One-Way ANOVA

PSYC 2020-A01 / PSYC 6022-A01 | 2025-11-21 | Lab 14

Jessica Helmer

Outline

- Assignment 13 Review
- One-Way ANOVA

Learning objectives:

R: One-Way ANOVA

Housekeeping

Our very last lab!!

I will not have office hours 11/26 (institute holiday), and 12/03 will be my final office hours (might be virtual only; will send out an announcement if so).

- Still happy to other schedule time if anyone has any questions or concerns

Extra Credit Opportunity: If class gets to 85% completion on CIOS, I will add 1 point (10% of a lab assignment!) to everyone's grade.

- I very much appreciate your feedback on this lab!
- CIOS should open 11/24

Assignment 13 Review

Great work!

Remember the formula syntax: `outcome (DV) ~ predictor (IV)` or `outcome + predictor_1 + predictor_2 + ... + predictor_p`

- Think carefully about which variable is your outcome and which is (are) your predictor(s)

ANOVA

Analysis of Variance (ANOVA)

Analysis of Variance (ANOVA)

Method of understanding how categorical variables are related to a continuous outcome of interest

Compares *between groups* to *within groups* variance to jointly compare multiple means

ANOVA: Partitioning Variance

Total Sum of Squared Errors: error term around the grand mean

$$SS_{total} = \sum (x_i - \bar{x})^2$$

If we have outcomes separated into different groups (e.g., control, treatment A, treatment B), can partition SS_{total} into *between groups sums of squares* SS_B and *within groups sums of squares* SS_W

$$SS_{total} = SS_B + SS_W$$

So, with j indexing J total groups and i indexing I people within a group, we can write SS_{total} as…

ANOVA: Partitioning Variance

Total Sum of Squared Errors: error term around the grand mean

$$SS_{total} = \sum_{j=1}^{J} \sum_{i=1}^{I} (x_{ij} - \bar{x})^2$$

Between-Group Sum of Squares: error between the group mean and the grand mean

$$SS_B = \sum_{j=1}^{J} n_j (\bar{x_j} - \bar{x})^2$$

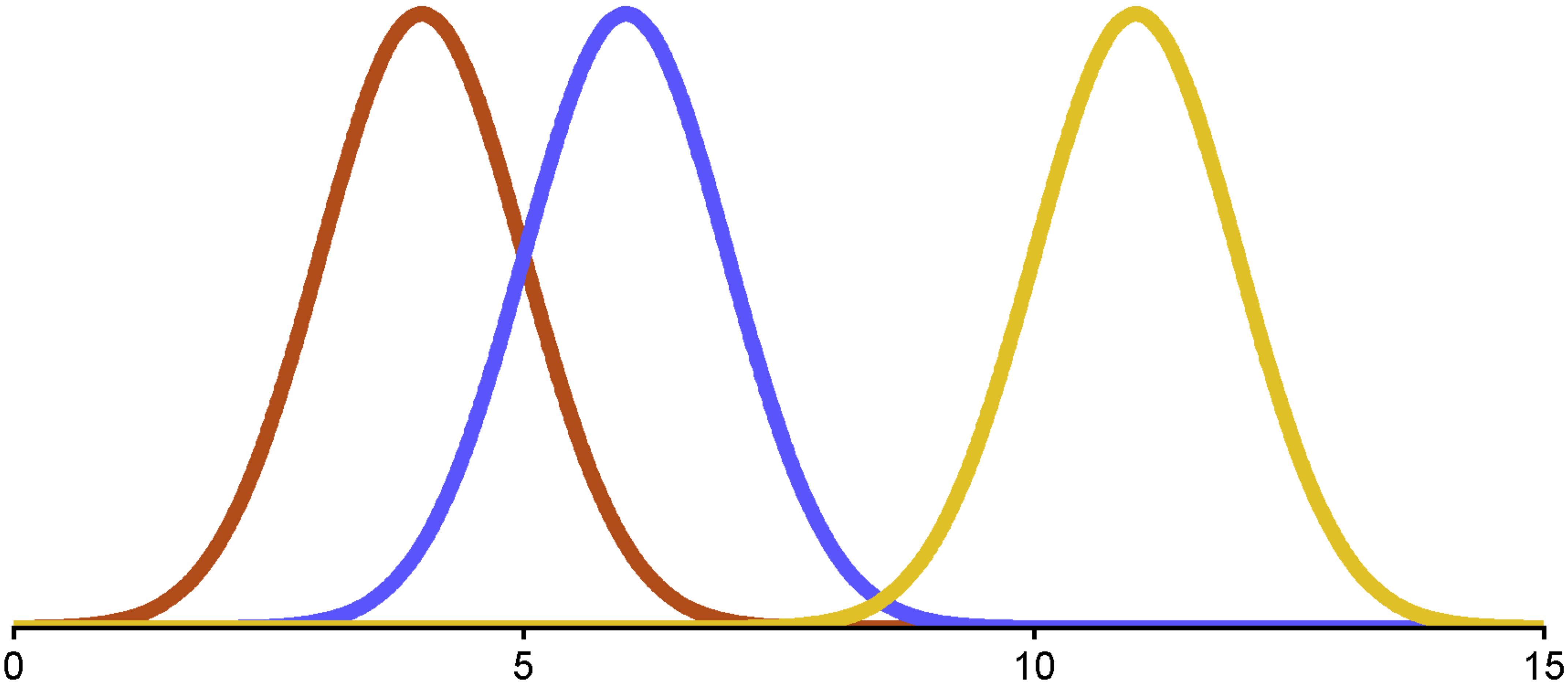
Within-Group Sum of Squares: error between observations and their group mean

$$SS_W = \sum_{j=1}^{J} \sum_{i=1}^{I} (x_{ij} - \bar{x_j})^2$$

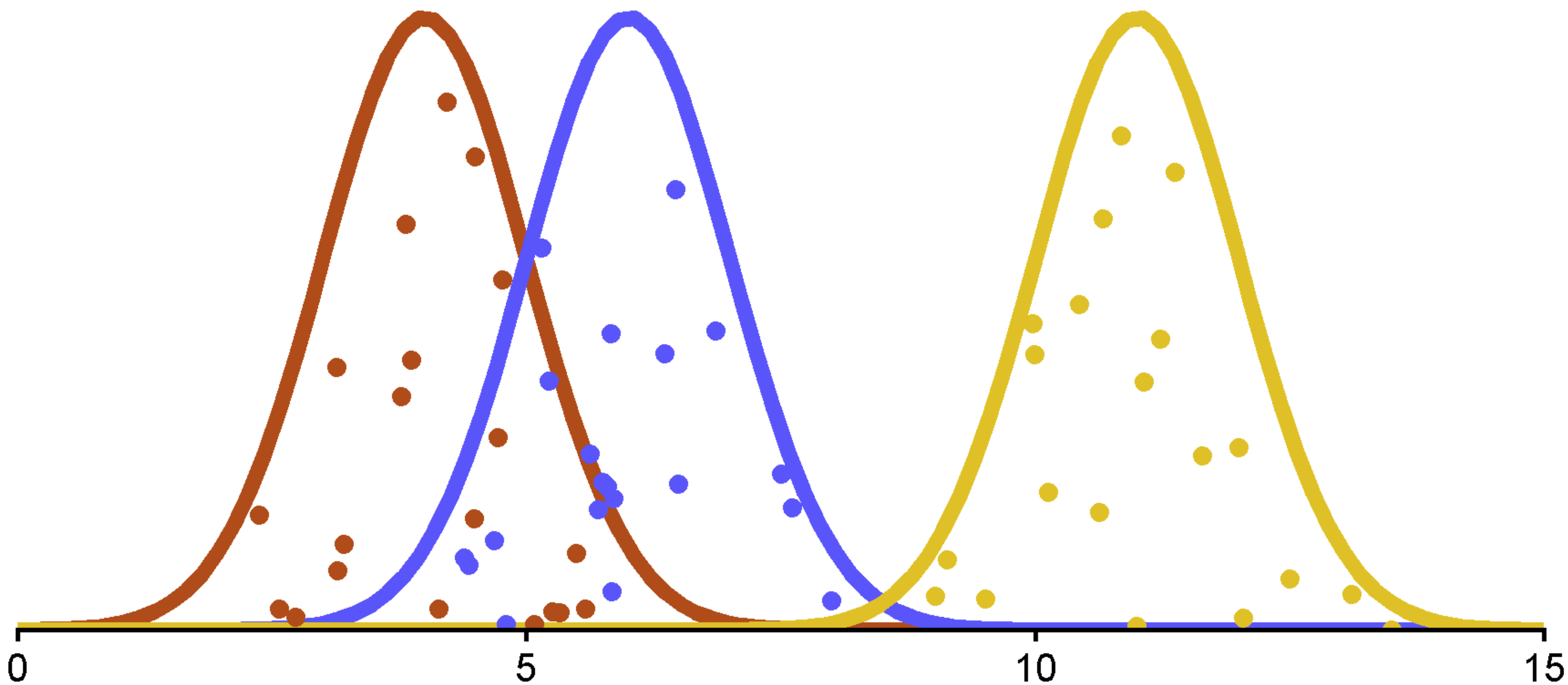
ANOVA: Partitioning Variance

Let’s say we have three groups and some outcome variable

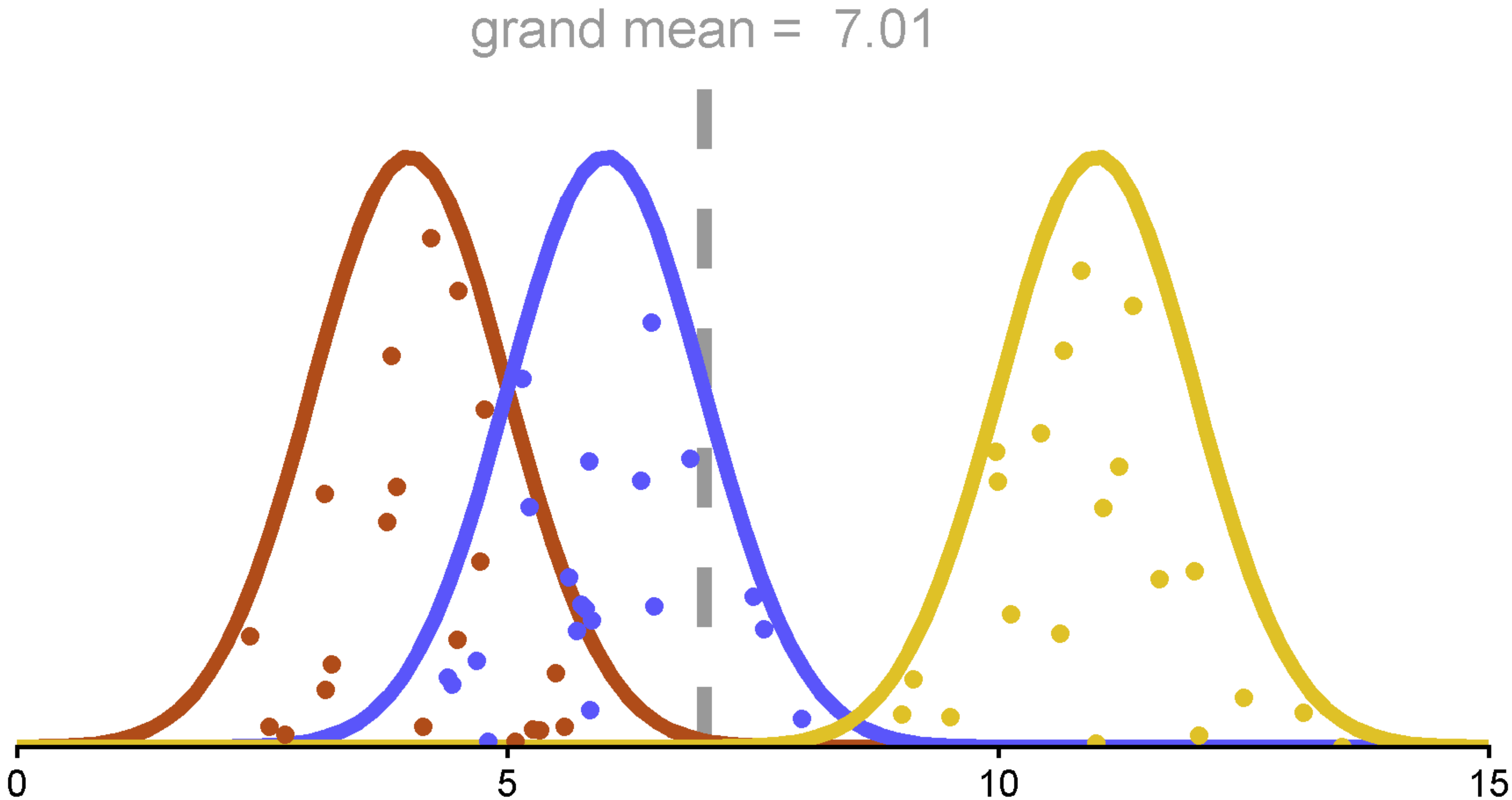
ANOVA: Partitioning Variance



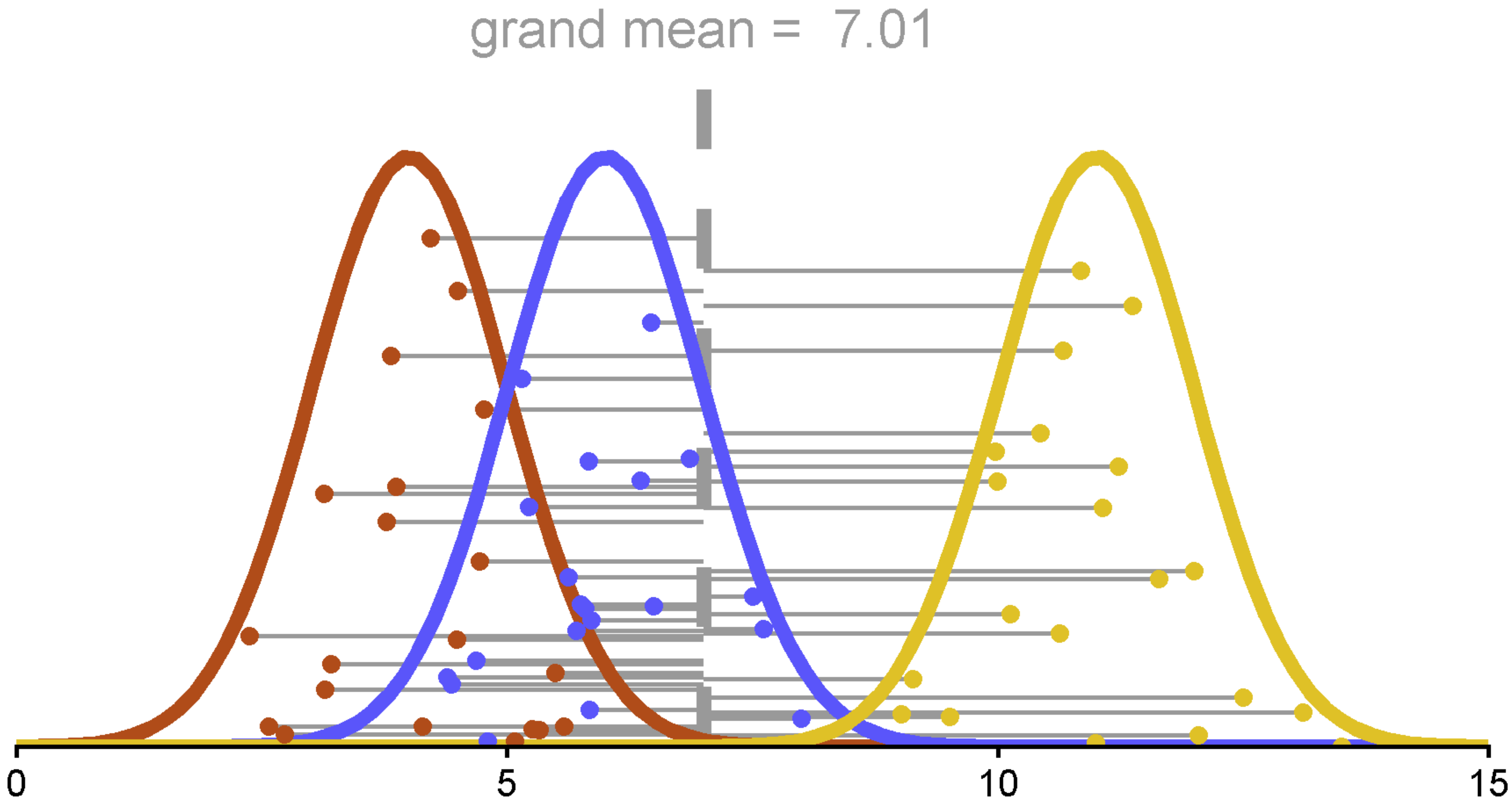
ANOVA: Partitioning Variance



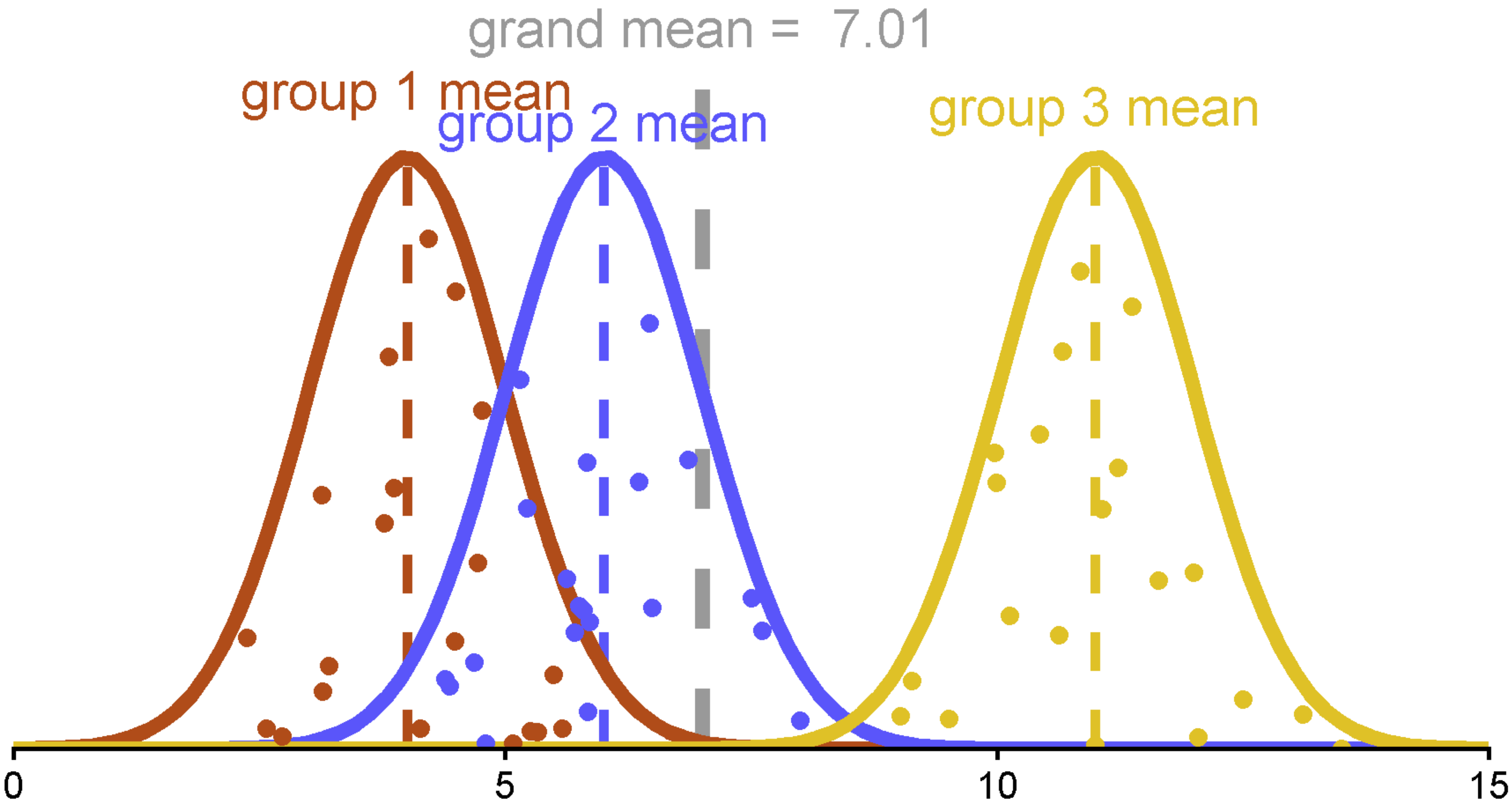
ANOVA: Partitioning Variance



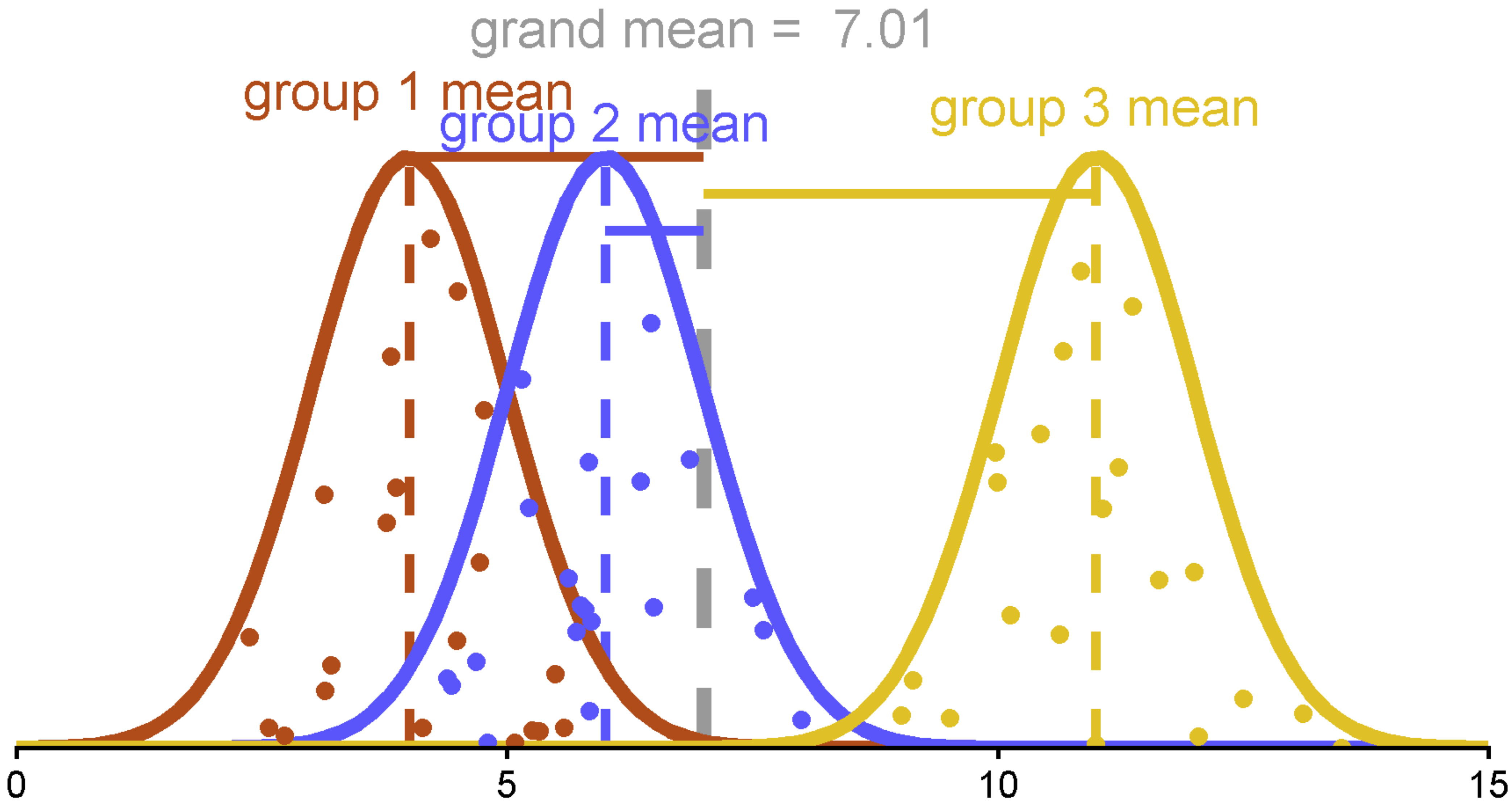
ANOVA: Partitioning Variance: SS_{total}



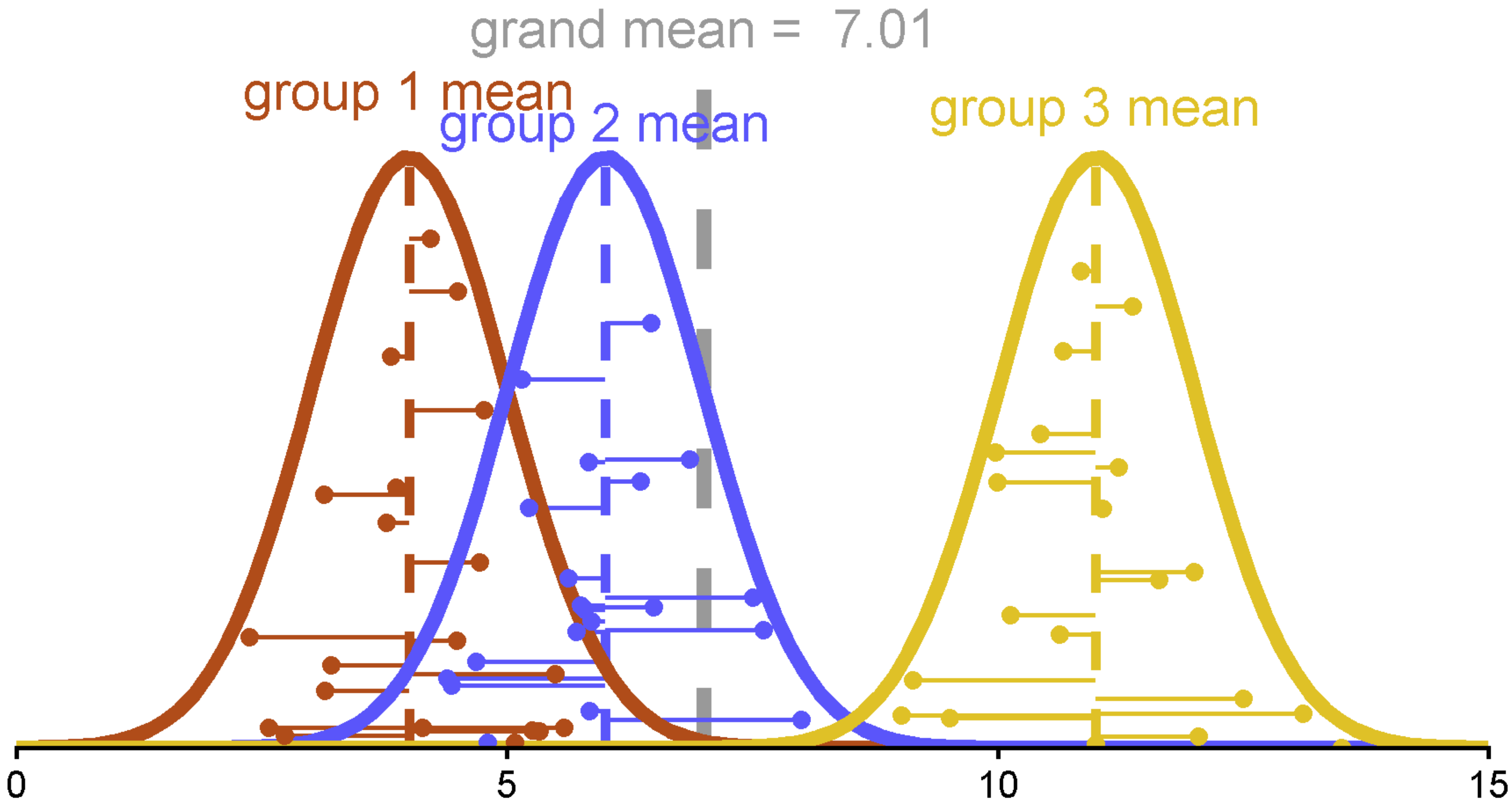
ANOVA: Partitioning Variance



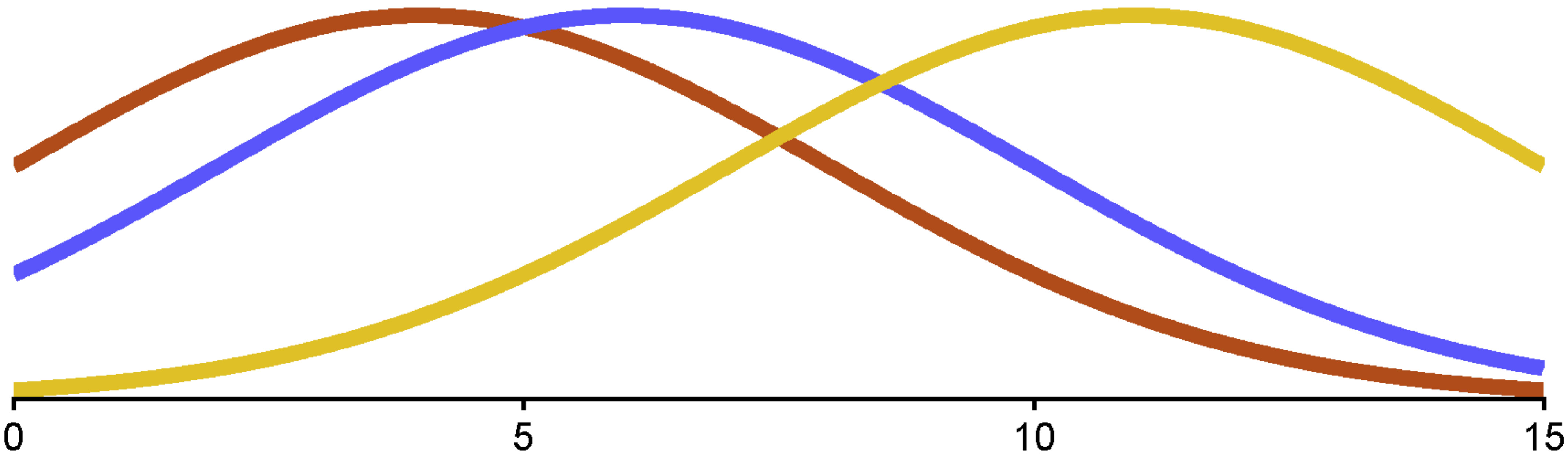
ANOVA: Partitioning Variance: SS_B



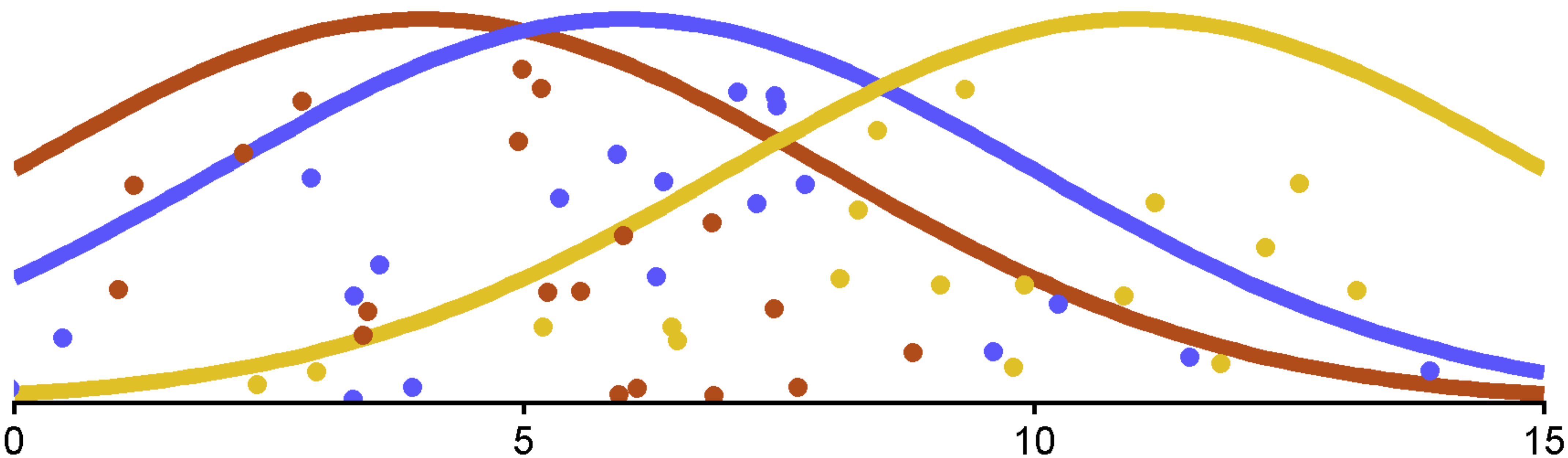
ANOVA: Partitioning Variance: SS_W



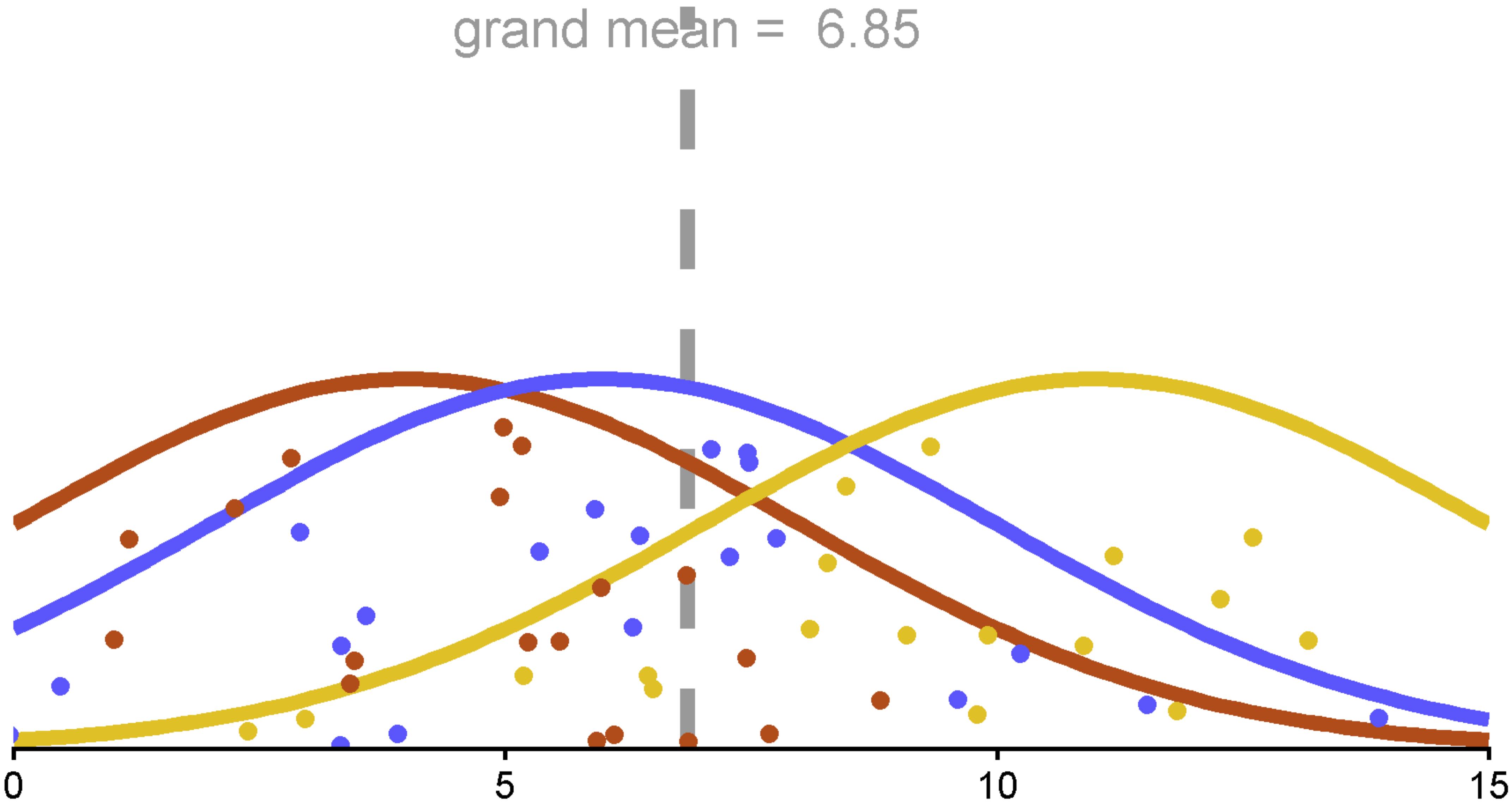
ANOVA: Partitioning Variance



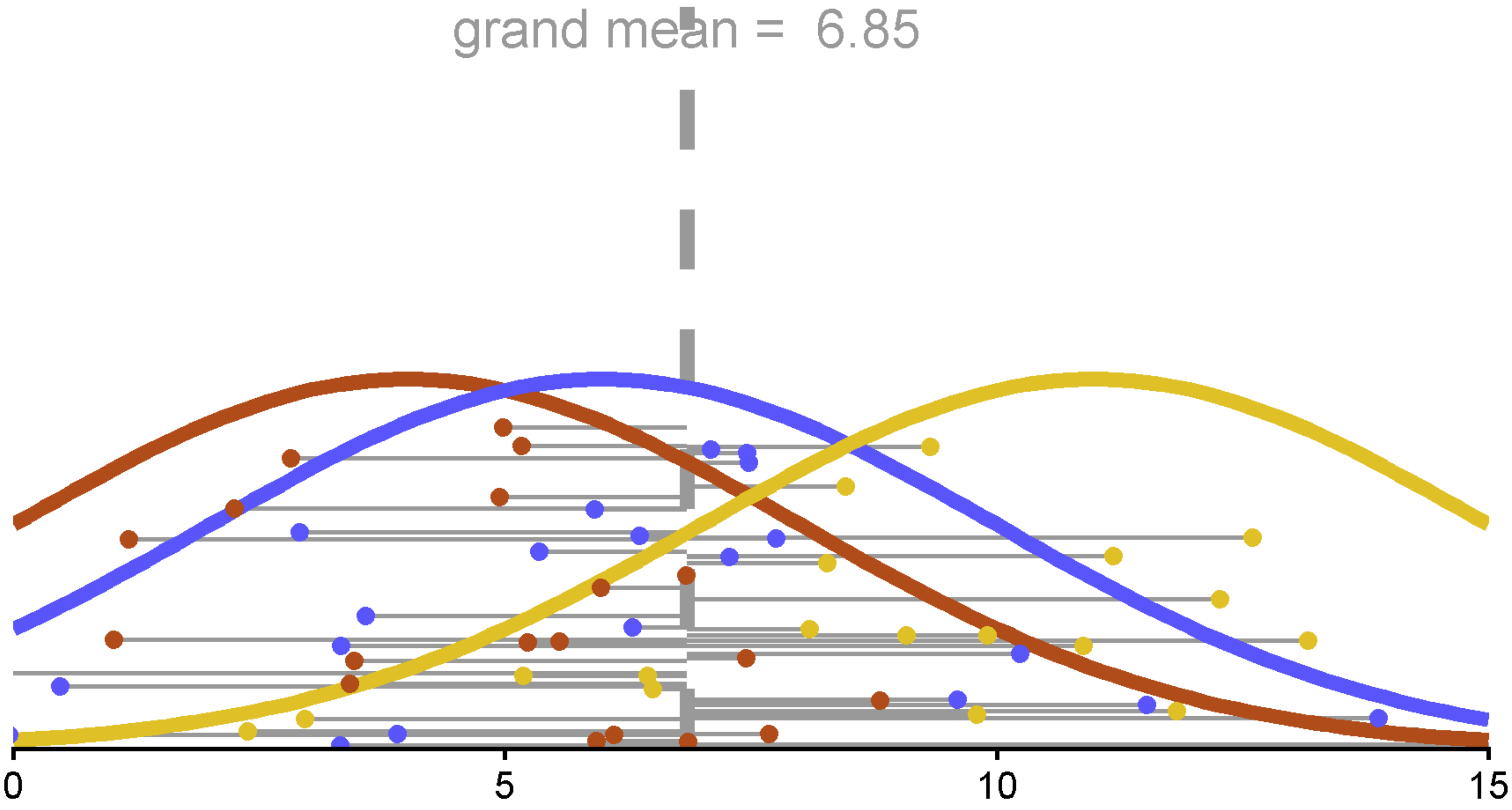
ANOVA: Partitioning Variance



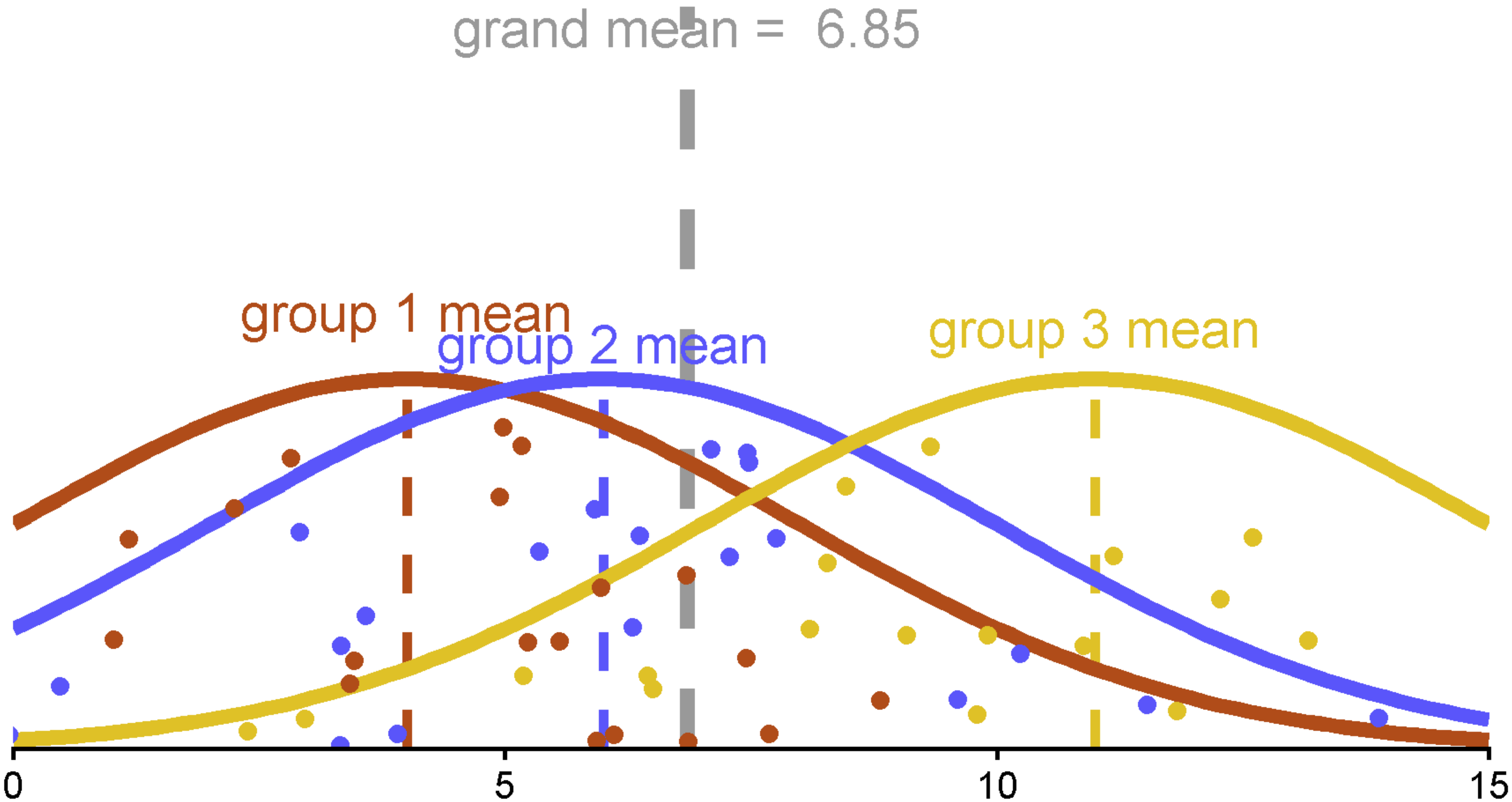
ANOVA: Partitioning Variance



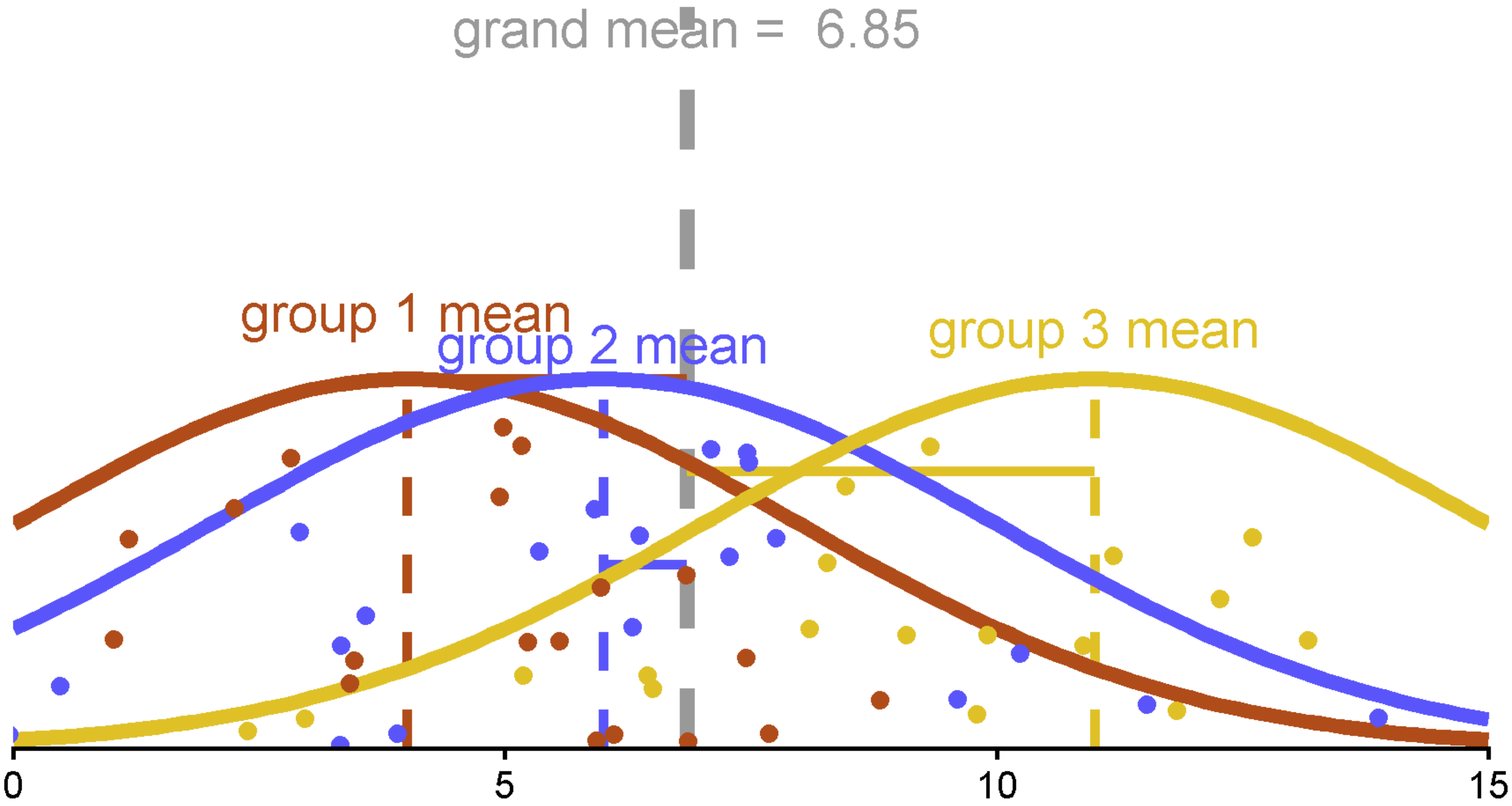
ANOVA: Partitioning Variance: SS_{total}



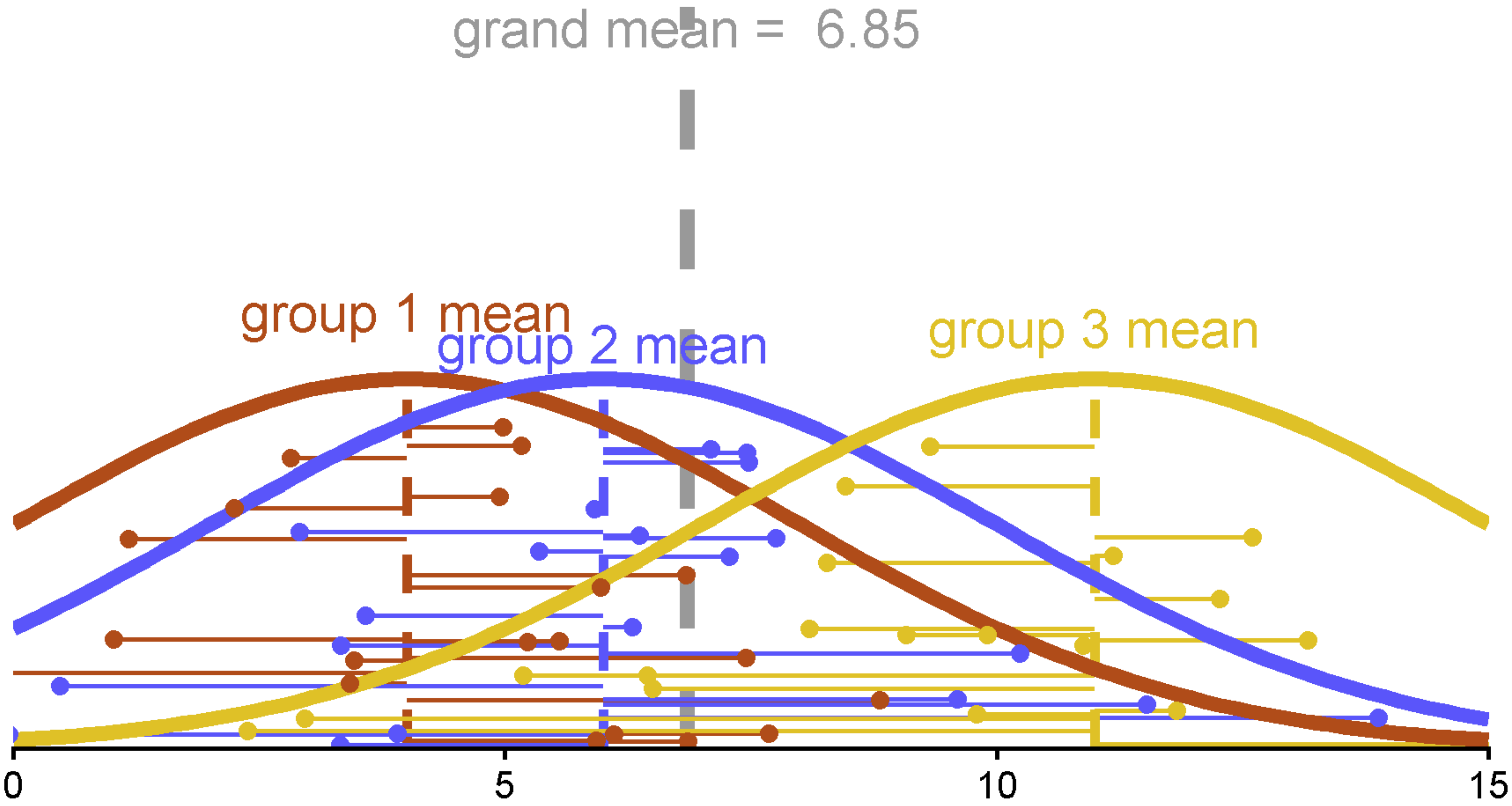
ANOVA: Partitioning Variance



ANOVA: Partitioning Variance: SS_B



ANOVA: Partitioning Variance: SS_W



ANOVA: Partitioning Variance

Ratio of SS_B to SS_W

If $SS_B >> SS_W$, good model. Easy to predict differences based on groups.

If $SS_B \approx SS_W$, less good model. Not easy to predict differences based on groups.

$$F \propto \frac{\text{between-group variance}}{\text{within-group variance}}$$

ANOVA: F

Mean Square Between

$$MS_B = \frac{SS_B}{J-1} = \frac{\sum_{j=1}^{J} n_j(\bar{x}_j - \bar{x})^2}{J-1}$$

Mean Square Within

$$MS_W = \frac{SS_W}{N-J} = \frac{\sum_{j=1}^{J} \sum_{i=1}^{I} (x_{ij} - \bar{x})^2}{N-J}$$

$$F = \frac{MS_B}{MS_W}$$

ANOVA: Hypotheses

H_0: $\sigma_B^2 = 0$

H_1: $\sigma_B^2 \neq 0$

Equivalent to:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_J$

H_1: not H_0

ANOVA: R function

`aov(y ~ grouping_var, data = data)`

- `y` = outcome
- `grouping_var` = name of grouping variable
- `data` = dataframe that includes `grouping_var` and `y`

Very similar syntax to `lm(y ~ predictor)` (and you can think of it similarly)

`summary(model)`

ANOVA: Example

```r
library(palmerpenguins)

summary(penguins |> select(species, flipper_length_mm))
```

```
      species    flipper_length_mm
 Adelie   :152   Min.   :172.0
 Chinstrap: 68   1st Qu.:190.0
 Gentoo   :124   Median :197.0
                 Mean   :200.9
                 3rd Qu.:213.0
                 Max.   :231.0
                 NA's   :2
```

Does flipper length vary by penguin species?

ANOVA: Example

```r
flipper_model <- aov(flipper_length_mm ~ species, data = penguins)
summary(flipper_model)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)    
species       2  52473   26237   594.8 <2e-16 ***
Residuals   339  14953      44                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
2 observations deleted due to missingness
```

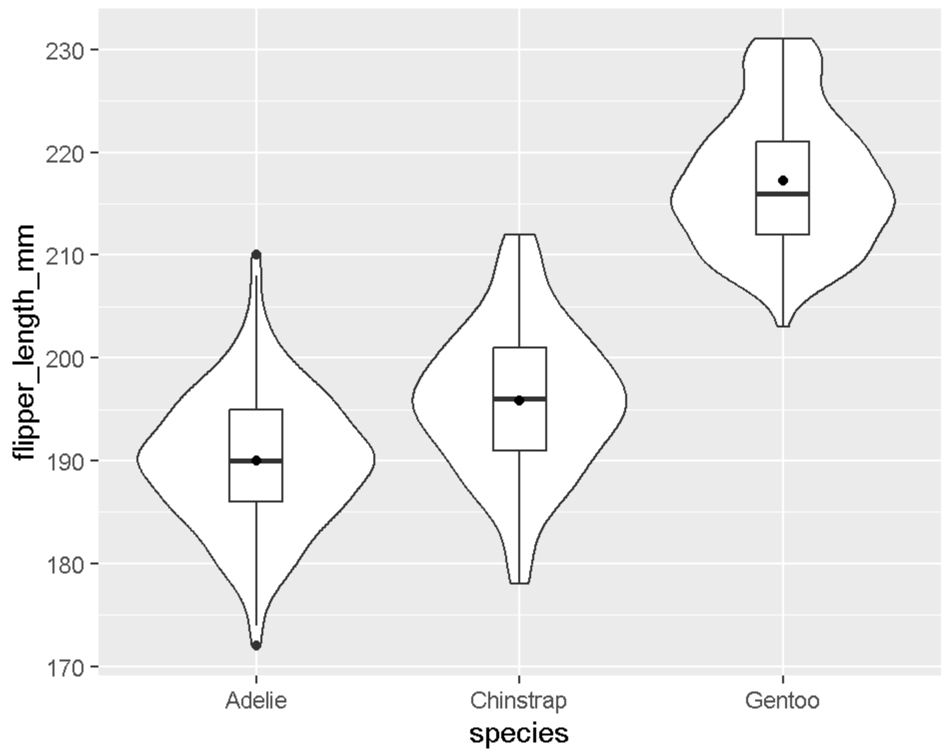
Huge F-statistic!

We reject the null in favor of the alternative that there is significant variation between penguin species in flipper length

ANOVA: Example

Good practice to plot the distributions / means to see the patterns!

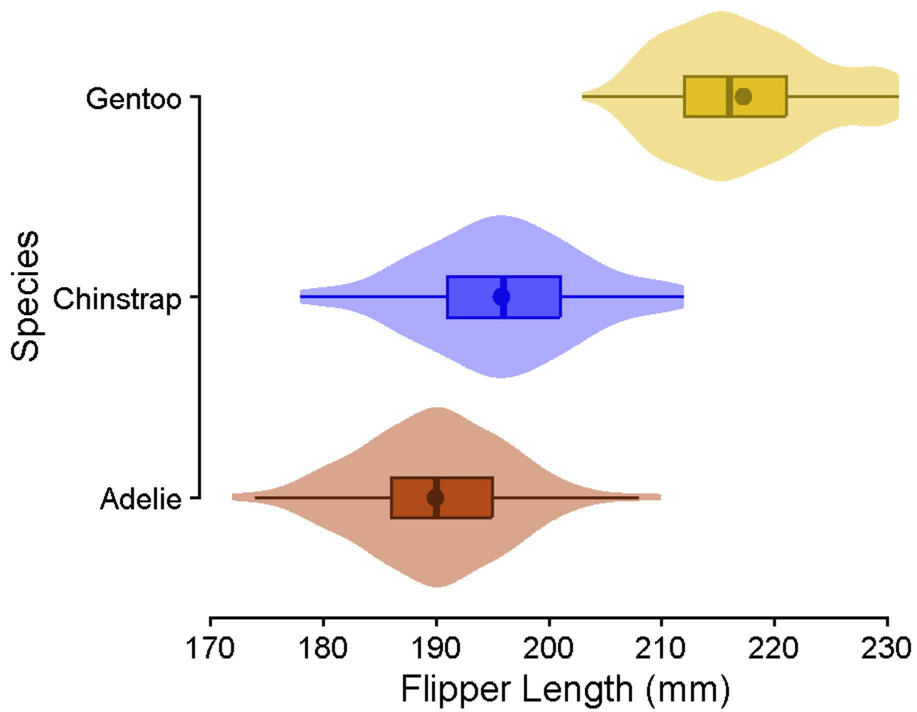
Plot Code



ANOVA: Example

Sprucing it up a little…

Plot | Code



Assignment 14

Thank you all for a great semester! Don’t forget to complete the CIOS when it comes out!