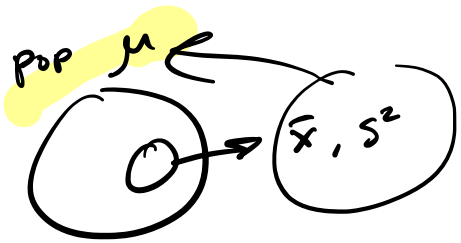
STAT 515 fa 2023 Lec 17 slides

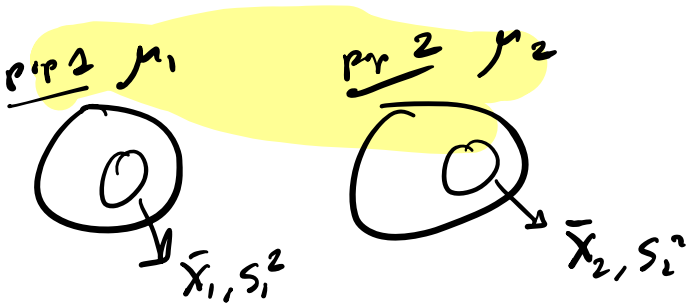
Comparative experiments and analysis of variance



Karl B. Gregory

University of South Carolina

$H_0: \mu_1 = \mu_2 = \cdots = \mu_K$



These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Comparative experiments randomly assign subjects to different treatments.

Observational studies compare subjects existing in different circumstances.

Exercise: Experimental or observational?

1. Randomly assign plant clones to different drought conditions and measure CO_2 uptake.
2. Compare performance in school of children from different backgrounds.
3. Randomly assign tracts of a field to different fertilizers and compare yields.
4. Compare recycling habits of college students in Greenville and Columbia.

Observational studies are beset with the problem of *confounding variables*.

Confounding variable: An unrecorded property/circumstance associated with the outcome of interest as well as with a property/circumstance measured in the study.

Example: Family income and grades in school of children.

Is hours watching TV a confounding variable?

- Is hours watching TV associated with grades in school?
- Is hours watching TV associated with family income?

If yes to both, hours watching TV would be a confounder if ignored in the study.

The random assignment in comparative experiments breaks associations between measured and unmeasured variables, eliminating the problem of confounding variables.

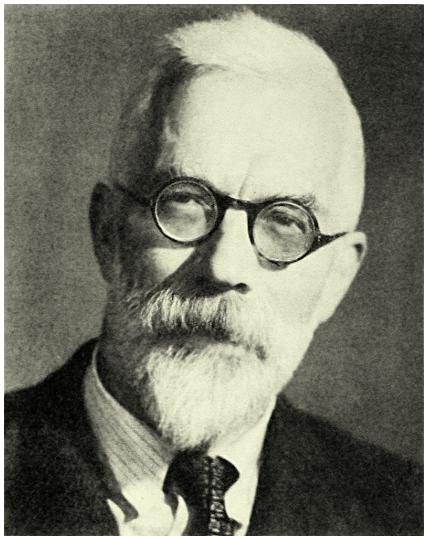
Observational studies cannot establish causation—only association.
Comparative experiments *can* establish causation.

Vocabulary for comparative experiments

- *Treatment*: A condition imposed by the investigator.
- *Experimental unit (EU)*: Each subject in the study—person, animal, etc.
- *Response*: Outcome measured on each EU after treatment applied.

Example: How to package a steak? Twelve steaks assigned to four different packagings (three to each) and bacteria per cm^2 recorded after nine days [1].

Steak	Packaging	log(# bact/cm^2)	Steak	Packaging	log(# bact/cm^2)
1	Commercial	7.66	10	Mixed Gas	7.41
6	Commercial	6.98	9	Mixed Gas	7.33
7	Commercial	7.80	2	Mixed Gas	7.04
12	Vacuum	5.26	8	CO_2	3.51
5	Vacuum	5.44	4	CO_2	2.91
3	Vacuum	5.80	11	CO_2	3.66



Fisher

F-distribution

Y_{ij}

Example (cont): Here are the treatment means. How can we compare them?

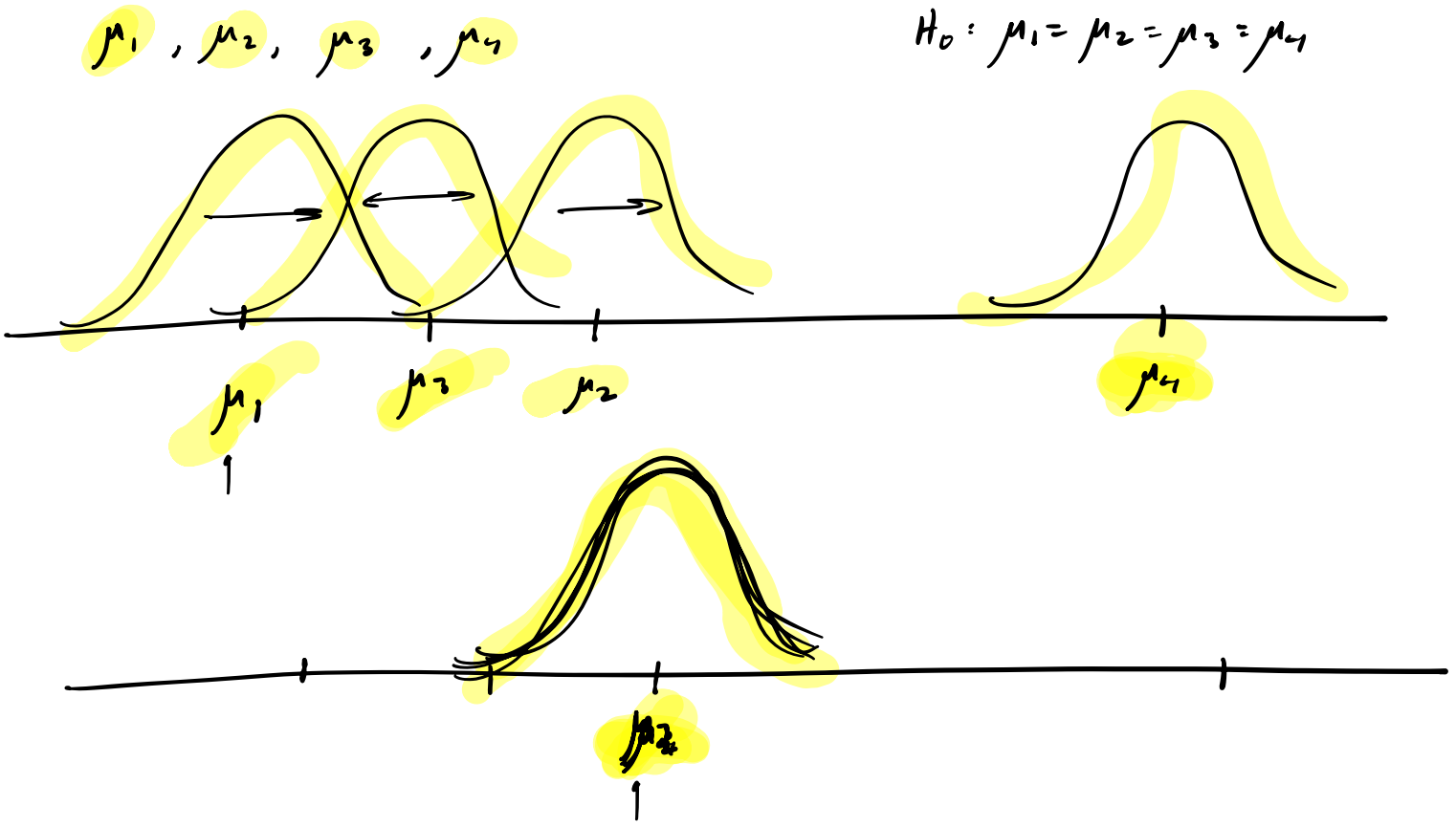
Packaging	mean of log(# bact/cm^2)
Commercial	7.48
Vaccuum	5.50
Mixed Gas	7.26
CO_2	3.36

$\bar{Y}_{1\cdot} = \frac{1}{3}(Y_{11} + Y_{12} + Y_{13})$

$\bar{Y}_{2\cdot}$

$\bar{Y}_{3\cdot}$

$\bar{Y}_{4\cdot}$



$\bar{Y}_{..}$ = Overall mean.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$K = 4$

$N = 12$

$n_1 = 3$

$n_2 = 3$

$n_3 = 3$

$n_4 = 3$

$7.48 = \frac{1}{3}(Y_{11} + Y_{12} + Y_{13}) = \bar{Y}_{1.}$

$5.50 = \frac{1}{3}(Y_{21} + Y_{22} + Y_{23}) = \bar{Y}_{2.}$

$7.26 = \bar{Y}_{3.}$

$3.36 = \bar{Y}_{4.}$

Steak	Packaging	log(# bact/cm^2)	Steak	Packaging	log(# bact/cm^2)
1	Commercial	$Y_{11} = 7.66$	10	Mixed Gas	$Y_{21} = 7.41$
6	Commercial	$Y_{12} = 6.98$ (7.48)	9	Mixed Gas	$Y_{22} = 7.33$
7	Commercial	$Y_{13} = 7.80$	2	Mixed Gas	$Y_{23} = 7.04$
12	Vacuum	$Y_{21} = 5.26$	8	CO_2	$Y_{41} = 3.51$
5	Vacuum	$Y_{22} = 5.44$	4	CO_2	$Y_{42} = 2.91$
3	Vacuum	$Y_{23} = 5.80$ (5.50)	11	CO_2	$Y_{43} = 3.66$

(7.26)

Let

- K be the number of treatments.
- $n_1, \ldots, n_K$ be the numbers of EUs assigned to the treatments.
- $N = n_1 + \ldots + n_K$ be the total number of EUs.
- Y_{ij}, $j = 1, \ldots, n_i$, $i = 1, \ldots, K$ be response for EU j in treatment group i.

Cell-means or one-way ANOVA model

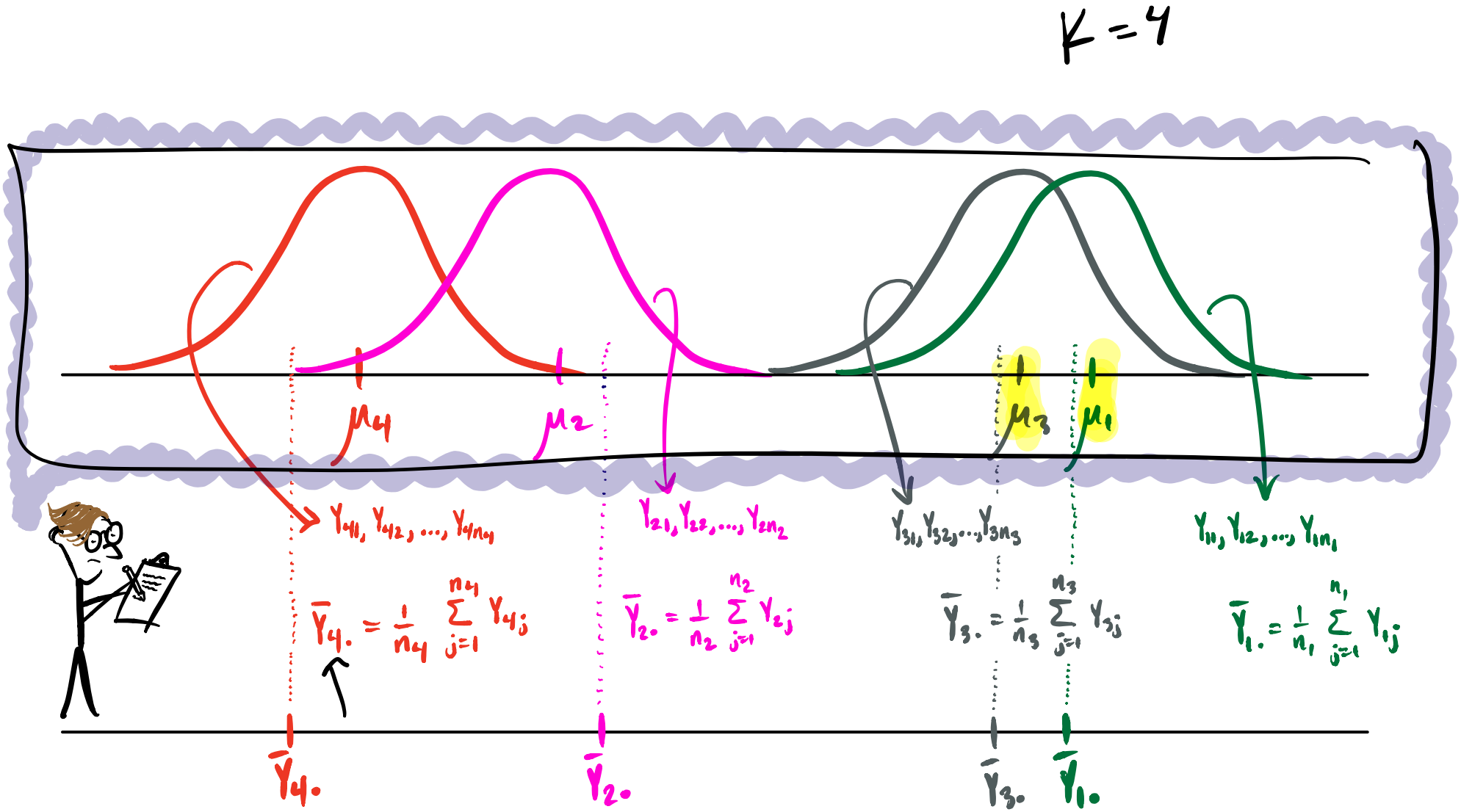
Assume

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad j = 1, \ldots, n_i, \quad i = 1, \ldots, K,$$

where

- $\mu_1, \ldots, \mu_K$ are the population means for treatments $1, \ldots, K$.
- $\{\varepsilon_{ij} : j = 1, \ldots, n_i, i = 1, \ldots, K\} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2_\varepsilon)$.

.



Estimate $\mu_1, \ldots, \mu_K$ with treatment means $\bar{Y}_{i\cdot} = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$, $i = 1, \ldots, K$.

$K = 3$ $\mu_1 = \mu_2 = \mu_3$

$\mu_1 \neq \mu_2 \neq \mu_3$

$\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 2$

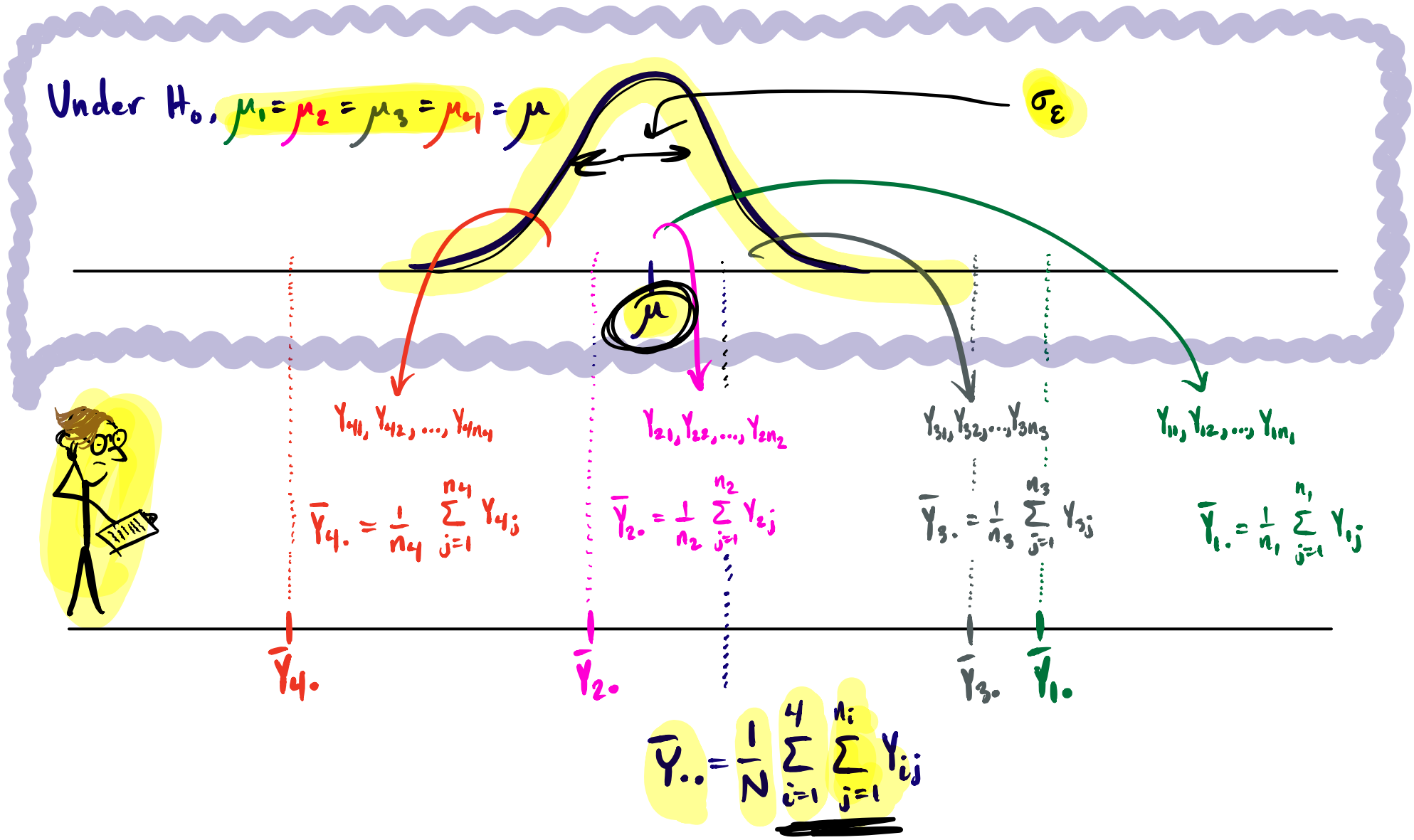
Research question: Do/does any of the treatments affect the response?

Central hypotheses in cell means model

H_0: $\mu_1 = \cdots = \mu_K$ (The treatments have no effect)

H_1: $\mu_i \neq \mu_{i'}$ for some $i \neq i'$, i.e. not all treatment means are equal

To build a test statistic, we look at the spread of $\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{K\cdot}$.



Estimate overall mean with $\bar{Y}_{\cdot\cdot} = N^{-1} \sum_{i=1}^{K} \sum_{j=1}^{n_i} Y_{ij}$.

$\bar{Y}_{1\cdot}$ $\bar{Y}_{2\cdot}$ $\bar{Y}_{3\cdot}$ $\bar{Y}_{4\cdot}$

Under H_0: all data come from 1 single distribution.

Then there is only one mean instead of K different means.

$$\frac{(\bar{Y}_{1\cdot} - \bar{Y}_{\cdot\cdot})^2}{\hat{\sigma}^2/n_1} + \frac{(\bar{Y}_{2\cdot} - \bar{Y}_{\cdot\cdot})^2}{\hat{\sigma}^2/n_2} + \frac{(\bar{Y}_{3\cdot} - \bar{Y}_{\cdot\cdot})^2}{\hat{\sigma}^2/n_3} + \frac{(\bar{Y}_{4\cdot} - \bar{Y}_{\cdot\cdot})^2}{\hat{\sigma}^2/n_4} \sim F \text{ dist}$$

Measures how spread out the treatment means are.

A preliminary test statistic

Under the cell means model, under H_0: $\mu_1 = \cdots = \mu_K$, we have

$$\sum_{i=1}^{K} \left(\frac{\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}}{\sigma_\varepsilon / \sqrt{n_i}} \right)^2 \sim \chi^2_{K-1}.$$

A _____(larger/smaller) value of this casts _____(more/less) doubt on H_0.

Note that σ^2_ε is unknown, so we cannot compute this.

With the *residuals* $\hat{\varepsilon}_{ij} = Y_{ij} - \bar{Y}_{i\cdot}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, K$, use estimator

$$\hat{\sigma}^2_\varepsilon = \frac{1}{N-K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} \hat{\varepsilon}^2_{ij}.$$

Example (cont): This table includes the residuals from the steak experiment.

Steak	Packaging	log(# bact/cm^2)	$\bar{Y}_i$	$\hat{\varepsilon}_{ij}$
1	Commercial	7.66	7.48	0.18
6	Commercial	6.98	7.48	−0.50
7	Commercial	7.80	7.48	0.32
12	Vacuum	5.26	5.50	−0.24
5	Vacuum	5.44	5.50	−0.06
3	Vacuum	5.80	5.50	0.30
10	Mixed Gas	7.41	7.26	0.15
9	Mixed Gas	7.33	7.26	0.07
2	Mixed Gas	7.04	7.26	−0.22
8	CO_2	3.51	3.36	0.15
4	CO_2	2.91	3.36	−0.45
11	CO_2	3.66	3.36	0.30

residuals

12

12 − 4 = 8 df

We get

$$\hat{\sigma}^2_\varepsilon = \frac{1}{12-4}\left[(0.18)^2 + (-0.50)^2 + \cdots + (0.30)^2\right] = 0.11585.$$

A test statistic

Under the cell means model, under H_0: $\mu_1 = \cdots = \mu_K$, we have

$$F_{\text{test}} = \frac{1}{K-1} \sum_{i=1}^{K} \left(\frac{\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}}{\hat{\sigma}_\varepsilon / \sqrt{n_i}} \right)^2 \sim F_{K-1, N-K}.$$

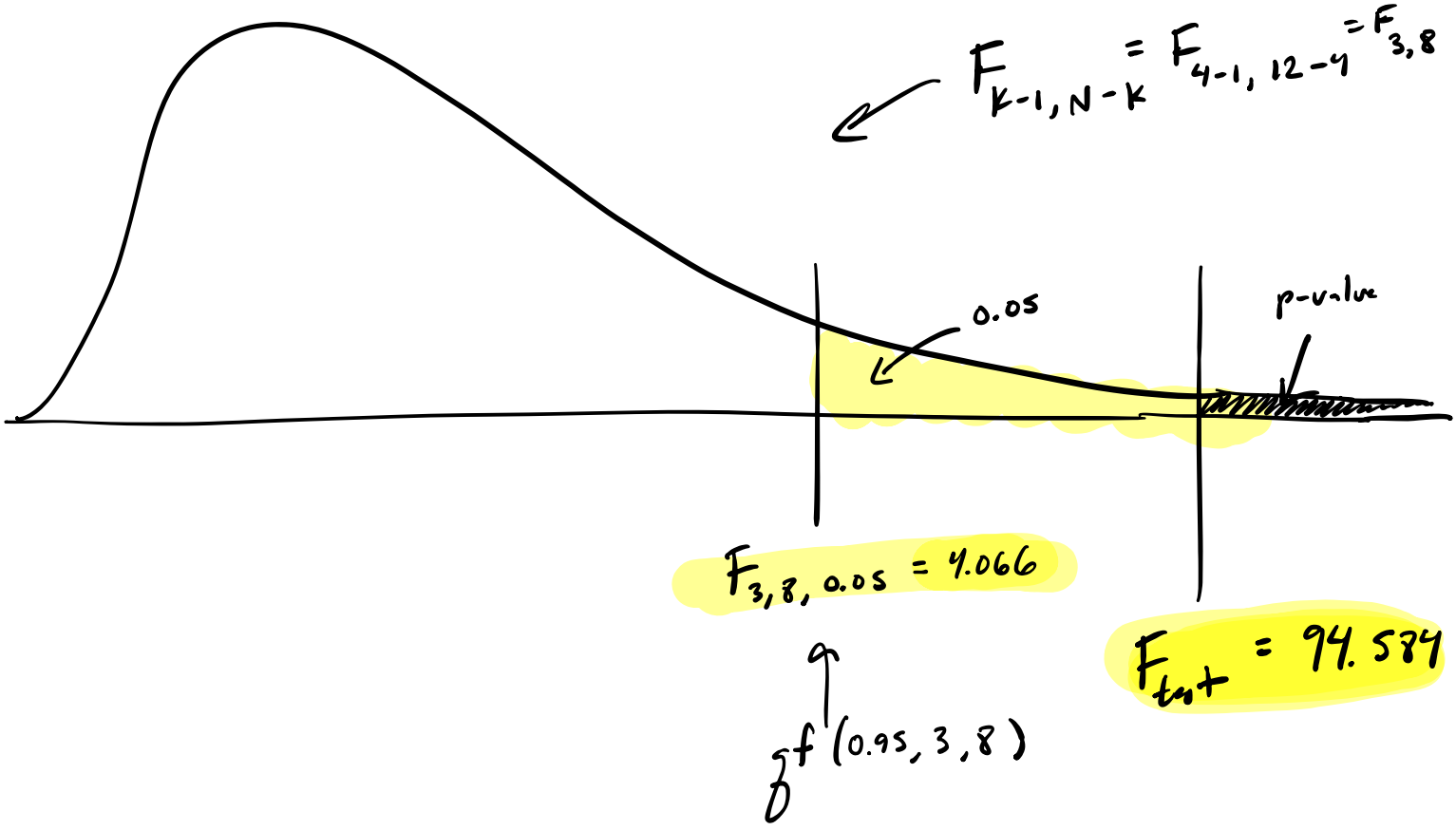
In the above, $F_{K-1,N-K}$ denotes the F-distribution with numerator df $K-1$ and denominator df $N-K$ (next slide).

Test H_0: $\mu_1 = \cdots = \mu_K$ versus alternative at significance level α with criterion

$$\text{Reject } H_0 \text{ if } F_{\text{test}} > F_{K-1,N-K,\alpha}.$$

The p-value is $P(F > F_{\text{test}})$, where $F \sim F_{K-1,N-K}$.

stack



$$p\text{-val} = 1 - pf(94.584, 3, 8) = 1.38 \times 10^{-6}$$

The F-distributions

The F-distribution with num. df $\nu_1 > 0$ and den. df $\nu_2 > 0$ has pdf given by

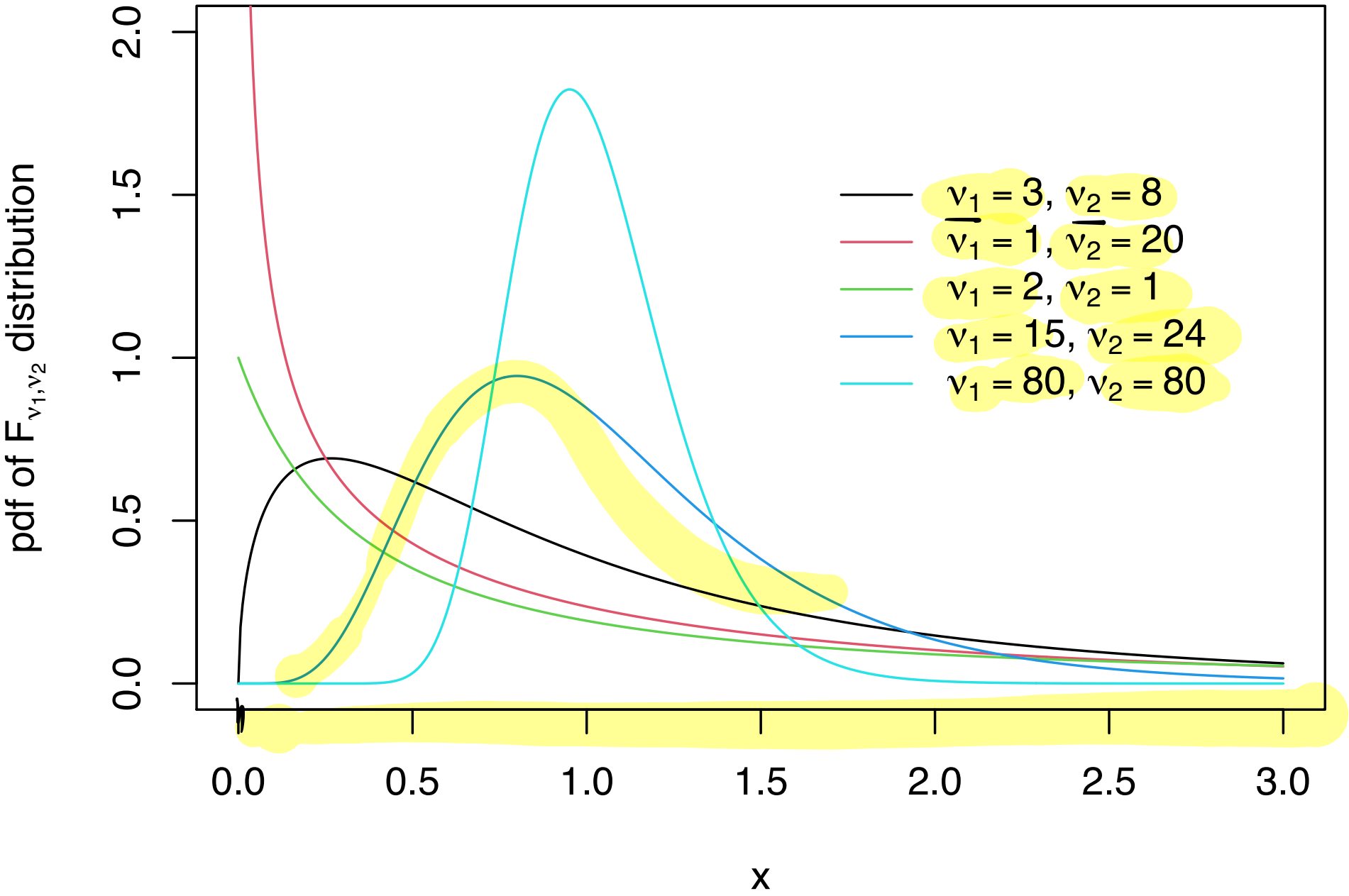
$$f(x) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1} \left(1+\frac{\nu_1}{\nu_2}x\right)^{-\frac{\nu_1+\nu_2}{2}}, \quad x > 0.$$

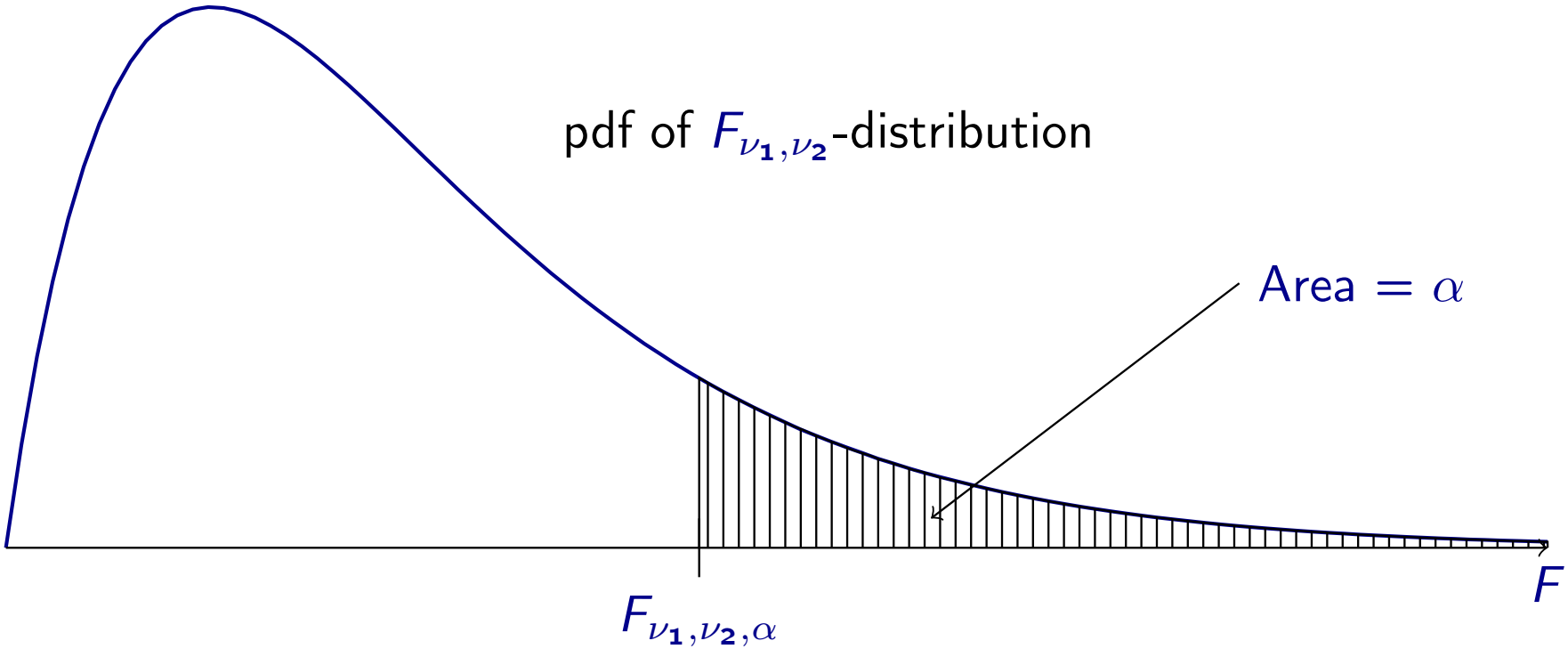
We write $X \sim F_{\nu_1,\nu_2}$.

F-distributed rv as ratio of chi-squared rvs

If $W_1 \sim \chi^2_{\nu_1}$ and $W_2 \sim \chi^2_{\nu_2}$ are independent, then

$$\frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1,\nu_2}.$$


pdf of F_{ν_1,ν_2} distribution
0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x
$\nu_1 = 3, \nu_2 = 8$
$\nu_1 = 1, \nu_2 = 20$
$\nu_1 = 2, \nu_2 = 1$
$\nu_1 = 15, \nu_2 = 24$
$\nu_1 = 80, \nu_2 = 80$



Can use function `qf()` to look up the values, e.g.

$$F_{3,8,0.05} = \texttt{qf(.95,3,8)} = 4.066181$$
$$F_{3,8,0.01} = \texttt{qf(.99,3,8)} = 7.590992$$

Can get area under the curve to the left with the `pf()` function.

Exercise: Compute the test statistic F_{test} for the steak data and consider

$$H_0\text{: } \mu_1 = \mu_2 = \mu_3 = \mu_4 \text{ vs } H_1\text{: Not all means equal.}$$

1. State whether you reject H_0 at the $\alpha = 0.05$ significance level.
2. Compute the p-value.

Analysis of variance (ANOVA): Decomposition of the variability in Y_{ij} into

1. *Between-treatment variation*: Variability due to treatment effects.
2. *Within-treatment variation*: Variability due to differences among EUs.

SS = sum of squares

$$SS_{Total} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2 \qquad \text{(Total variation)}$$

$$SS_{Treatment} = \sum_{i=1}^{K} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 \qquad \text{(Between-treatment)}$$

$$SS_{Error} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\hat{\varepsilon}_{ij}^2 \qquad \text{(Within-treatment)}$$

We have

$$\underbrace{SS_{Total}}_{\text{Total}} = \underbrace{SS_{Treatment}}_{\text{Between}} + \underbrace{SS_{Error}}_{\text{Within}}.$$

Sampling distributions of scaled sums of squares

Under the cell means model under H_0: $\mu_1 = \cdots = \mu_K$, we have

$$SS_{\text{Total}}/\sigma_\varepsilon^2 \sim \chi^2_{N-1}$$
$$SS_{\text{Treatment}}/\sigma_\varepsilon^2 \sim \chi^2_{K-1}$$
$$SS_{\text{Error}}/\sigma_\varepsilon^2 \sim \chi^2_{N-K}.$$

Also define

MS = Mean square

$$MS_{\text{Treatment}} = SS_{\text{Treatment}}/(K-1)$$
$$MS_{\text{Error}} = SS_{\text{Error}}/(N-K).$$

$H_0: \mu_1 = \cdots = \mu_K$

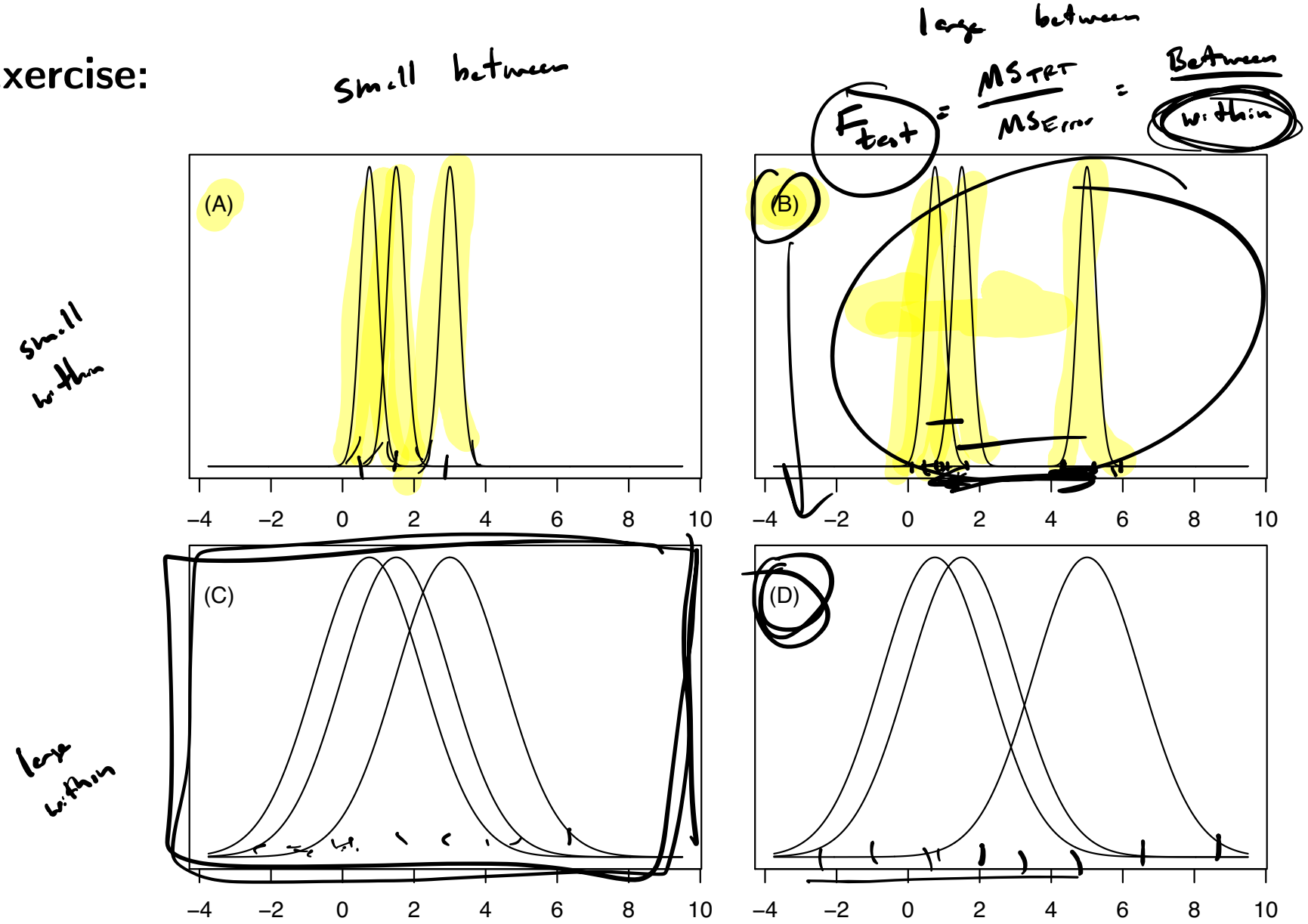
H_1: Not all μs the same

Exercise: Show that

$$F_{\text{test}} = \frac{1}{K-1}\sum_{i=1}^{K}\left(\frac{\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}}{\hat{\sigma}_\varepsilon/\sqrt{n_i}}\right)^2 = \frac{MS_{\text{Treatment}}}{MS_{\text{Error}}}.$$

$= \dfrac{\text{between dot variation}}{\text{within dot variation}}$

Exercise:



i) Largest F_{test}? ii) Smallest? iii) Two with larger $\text{MS}_{\text{Treatment}}$? iv) Larger MS_{Error}?

$H_0: \mu_1 = \cdots = \mu_K$

$H_1:$ Not all equal

The *ANOVA table* is a table presenting all of these values:

Source	Sum of Sq	df	Mean Sq	F	p-value
Treatment	$SS_{\text{Treatment}}$	$K-1$	$MS_{\text{Treatment}}$	F_{test}	$P(F > F_{\text{test}})$
Error	SS_{Error}	$N-K$	MS_{Error}		where $F \sim F_{K-1,N-K}$
Total	SS_{Total}	$N-1$			

R omits these

Exercise: Get the ANOVA table for the steaks data using `lm()` and `anova()`.

```
# read in the data and format it for ANOVA:

bacteria <- c(7.66,6.98,7.80,
              5.26,5.44,5.80,
              7.41,7.33,7.04,
              3.51,2.91,3.66)

packaging <- c(rep("Commercial",3),
               rep("Vacuum",3),
               rep("Mixed Gas",3),
               rep("C02",3))

packaging <- as.factor(packaging)

# estimate model with lm() function and retrieve ANOVA table:

model <- lm(bacteria ~ packaging)
anova(model)
```

Consider the assumptions of the model

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, K,$$

where $\{\varepsilon_{ij} : j = 1, \ldots, n_i, i = 1, \ldots, K\} \overset{\text{ind}}{\sim} \text{Normal}(0, \sigma^2_\varepsilon)$.

(A.1) The responses are Normally distributed around the treatment means.

To check: Look at a QQ plot of the residuals.

(A.2) The responses have the same variance in all treatment groups.

To check: Look at the residuals versus fitted values plot.

(A.3) The responses are independent from each other.

Cannot check: Trust the random assignment of EUs to treatments.

Use `plot()` on the output of `lm()`.

```
plot( lm( y ~ as.factor(x) ) )
```

Exercise: Check the diagnostic plots for the steaks example.


R. O. Kuehl.
Design of Experiments: Statistical Principles of Research Design and Analysis.

Duxbury/Thomson Learning, 2000.
Google-Books-ID: mIV2QgAACAAJ.