Intro to NHANES and sampling

EXPERIMENTAL DESIGN IN R

Kaelen Medeiros
Data Scientist

datacamp

Intro to NHANES dataset

NHANES = National Health and Nutrition Examination Survey

- Conducted by the National Center for Health Statistics (NCHS), a division of the Centers for Disease Control (CDC)
- Data collected a variety of ways, including interviews & a physical exam
- Questions cover medical, dental, socioeconomic, dietary, and general health-related conditions

Intro to sampling

Probability Sampling: probability is used to select the sample (in various ways)

Non-probability Sampling: probability is not used to select the sample

- Voluntary response: whoever agrees to respond is the sample
- Convenience sampling: subjects convenient to the researcher are chosen.

Sampling

Simple Random Sampling (SRS)

```
sample()
```

Stratified Sampling

```
dataset %>%
   group_by(strata_variable) %>%
   sample_n()
```

Cluster Sampling

```
cluster(dataset,
        cluster_var_name,
        number_to_select,
        method = "option")
```

Systematic Sampling

Multi-stage Sampling

Let's practice!

EXPERIMENTAL DESIGN IN R

Randomized Complete Block Designs (RCBD)

EXPERIMENTAL DESIGN IN R

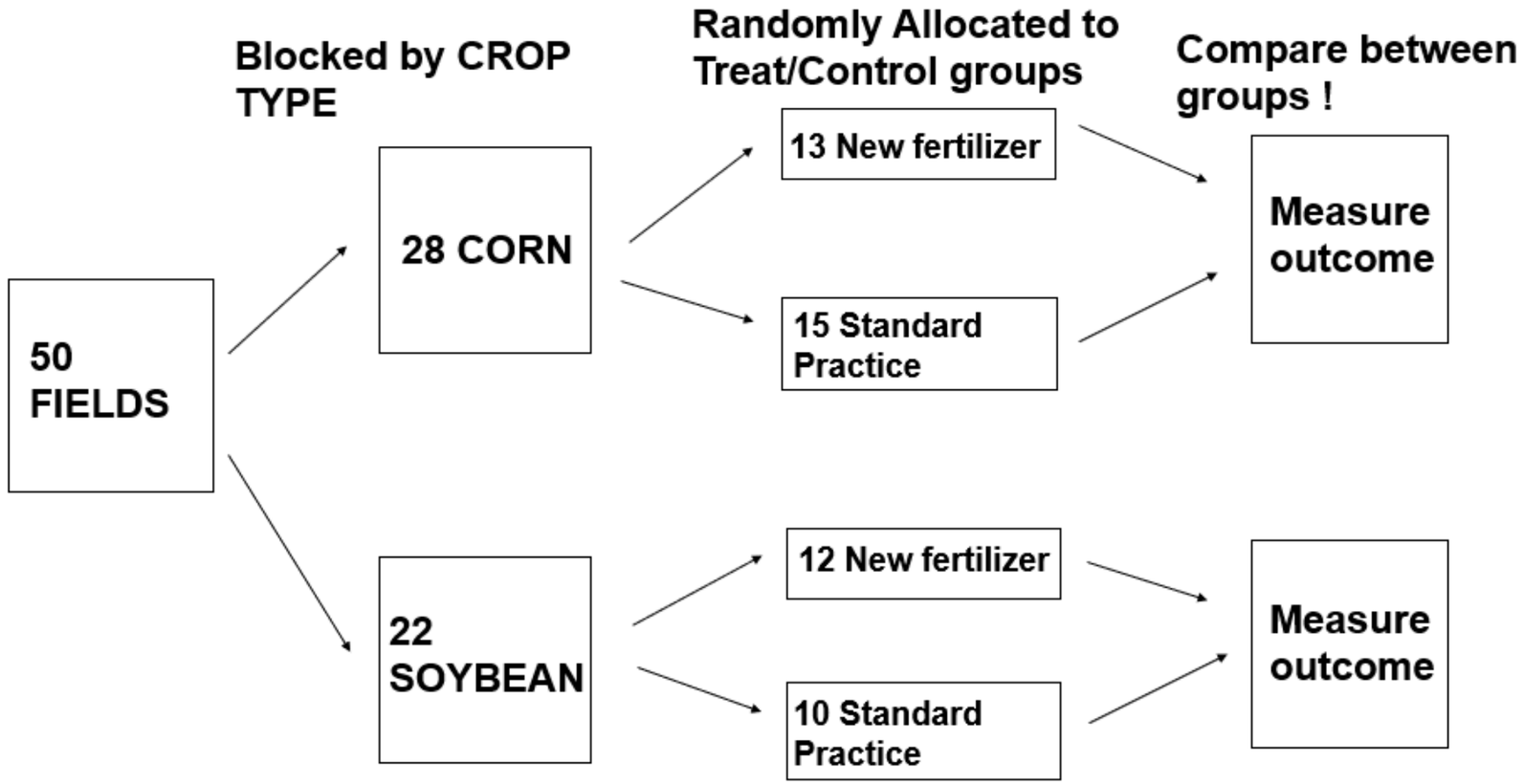
Kaelen Medeiros
Data Scientist

datacamp

RCBDs

- **Randomized** - the treatment is assigned randomly inside each block
- **Complete** - each treatment is used the same number of times in every block
- **Block** - experimental groups are blocked to be similar (e.g. by sex)
- **Design** - this is your experiment!

RCBD workflow



agricolae

```
library(agricolae)
trt <- letters[1:4]
rep <- 4
design.rcbd <- design.rcbd(trt,
                           r = rep,
                           seed = 42,
                           serie = 0)
design.rcbd$sketch
```

```
     [,1] [,2] [,3] [,4]
[1,] "d"  "c"  "a"  "b"
[2,] "b"  "a"  "d"  "c"
[3,] "d"  "a"  "b"  "c"
[4,] "a"  "b"  "d"  "c"
```

Let's practice!

EXPERIMENTAL DESIGN IN R

Balanced Incomplete Block Designs (BIBD)

EXPERIMENTAL DESIGN IN R

Kaelen Medeiros
Data Scientist

datacamp

Balanced Incomplete Block Designs

- **Balanced** - each pair of treatments occur together in a block an equal number of times
- **Incomplete** - not every treatment will appear in every block
- **Block** - experimental groups are blocked to be similar (e.g. by sex)
- **Design** - this is your experiment!

Is there a BIBD?

- Let:
  - **t** = # of treatments
  - **k** = # of treatments per block
  - **r** = # replications $\lambda = r \times \frac{(k-1)}{t-1}$

BIBD examples (1)

$$\lambda = r \times \frac{(k-1)}{t-1}$$

```
lambda = 2 * (2-1) / 4-1
lambda = 2/3 # = 0.667
```

Invalid BIBD

A	B	C	D
A	C	D	A
A	B	D	A
B	C	D	A

BIBD examples (2)

$$\lambda = r \times \frac{(k-1)}{t-1}$$

```
lambda = 3 * (3-1) / 4-1


lambda = 6/3 # = 2
```

Valid BIBD

B	C	A	D
D	A	C	B
B	C	A	D
A	B	D	C

Let's practice!

EXPERIMENTAL DESIGN IN R