simon-5502-05-slides

2025-02-18

Topics to be covered

- What you will learn
  - Mathematical model of interactions
  - Interactions in multi-factor analysis of variance
  - Interactions in analysis of covariance
  - Interactions in multiple linear regression

Mathematical model, 1

- $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
  - i=1,…,a j=1,…,b, k=1,…,n
- If 1 is the reference category
  - $\alpha_1 = 0$
  - $\beta_1 = 0$
  - $(\alpha\beta)_{1j} = 0$
  - $(\alpha\beta)_{i1} = 0$

Speaker notes

You may see papers or books that present the mathematical model for an interaction. The model I present is a balanced model with the first category having levels one through a, the second category having levels one through b and for each combination of categories there are n observations.

If you set the first level as the reference category for each category, then you need to set some of these parameters to zero.

Mathematical model, 2

- $SS_A = \Sigma_i nb(\bar{Y}_{i..} - \bar{Y}_{...})^2$
- $SS_B = \Sigma_i na(\bar{Y}_{.j.} - \bar{Y}_{...})^2$
- $SS_{AB} = \Sigma_i \Sigma_j n(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{ij.})^2$
- $SS_E = \Sigma_i \Sigma_j \Sigma_k (Y_{ijk} - \bar{Y}_{ij.})^2$
- $SS_T = \Sigma_i \Sigma_j \Sigma_k (Y_{ijk} - \bar{Y}_{...})^2$

Speaker notes

The dot notation may be a bit confusing until you get used to it, but $\bar{Y}_{i..}$ is the average within the ith group, averaging across the subscripts j and k. $\bar{Y}_{.j.}$ is the average within the ith group, averaging across the subscripts i and k. $\bar{Y}_{ij.}$ is the average within the combination of the ith group and the jth group, averaging across the subscript k. Finally, $\bar{Y}_{...}$ is an overall mean and the average across all three subscripts.

Test for an interaction

- SS_{AB} has (a-1)(b-1) degrees of freedom
- SS_E has ab(n-1) degrees of freedom
- Accept H_0 if $F = \frac{MS_{AB}}{MS_E}$ is close to one
  - In R, fit a model without an interaction
  - Compare to a model with interaction
  - Using the anova function

Speaker notes

The formal test for an interaction uses an F ratio and you accept the null hypothesis if that F ratio is close to one. You would reject the null hypothesis if the F ratio is much larger than one.

It is not easy to get R to display all the sums of squares and mean squares that I defined above. Instead, compute two models-one without an interaction and one with an interaction. Compare those two models using the anova function.

Break #1

- What you have learned
  - Mathematical model of interactions
- What's coming next
  - Interactions in multi-factor analysis of variance

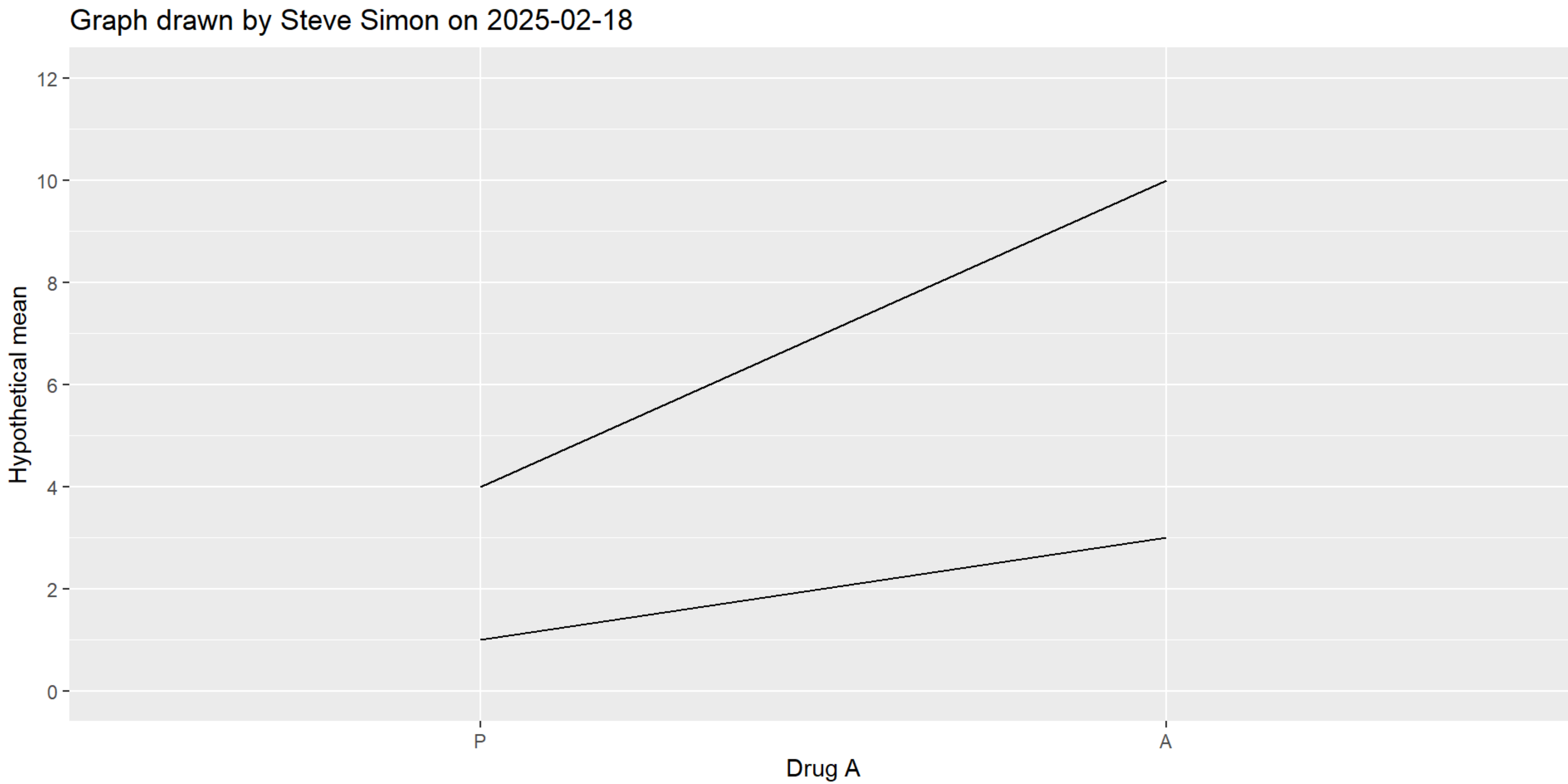
Hypothetical applications

- Study with two drugs and two placebos
- Four treatment options
  - Placebo plus placebo
  - Drug A plus placebo
  - Placebo plus Drug B
  - Drug A plus Drug B

Listing of hypothetical data

```
# A tibble: 8 × 3
  a     b         y
  <fct> <fct> <dbl>
1 P     P       0.9
2 P     P       1.1
3 A     P       2.9
4 A     P       3.1
5 P     B       3.9
6 P     B       4.1
7 A     B       9.9
8 A     B      10.1
```

Line plot of hypothetical data, 1



Create indicator variables

```
# A tibble: 8 × 6
  a     b       i_a   i_b  i_ab     y
  <fct> <fct> <dbl> <dbl> <dbl> <dbl>
1 P     P         0     0     0   0.9
2 P     P         0     0     0   1.1
3 A     P         1     0     0   2.9
4 A     P         1     0     0   3.1
5 P     B         0     1     0   3.9
6 P     B         0     1     0   4.1
7 A     B         1     1     1   9.9
8 A     B         1     1     1  10.1
```

Interaction model, 1

- `lm(y=a+b+a:b, data=hyp)`
- `lm(y=a*b, data=hyp)`
- `lm(y=i_a+i_b+i_ab, data=hyp_1)`

```
# A tibble: 4 × 2
  term        estimate
  <chr>          <dbl>
1 (Intercept)     1.00
2 i_a             2
3 i_b             3
4 i_ab            4
```

Interaction model, 2

- (Intercept): estimated average effort of two placebos
- i_a: estimated average change due to drug A alone
- i_b: estimated average change due to drug B alone
- i_ab: estimated synergistic effect of both drugs

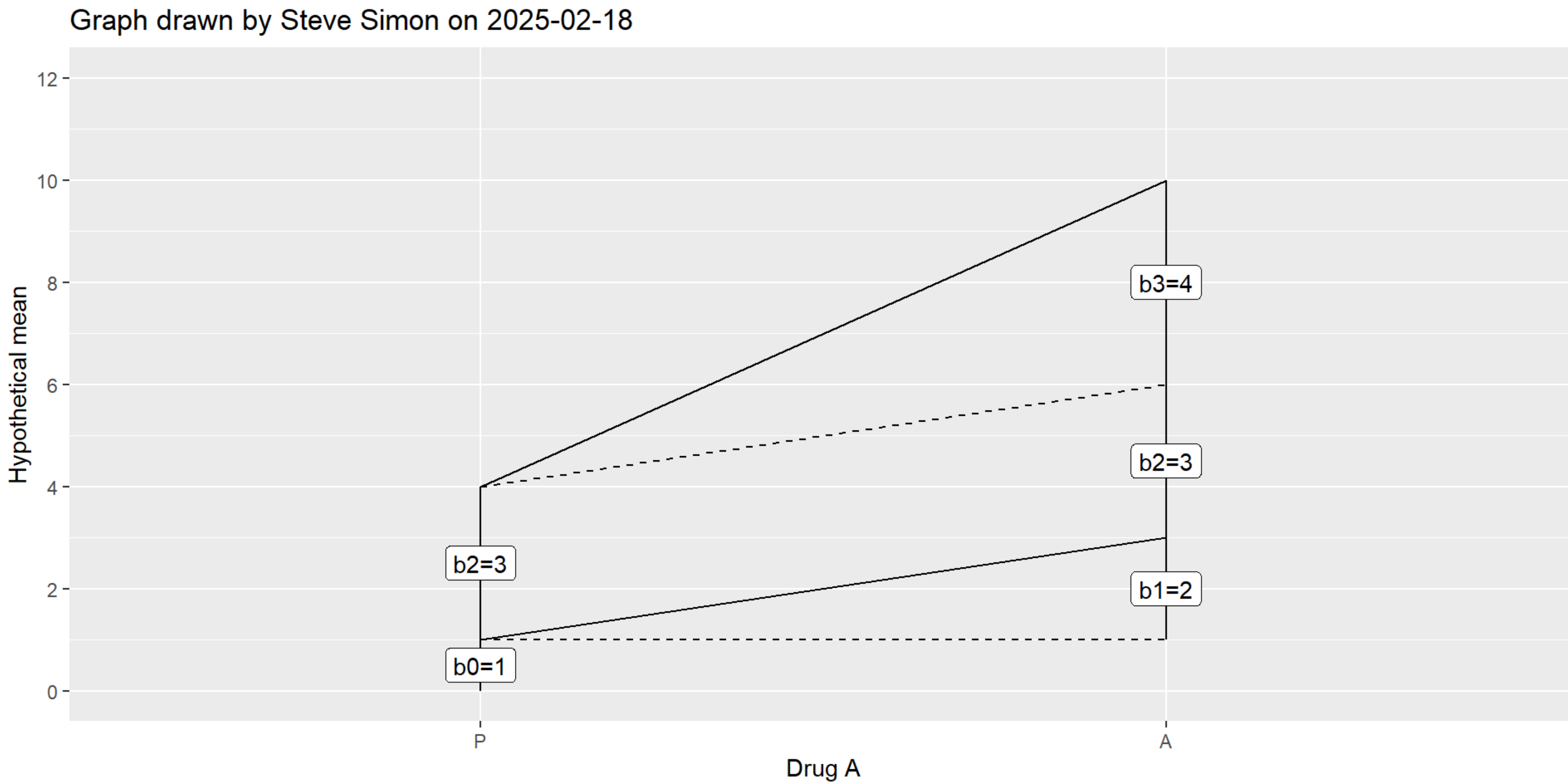
Speaker notes

The intercept is the estimated average value of Y when all of the indicator variables are equal to zero. This is the average in the P P group. The slope associated with i_a is the estimated average change when moving from placebo to drug A, holding drug B constant. The slope associated with i_b is the estimated average change when moving from placebo to drug B, holding drug A constant. The slope associated with i_ab is the interaction, a measure of how much more the effect of the two drugs combined compared to the effect of drug A along plus drug B alone.

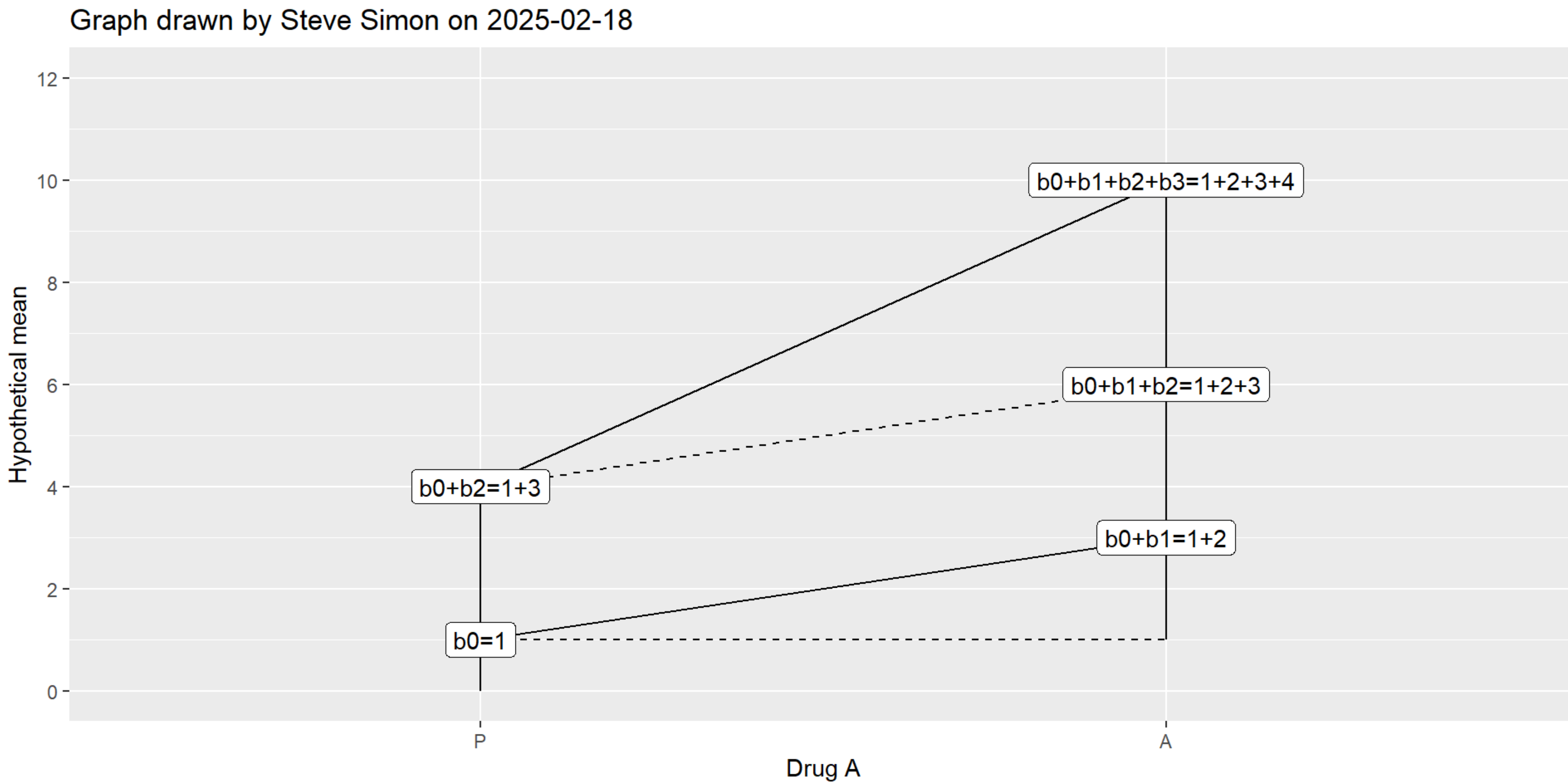
Interaction model, 3

- `P P mean = 1`
- `A P mean = 1 + 2`
- `P B mean = 1     + 3`
- `A B mean = 1 + 2 + 3 + 4`

Line plot of hypothetical data, 2



Line plot of hypothetical data, 3



Study of three exercise programs

- Swimming
- Jogging
- Reading (control)

Descriptive statistics for exercise program

```
# A tibble: 3 × 2
  prog     pct
  <fct>    <glue>
1 Reading  33% (300/900)
2 Jogging  33% (300/900)
3 Swimming 33% (300/900)
```

Descriptive statistics for hours

```
# A tibble: 1 × 4
  hours_mean hours_sd hours_min hours_max
       <dbl>    <dbl>     <dbl>     <dbl>
1       2.00    0.495     0.175      4.07
```

Descriptive statistics for gender

```
# A tibble: 2 × 2
  gender pct
  <fct>  <glue>
1 Male   50% (450/900)
2 Female 50% (450/900)
```

Descriptive statistics for effort

- 0 = no effort, 50=maximum effort

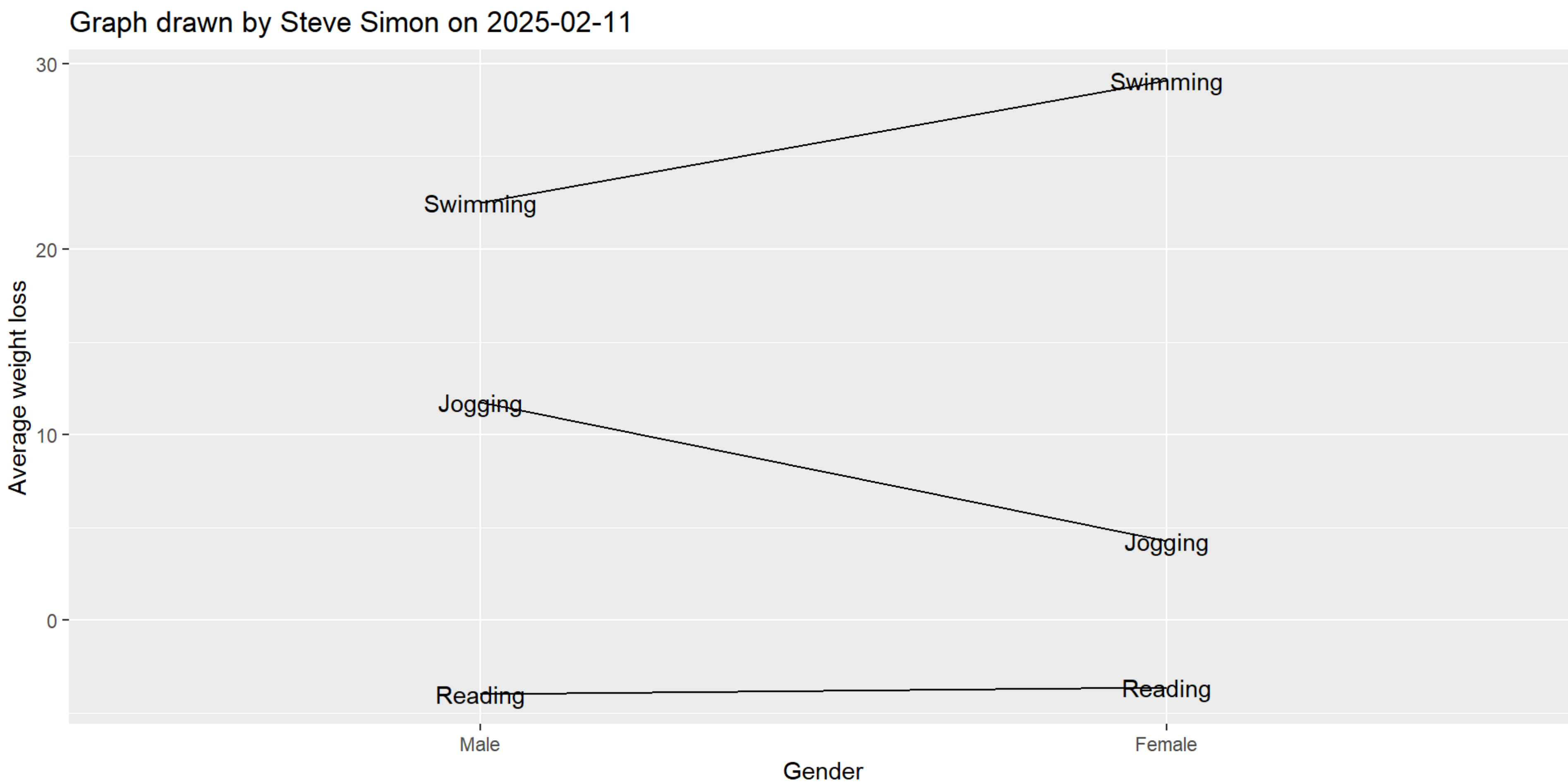
```
# A tibble: 1 × 4
  effort_mean effort_sd effort_min effort_max
        <dbl>     <dbl>      <dbl>      <dbl>
1        29.7      5.14       12.9       44.1
```

Descriptive statistics for loss

- A negative value is a weight gain

```
# A tibble: 1 × 4
  loss_mean loss_sd loss_min loss_max
      <dbl>   <dbl>    <dbl>    <dbl>
1      10.0    14.1    -17.1     54.2
```

Line plot, weight loss by program and gender



Interaction model

- Reference categories: male and reading

```
# A tibble: 6 × 5
  term                         estimate std.error statistic p.value
  <chr>                           <dbl>     <dbl>     <dbl> <glue>
1 (Intercept)                    -3.96      0.532    -7.43  p < 0.001
2 genderFemale                    0.335     0.753     0.446 p = 0.656
3 progJogging                    15.7       0.753    20.9   p < 0.001
4 progSwimming                   26.5       0.753    35.2   p < 0.001
5 genderFemale:progJogging       -7.82      1.06     -7.35  p < 0.001
6 genderFemale:progSwimming       6.26      1.06      5.88  p < 0.001
```

Live demo, Interactions in multi-factor analysis of variance

Break #2

- What you have learned
  - Interactions in multi-factor analysis of variance
- What's coming next
  - Interactions in analysis of covariance

Interactions in analysis of covariance, 1

Interactions in analysis of covariance, 2

Coefficients and interpretation

```
NULL
```

Live demo, Interactions in analysis of covariance

Break #3

- What you have learned
  - Interactions in analysis of covariance
- What's coming next
  - Interactions in multiple linear regression

Interactions in multiple linear regression

- Center your continuous predictors

```
# A tibble: 4 × 5
  term          estimate std.error statistic p.value
  <chr>            <dbl>     <dbl>     <dbl> <glue>
1 (Intercept)    16.8      0.436      38.6   p < 0.001
2 hours           6.87     0.926       7.42 p < 0.001
3 effort          0.890    0.0855     10.4   p < 0.001
4 hours:effort    0.373    0.187       2.00 p = 0.046
```

Summary

- What you have learned
  - Mathematical model of interactions
  - Interactions in multi-factor analysis of variance
  - Interactions in analysis of covariance
  - Interactions in multiple linear regression