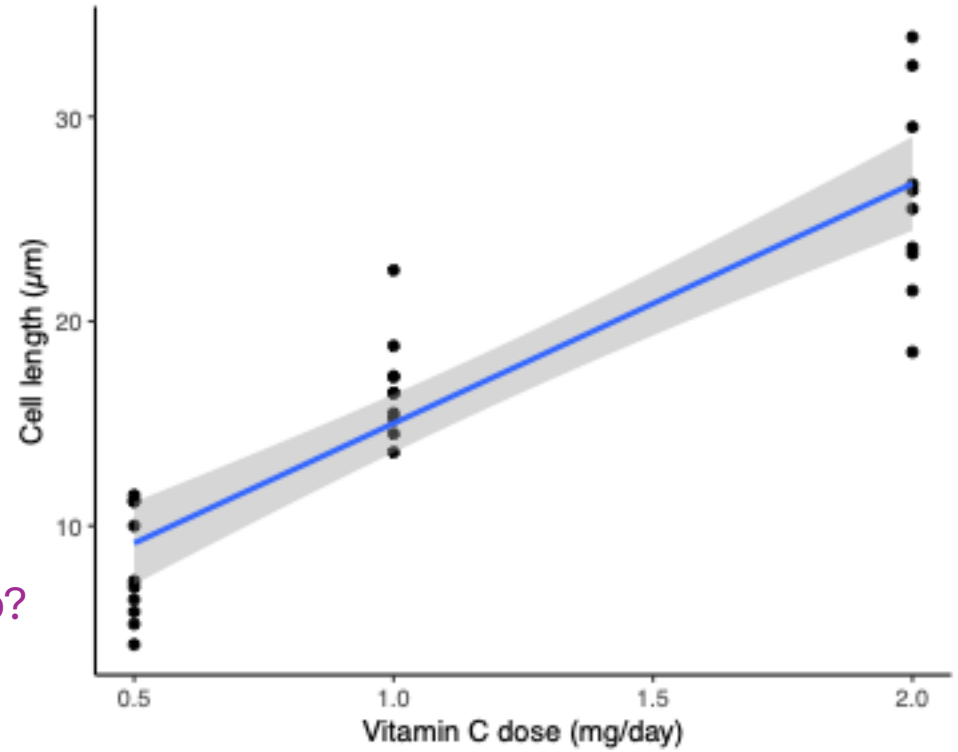
Crampton, E. W. (1947).

Example Experiment – Do the analysis

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

- Two variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.427   2.309   0.0285 *
dose          11.716      1.079  10.860 1.51e-11 ***
---
```



Crampton, E. W. (1947).

  - Does the line have a slope **significantly** different from zero?
  - What's the slope of the line?
  - What's the intercept of the line?
  - What's the uncertainty estimate on any of the above?

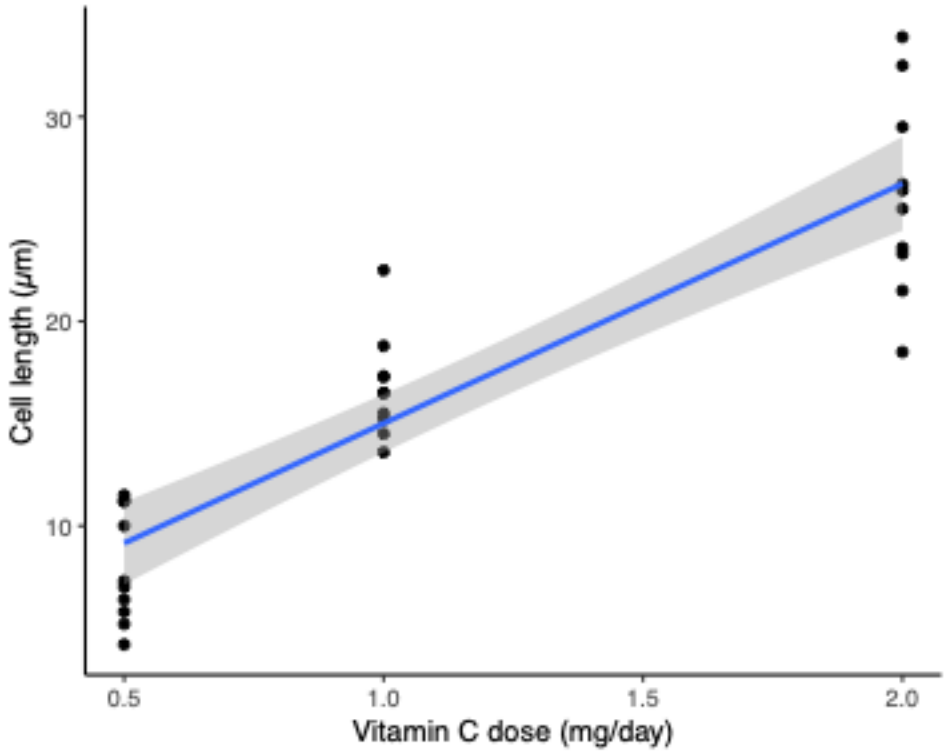
- Questions
  - Does cell length depend on vitamin C dose? Yes, $P = 1.5 \times 10^{-11}$
  - How much does cell length change with dose? 12 μm/(mg/day)
  - It would be cruel to deprive them of vitamin C entirely – Predict how much growth without any? 3.3 μm
  - How sure are we of the answers? 12 ± 1.1 μm/(mg/day) and 3.3 ± 1.4 μm SE gives grey area in plot

Example Experiment – Report it

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

- Two variables:
  - Cell length (numeric, continuous)
  - Vitamin C dose (numeric, continuous)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.295      1.427   2.309   0.0285 *
dose          11.716      1.079  10.860 1.51e-11 ***
---
```



Crampton, E. W. (1947).

- Get the right units
- Give the uncertainty
- 2 or 3 significant figures sufficient – only report what's needed
- Describe it as a **linear regression**,
  - Only one explanatory variable, if added more would be a **multiple regression**
  - or just a **linear model** (many other sorts too)
- Questions
  - Does cell length depend on vitamin C dose? Yes, $P = 1.5 \times 10^{-11}$
  - How much does cell length change with dose? 12 μm/(mg/day)
  - It would be cruel to deprive them of vitamin C entirely – Predict how much growth without any? 3.3 μm
  - How sure are we of the answers? 12 ± 1.1 μm/(mg/day) and 3.3 ± 1.4 μm SE gives grey area in plot

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