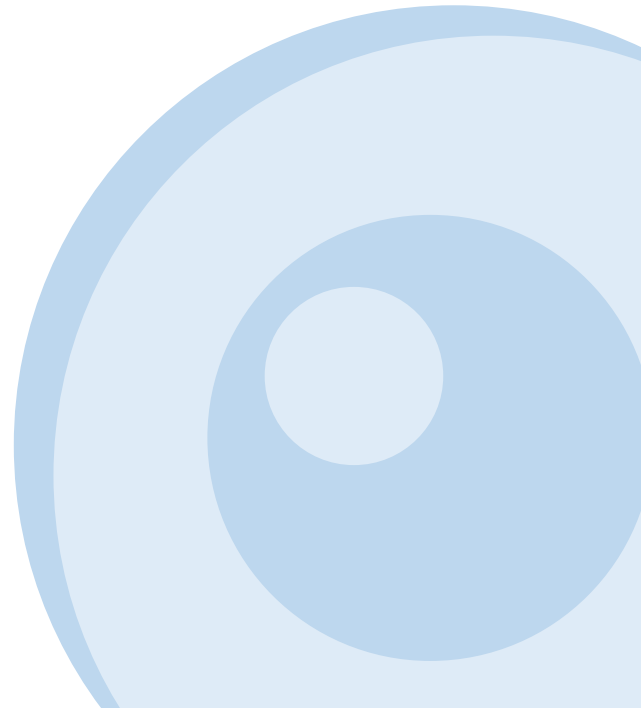


PSYC 193R:

Data Analysis & Visualization with R

Lecture 11: One-way ANOVA

Jonas S.-H. Lau [silau -et- ucsd.edu]
psyc193r.ucsd.edu



Midterm

Your midterm grade:
what you did in class + 85% of your corrections

- If you didn't do ANYTHING in-class, the highest score you can get is 92%
- If you got everything in-class, the highest score you can get is 108%



Midterm (logic for *take-home* corrections)

- Metacognition problem
- The goal of the midterm is to help you consolidate your knowledge
 - An opportunity for correct will help achieve that too
- A little help you with your grades
- Grades should be distributed fairly
- You're treated as a statistics consultant (*a.k.a.* an adult)



Midterm (rules)

- It is optional to make corrections:
 - If you get your paper back but do not return it, your grade == 0
- Absolutely NO DISCUSSION with fellow students
- Use a different color to work on your corrections
 - Only your corrections will be graded
- All the corrections should be made in the paper
(R Notebook won't be graded, but you can still upload it)
- Deadline to submit your corrections is next Thursday (2/22) in class
- Never post midterm paper online / share with other people
- No homework assignment this week

To prepare for the final

- Your final will definitely be in-class only
- Make sure you know the differences between tests
 - And the data structure that support each type of test
- You should have good enough class notes
- Homework should be adequately commented

- READ the textbook for basic statistics
 - E.g., reporting Z, t, F statistics, etc



Tricks in the Fitting Room

It is an open secret that retailers play "tricks" on shoppers that would lead to more sales. The same is true for grocery stores, electronics shops, clothing or online retailers.

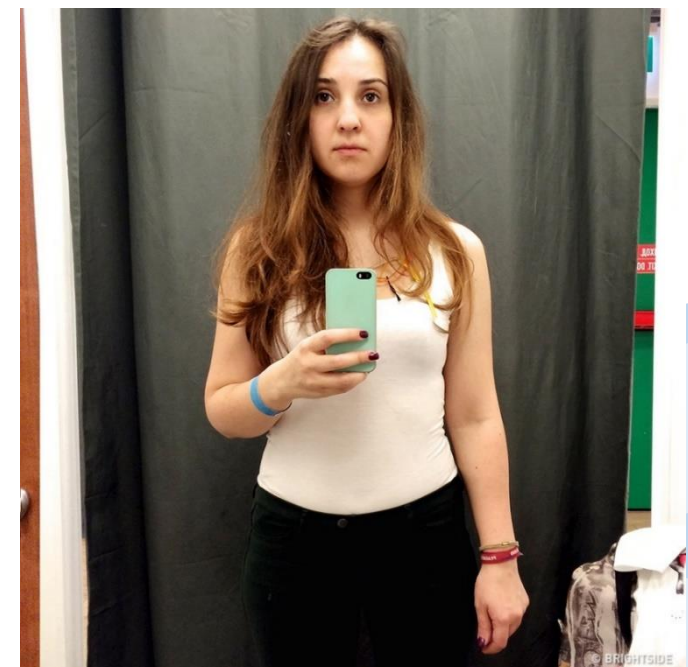
Regular Indoor Lighting



Harsh Overhead Lighting



Soft Vertical Lighting

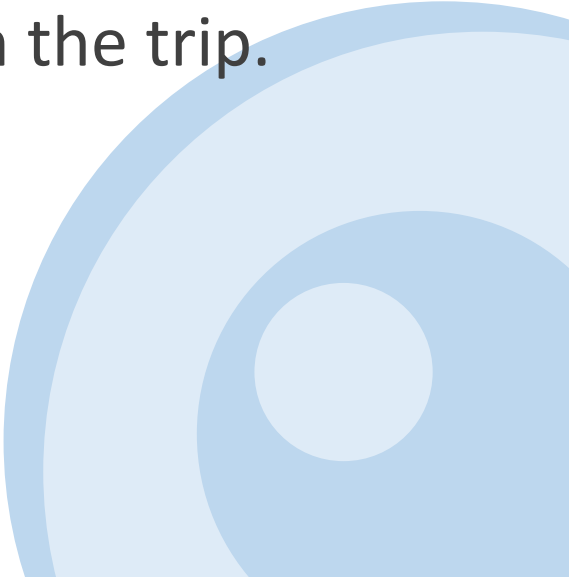


Tricks in the Fitting Room

A clothing retailer was in the process of redesigning their UTC store. They wanted to know if different kinds of lighting affect sales.

They installed **three kinds of lightings in the 3 different fitting rooms**. For a weekend, shoppers were randomly assigned into one of the three rooms when they requested to try something on. After they were done fitting, the clothes were marked as 0, 1, or 2, representing the rooms they were in.

The cashiers registered the **amount of money they spent** on the trip.



One-way ANOVA

One-way = One Factor = 1 Independent Variable
(in this case, we have 1 IV and 3 levels within the IV)

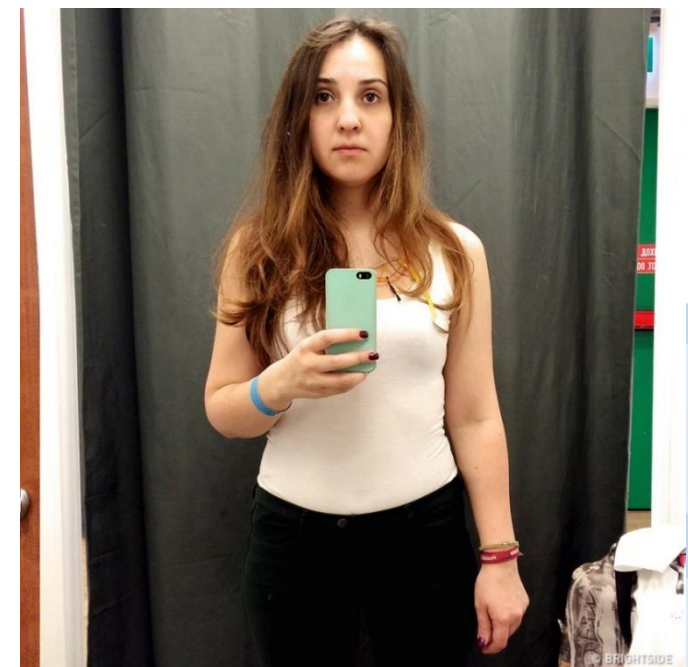
Regular Indoor Lighting



Harsh Overhead Lighting



Soft Vertical Lighting



One-way ANOVA

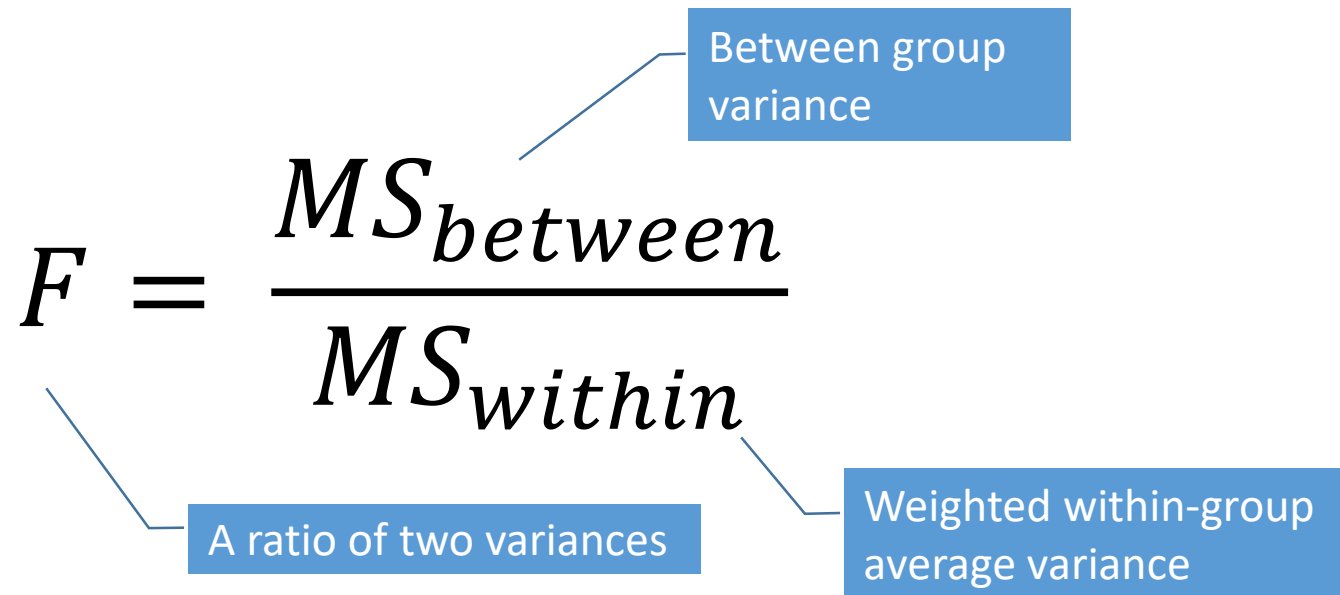
One-way = One Factor = 1 Independent Variable
(in this case, we have 1 IV and 3 levels within the IV)

$$F = \frac{MS_{between}}{MS_{within}}$$

Between group variance

A ratio of two variances

Weighted within-group average variance

The diagram shows the F-ratio formula $F = \frac{MS_{between}}{MS_{within}}$. Three blue callout boxes with white text provide context: 'Between group variance' points to the numerator $MS_{between}$; 'A ratio of two variances' points to the entire fraction; and 'Weighted within-group average variance' points to the denominator MS_{within} . The labels are connected to the formula by thin blue lines.

One-way ANOVA

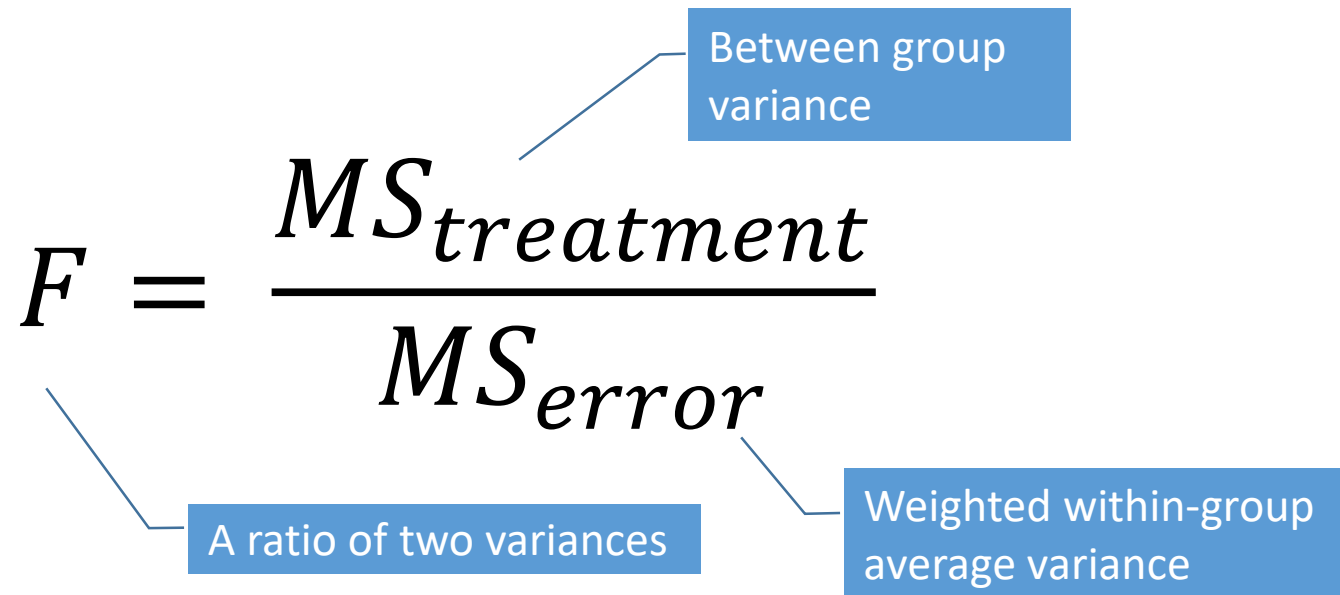
One-way = One Factor = 1 Independent Variable
(in this case, we have 1 IV and 3 levels within the IV)

$$F = \frac{MS_{treatment}}{MS_{error}}$$

Between group variance

A ratio of two variances

Weighted within-group average variance

The diagram shows the F-ratio formula $F = \frac{MS_{treatment}}{MS_{error}}$. Three blue callout boxes with white text provide context: 'Between group variance' points to the numerator $MS_{treatment}$; 'A ratio of two variances' points to the entire fraction; and 'Weighted within-group average variance' points to the denominator MS_{error} . The labels are connected to the formula by thin blue lines.

One-way ANOVA

One-way = One Factor = 1 Independent Variable
(in this case, we have 1 IV and 3 levels within the IV)

Regular Indoor Lighting

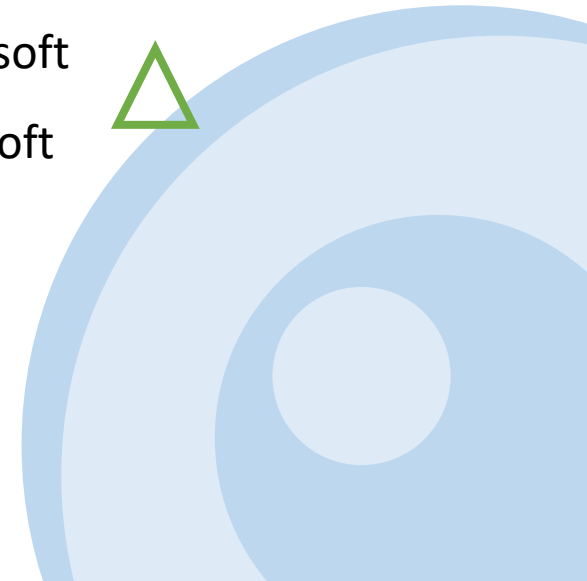
M_{regular}
 S_{regular} 

Harsh Overhead Lighting

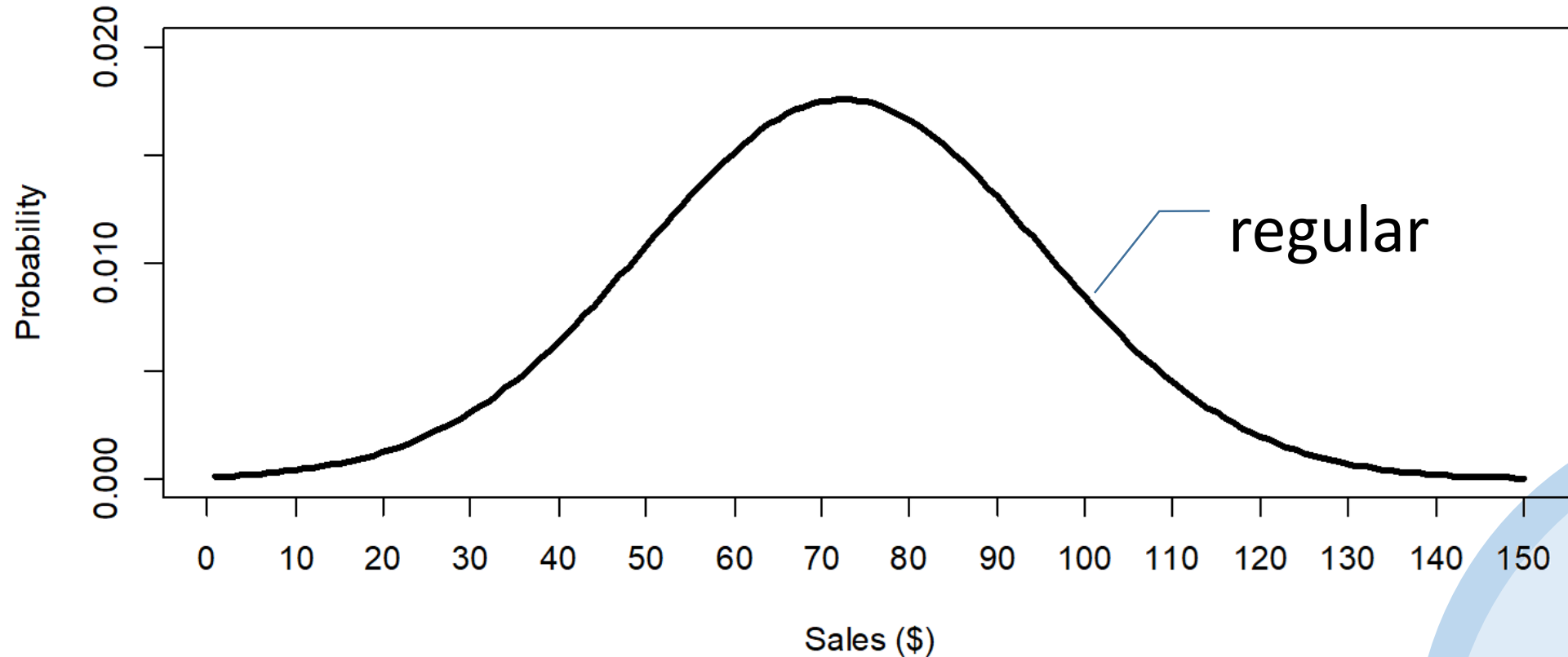
M_{harsh}
 S_{harsh} 

Soft Vertical Lighting

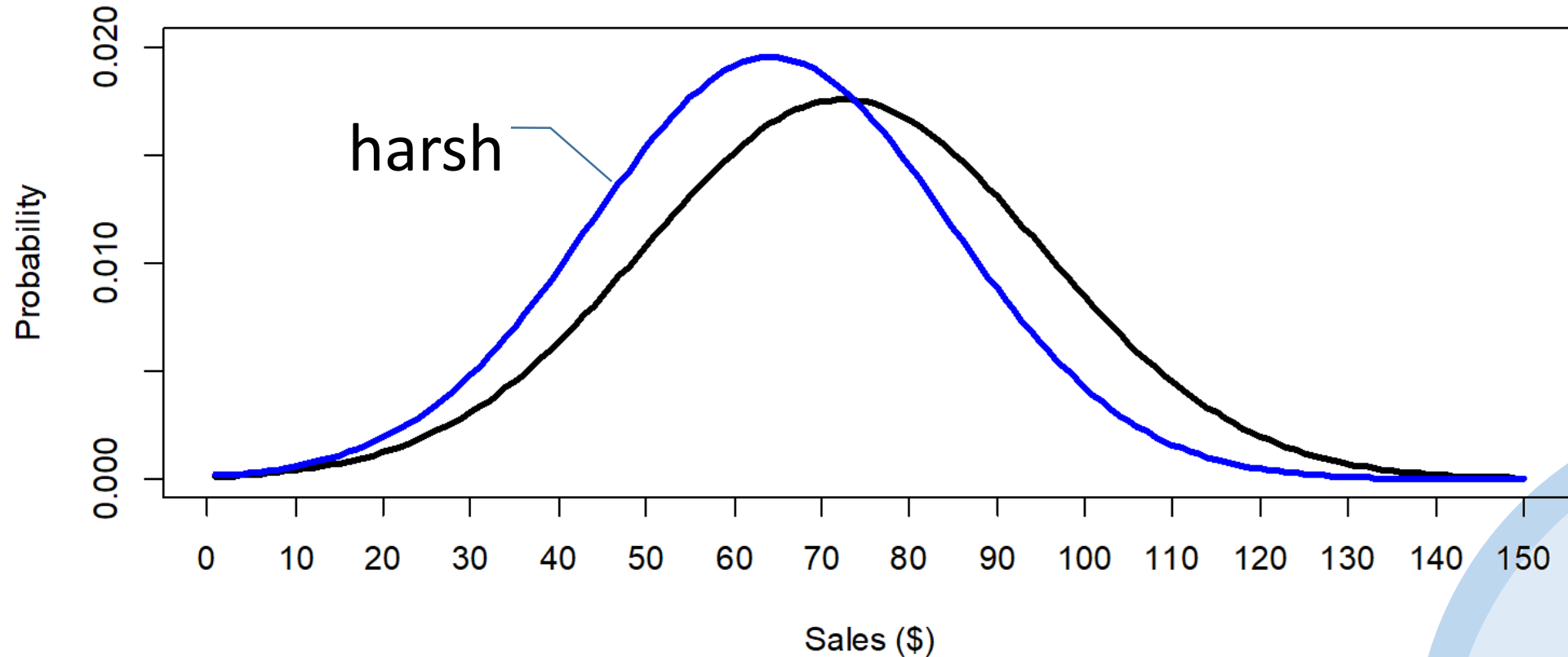
M_{soft}
 S_{soft} 



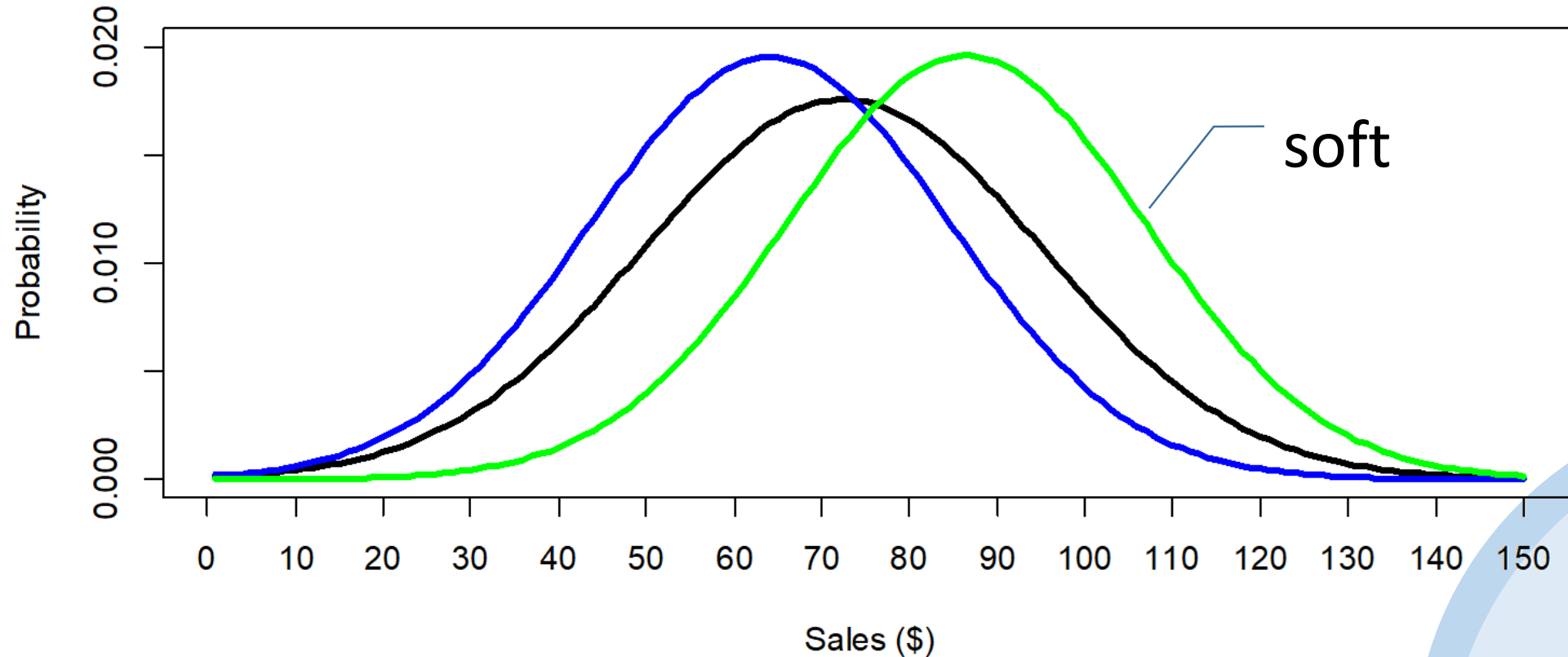
MS_{within} : Average dispersion within a condition



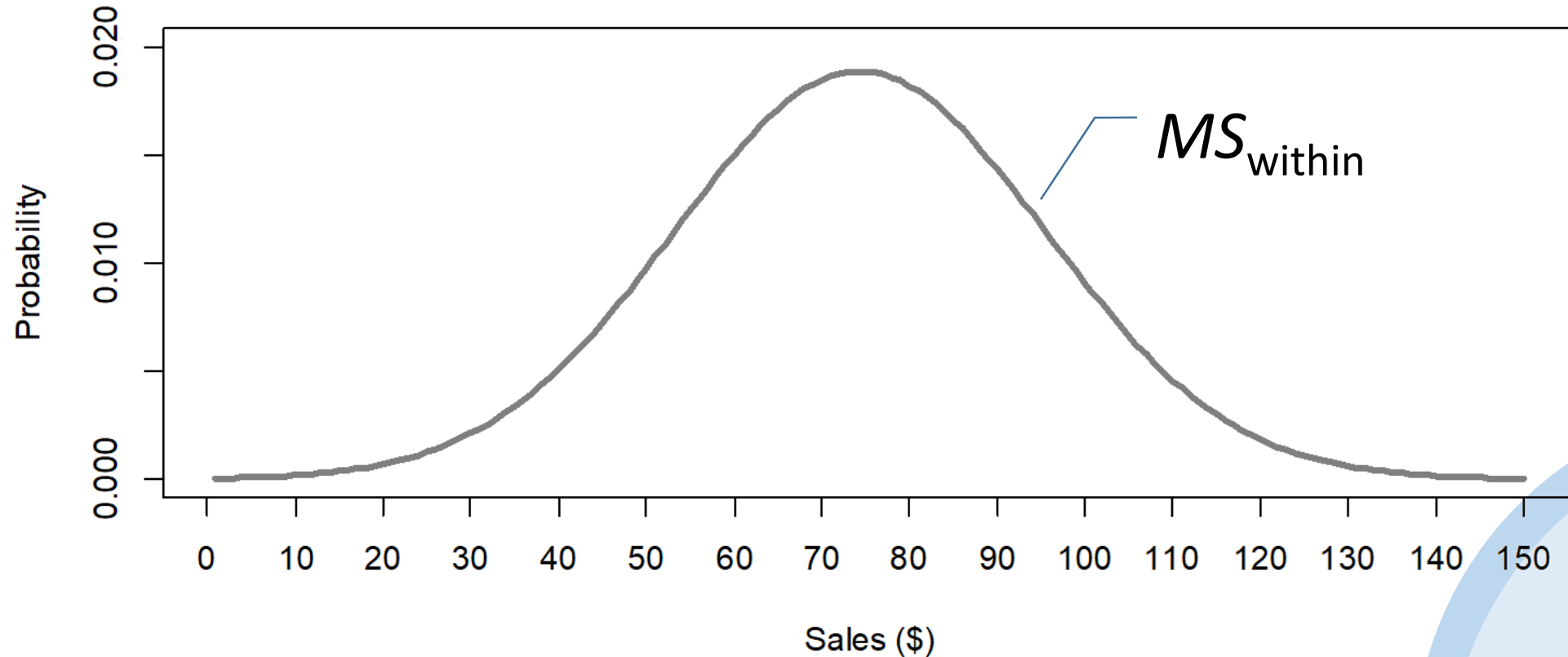
MS_{within} : Average dispersion within a condition



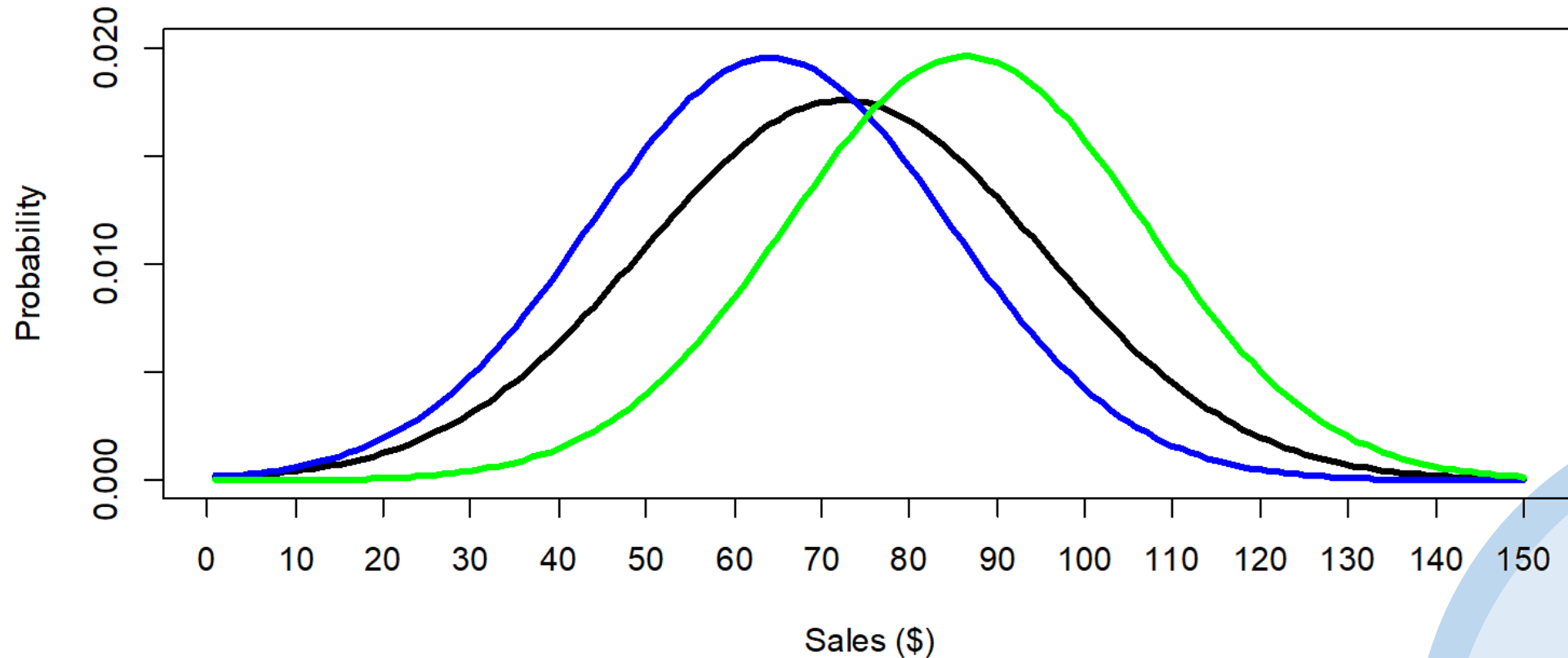
MS_{within} : Average dispersion within a condition



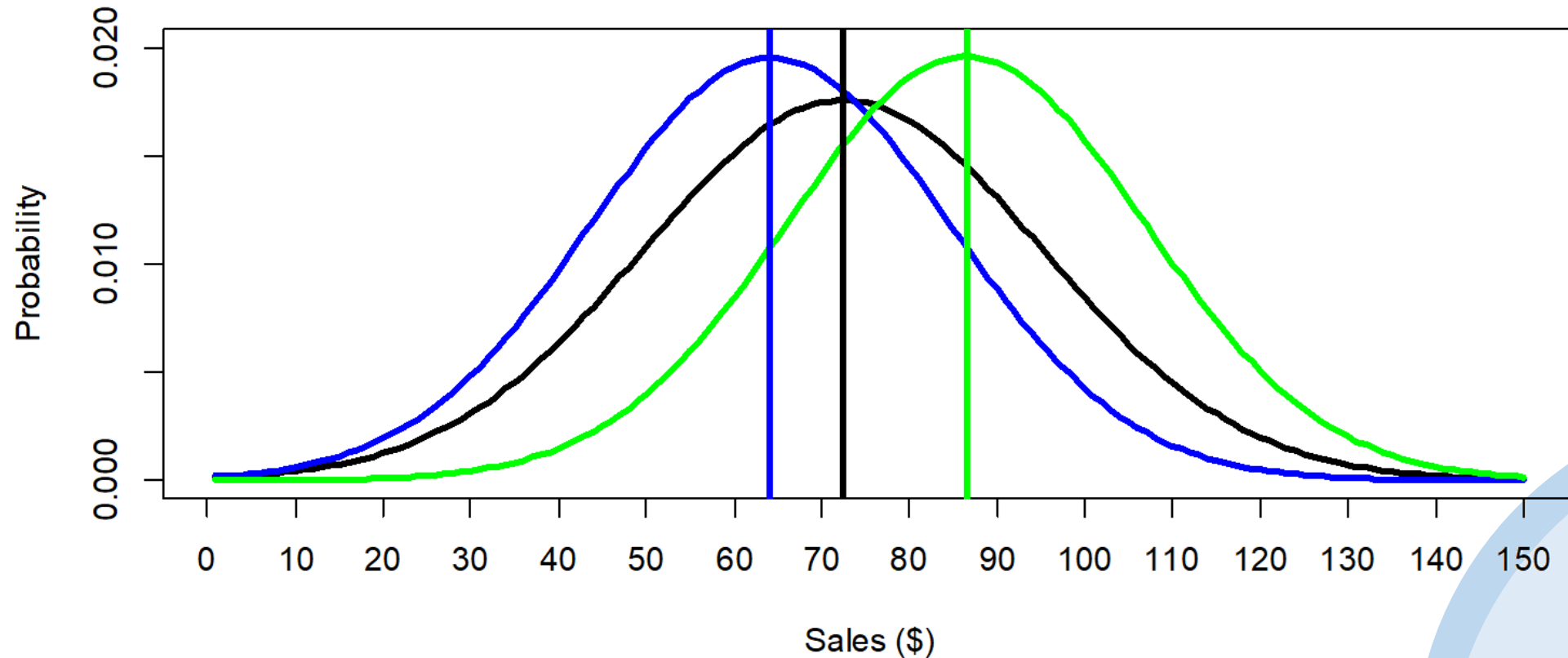
MS_{within} : Average dispersion within a condition



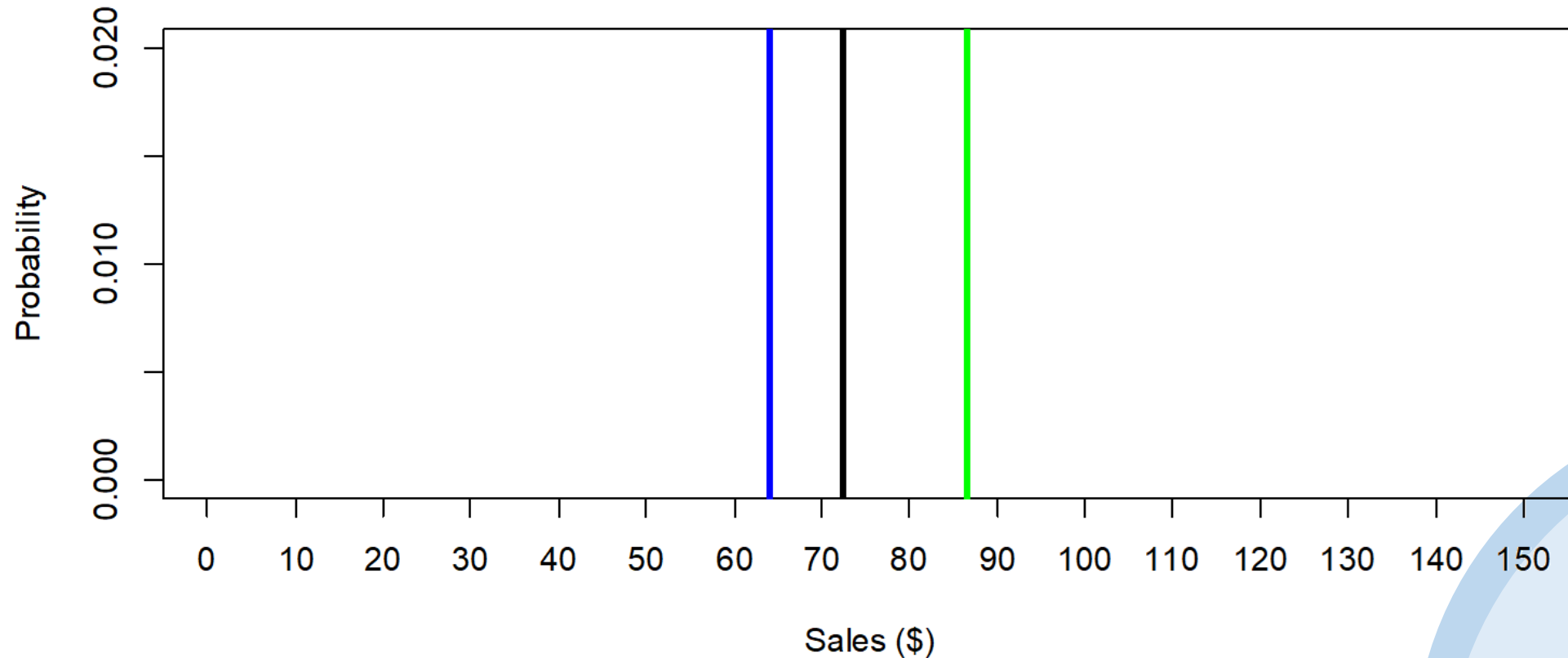
MS_{between} : Variations between conditions



MS_{between} : Variations between conditions

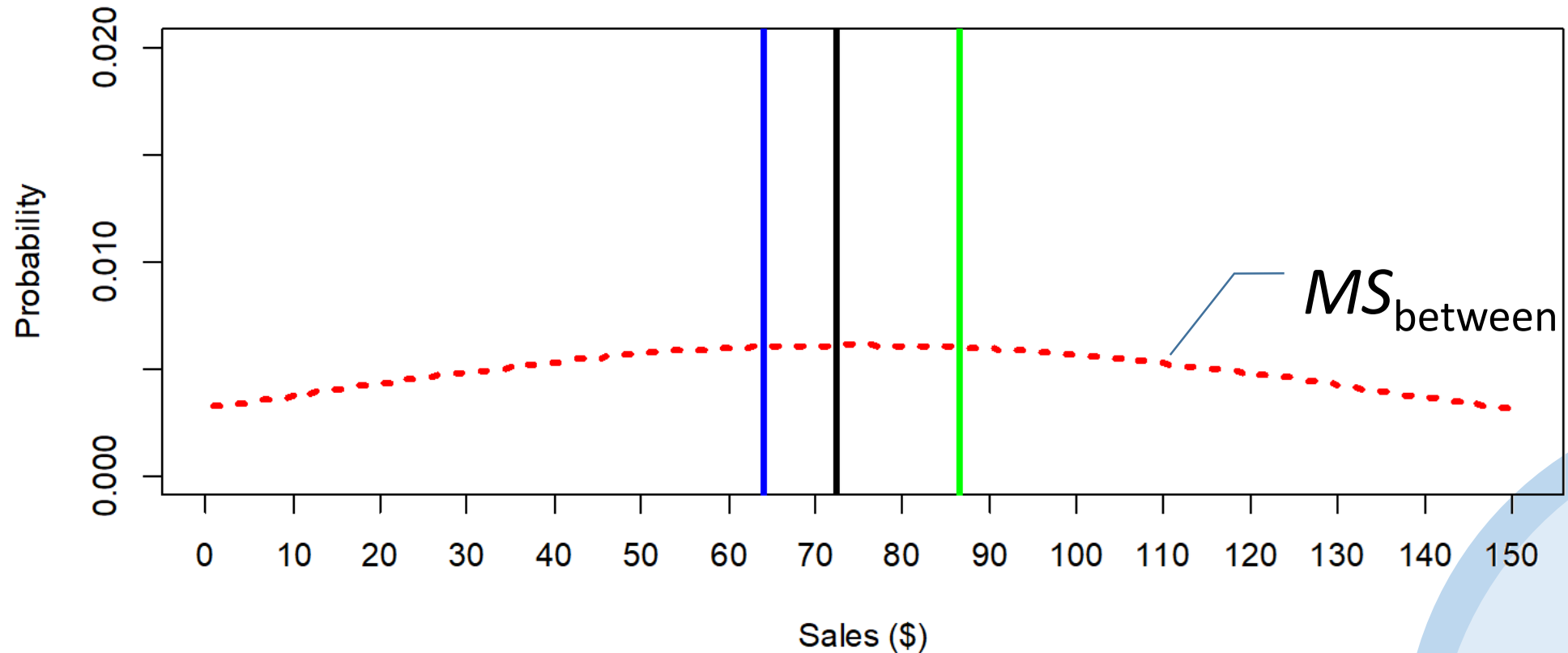


MS_{between} : Variations between conditions



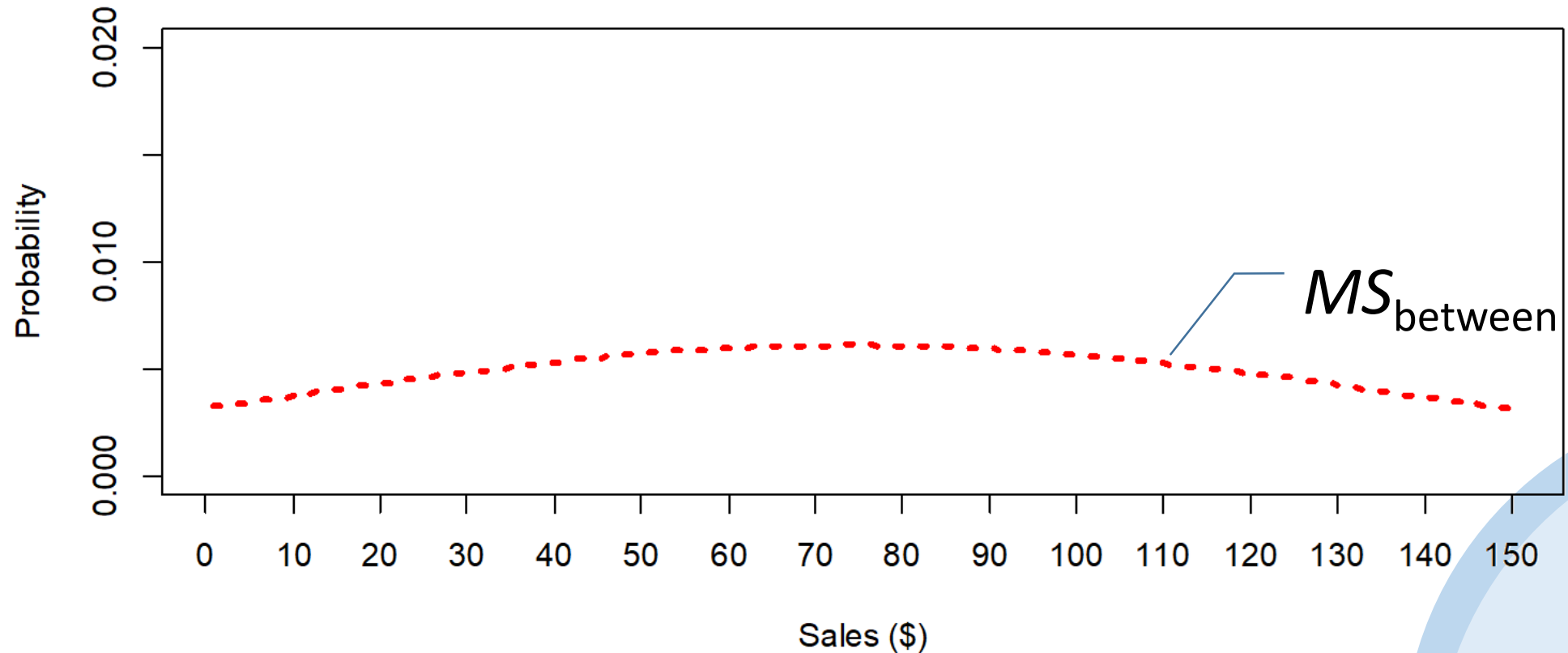
$$n_{\text{regular}} = 31, n_{\text{harsh}} = 32, n_{\text{soft}} = 35$$

MS_{between} : Variations between conditions

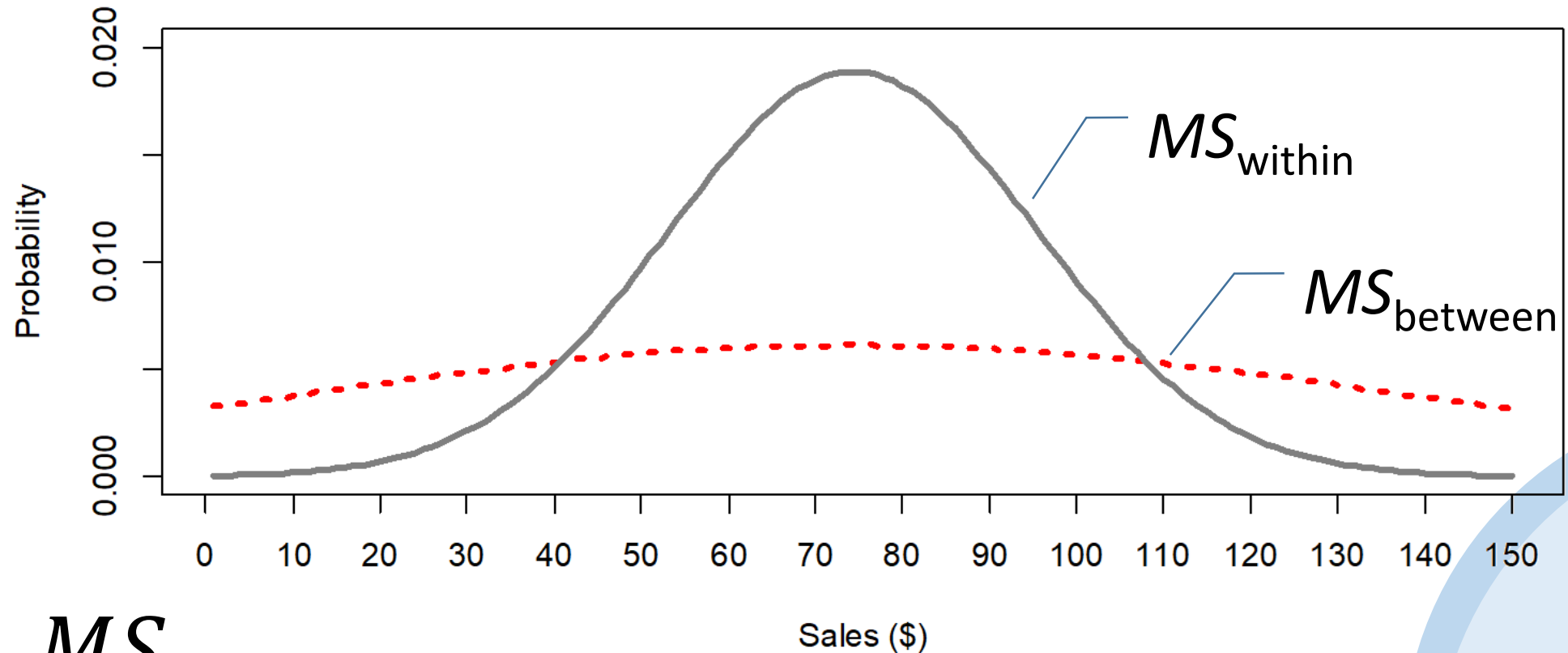


$$n_{\text{regular}} = 31, n_{\text{harsh}} = 32, n_{\text{soft}} = 35$$

MS_{between} : Variations between conditions

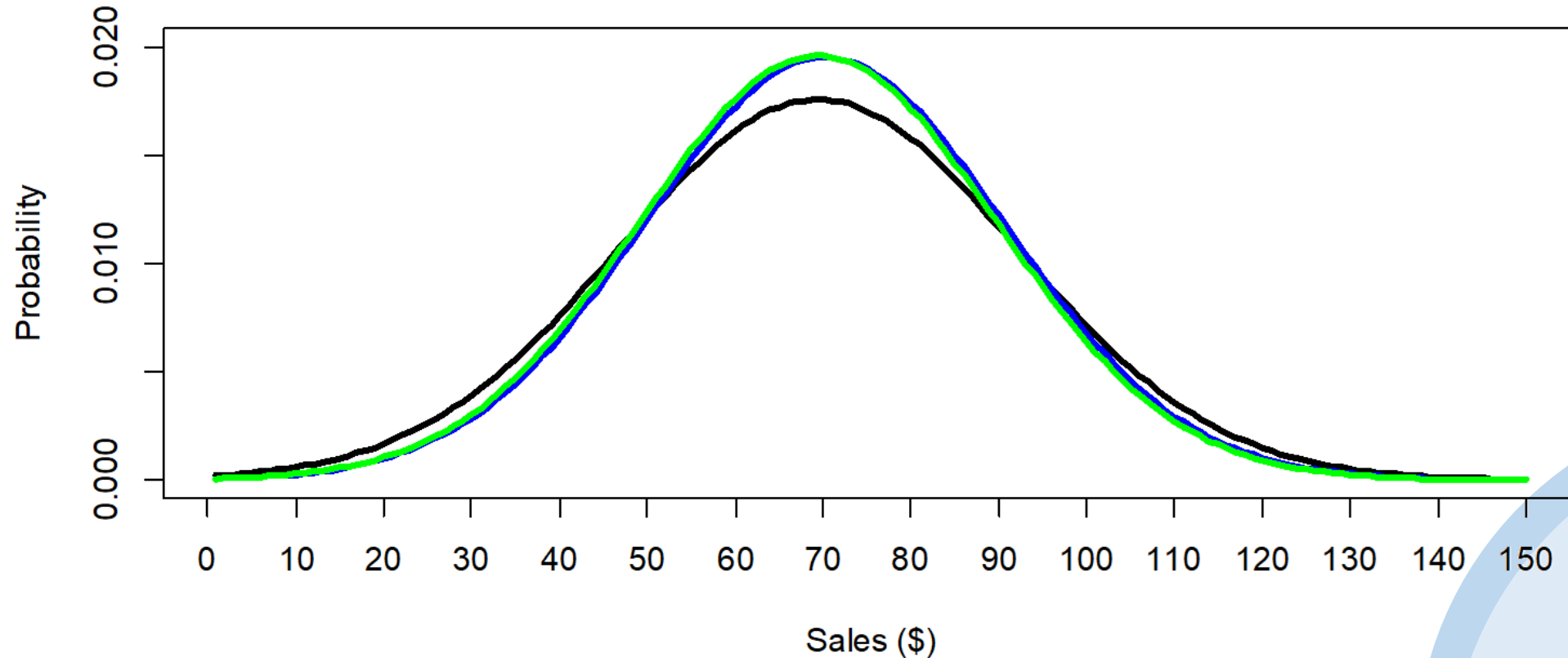


MS_{between} : Variations between conditions

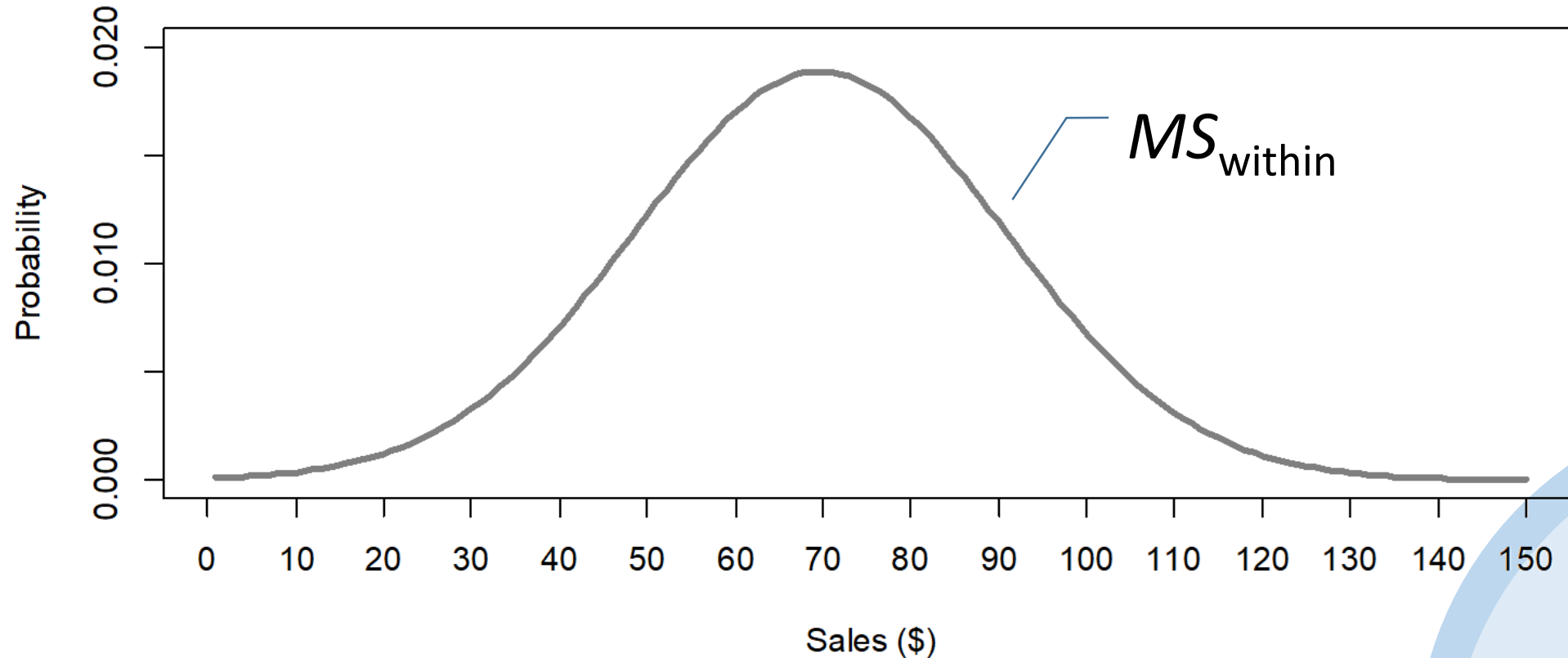


$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}}$$

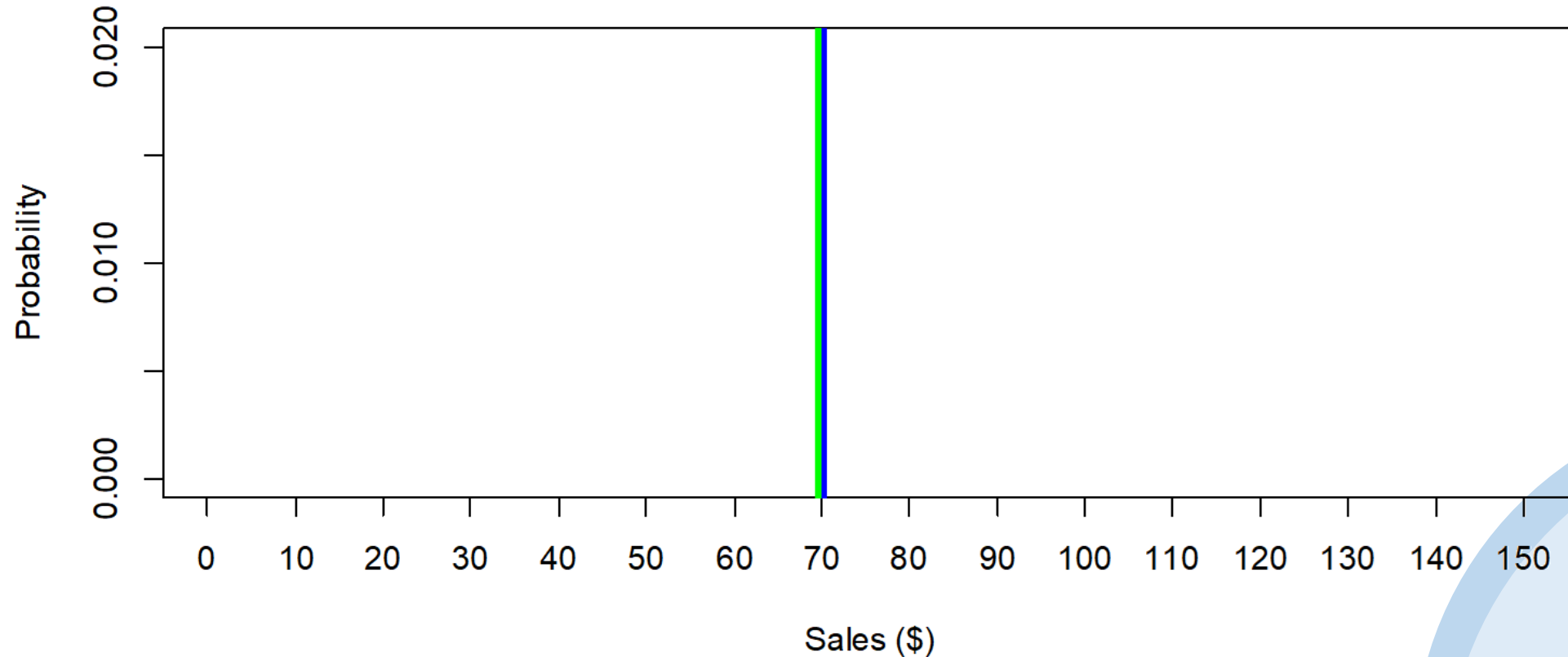
Scenario 2: Small between condition variations



Scenario 2: Small between condition variations

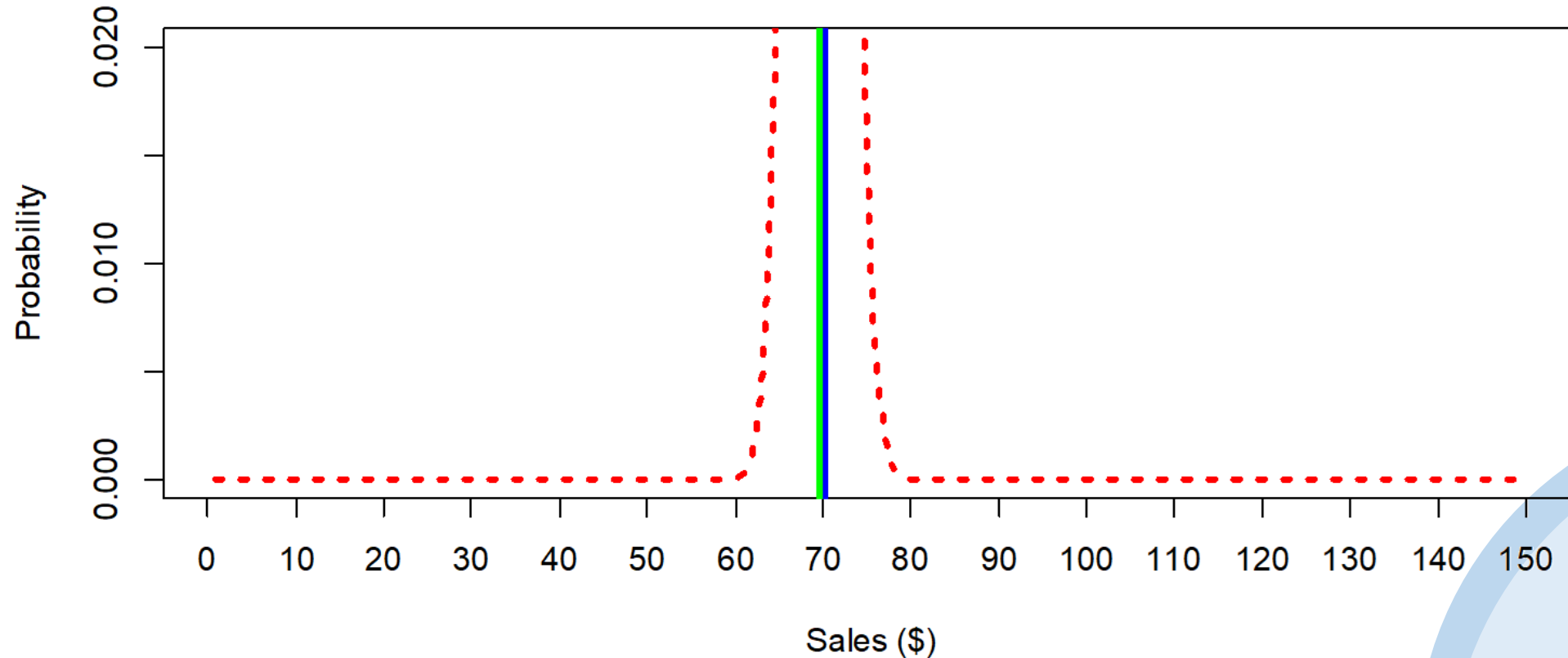


Scenario 2: Small between condition variations



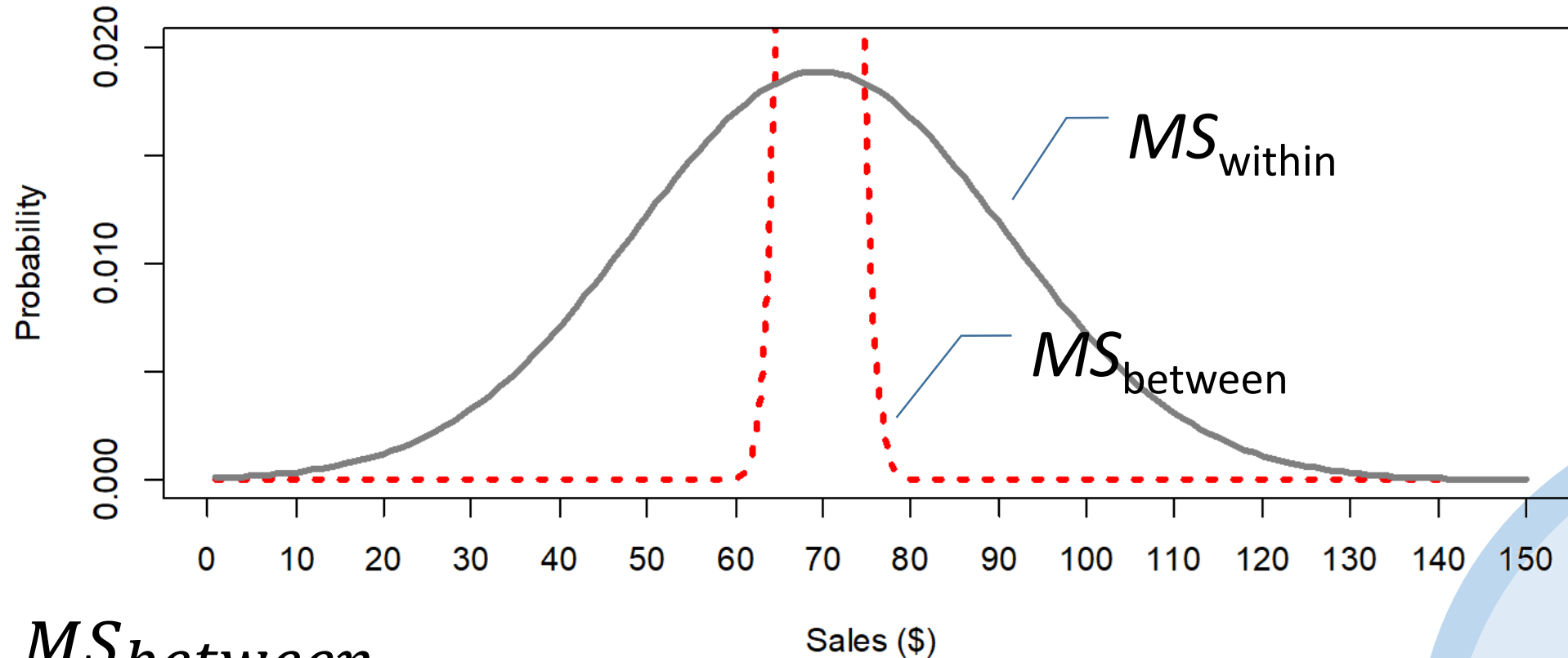
$$n_{\text{regular}} = 31, n_{\text{harsh}} = 32, n_{\text{soft}} = 35$$

Scenario 2: Small between condition variations



$$n_{\text{regular}} = 31, n_{\text{harsh}} = 32, n_{\text{soft}} = 35$$

Scenario 2: Small between condition variations



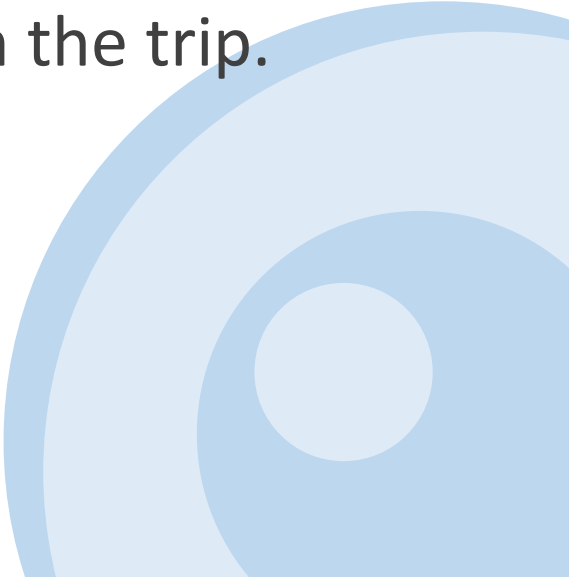
$$F = \frac{MS_{between}}{MS_{within}} \approx 0$$

Tricks in the Fitting Room

A clothing retailer was in the process of redesigning their UTC store. They wanted to know if different kinds of lighting affect sales.

They installed **three kinds of lightings in the 3 different fitting rooms**. For a weekend, shoppers were randomly assigned into one of the three rooms when they requested to try something on. After they were done fitting, the clothes were marked as 0, 1, or 2, representing the rooms they were in.

The cashiers registered the **amount of money they spent** on the trip.

A decorative graphic in the bottom right corner consisting of three concentric circles in shades of blue, with the innermost circle being the lightest and the outermost being the darkest.

Tricks in the Fitting Room: The Design

- What was the research question?
- What was the target population?
- What were the IV(s) and DV(s)?
- How were the IV(s) and DV(s) operationalized?
(also, scale of measurement?)
- Was it a between / within-subjects design?
- What are the expected results?
If the manipulation does not work:
If the manipulation works:



Import Data

- Download the data file:
https://psyc193r.ucsd.edu/data/psyc193r_lighting_data.csv
- Tell R where your data file is:
Session → Set Working Directory → Choose Directory
`setwd()`
- Check the *comma separated values* (csv) file with MS Excel to ensure correct format
- Import the file into R as a data frame, saving it to “lightingData” variable
`... = read.csv("file_name.csv")`

Data Formatting / Processing / Cleaning

- Factorize shopperId

```
lightingData$shopperId = ...
```

- Factorize condition, and label the three levels

```
lightingData$condition = factor(lightingData$condition, labels =  
  c("...", "...", "..."))
```

- Coerce money to a numeric variable

```
lightingData$money = ...
```

Summary statistics

- Use aggregate to find the descriptive statistics for each group

- Means

```
aggregate( ____ ~ ____, data = ____, FUN = ____)
```

- Standard Deviations

```
aggregate( ____ ~ ____, data = ____, FUN = ____)
```

- Numbers of participants

```
aggregate( ____ ~ ____, data = ____, FUN = ____)
```



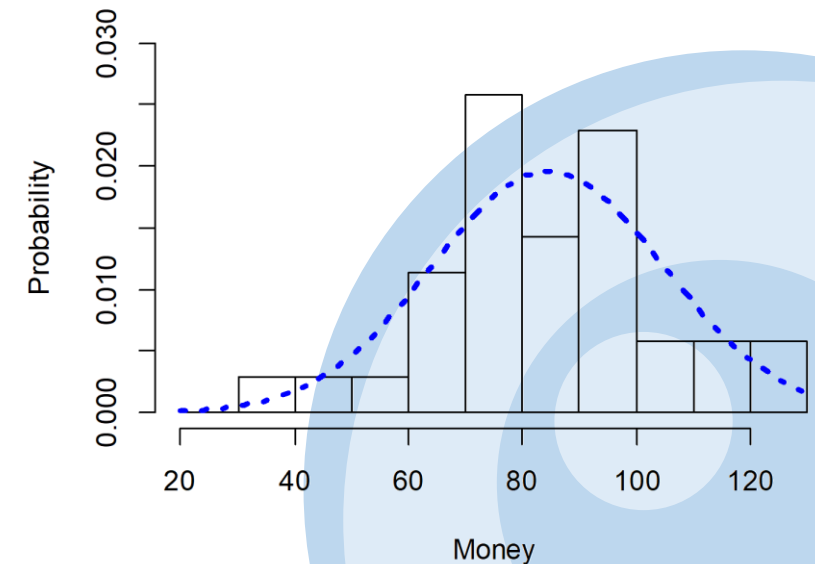
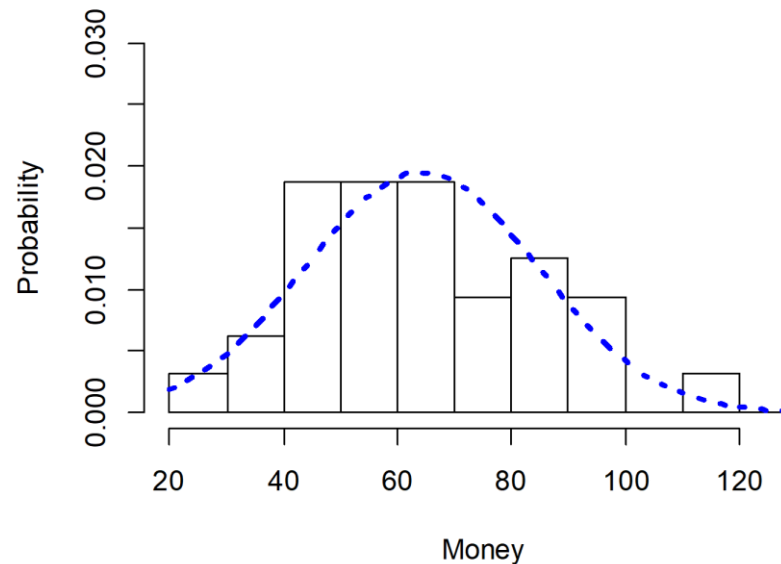
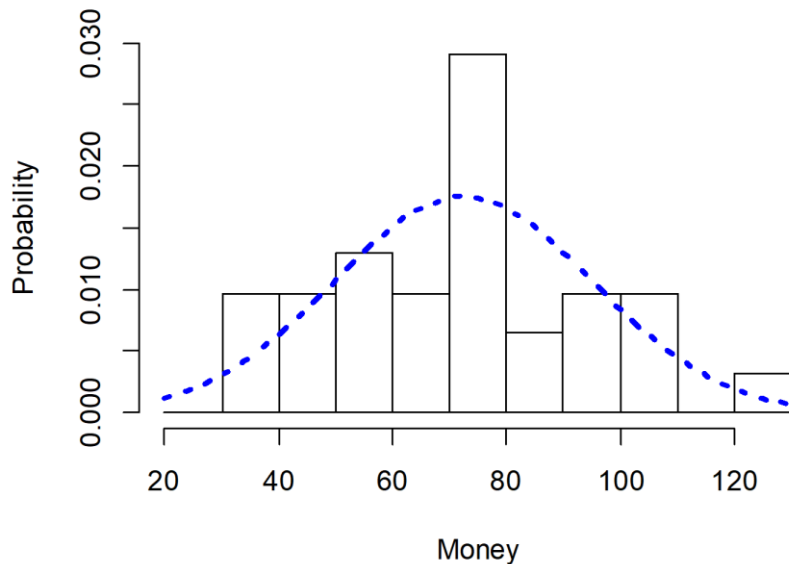
Checking Assumptions

- Normality Assumptions → $n = 30$

- Outliers

Plot a histogram for each group

```
hist(lightingData[ lightingData$__ == "___", ]$money)  
hist(lightingData[ lightingData$__ == "___", ]$money)  
hist(lightingData[ lightingData$__ == "___", ]$money)
```



Checking Assumptions

- Normality Assumptions → $n = 30$

- Outliers

Plot a histogram for each group

```
hist(LightingData[ LightingData$___ == "___", ]$money)
hist(LightingData[ LightingData$___ == "___", ]$money)
hist(LightingData[ LightingData$___ == "___", ]$money)
```

- Equal Variance Assumption $K^2(df) = \underline{\hspace{1cm}}$, $p = \underline{\hspace{1cm}}$

```
bartlett.test( ___ ~ ___ , data = LightingData)
```

Bartlett test of homogeneity of variances

data: money by condition

Bartlett's K-squared = 0.47868, df = 2, p-value = 0.7871

Performing one-way ANOVA: `aov()` & `summary()`

We want to know if there is a difference on ____ in different ____ .

Step 1: Create an ANOVA object (save it as _____.aov variable)

```
lighting.aov = aov( _____ ~ _____, data = lightingData)
```

Step 2: Print out the summary table

```
summary(lighting.aov)
```



Performing one-way ANOVA: `aov()` & `summary()`

We want to know if there is a difference on ____ in different ____ .

Step 1: Create an ANOVA object (save it as _____.aov variable)

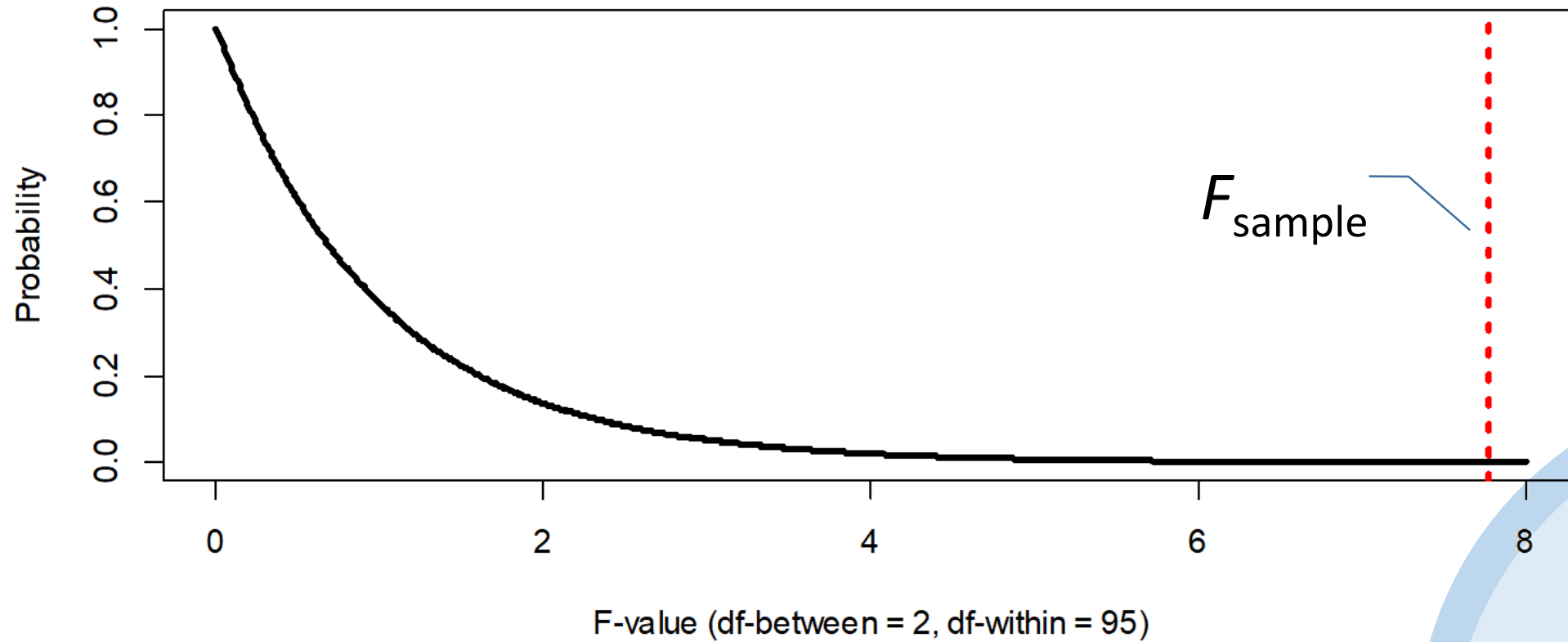
```
lighting.aov = aov( money ~ condition, data = lightingData)
```

Step 2: Print out the summary table

```
summary(lighting.aov)
```

```
              Df Sum Sq Mean Sq F value    Pr(>F)
condition      2   6973    3486   7.767 0.000751 ***
Residuals    95  42641     449
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F distribution



Hypothesis Testing – Step-by-step

- Step 0: Gather all the information you need

- Check assumptions
- Data cleaning

- Step 1: Set up the hypotheses

- H_0 (null hypothesis)
- H_1 (Research hypothesis)

- Step 2: Create an `aov()` object

- Step 3: Use `summary()` to obtain sample F score and p value

- Step 4: Decision & Conclusion

(compare p against α)

check: $df_{\text{between}} = k - 1$; $df_{\text{within}} = N - k$

kurtosis of sampling distribution ?

but $n's \geq 30$ ✓

H_0 :

H_1 :

...`aov` =

See `summary()` output

Report sample means, standard deviations,
sample sizes,

$F(df_{\text{between}}, df_{\text{within}}) = \underline{\hspace{1cm}}$, $p = \underline{\hspace{1cm}}$
decision/ conclusion