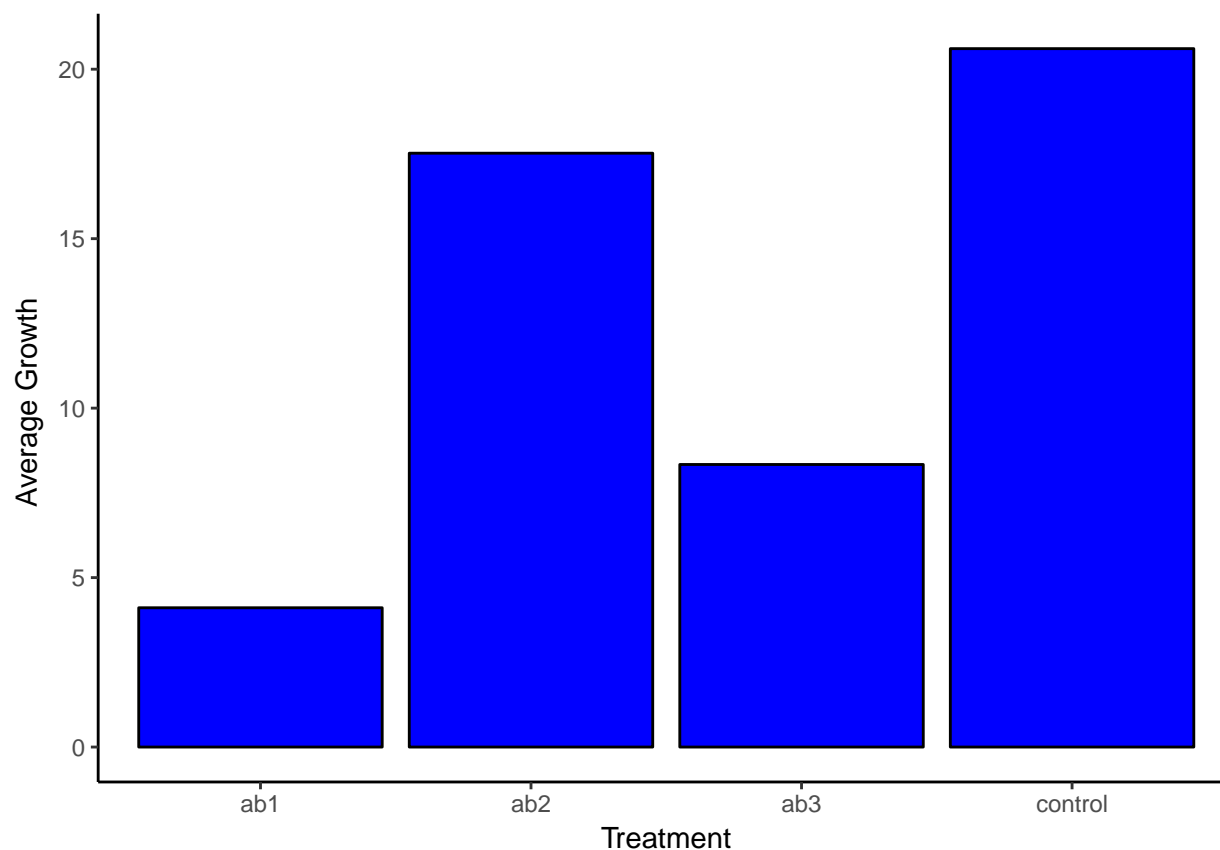# AntibioticsANOVA

A student has conducted an experiment evaluating the effect of three different antibiotics — **ab1**, **ab2**, and **ab3** — on the growth of *E. coli* in lab cultures. The results are presented in graphical format below:



We also include charts summarizing the data: one for the overall dataset, one for each type of treatment, and one for the control. Below is the overall summary:

## Factors are dropped from the summary

|        | N  | Mean  | SD   | Min  | Q1  | Median | Q3    | Max   |
|--------|----|-------|------|------|-----|--------|-------|-------|
| growth | 16 | 12.64 | 7.25 | 2.88 | 5.8 | 14.03  | 19.98 | 21.67 |

Below is the summary for **ab1**:

## Factors are dropped from the summary

|        | N | Mean | SD   | Min  | Q1   | Median | Q3   | Max  |
|--------|---|------|------|------|------|--------|------|------|
| growth | 4 | 4.11 | 1.89 | 2.88 | 3.06 | 3.31   | 5.16 | 6.93 |

Below is the summary for **ab2**:

## Factors are dropped from the summary

|        | N | Mean  | SD   | Min   | Q1    | Median | Q3    | Max   |
|--------|---|-------|------|-------|-------|--------|-------|-------|
| growth | 4 | 17.52 | 2.63 | 14.78 | 15.71 | 17.13  | 19.34 | 21.04 |

Below is the summary for **ab3**:

```
## Factors are dropped from the summary
```

|        | N | Mean | SD   | Min  | Q1   | Median | Q3    | Max   |
|--------|---|------|------|------|------|--------|-------|-------|
| growth | 4 | 8.34 | 3.64 | 4.67 | 5.82 | 7.69   | 10.85 | 13.28 |

Below is the summary for the **control**:

```
## Factors are dropped from the summary
```

|        | N | Mean  | SD   | Min   | Q1    | Median | Q3    | Max   |
|--------|---|-------|------|-------|-------|--------|-------|-------|
| growth | 4 | 20.61 | 0.91 | 19.46 | 19.98 | 20.65  | 21.23 | 21.67 |

By examining the data, we might expect **ab1** and **ab3** to be significantly different from the control group; **ab2** does not appear significantly different. Thus we should expect that our ANOVA will return that the means of at least two groups differ. Let $\mu_0$ represent the mean growth of the control group, $\mu_1$ represent the mean of the group treated by **ab1**, $\mu_2$ represent the mean of the group treated by **ab2**, and $\mu_3$ represent the mean of the group treated by **ab3**. Then we have our hypotheses:

**Null Hypothesis** ($H_0$)**:** There is not a significant difference in the means of the data: $\mu_0 = \mu_1 = \mu_2 = \mu_3$.

**Alternate Hypothesis** ($H_1$)**:** At least one of the antibiotics **ab1**, **ab2**, **ab3** has a mean significantly different from the control group. That is, for some $i \in \{1, 2, 3\}$ we have $\mu_i \neq \mu_0$.

To test these hypotheses we ran an ANOVA-design linear model, and compared the results using a likelihood ratio test. Let $G$ denote growth, let $X_i$ denote treatment by **abi** for $i \in \{1, 2, 3\}$, and let $\epsilon$ denote a normally distributed error term. We fit a null model of the form $G = \beta_0$ to this data, optimized for the value of $\beta_0$ which minimized the negative log likelihood, and then stored that negative log likelihood value.

For our alternate model, we fit a generalized linear model of the form

$$G = g(X_1, X_2, X_3,) = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \epsilon$$

Note that this is a function $g : \mathbb{R}^3 \to \mathbb{R}$ so we defined our domain as a subset of $\mathbb{R}^3$. $X_1$ took on the values $(0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$, $X_2$ took on the values $(0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0)$, and $X_3$ took on the values $(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1)$. The values of 1 in these vectors correspond to the values of $G = g(X_1, X_2, X_3)$ for which the treatments **ab1**, **ab2**, and **ab3** were affected, respectively.

We optimized the model $g(X_1, X_2, X_3)$ for values of the $\beta_i$ which minimized the negative log likelihood and then stored this value. Then we found $D$, which is twice the difference of our two stored negative log likelihood values. Using a one-tailed $\chi^2$ distribution with 3 degrees of freedom — the difference in the amount of parameters between the models — we found the p-value associated with a $\chi$-score of $D$. The p-value was approximately $2.97 \cdot 10^{-8}$.

As the p-value is smaller than 0.05, we reject the null hypothesis and conclude that there is a significant difference between at least two of the means associated with different treatment types.