ABOUT

This notebook illustrates how to run an Analysis of Variance (ANOVA) in R. It is used as a companion for a lecture in the CS-411 course "Digital Education". Patrick Jermann, CEDE, EPFL

TOOLS TO RUN THIS NOTEBOOK

SIMPLE

Jupyter Notebooks with R kernel

- http://noto.epfl.ch
- Does not require any installation on your machine

MORE INVOLVED

R and Rstudio IDE

- https://rstudio.com/products/rstudio/download/#download
- Requires installation of the R language and Rstudio editor.

Alternatively, you can do the analyses in Python in NOTO or in your favourite computing environment, but I provide examples in R

REFERENCES

Seltman, H. J. (2012). Experimental design and analysis.

http://www.stat.cmu.edu/~hseltman/309/Book/

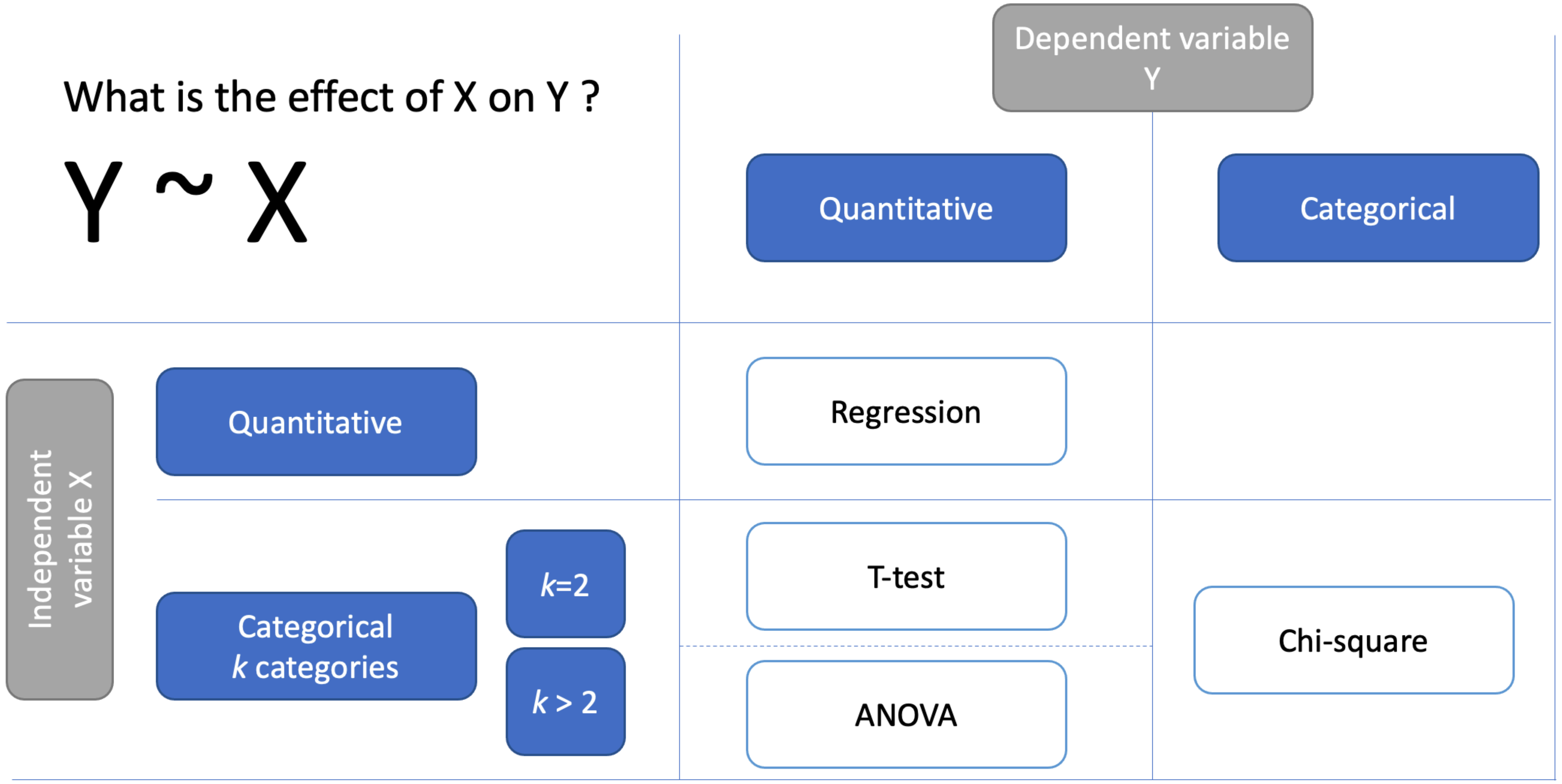
- t-test: chapter 6
- ANOVA: chapter 7
- Regression: chapter 10
- Chi-square: chapter 16

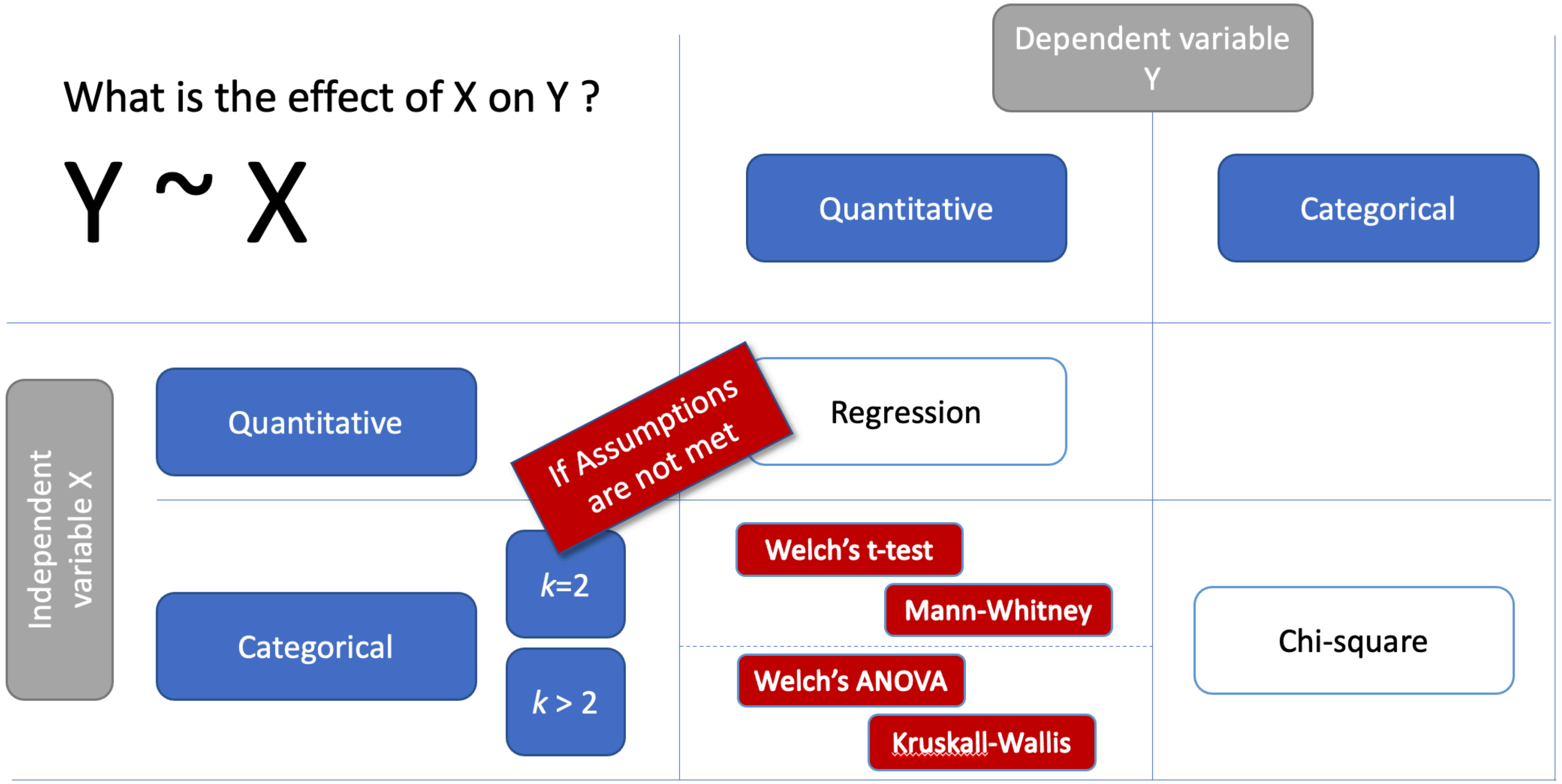
Jose, P. E. (2013). Doing statistical mediation and moderation. Guilford Press.
https://books.google.ch/books?id=aJFcO81Ro-0C&printsec=copyright&redir_esc=y#v=onepage&q&f=false

- Basic Mediation: chapter 3
- Basic Moderation: chapter 5

A CHEAT-SHEET TO DETERMINE WHICH TEST TO USE

Different statisticatl tests are appropriate depending on the type of Independent and Dependent variables and on the type of hypotheses we want to test.




What is the effect of X on Y ?
Y ~ X
Dependent variable Y
Quantitative
Categorical
Independent variable X
Quantitative
Categorical
If Assumptions are not met
Regression
k=2
k > 2
Welch's t-test
Mann-Whitney
Welch's ANOVA
Kruskall-Wallis
Chi-square

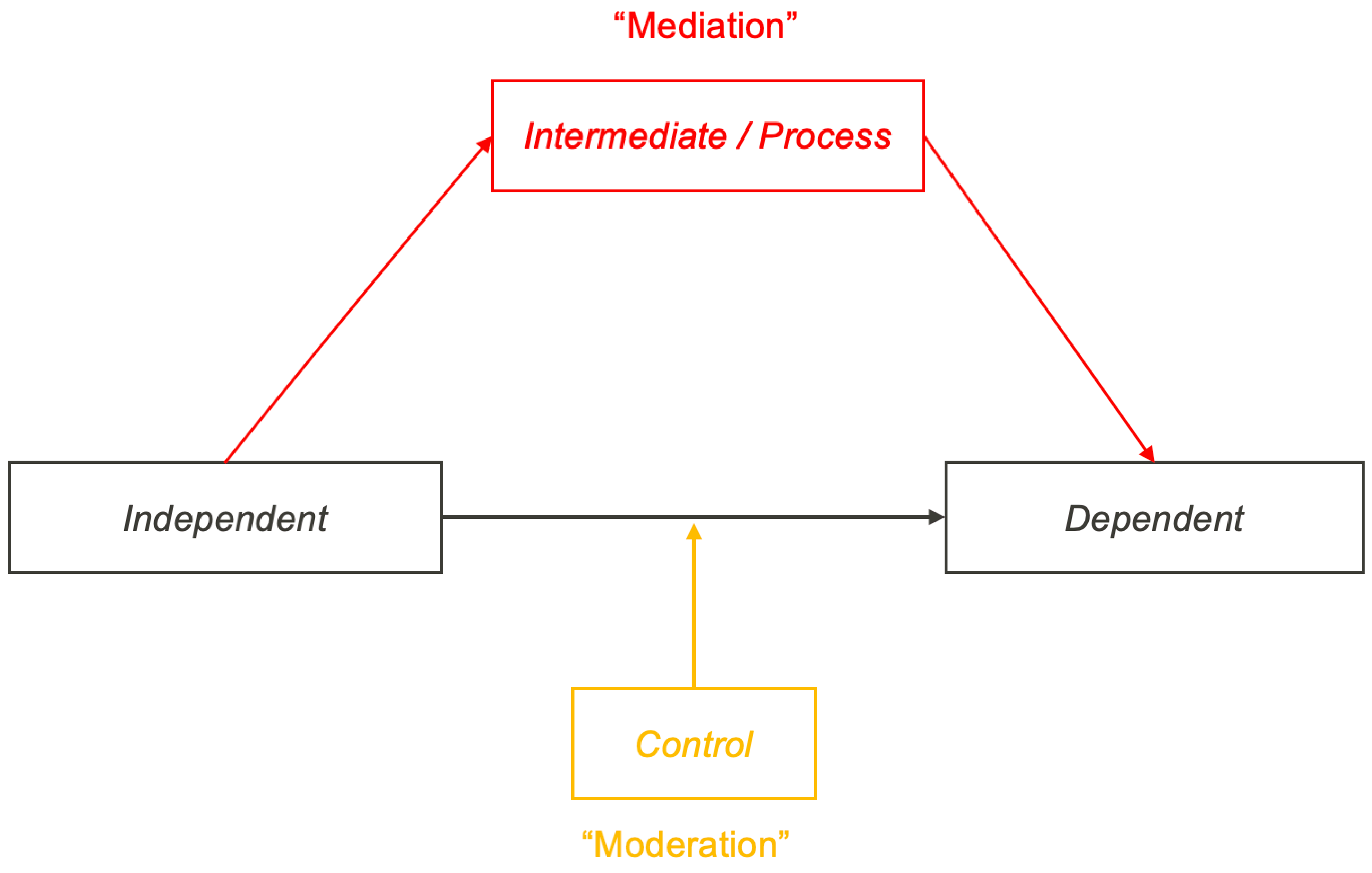
MEDIATION

- explains how or why an intervention works
- mediator explains all or part of the treatment's impact on an intended outcome
- is an intermediate outcome that is measured or observed after the onset of the intervention. E.g. fidelity of application, how many questions were asked ?

MODERATION

- explains for whom the intervention benefits or what conditions must exist for the intervention to be effective.
- a factor that reflects who is most affected by the treatment
- a factor that exists prior to the introduction of an intervention

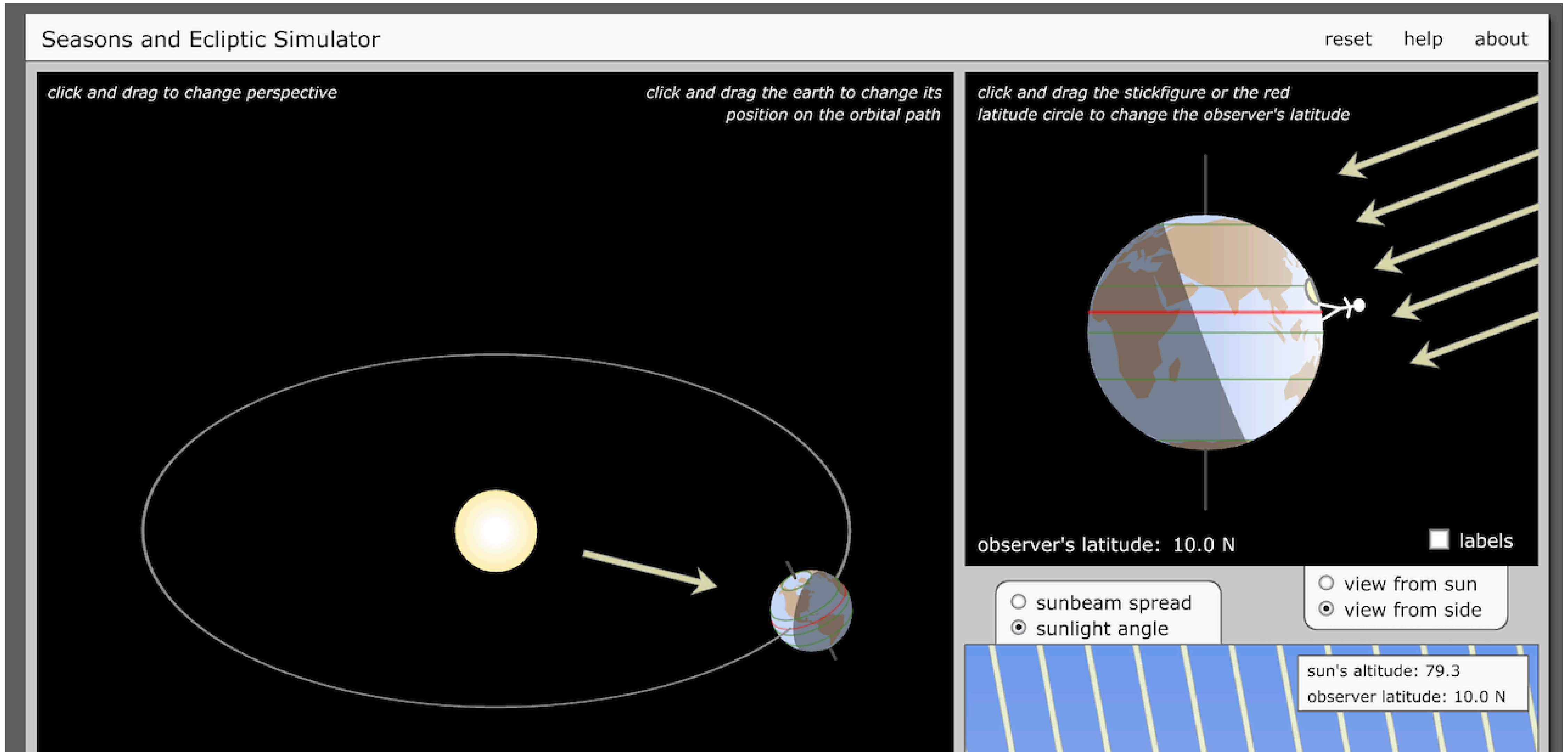
Eg. student characteristics, such as special education status, gender, ...


"Mediation"
Intermediate / Process
Independent
Dependent
Control
"Moderation"

EXPERIMENT (IPS VS PSI)

In this **imaginary** experiment, we are studying the effect of the order of instruction and problem-solving (independent variable) on learning (dependent variable) how the position of the earth relative to the sun influences seasons.

Participants used a simulation (https://astro.unl.edu/classaction/animations/coordsmotion/eclipticsimulator.html) during the problem-solving phase and watched a video during the instruction phase.



PARTICIPANTS

The sample consisted of N=200 participants.

INDEPENDENT VARIABLE

Order of instruction The independent variable has two modalities (also called conditions):

- I-PS : instruction followed by problem-solving
- PS-I : problem-solving followed by instruction

Participants were *randomly* assigned to one of the experimental conditions.

DEPENDENT VARIABLE

Learning gain. Participants completed a 10 question *pre-test* before starting the experiment. The pre-test was a series of questions about their understanding of the sun-earth relative positions. After the experiment, particpants completed a 10 question *post-test* with similar questions as the pre-test. The learning gain was computed as :

$$learning.gain = post.test - pre.test$$

another possibilty would be the relative learning gain

$$rel.gain = \frac{post.test - pre.test}{max - pre.test}$$

CONTROL VARIABLES

Age group. Participants were recruited among highschool students who are interested in following studies at EPFL (kids), students doing their bachelor as well as alumni who are active professionally (professionals).

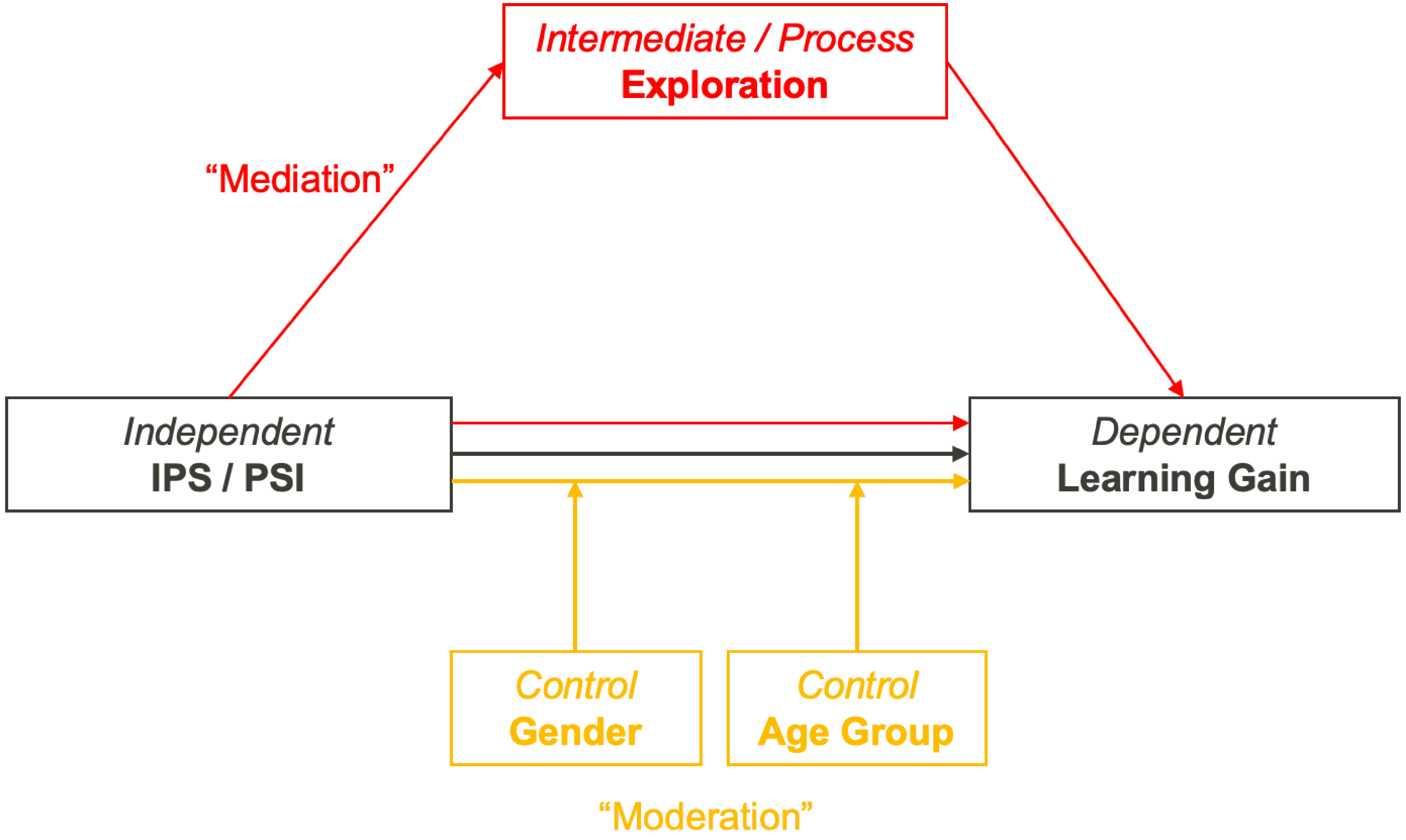
Young learners (e.g., second to fifth graders) may have insufficient prior knowledge about cognitive and metacognitive learning strategies to generate multiple solutions during initial problem solving

Gender. Experimenters also asked for the gender of the participants, either Male (M) or Female (F).

Self-regulation skills. Participants also filled in a questionnaire about their self.regulation skills by using the Learning Companion (https://companion.epfl.ch)

INTERMEDIATE / PROCESS VARIABLES

Solutions. The simulation system logged every simulation run and counted how often students used the simulation to generate a potential solution.



Intermediate / Process
Exploration
"Mediation"
Independent
IPS / PSI
Dependent
Learning Gain
Control
Gender
Control
Age Group
"Moderation"

DATASET

This dataset was generated to illustrate basic statistical techniques like ANOVA and regression as well lightly more advanced techniques like mediation and moderation. However, we tried as much as possible to implement variations compatible with insights found in the literature about Productive Failure:

Sinha, T., & Kapur, M. (2021). When Problem Solving Followed by Instruction Works: Evidence for Productive Failure. Review of Educational Research, 91(5), 761–798. https://doi.org/10.3102/00346543211019105

ANALYSIS

LOADING DATA

In [2]:

```
# Here we read the data and then convert the factors to the right labels
df <- suppressMessages(read_delim(file = "dataset.csv", delim = ",") %>%
  mutate(
    condition = factor(condition, labels = c("IPS", "PSI")),
    gender = factor(gender, labels = c("M", "F")),
    age.group = factor(age.group,
      labels = c("kids", "students", "professionals")
    )
  ))

head(df)
```

A tibble: 6 x 6

condition	gender	age.group	solutions	self.regulation	learning
<fct>	<fct>	<fct>	<dbl>	<dbl>	<dbl>
PSI	F	kids	20	6.4129659	1.1447316
PSI	F	students	20	5.4942910	2.3840504
PSI	F	professionals	24	10.1754505	-0.3823993
PSI	M	kids	12	7.6805230	0.2294454
PSI	M	kids	9	0.4995889	0.4921680
IPS	M	kids	5	6.8204271	-1.3863676

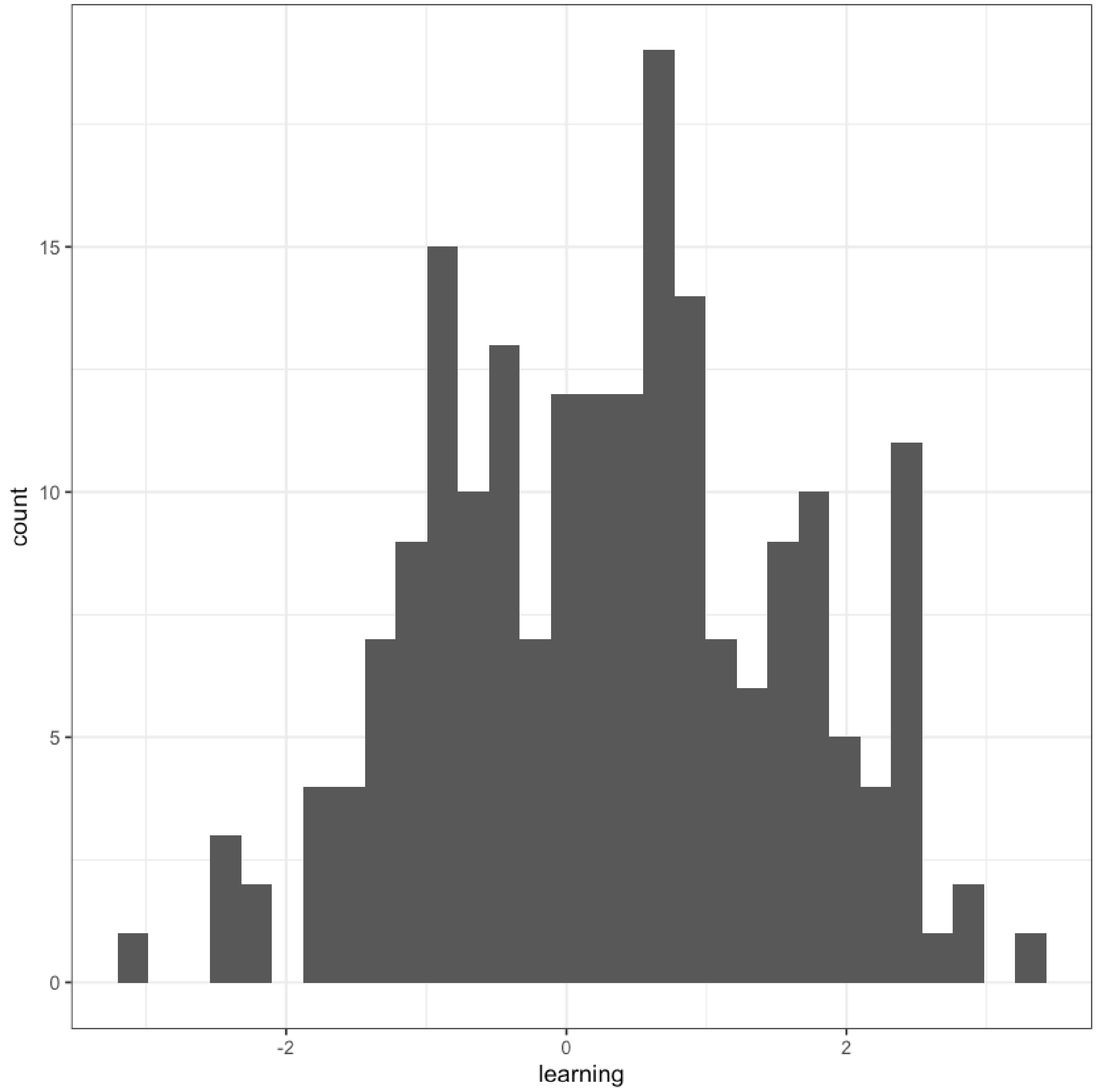
DESCRIPTIVES

In [3]:
```
summary(df)
```

```
 condition gender            age.group     solutions      self.regulation       l
 IPS:102   M: 95   kids           :62   Min.   : 0.00   Min.   :-5.984   Min.
 PSI: 98   F:105   students       :73   1st Qu.:10.00   1st Qu.: 4.301   1st
                   professionals:65   Median :15.00   Median : 8.151   Medi
                                        Mean   :14.91   Mean   : 8.392   Mean
                                        3rd Qu.:19.00   3rd Qu.:12.184   3rd
                                        Max.   :31.00   Max.   :30.920   Max.
```

DEPENDENT VARIABLE: LEARNING GAIN

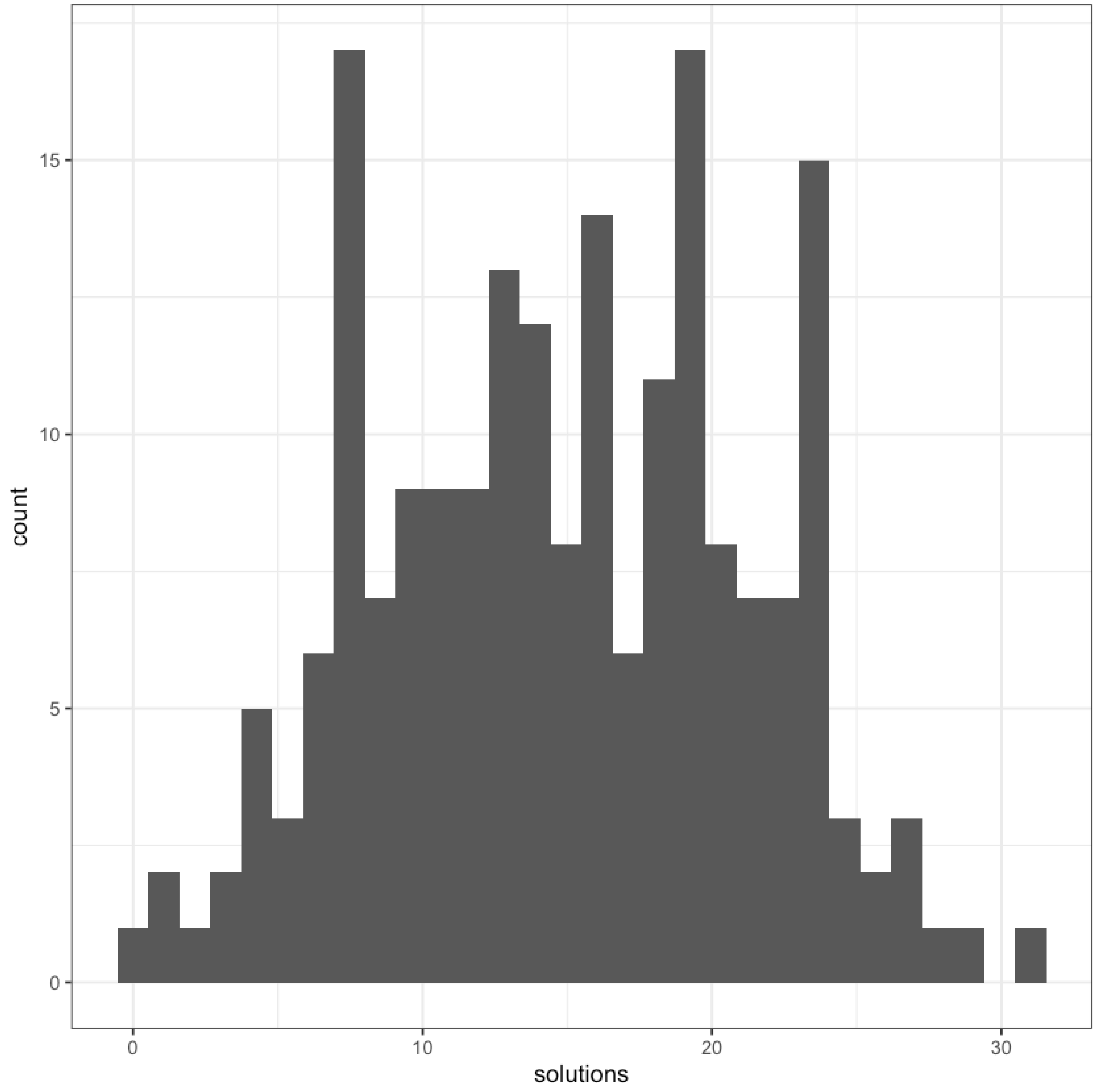
In [4]:
```
df %>%
  ggplot(aes(x=learning)) +
  geom_histogram(bins=30) +
  theme_bw()
```



INTERMEDIATE VARIABLE: SOLUTIONS

In [5]:

```
df %>%
  ggplot(aes(x=solutions)) +
  geom_histogram(bins=30) +
  theme_bw()
```



CONTROL VARIABLES

GENDER AND CONDITION

In [6]:

```
#library(janitor) # Gives tabyl

#df %>% tabyl(condition, gender)
#df %>% tabyl(condition, gender) %>% chisq.test()
table(df$condition, df$gender)
```

```
      M  F
IPS  51 51
PSI  44 54
```

AGE GROUP AND CONDITION

In [7]:
```
#df %>% tabyl(condition, age.group)
#df %>% tabyl(condition, age.group) %>% chisq.test()
table(df$condition, df$age.group)
```

```
      kids  students  professionals
IPS     35        35             32
PSI     27        38             33
```

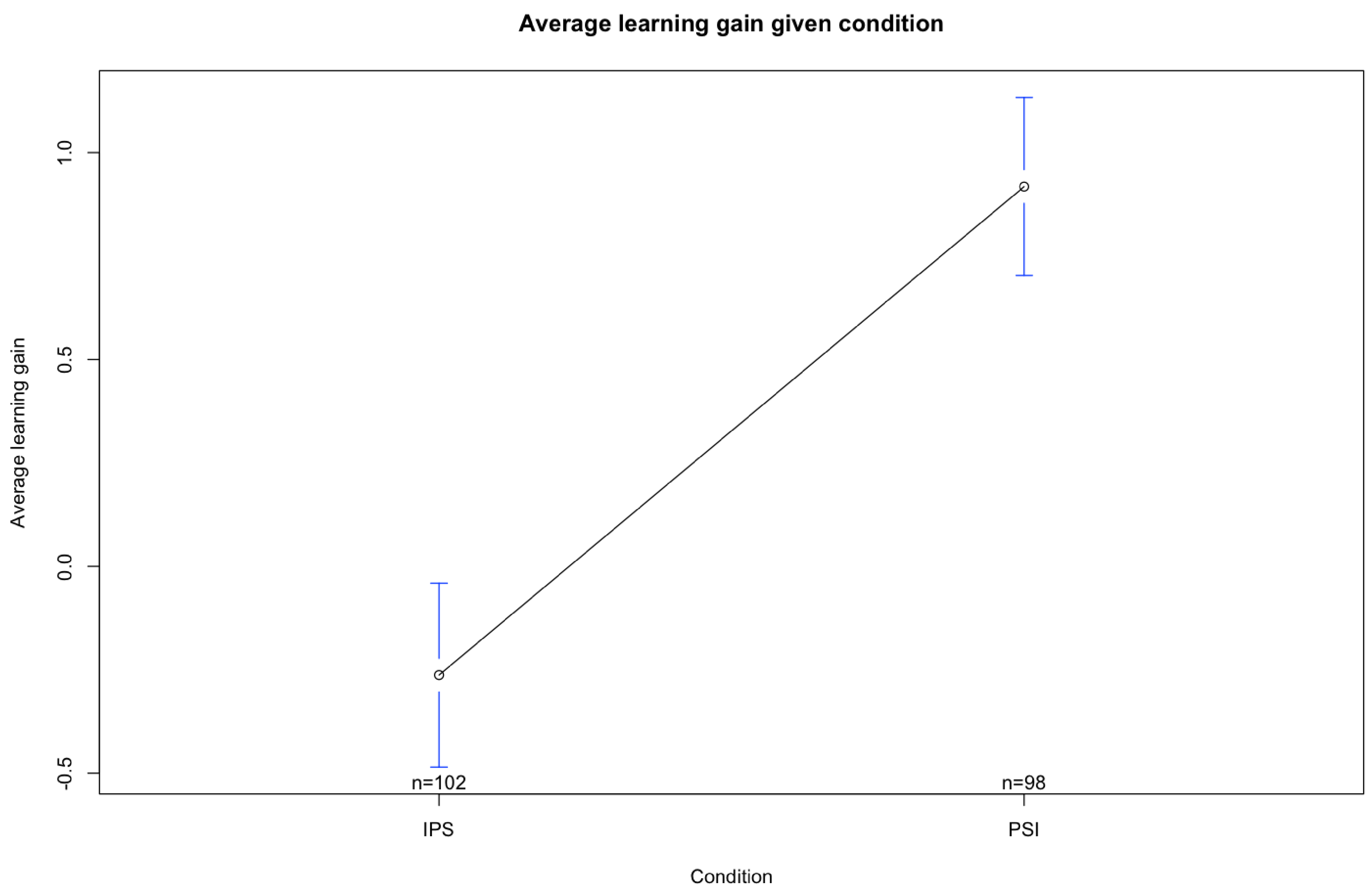
AGE GROUP AND GENDER

In [8]:

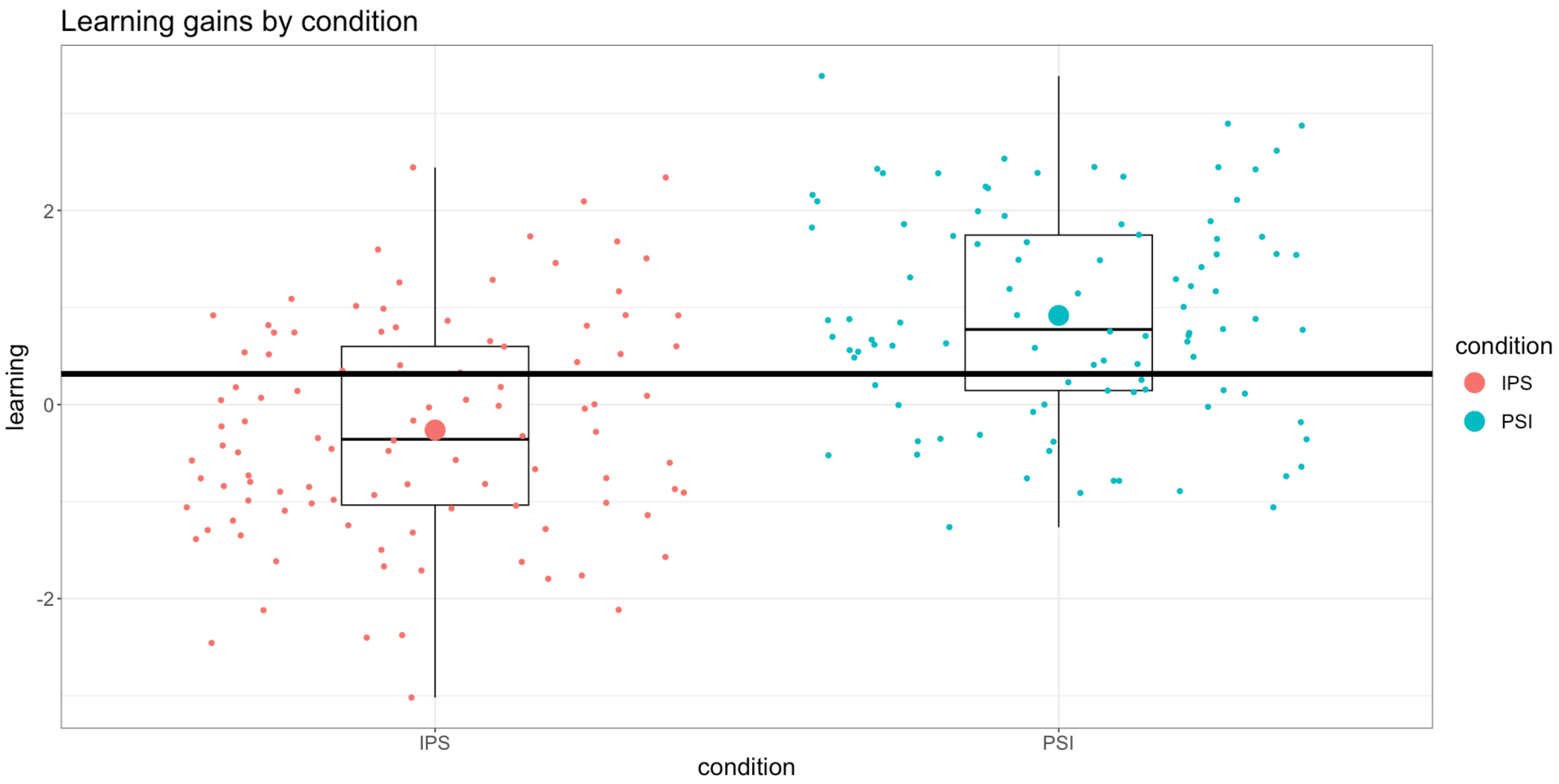
```
#df %>% tabyl(gender, age.group)
#df %>% tabyl(gender, age.group) %>% chisq.test()
table(df$gender, df$age.group)
```

```
   kids students professionals
M   30       37            28
F   32       36            37
```

QUESTION 1: DOES THE EXPERIMENTAL TREATMENT AFFECT LEARNING ?



Looking at the means and confidence intervals is a good first step, but we need to do a statistical test to determine if the difference is statistically significant. The ANOVA test relies on a comparison of variances between groups and within groups. The following chart illustrates these variances



ANOVA WITH ONE FACTOR

ANOVA compares the variation "between" the groups and "within" the groups based on their ratio. It assumes that the measured variable is normally distributed in each group and that the variance is the same in each group.

$$F = \frac{MeanSquares_{between}}{MeanSqares_{within}}$$

The sample variance (Mean Sum of Squares) is computed as the Sums of Squares divided by the Degrees of freedom.

$$F = \frac{SS_{between}/df_{between}}{SS_{within}/df_{within}}$$

If F is larger than 1, the differences between the groups are more important than the differences inside the groups.

TOTAL VARIANCE: BETWEEN AND WITHIN GROUPS

Variance is a measure of "spread" based on the average squared deviation from the mean.

$$SS_{total} = \sum_{i=1,j=1}^{n_i,k} (Y_{ij} - \bar{Y})^2$$

MEAN SUMS OF SQUARE BETWEEN GROUPS

$$SS_{between} = \sum_{i=1}^{k} n_i(\bar{Y_i} - \bar{Y})^2$$

Where k is the number of groups $\bar{Y_i}$ is the mean for group i and $\bar{Y}$ is the grand mean. We multiply by n_i because we account for the difference between the group mean and the global mean for each observation.

Finally, we divide by the degrees of freedom: $MS_{Between} = SS_{between}/df_{between}$ where the $df_{between}$ is $k - 1$.

MEAN SUMS OF SQUARE WITHIN GROUPS

For each group i we have $SS_i = \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$ which is essentially the difference between each observation j of the group and the mean for that group.

To obtain the Mean sums of squares, we add up the SS_i for each group:
$SS_{within} = \sum_{i=1}^{k} SS_i$, we divide by the degrees of freedom :
$MS_{within} = SS_{within}/df_{within}$ where the degrees of freedom are is the sum of the degrees of freedom for each subgroup.
$df_{within} = \sum_{i=1}^{k} df_i = \sum_{i=1}^{k} (n_i - 1) = N - k$

COMPUTING VARIANCES BY HAND IN R

In [11]:
```
Y_bar = mean(df$learning) # The mean of learning for all subjects
Y_s = sd(df$learning) # The standard deviation of the learning for all subjects
```

COMPUTING SSTOTAL

In [12]:
```
ss_total = sum((df$learning - Y_bar)^2)
ss_total
```

310.564206640921

COMPUTING SSB

In [13]:

```
between_ss = function(x) {
    sum(length(x)*(mean(x) - Y_bar)^2)
}

ss_between = sum(tapply(df$learning, df$condition, between_ss))
ss_between

mss_between = ss_between / 1 # (k groups - 1)
mss_between
```

69.6619314204476

69.6619314204476

COMPUTING SSW

In [14]:
```
within_ss = function(x) {
    sum((x - mean(x))^2)
}
ss_within = sum(tapply(df$learning, df$condition, within_ss))
ss_within

mss_within = ss_within / (length(df$learning) - 2) # N - k groups
mss_within
```

240.902275220474

1.21667815767916

THE F-RATIO

$$F = \frac{MeanSquares_{between}}{MeanSqares_{within}} = \frac{69.66193}{1.216678} = 57.25584$$

In [15]:

```
# Compute the ratio of the between group variance to the within group variance
F = mss_between / mss_within
F
```

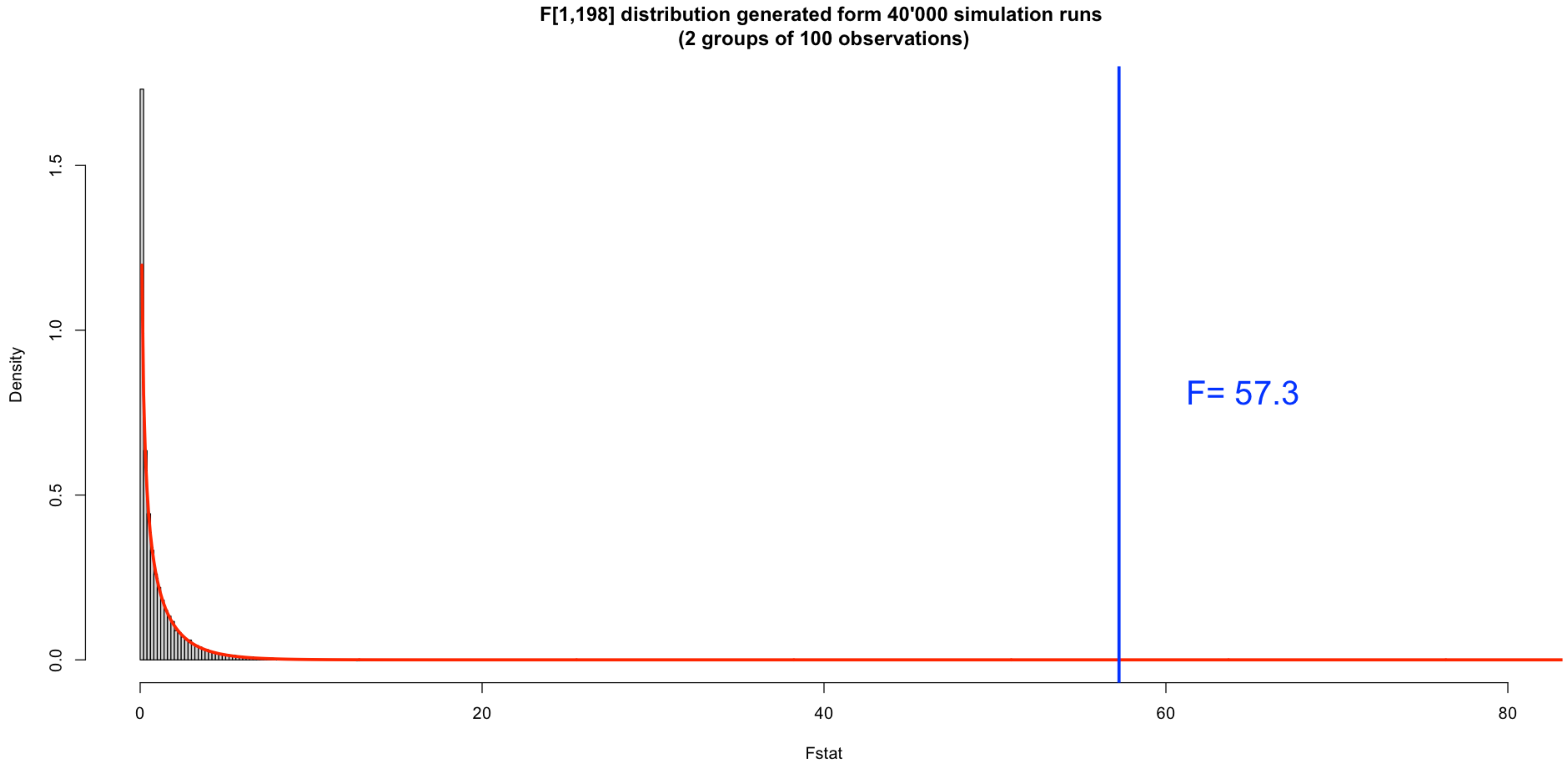
57.2558412270087

This F-ratio (computed from our experimental groups) is to be compared with a F distribution parametrised with df=1 (2 groups - 1) and df= 198 (200 subjects - 2 groups).

The theoretical F distribution corresponds to the F-ratios that would be obtained when:

- two samples are drawn from two populations with means μ_1 and μ_2
- two populations have the same mean: $\mu_1 = \mu_2$. This corresponds to our *null hypothesis*.
- three populations have the same variance: σ^2

We generated 40000 runs of a simulation that draws two samples of 100 observations from a normal population with the same mean and variance. For each randomly generated example, we computed the F-ratio. Here is the distribution of these 40000 F-ratios and the corresponding F[1,198] distribution.



THE ANOVA TEST

Given our observed F-ratio and the theoretical F-distribution for 2 groups (df1 = 2 groups - 1 = 1) of 100 observations (df2 = 200 observations - 2 groups = 198), we now can perform our test.

Under the "Null" hypothesis for the ANOVA, the F-ratio for 2 groups and a sample size of 200, which have the **same mean** and **same variance**, follows a F-distribution with [1,198] degrees of freedom.

- $H_0 : \mu_1 = \mu_2 = \ldots = \mu_n$

The "Alternative" hypothesis is that:

- $H_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_n$

HOW TO DECIDE WHETHER OUR F-RATIO IS "FOLLOWING" THE F-DISTRIBUTION ?

P-VALUE

Our experiment produced a F-ratio which is rather extreme: there are only 0.00000000000014% of the theoretical F-ratios for such experiments that would be larger than the value we observed. This proportion is called the *p-value*: what is the probability to have drawn samples for our experiment which would produce a F-ratio larger than 57.3 It corresponds to the area under the curve to the right of F = 57.3

In [17]:
```
# pf gives the probability of getting an F value greater than F
p.value = pf(F, 1, 198, lower.tail = FALSE)
p.value
```

1.4115138248261e-12

In social sciences, it is commonly accepted that to reject the null hypothesis, i.e. to say that our F-ratio does probably not stem from the theoretical F-distribition, it has to come from the 5% most extreme values. *This is called the alpha level*, written $\alpha = 0.05$.

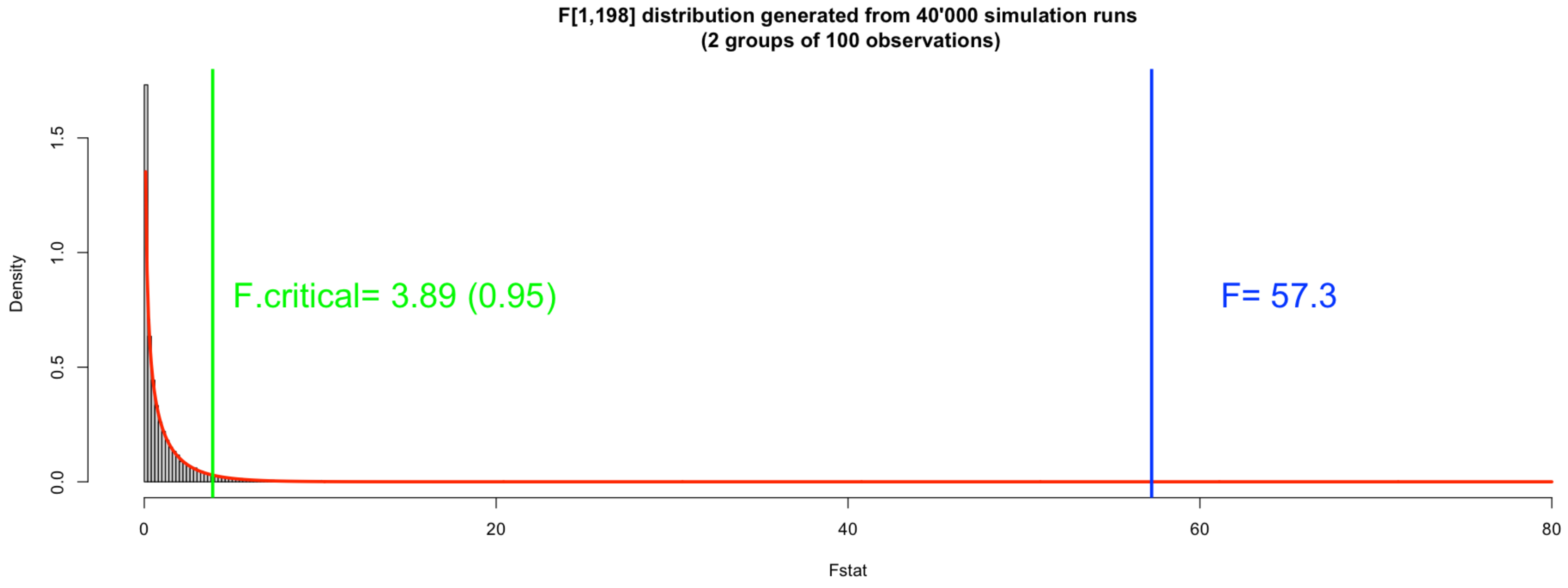
In our example $p = 1.411514e - 12 <<< \alpha = 0.05$ and hence we **reject the Null hypothesis**. Therefore we conclude that the two samples do not belong to two populations with the same means.

CRITICAL VALUE

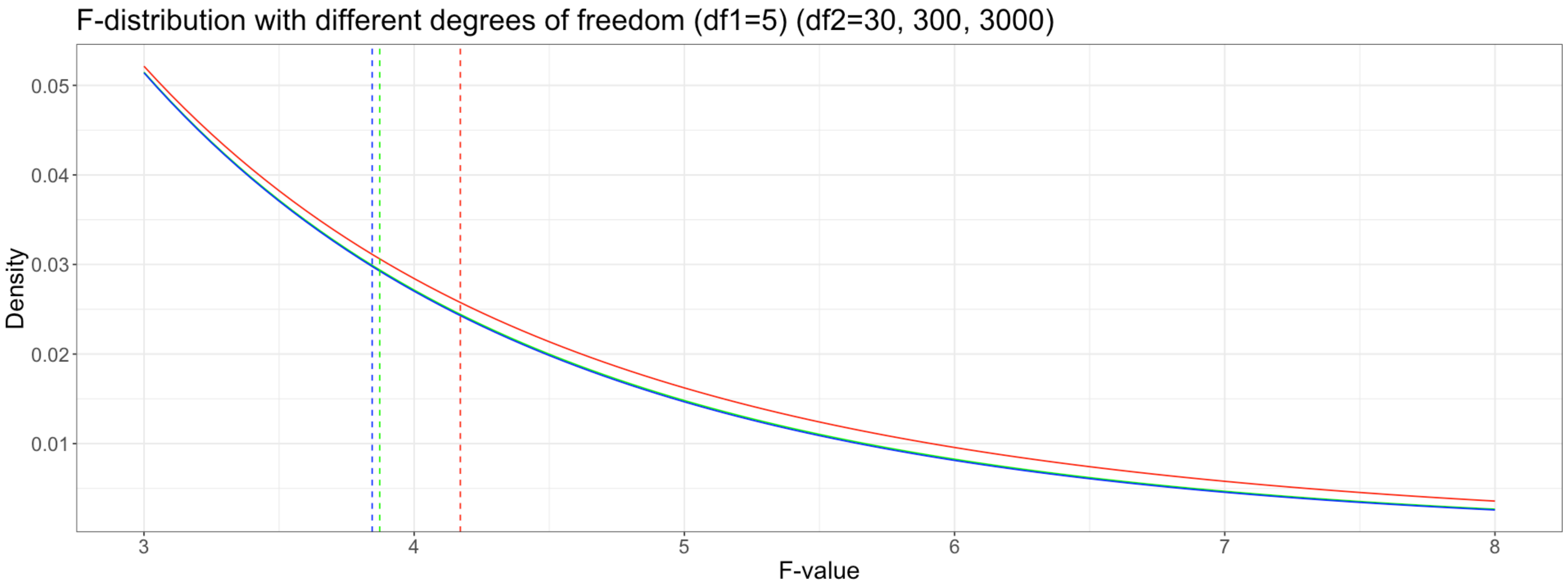
What is the F-ratio above which we can reject the Null hypothesis ? This value is called the *critical value* and corresponds to the F value for a probablity of $1 - \alpha$, i.e. 0.95.

In [18]:
```
alpha = 0.05
F.critical = qf(1-alpha, df1=1, df2=198)
F.critical
```

3.88885293289187



The degrees of freedom affect the shape of the F-distribution. The second degree of freedom change the shape of the tails. Less observations lead to thicker tails and therefore to larger critical values.



In [21]:

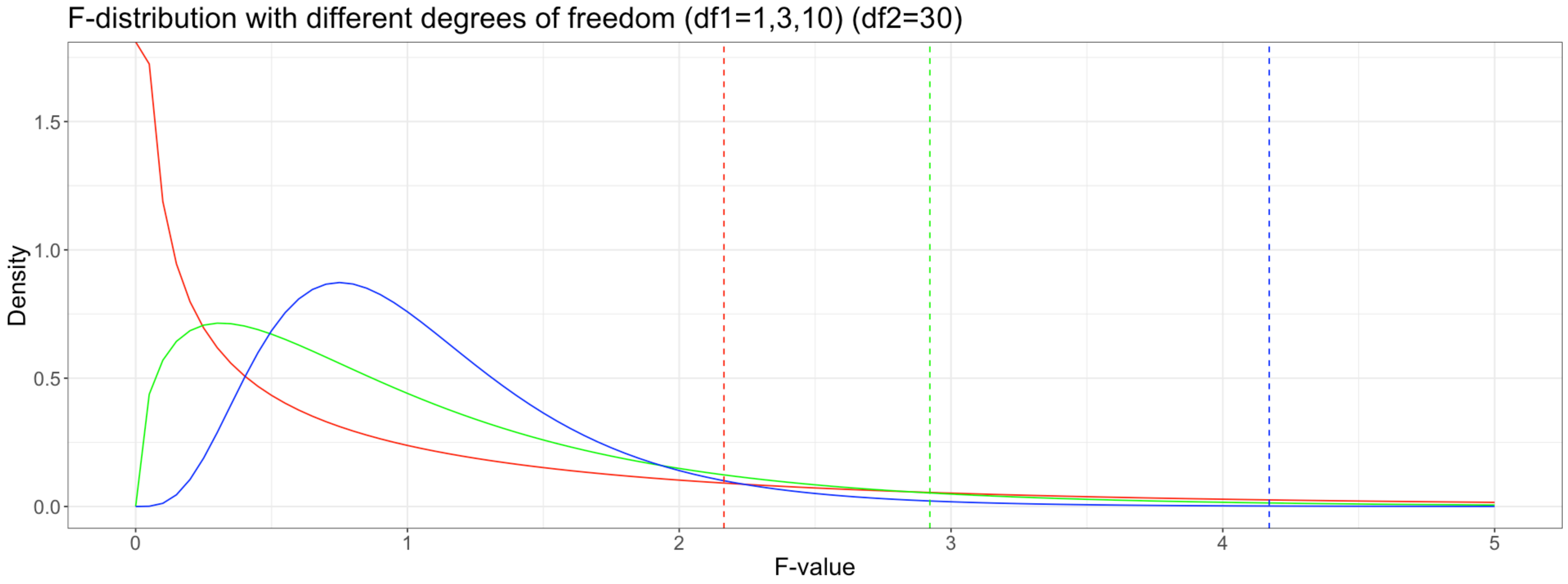
```
alpha = 0.05
F.critical = qf(1-alpha, df1=1, df2=30)
F.critical

alpha = 0.05
F.critical = qf(1-alpha, df1=1, df2=300)
F.critical

alpha = 0.05
F.critical = qf(1-alpha, df1=1, df2=3000)
F.critical
```

4.17087678576669

3.87264226173045

3.84456038697256

The degrees of freedom affect the shape of the F-distribution. The first degree of freedom change the shape of the curve.



ANOVA IN R

Steps:

- Step 1: Build a linear model with DV ~ IV `model = lm(DV ~IV)`:
- Step 2: Calculates type-II or type-III analysis-of-variance tables `Anova(model)`
- Step 3: Check assumptions
  - Normality
  - Homoscedasticity

STEP 1 : BUILD A LINEAR MODEL

NB: specify the contrasts that are used for the linear model as "contr.sum", which is not the default in R.

In [23]:

```
model.0 <- lm(learning ~ condition,
              contrasts=list(condition=contr.sum),
              data=df)
summary(model.0)
```

```
Call:
lm(formula = learning ~ condition, data = df, contrasts = list(condition =

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7555 -0.7754 -0.1243  0.8605  2.7054 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.32767    0.07801   4.200 4.03e-05 ***
condition1  -0.59030    0.07801  -7.567 1.41e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.103 on 198 degrees of freedom
Multiple R-squared:  0.2243,	Adjusted R-squared:  0.2204 
F-statistic: 57.26 on 1 and 198 DF,  p-value: 1.412e-12
```

STEP 2: LOOK AT THE ANOVA "INTERPRETATION" OF THE MODEL

In [24]:
```
library(car) # load library car first.
Anova(model.0, type="II")
```

A anova: 2 x 4

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
condition	69.66193	1	57.25584	1.411514e-12
Residuals	240.90228	198	NA	NA

STEP 3: CHECK ASSUMPTIONS

CHECKING NORMALITY ASSUMPTIONS

I present three methods to check the normality of the residuals for our linear model.

- The Shapiro Wilks test (available as `shapiro.test()`)
- The Kolmogorov-Smirnov test (available as `ks.test()`)
- a visual inspection test.

SHAPIRO.TEST : TESTING NORMALITY OF RESIDUALS IN EACH GROUP

The Shapiro-Wilks test allows to test whether a variable is normally distributed.

H_0: The sample is normally distributed.

H_1: The sample *is not* normally distributed.

In [25]:
```
shapiro.test(model.0$residuals)
```

```
	Shapiro-Wilk normality test

data:  model.0$residuals
W = 0.9902, p-value = 0.1909
```

The p-value is larger than 0.05 and therefore we cannot reject the Null hypothesis. According to this test, the residuals from our model are normally distributed.

The `shapiro.test()` is very sensitive to deviations from normality, especially if the sample size is large. Textbooks usually recommend checking the normality assumption visually (with qq plots) rather than through tests.

KOLMOGOROV-SMIRNOV TEST: TESTING NORMALITY OF RESIDUALS

The Kolmogorov-Smirnov test allows to test whether two samples were drawn from the same distribution. This allows to compare our observations with a sample that follows a normal distibution with the same mean and standard deviation. This test is preferred to the Shapiro Wilks test for large samples.

H_0 : The two samples stem from the same distribution

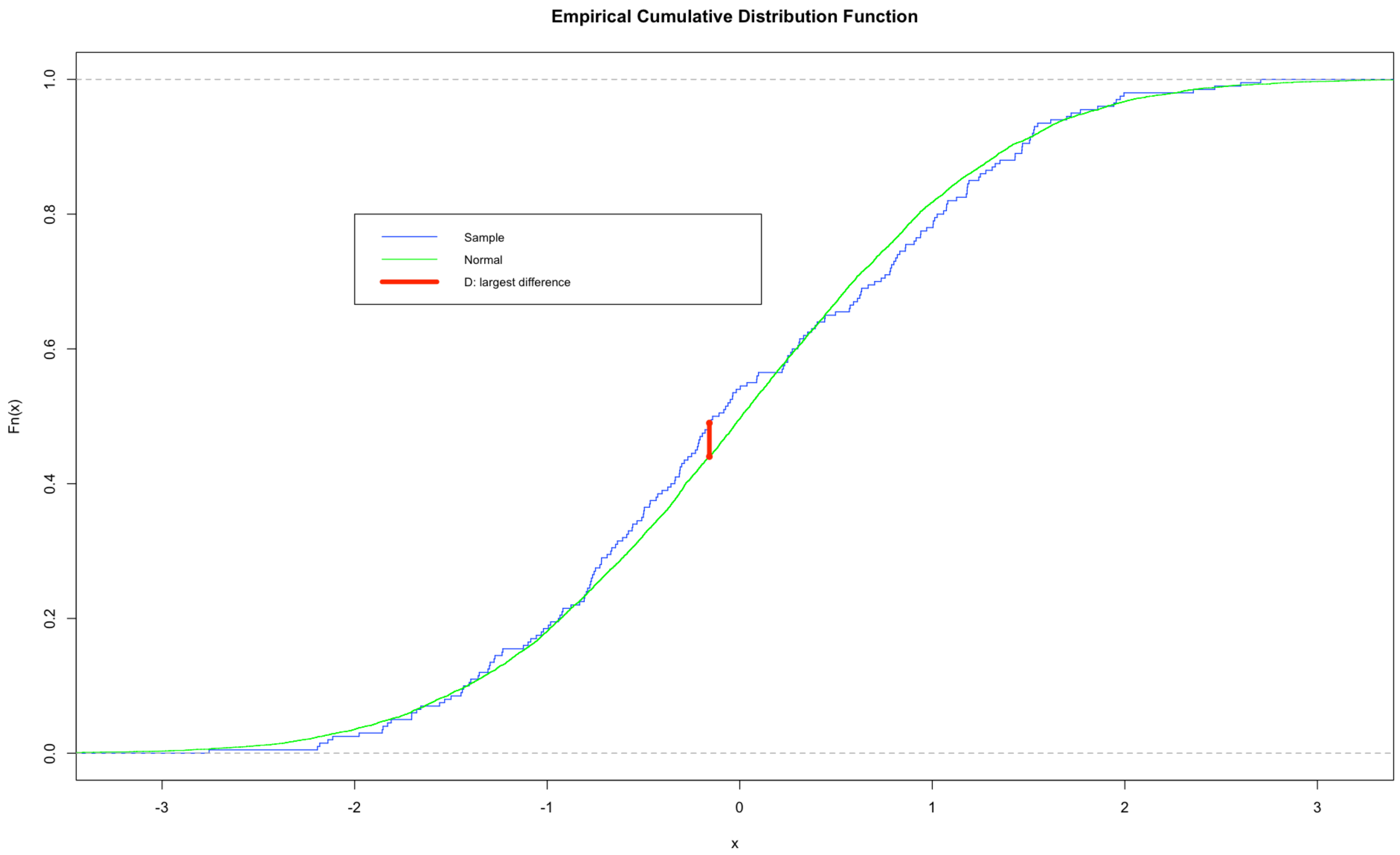
H_1 : The two samples *do not* stem from the same distribution

In [26]:

```
x <- model.0$residuals
ks.test(x, "pnorm", mean(x, na.rm = T), sd(x, na.rm = T))
```

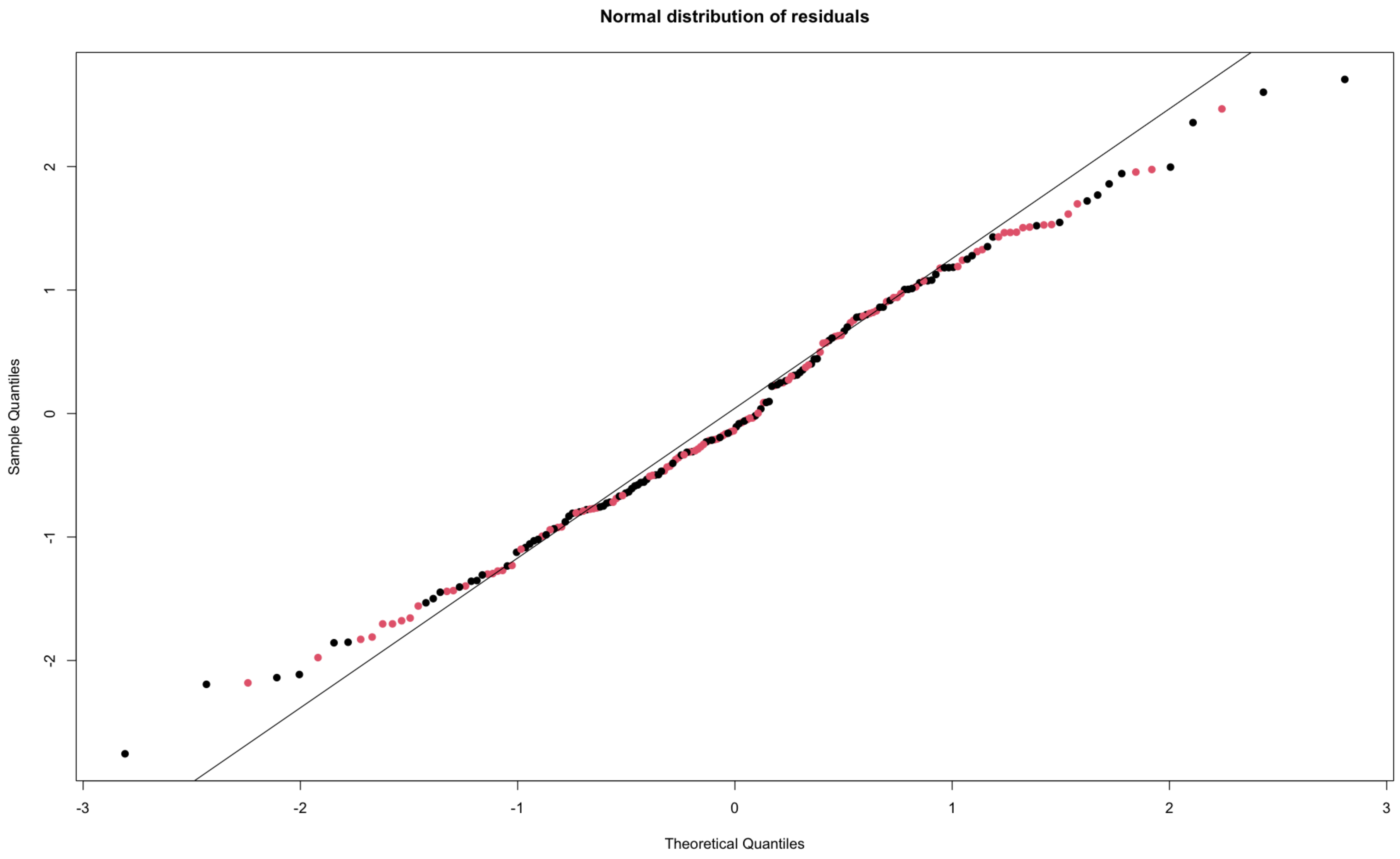
```
        Asymptotic one-sample Kolmogorov-Smirnov test

data:  x
D = 0.050941, p-value = 0.677
alternative hypothesis: two-sided
```

The p-value is larger than 0.05, we therefore cannot reject H_0 and hence conclude that it is likely that the residuals follow a normal distribution.


Empirical Cumulative Distribution Function
Sample
Normal
D: largest difference
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3
x

VISUALLY CHECKING NORMALITY OF RESIDUALS

In [29]:
```
options(repr.plot.width=16, repr.plot.height=10)
qqnorm(model.0$residuals, col=df$condition, pch=19,
       main="Normal distribution of residuals")
qqline(model.0$residuals)
```



CHECKING THE HOMOSCEDASTICITY OF THE RESIDUALS

We would like to have the same variance of residuals across groups. This means that the model explains similarly well observations from both groups. If this was not the case, we'd have for example very similar errors for all observations in the IPS group and a larger variation of errors in the PSI group. This would indicate that there is something "wrong" in the measured data, e.g. all individuals from IPS have the same learning gain, whereas individuals from the PSI group have a spread of learning gains.

Equality of variances can be tested with the `bartlett.test()` in R.

H_0: The variances are the same in the groups

H_1: The variances *are not* the same in the groups

In [30]:
```
bartlett.test(residuals(model.0) ~ df$condition)
```

```
	Bartlett test of homogeneity of variances

data:  residuals(model.0) by df$condition
Bartlett's K-squared = 0.27535, df = 1, p-value = 0.5998
```

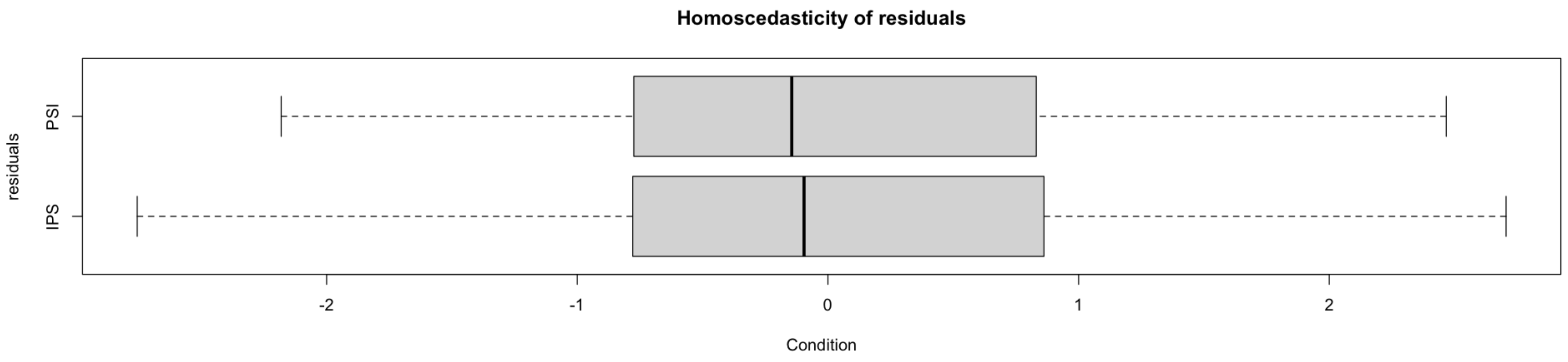
In our case, the p-value is much larger than .05 which does not allow us to reject the null hypothesis H_0. Hence we conclude that the variances are equal in both groups.

VISUAL INSPECTION OF EQUAL VARIANCES

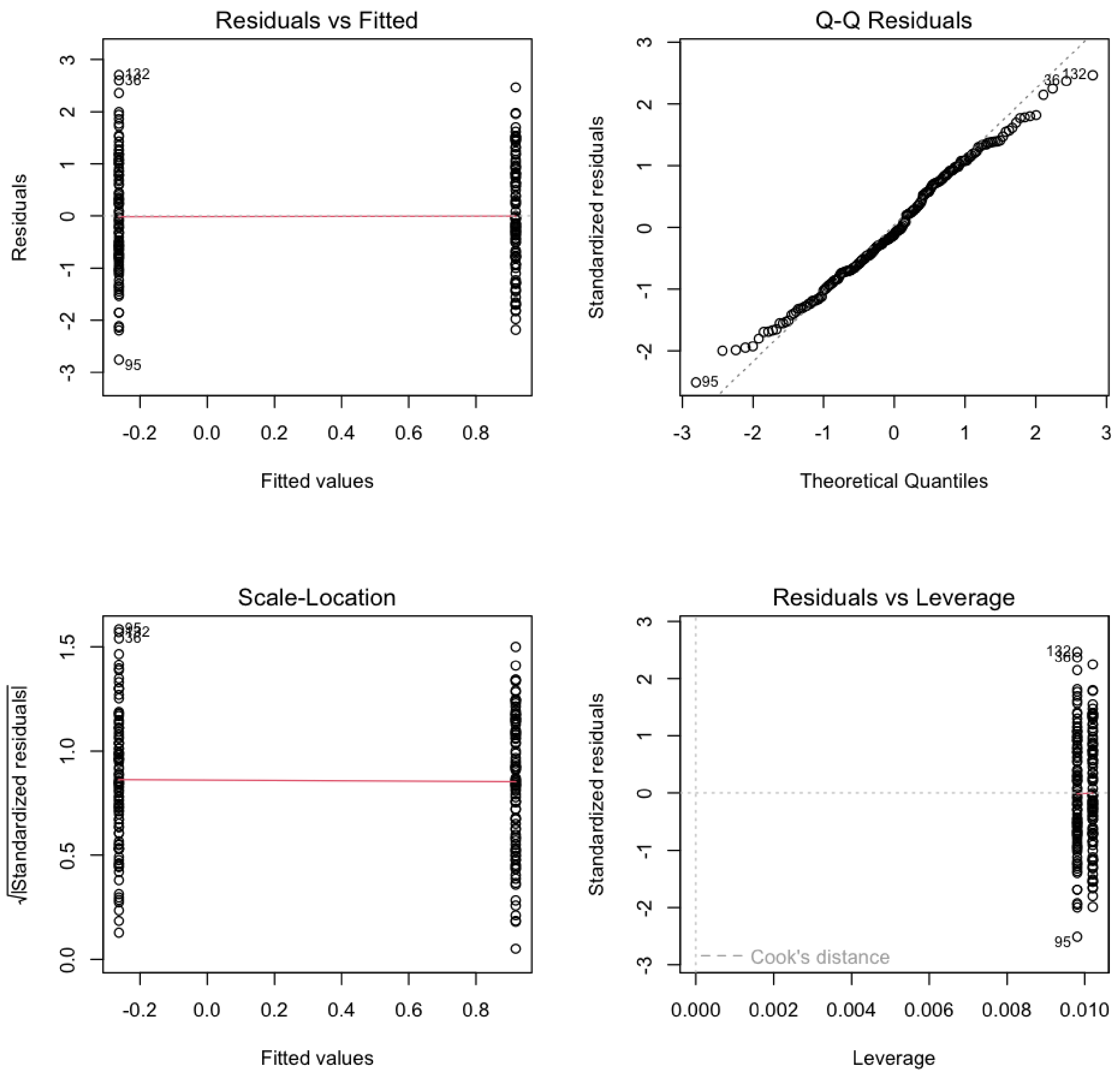
An alternative was to check for equality of variances conists of plotting boxplots of the residuals. If the shape of the boxplots is more or less the same, the variances are more or less equivalent.

In [31]:

```
options(repr.plot.width=16, repr.plot.height=4)
boxplot(model.0$residuals ~ df$condition,
        main="Homoscedasticity of residuals",
        ylab="residuals",
        xlab="Condition", horizontal=TRUE)
```



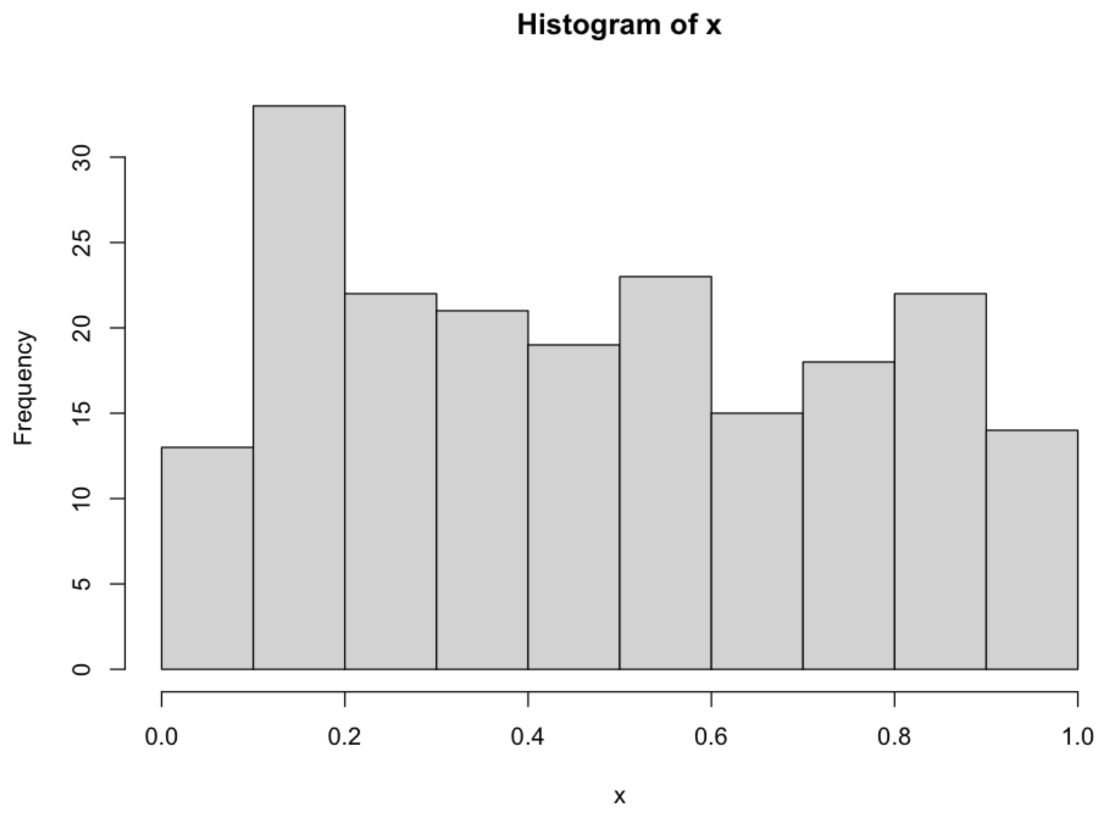
In [33]:
```
par(mfrow = c(2, 2))
plot(model.0)
```

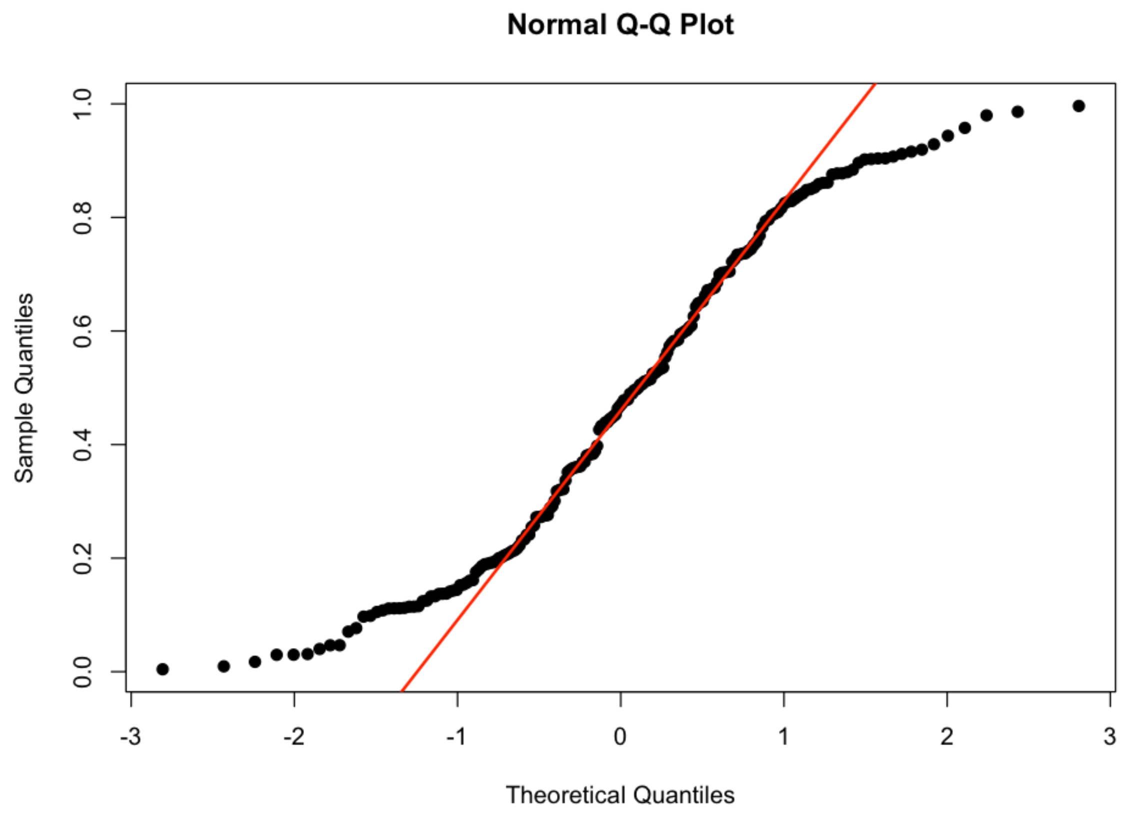


HOW WOULD IT LOOK WITH NON-NORMAL VARIABLES ?

In [34]:
```
x <- runif(200)
shapiro.test(x)
```

```
        Shapiro-Wilk normality test

data:  x
W = 0.94965, p-value = 1.746e-06
```



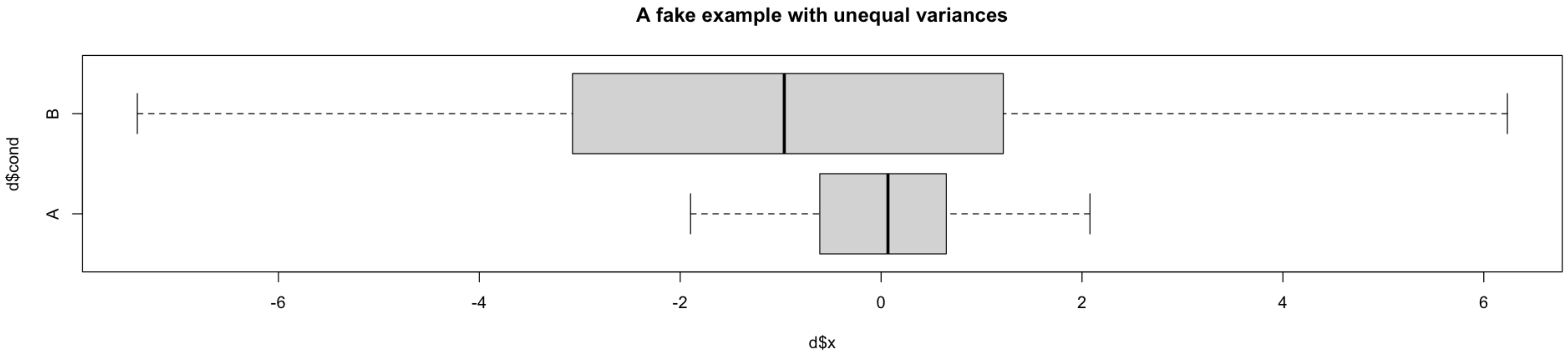


HOW WOULD IT LOOK WITH UNEQUAL VARIANCES ?

In [37]:

```
options(repr.plot.width=16, repr.plot.height=4)
bartlett.test(d$x ~ d$cond)
boxplot(d$x ~ d$cond, horizontal=TRUE, main="A fake example with unequal varian
```

```
	Bartlett test of homogeneity of variances

data:  d$x by d$cond
Bartlett's K-squared = 129.17, df = 1, p-value < 2.2e-16
```



WHAT IF ASSUMPTIONS ARE NOT MET ?

Normality: ANOVA is said to be pretty robust against deviations of normality, which means that the validity of p-values are not too much affected by skew (the distribution is asymetric) or kurtosis (the distribution is too heavy or too light tailed).

=> Data Transformation. Trying to transform the dependent variable so that the distribution approaches normality, by taking 1/x, log(x) or sqrt(x).

=> Using a non-parametric equivalent for ANOVA: Kruskall-Wallis rank test.

Equality of variance: Deviations for the equality of variance have most impact on the result of the ANOVA if the group sizes are unequal.

=> Using the Welch correction for `oneway.test()` by specifying `var.equal=FALSE`.

RUNNING A NON-PARAMETRIC KRUSKALL-WALLIS AS AN ALTERNATIVE

The principle for the Kurskall Wallis test is very similar to the idea behind ANOVA. The difference is that rather than using the raw scores, the Kruskall-Wallis test relies on **ranks**. This test does **not make assumptions** about the distribution of the residuals, nor about the variances.

H_0: The mean ranks of the groups are the same.

H_1: The mean ranks of the groups *are not* the same.

The decision variable:

$$H = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r_{i\cdot}} - \bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\bar{r_{ij}} - \bar{r})^2} \sim \chi^2_{[g-1]}$$

In [38]:
```
kruskal.test(learning ~ condition, data=df)
```

```
	Kruskal-Wallis rank sum test

data:  learning by condition
Kruskal-Wallis chi-squared = 43.443, df = 1, p-value = 4.365e-11
```

From the results of the test we see that we can reject the Null hypothesis ($p < .05$) and therefore conclude that the mean ranks are different among the two groups.

ANOVA WITH 2 FACTORS

We now add a control variable (`age.group`) as a new factor to the ANOVA. This introduces the possibility for interaction between variables.

In order to test for a potential moderation effect (the effect of the condition varies depending on another variable), we include interaction effects in the linear model.

The total variance is now decomposed into:

$$SSTotal = SSFactor1 + SSFactor2 + SSInteraction + SSWithin$$

The degrees of freedom for an interaction effect between 2 variables with k and m levels are $(k-1)(m-1)$, with condition and gender: $(2-1)(2-1) = 1$ and with condition and age group $(2-1)(3-1) = 2$.

In the specification of the model, the interaction between 2 factors is written with a column as in `condition:age.group`.

In [39]:

```
model.2 <- lm(learning ~
                condition +
                age.group +
                condition:age.group,
              contrasts=list(condition=contr.sum, age.group=contr.sum),
              data=df)
summary(model.2)
```

```
Call:
lm(formula = learning ~ condition + age.group + condition:age.group,
    data = df, contrasts = list(condition = contr.sum, age.group = contr.su

Residuals:
     Min       1Q   Median       3Q      Max
-2.55457 -0.74262 -0.03032  0.87658  2.24990

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            0.31422    0.07583   4.144 5.10e-05 ***
condition1            -0.56384    0.07583  -7.435 3.27e-12 ***
age.group1            -0.33434    0.10940  -3.056  0.00256 **
age.group2             0.16601    0.10465   1.586  0.11428
condition1:age.group1 -0.02690    0.10940  -0.246  0.80599
condition1:age.group2 -0.24722    0.10465  -2.362  0.01915 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.066 on 194 degrees of freedom
Multiple R-squared:  0.2897,    Adjusted R-squared:  0.2714
F-statistic: 15.82 on 5 and 194 DF,  p-value: 4.628e-13
```

In [40]:
```
Anova(model.2, type="III")
```

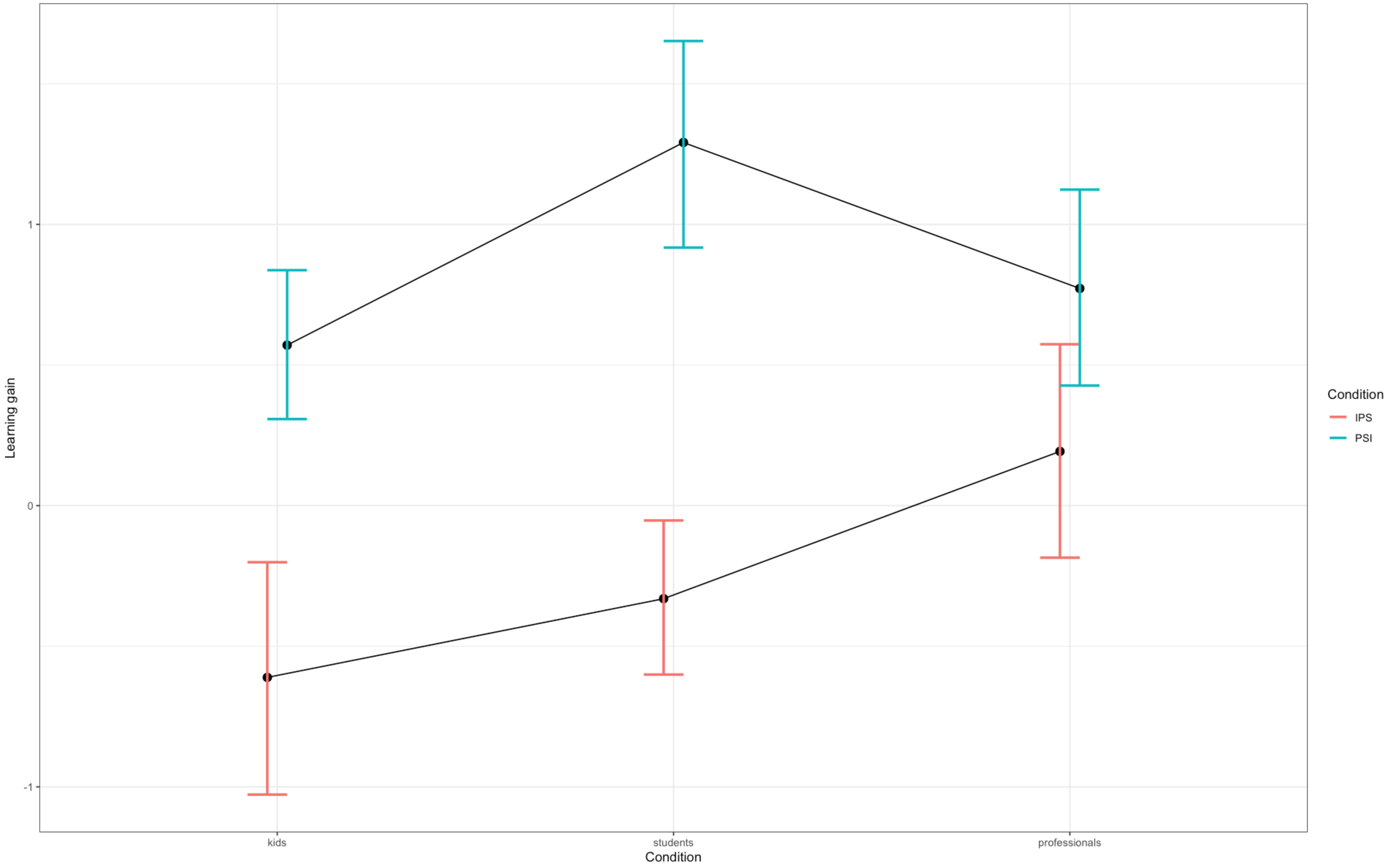
A anova: 5 x 4

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
(Intercept)	19.525149	1	17.170622	5.099499e-05
condition	62.866648	1	55.285595	3.267421e-12
age.group	10.629625	2	4.673902	1.041208e-02
condition:age.group	9.364644	2	4.117683	1.772672e-02
Residuals	220.602306	194	NA	NA

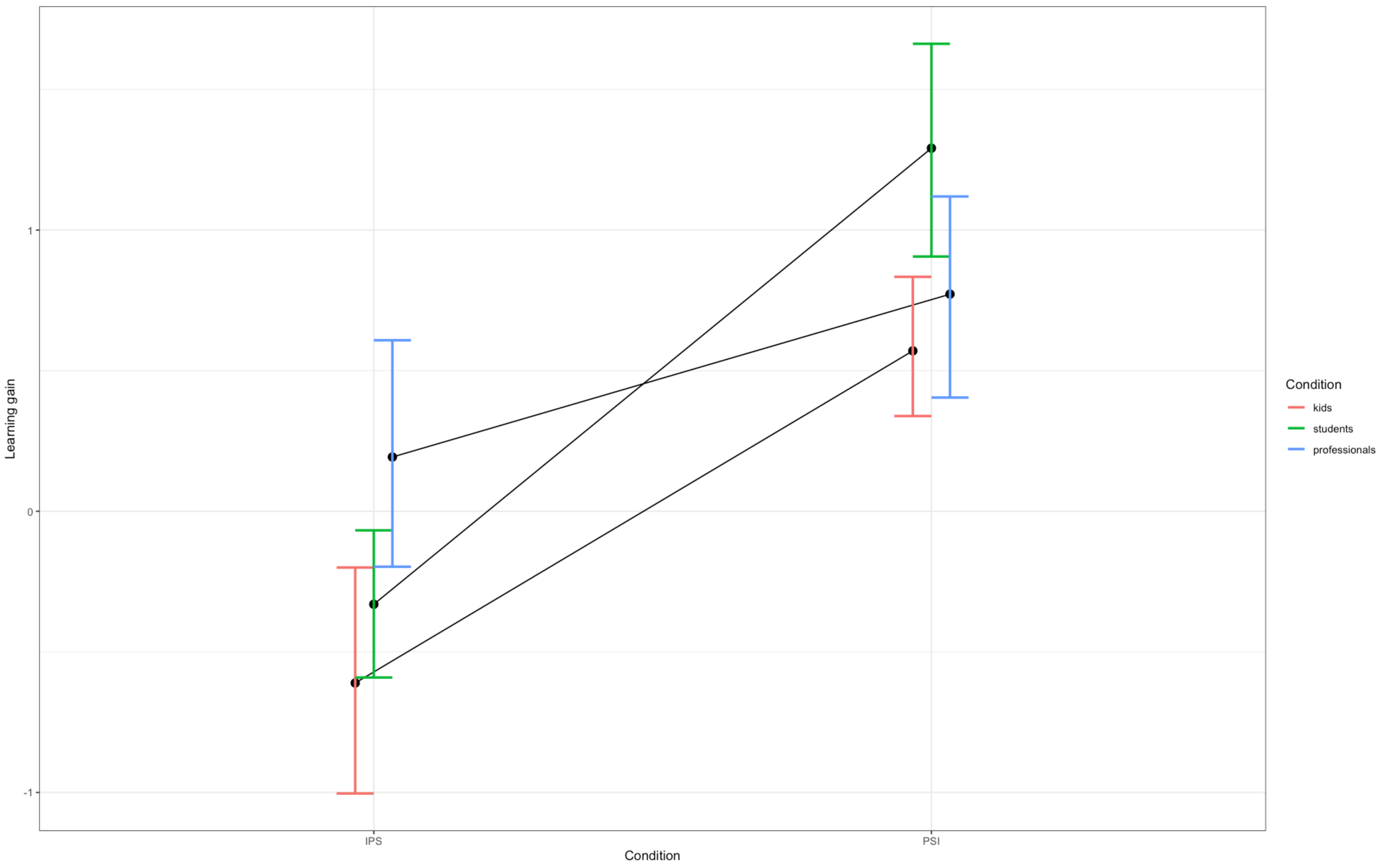
IS THERE AN INTERACTION ?

- When there is an interaction, we use **type III sums of squares** and don't interpret main effects.
- If there are no interactions, switch to a model that only includes main effects and use **type II sums of squares**.

AGE GROUP AND CONDITION



```
Warning message:
"Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead."
```



REPORTING THE ANOVA WITH INTERACTION

A two way ANOVA was conducted with the experimental condition, and the control variable age group. We tested for interactions between the condition and the control variables. There was a significant interaction effect between condition and age group ($F[2,194]=4.1177$, $p = .0104$). Inspection of the graphical patterns of the means indicates that the PSI condition worked especially well for students in comparison with kids and professionals.

GENDER AND CONDITION

Let's do the same analysis with the control variable `gender`. We start with a model that contains the interaction term (type III). Since there is no interactions between the factors, we re-run the model without interaction and use type II sums of squares.

In [43]:
```
model.with.interaction <- lm(learning ~ condition + gender + condition:gender,
                contrasts=list(condition=contr.sum, gender=contr.sum),
                data=df)

Anova(model.with.interaction, type="III")
```

A anova: 5 x 4

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
(Intercept)	20.3263220	1	16.7323913	6.281746e-05
condition	67.4350612	1	55.5117561	2.899331e-12
gender	2.4509083	1	2.0175591	1.570777e-01
condition:gender	0.4009976	1	0.3300966	5.662611e-01
Residuals	238.0986104	196	NA	NA

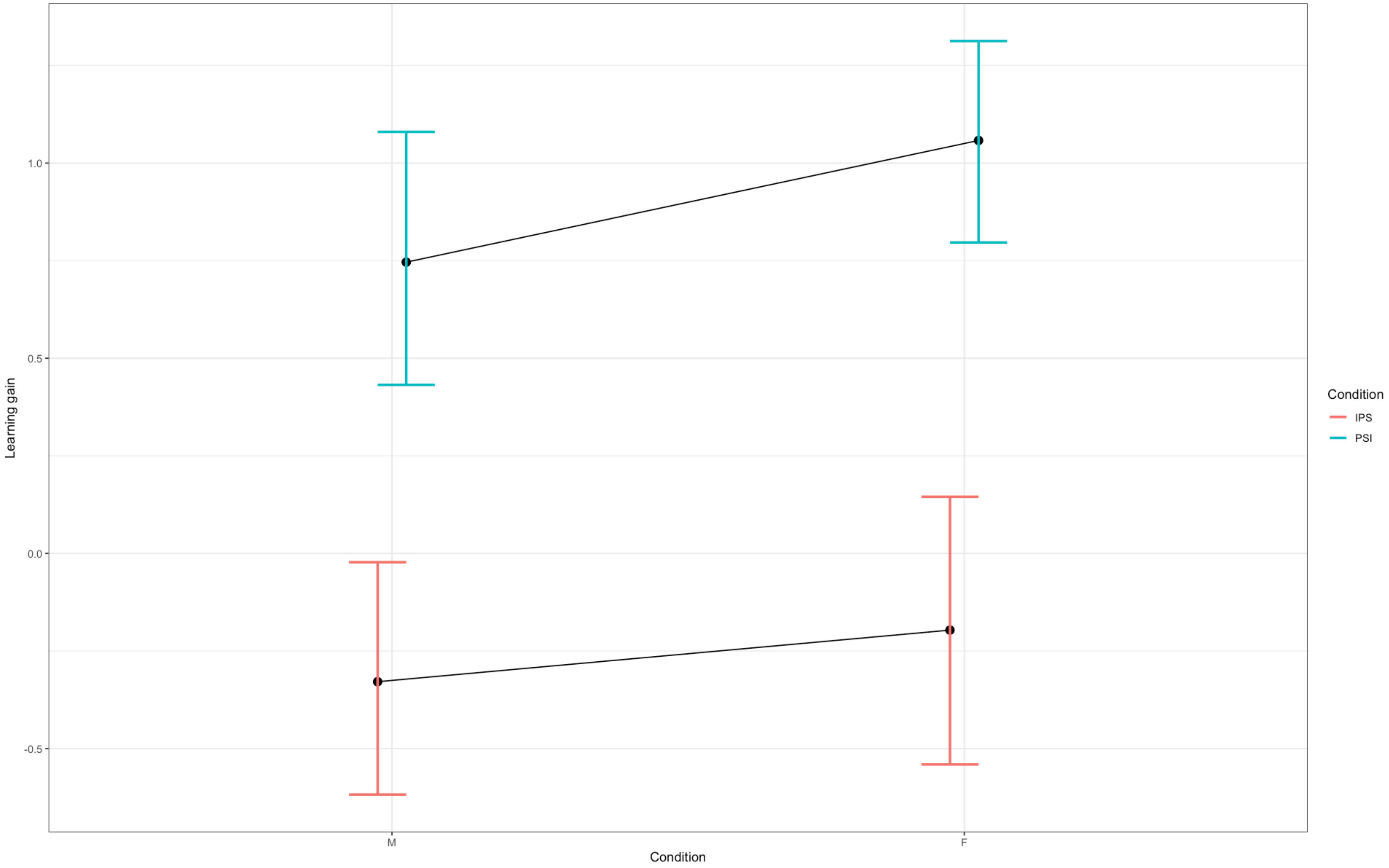
In [44]:

```
model.without.interaction <- lm(learning ~ condition + gender,
                contrasts=list(condition=contr.sum, gender=contr.sum),
                data=df)

Anova(model.without.interaction, type="II")
```

A anova: 3 x 4

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
condition	68.166682	1	56.305486	2.085951e-12
gender	2.402667	1	1.984596	1.604834e-01
Residuals	238.499608	197	NA	NA

GENDER



REPORTING THE ANOVA WITHOUT INTERACTION

In [46]:

```
tapply(df$learning, df$condition, mean)
tapply(df$learning, df$condition, sd)
```

IPS: -0.262625063017639 **PSI:** 0.91796637641757

IPS: 1.13120725443642 **PSI:** 1.07290750612109

A two way ANOVA was conducted with the experimental condition, and the control variable gender. There was no interaction effect between condition and gender. There is a main effect of the experimental condition ($F[1,197]=56.306$, $p<.000$). The subjects in the PSI group had a larger learning gain ($M=0.918$, $sd=1.07$) than the subjects in the IPS group ($M=-0.262$, $sd=1.13$). There was no main effect of gender ($F[1,197]=1.98$, $p > .05$).

In [ ]: