Intro to experimental design

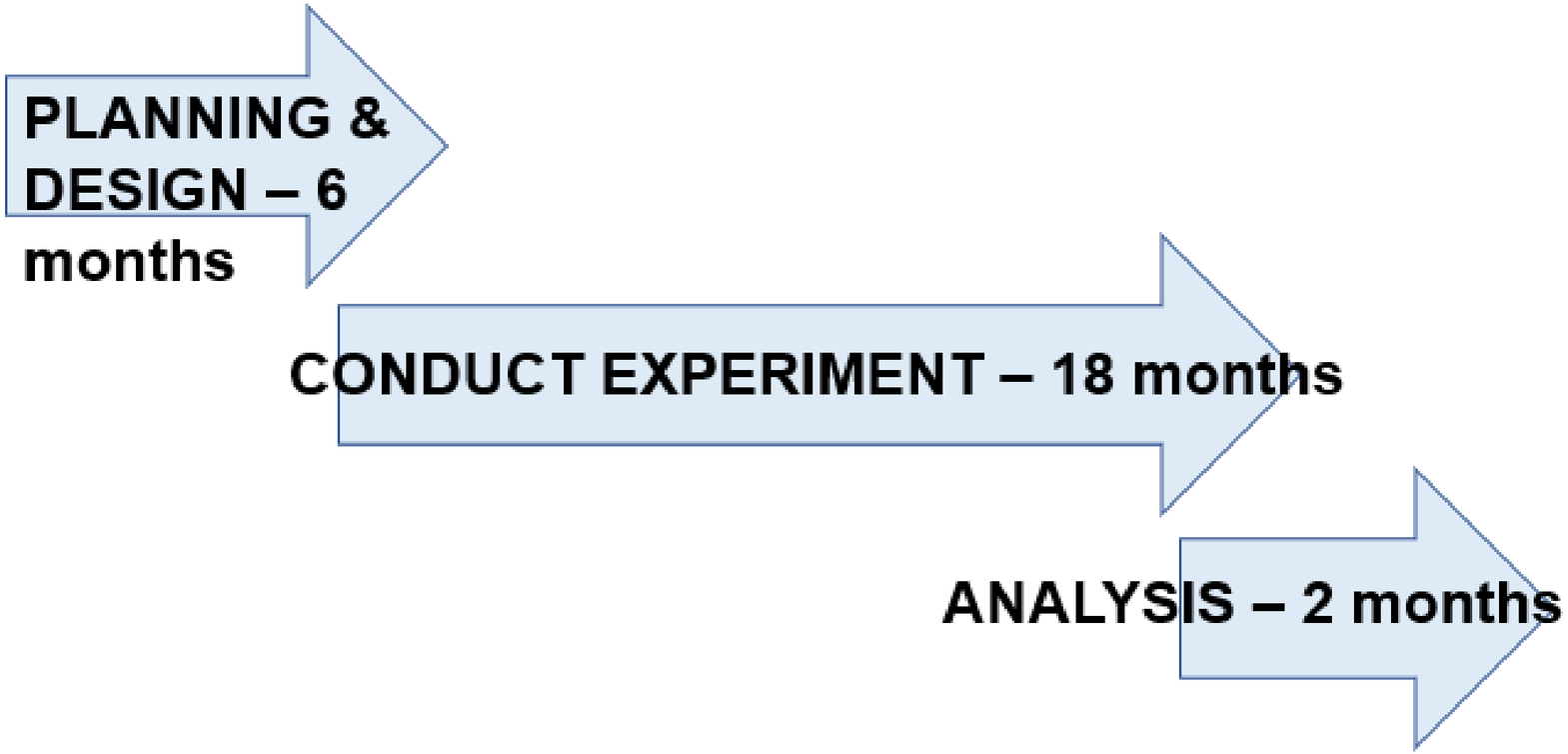
EXPERIMENTAL DESIGN IN R

Kaelen Medeiros
Data Scientist

datacamp

Steps of an experiment

- Planning
  - dependent variable = outcome
  - independent variable(s) = explanatory variables
- Design
- Analysis


PLANNING & DESIGN – 6 months
CONDUCT EXPERIMENT – 18 months
ANALYSIS – 2 months

Key components of an experiment

- Randomization
- Replication
- Blocking

Randomization

- Evenly distributes any variability in outcome due to outside factors across treatment groups
- Example:
  - double-blind medical trials
  - neither patient nor doctor knows which group has been assigned
  - group assignment is made randomly by 3rd party

Replication

- Must repeat an experiment to fully assess variability
- If we only conduct a drug efficacy experiment on one person, how can we properly generalize those results? (We can't!)

Blocking

- Helps control variability by making treatment groups more alike
- Inside of groups, differences will be minimal. Across groups, differences will be larger.
- One example is blocking treatment groups by sex.

Let's practice!

EXPERIMENTAL DESIGN IN R

Hypothesis testing

EXPERIMENTAL DESIGN IN R

Kaelen Medeiros
Data Scientist

Breaking down hypothesis testing:

- **Null hypothesis**:
  - there is no change
  - no difference between groups
  - the mean, median, or observation = a number
- **Alternative hypothesis**:
  - there is a change
  - difference between groups
  - mean, median, or observation is >, <, or != to a number

Power and sample size

- **Power**: probability that the test correctly rejects the null hypothesis when the alternative hypothesis is true.
- **Effect size**: standardized measure of the difference you're trying to detect.
- **Sample size**: How many experimental units you need to survey to detect the desired difference at the desired power.

Power and sample size calculations

```
library(pwr)
pwr.anova.test(k = 3,
               n = 20,
               f = 0.2,
               sig.level = 0.05,
               power = NULL)
```

```
 Balanced one-way analysis of variance power calculation
              k = 3
              n = 20
              f = 0.2
      sig.level = 0.05
          power = 0.2521043
 NOTE: n is number in each group
```

Let's practice!

EXPERIMENTAL DESIGN IN R