# 2 Way ANOVA (Highlights from Sleuth3 Chapter 13)

**1 way vs 2 way ANOVA**

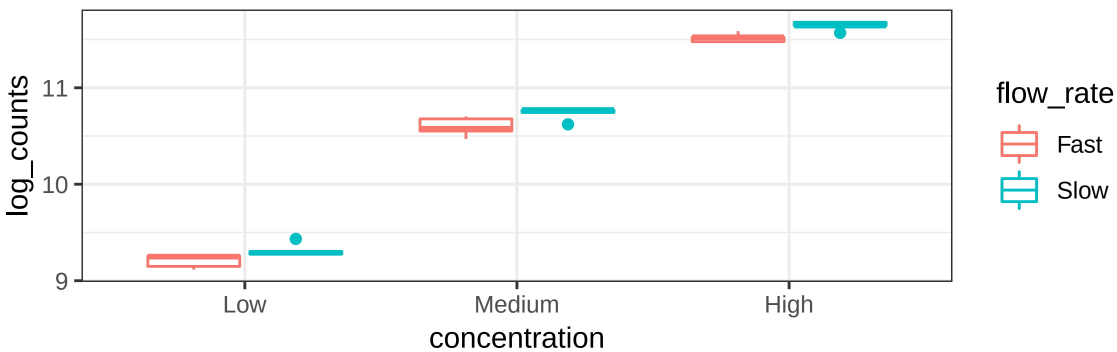
- 1 way ANOVA model
  - Quantitative response, categorical explanatory
  - Each group has its own mean, data are normally distributed around the group mean
- 2 way ANOVA model
  - Quantitative response, 2 categorical explanatory
  - Each combination of groups has its own mean, data are normally distributed around the mean for that combination of groups
- Actually everything is just a linear model

**Additive Model vs. Interactions Model**

**Additive Model:** `lm(response ~ explanatory1 + explanatory2, data = data)`

- Group mean = effect from one variable + ~~effect from second variable~~
- Visually: Difference in means for levels of variable 1 is constant across levels of variable 2
- Example: a calibration experiment was performed to explore the relationship between
  - the recorded `counts` from a gas chromatograph (response) – we use a log transformation to stabilize variance
  - the `concentration` of a compound in a mixture (Low, Medium, or High) and the `flow_rate` through the chromatograph (Slow or Fast)

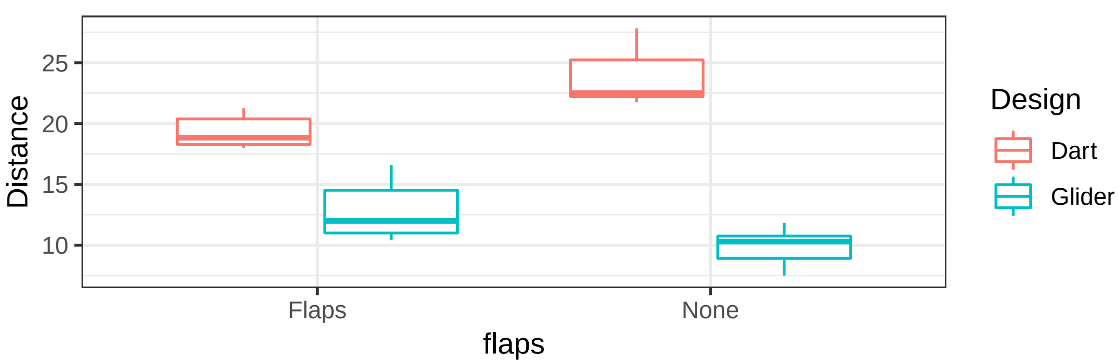
```
ggplot(data = chromatography, mapping = aes(x = concentration, y = log_counts, color = flow_rate)) +
  geom_boxplot() +
  theme_bw()
```




**Interactions Model:** `lm(response ~ explanatory1 * explanatory2, data = data)`

- Group mean = specific to combination of levels for variable 1 and 2
- Visually: Difference in means for levels of variable 1 varies across levels of variable 2
- Example: A motivated paper airplane thrower measured
  - The `Distance` travelled (response)
  - The `Design` (dart or glider) and whether or not flaps were put on the wings (Flaps or None)

```
ggplot(data = planes, mapping = aes(x = flaps, y = Distance, color = Design)) +
  geom_boxplot() +
  theme_bw()
```



### Fit group means on transformed scale

```r
lm_fit <- lm(log_counts ~ concentration + flow_rate, data = chromatography)
summary(lm_fit)
```

```
##
## Call:
## lm(formula = log_counts ~ concentration + flow_rate, data = chromatography)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.133752 -0.050017  0.004214  0.048191  0.108745
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          9.19898    0.02384 385.810  < 2e-16 ***
## concentrationMedium  1.40414    0.02920  48.084  < 2e-16 ***
## concentrationHigh    2.31775    0.02920  79.370  < 2e-16 ***
## flow_rateSlow        0.12576    0.02384   5.274 1.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0653 on 26 degrees of freedom
## Multiple R-squared:  0.996,  Adjusted R-squared:  0.9955
## F-statistic:  2140 on 3 and 26 DF,  p-value: < 2.2e-16
```

**1. Write down the estimated equation for the mean of `log(counts)` based on concentration and flow rate.**

**2. Express each of the following estimated means in terms of coefficient estimates above:**

$\hat{\mu}(log(\text{counts})|\text{concentration} = \text{"Low"}, \text{flow\_rate} = \text{"Fast"}) =$

$\hat{\mu}(log(\text{counts})|\text{concentration} = \text{"Medium"}, \text{flow\_rate} = \text{"Fast"}) =$

$\hat{\mu}(log(\text{counts})|\text{concentration} = \text{"High"}, \text{flow\_rate} = \text{"Fast"}) =$

$\hat{\mu}(log(\text{counts})|\text{concentration} = \text{"Low"}, \text{flow\_rate} = \text{"Slow"}) =$

$\hat{\mu}(log(\text{counts})|\text{concentration} = \text{"Medium"}, \text{flow\_rate} = \text{"Slow"}) =$

$\hat{\mu}(log(\text{counts})|\text{concentration} = \text{"High"}, \text{flow\_rate} = \text{"Slow"}) =$

**3. Conduct a test of the claim that for a given flow rate, the mean log counts is the same at all three concentrations.**

```r
lm_fit_reduced <- lm(log_counts ~ flow_rate, data = chromatography)
anova(lm_fit_reduced, lm_fit)
```

```
## Analysis of Variance Table
##
## Model 1: log_counts ~ flow_rate
## Model 2: log_counts ~ concentration + flow_rate
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     28 27.3717
## 2     26  0.1109  2    27.261 3196.8 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```