Research Design

08: factorial experiments; analysis of variance; issues with multiple testing

Scott Spencer | Columbia University

Initial questions?

factorial experiments

factorial experiments, **investigating multiple explanatory *factors* or sets of treatments**

factor

level

complete factorial experiment

fractional factorial experiment

main effects

interaction effects

$$
\begin{array}{c|c|c|c}
\ddots & a_0 & a_1 & \cdots \\
\hline
b_0 & Y_{a_0 b_0} & Y_{a_1 b_0} & \\
\hline
b_1 & Y_{a_0 b_1} & Y_{a_1 b_1} & \\
\hline
\vdots & & &
\end{array}
$$

$$\text{main effect } a = \frac{(a_1 b_0 - a_0 b_0) + (a_1 b_1 - a_0 b_1)}{2}$$

$$\text{interaction effect } ab = \frac{(a_1 b_1 - a_1 b_0) - (a_0 b_1 - a_0 b_0)}{2}$$

analysis of variance (ANOVA)

ANOVA, **generally — analyzing variation among three or more means**

ANOVA | This analysis compares ***variance ratios*** to determine whether or not significant ***differences exist among the means of several groups*** of observations, where ***each group follows a normal distribution***.

An analysis-of-variance extends the t-test, which is used to determine whether or not two means differ, to the case where there are ***three or more means***.

$$
\begin{array}{c|c|c|c}
\ddots & a_0 & a_1 & \dots \\
\hline
b_0 & Y_{a_0 b_0} & Y_{a_1 b_0} & \\
\hline
b_1 & Y_{a_0 b_1} & Y_{a_1 b_1} & \\
\hline
\vdots & & &
\end{array}
$$

$$\text{main effect } a = \frac{(a_1 b_0 - a_0 b_0) + (a_1 b_1 - a_0 b_1)}{2}$$

$$\text{interaction effect } ab = \frac{(a_1 b_1 - a_1 b_0) - (a_0 b_1 - a_0 b_0)}{2}$$

ANOVA, generally — analyzing variation among three or more means

ANOVA | This analysis compares ***variance ratios*** to determine whether or not significant ***differences exist among the means of several groups*** of observations, where ***each group follows a normal distribution***.

An analysis-of-variance extends the t-test, which is used to determine whether or not two means differ, to the case where there are ***three or more means***.

ASSUMPTIONS | inferences assume,

units are *independent, identically distributed*
variance is equal (*homoscedastic*) within each group
errors are *normally distributed*

ANOVA, the test — do differences exist in the means of groups not explained by sampling and variation?

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_k \quad , \quad H_a : \theta_i \neq \theta_j \text{ for some } i, j$$

NOTE | we don't always — *or usually* — believe there is zero effect, nor would we find that to be interesting.

ANOVA, **one-way analysis of variance**

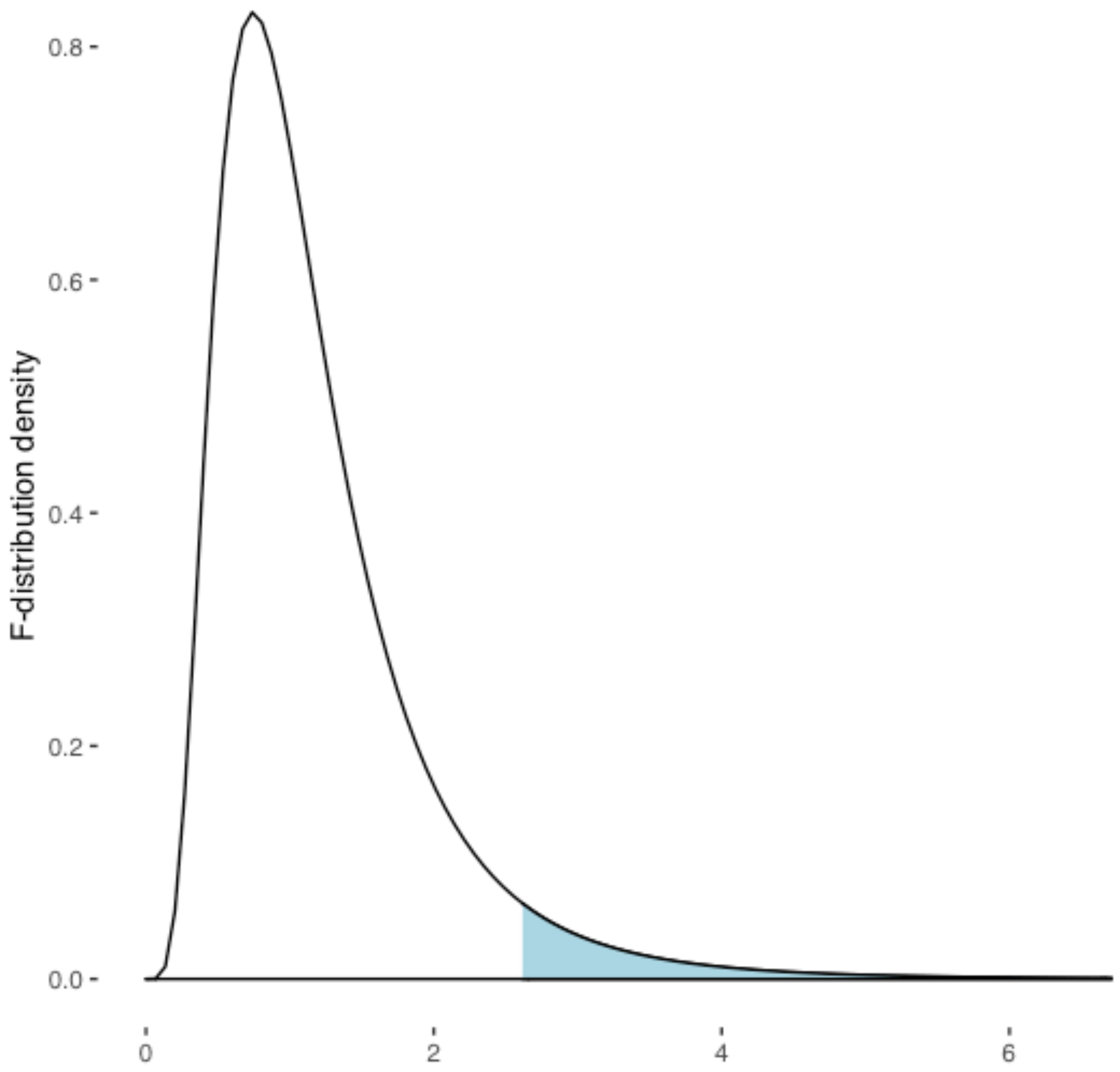
Source of variation	Degrees of freedom	Sum of squares	Mean square	F statistic
Between treatment groups	$k - 1$	$\text{SS}_\text{B} = \sum n_i(\bar{y}_i - \bar{y})^2$	$\text{MS}_\text{B} = \dfrac{\text{SS}_\text{B}}{k-1}$	$F = \dfrac{\text{MS}_\text{B}}{\text{MS}_\text{W}}$
Within treatment groups	$N - k$	$\text{SS}_\text{W} = \sum \sum (y_{ij} - \bar{y}_{i\cdot})^2$	$\text{MS}_\text{W} = \dfrac{\text{SS}_\text{W}}{N-k}$	
Total	$N - 1$	$\text{SS}_\text{T} = \sum \sum (y_{ij} - \bar{y})^2$		

ANOVA, test — where does the *F*-statistic fall in the *F*-distribution?

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_k \quad , \quad H_a : \theta_i \neq \theta_j \text{ for some } i, j$$

```
library(ggplot2)
library(ggthemes)

df1   <- 15
df2   <- 12
alpha <- 0.05
xmax  <- qf(0.001, df1, df2, lower.tail = FALSE)

ggplot() +
  theme_tufte(base_family = "sans") +
  stat_function(fun  = df,
                args = list(df1 = df1, df2 = df2),
                geom = "density",
                fill = "white",
                xlim = c(0, xmax) ) +
  stat_function(fun  = df,
                args = list(df1 = df1, df2 = df2),
                geom = "density",
                fill = "lightblue",
                xlim = c(qf(alpha, df1, df2, lower.tail = FALSE), xmax) ) +
  labs(y = "F-distribution Density")
```

```
F <- ___
P <- pf(F, df1, df2)
```



ANOVA, example from pre-lecture notes recoded

```
# example in pre-lecture notes
library(dplyr)
dat <- read.csv("Master Materials - Fall 2020/Lectures/quiz video and text data.csv",
header = TRUE)

# F-statistic
k <- nlevels( factor(dat$video) )
N <- nrow( dat )

SS_B <- dat %>%
  mutate(bar_quiz = mean(quiz)) %>%
  group_by(video) %>%
  summarise(sb = n() * (mean(quiz) - first(bar_quiz)) ^ 2) %>%
  ungroup() %>%
  summarise(SS_B = sum(sb)) %>% .$SS_B

SS_W <- dat %>%
  group_by(video) %>%
  mutate(sw = (quiz - mean(quiz)) ^ 2 ) %>%
  ungroup() %>%
  summarise(SS_W = sum(sw) ) %>%
  .$SS_W

Fstat <- ( SS_B / (k - 1) ) / ( SS_W / (N - k) )

# probability of this or greater variation in means from F-distribution
p <- pf(Fstat, df1 = k - 1, df2 = N - k, lower.tail = F)
```

Source of variation	Degrees of freedom	Sum of squares	Mean square	*F* statistic
Between treatment groups	$k-1$	$SS_B = \sum n_i(\bar{y}_i - \bar{y})^2$	$MS_B = \frac{SS_B}{k-1}$	$F = \frac{MS_B}{MS_W}$
Within treatment groups	$N-k$	$SS_W = \sum\sum (y_{ij} - \bar{y}_{i\cdot})^2$	$MS_W = \frac{SS_W}{N-k}$	
Total	$N-1$	$SS_T = \sum\sum (y_{ij} - \bar{y})^2$		

Compare with R function, which relies on a linear regression model:

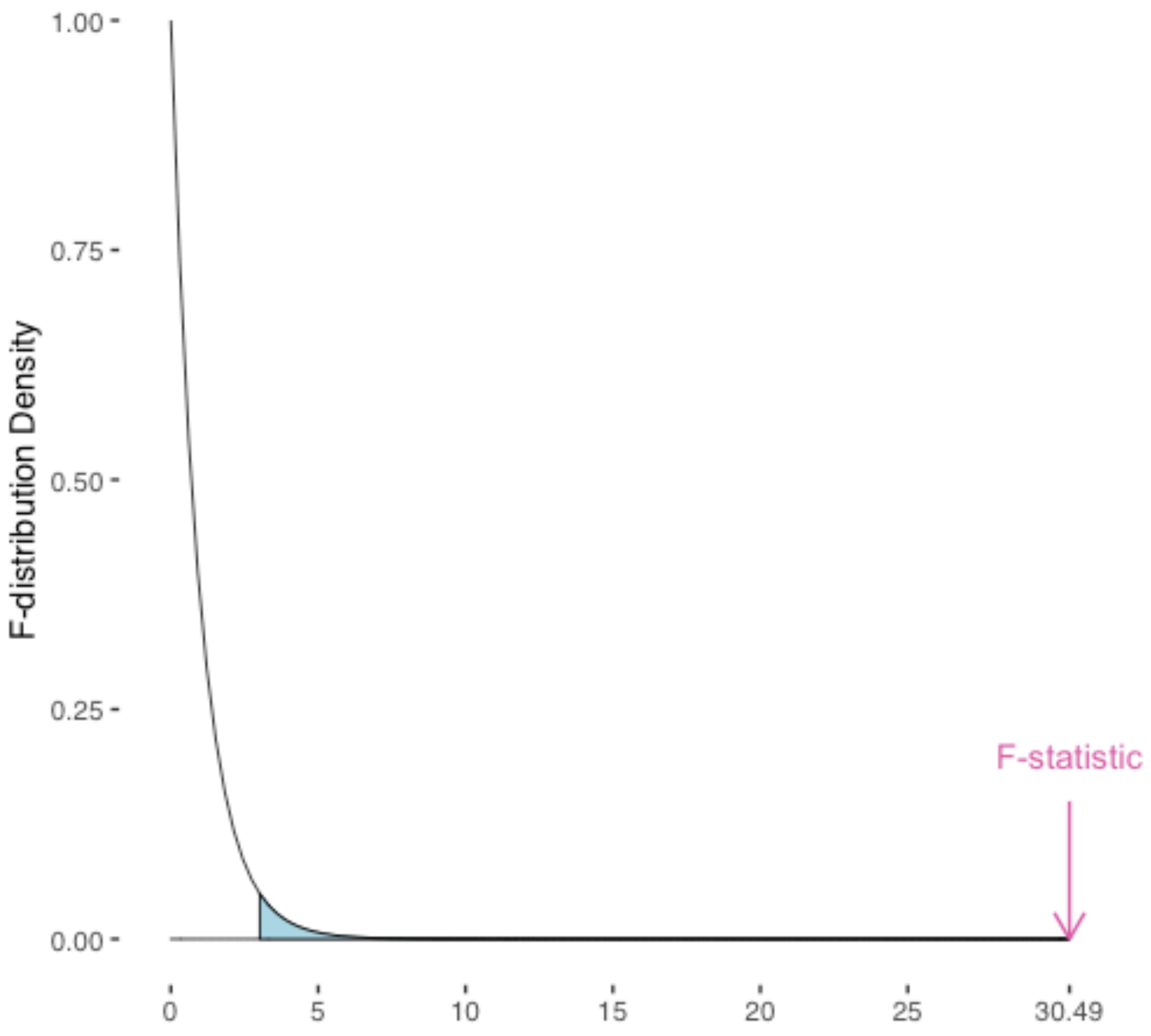
```
summary( aov(quiz ~ video, dat) )
```

ANOVA, example from pre-lecture notes recoded

```
# example in pre-lecture notes
library(dplyr)
dat <- read.csv("Master Materials - Fall 2020/Lectures/quiz video and text data.csv",
header = TRUE)

# F-statistic
k <- nlevels( factor(dat$video) )
N <- nrow( dat )

SS_B <- dat %>%
  mutate(bar_quiz = mean(quiz)) %>%
  group_by(video) %>%
  summarise(sb = n() * (mean(quiz) - first(bar_quiz)) ^ 2) %>%
  ungroup() %>%
  summarise(SS_B = sum(sb)) %>% .$SS_B

SS_W <- dat %>%
  group_by(video) %>%
  mutate(sw = (quiz - mean(quiz)) ^ 2 ) %>%
  ungroup() %>%
  summarise(SS_W = sum(sw) ) %>%
  .$SS_W

Fstat <- ( SS_B / (k - 1) ) / ( SS_W / (N - k) )

# probability of this or greater variation in means from F-distribution
p <- pf(Fstat, df1 = k - 1, df2 = N - k, lower.tail = F)
```

Compare with R function, which relies on a linear regression model:

```
summary( aov(quiz ~ video, dat) )
```



ANOVA, continuing example using two-way ANOVA with interaction

```
summary( aov(quiz ~ video + text + video:text, dat) )
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)    
video         2   2093  1046.5   39.61  < 2e-16 ***
text          1   1530  1529.8   57.90 1.80e-13 ***
video:text    2   1923   961.7   36.40 2.57e-15 ***
Residuals   424  11202    26.4                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

While ANOVA tests the probability of variation across all combinations of means if from a *F*-distribution …

… it provides no information on what is usually more important: effect sizes! Use regression for those.

multiple hypothesis testing

In any test, because we're using a significance level, α, we end up with ***false positives*** about that often.

As the tests multiply, so does the chance of getting false positives — dramatically.

 adjusting for compounded false positives — Bonferroni and other methods

For m number of tests, can adjust α significance level:

$$\alpha_{\text{adj}} = \frac{\alpha}{m}$$

but with many tests, Bonferroni's α_{adj} can also lead to ***inflated false negatives***. Other methods are available, including Tukey's *Honest Significant Difference* test:

```
TukeyHSD( aov(quiz ~ video + text + video:text, data = dat), conf.level = 0.95 )
```

References

Abelson, Robert P. *Statistics as Principled Argument*. Psychology Press, 1995.

Casella, George, and Roger L. Berger. "Analysis of Variance and Regression, Chp. 11." In *Statistical Inference*. 2nd ed. Australia; Pacific Grove, CA: Thomson Learning, 2002.

Cox, D. R., and N. Reid. "Factorial designs: basic ideas, Chp. 5." In *The Theory of the Design of Experiments*. Monographs on Statistics and Applied Probability 86. Boca Raton: Chapman & Hall/CRC, 2000.

Gelman, Andrew, Tue Tjur, Peter McCullagh, Joop Hox, Herbert Hoijtink, and Alan M. Zaslavsky. "*Analysis of Variance? Why It Is More Important than Ever. With Discussion and Rejoinder.*" The Annals of Statistics 33, no. 1 (February 2005): 1–53.

Gelman, Andrew, and Eric Loken. "*The Statistical Crisis in Science*." American Scientist 102 (November 2014): 1–6.