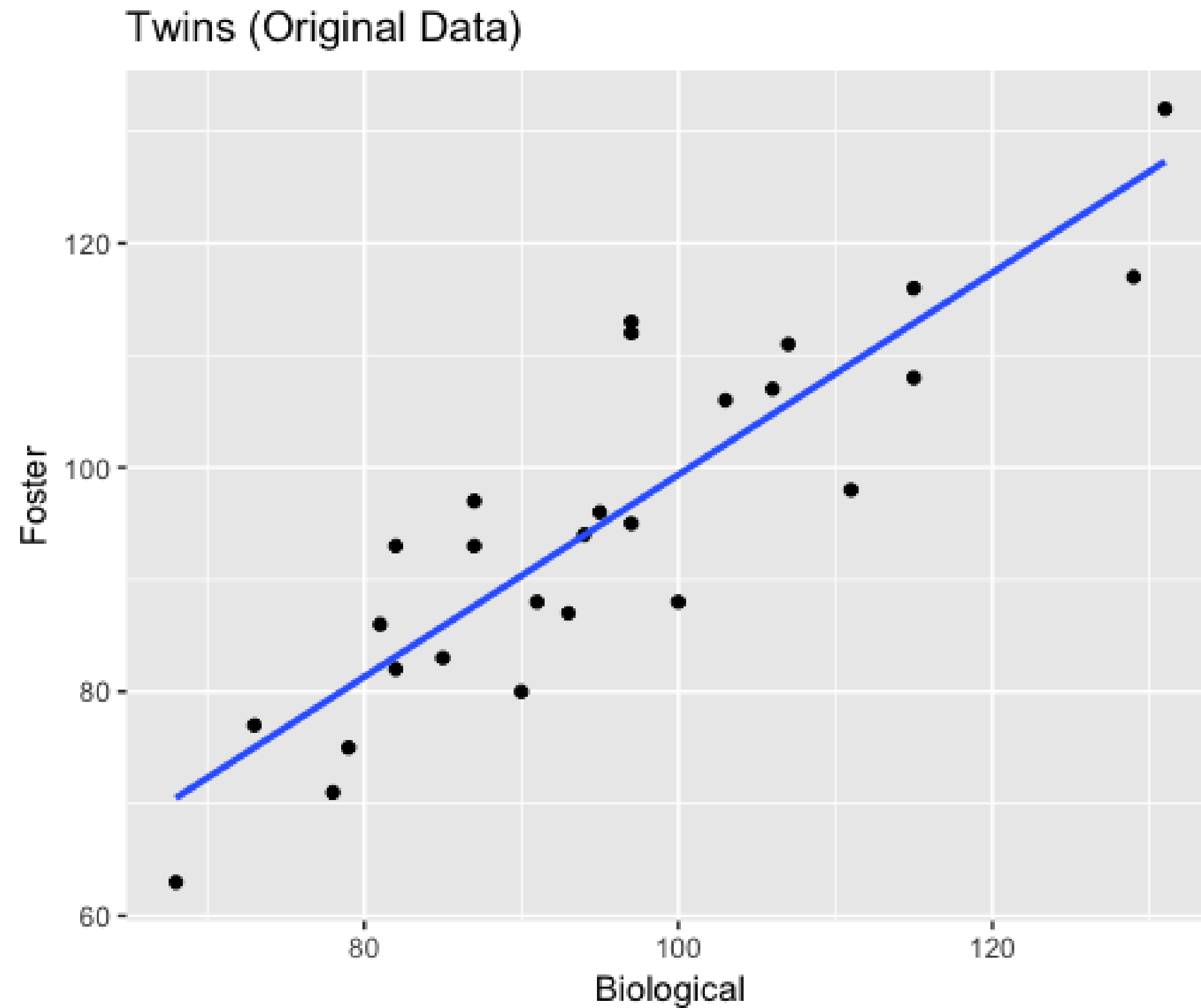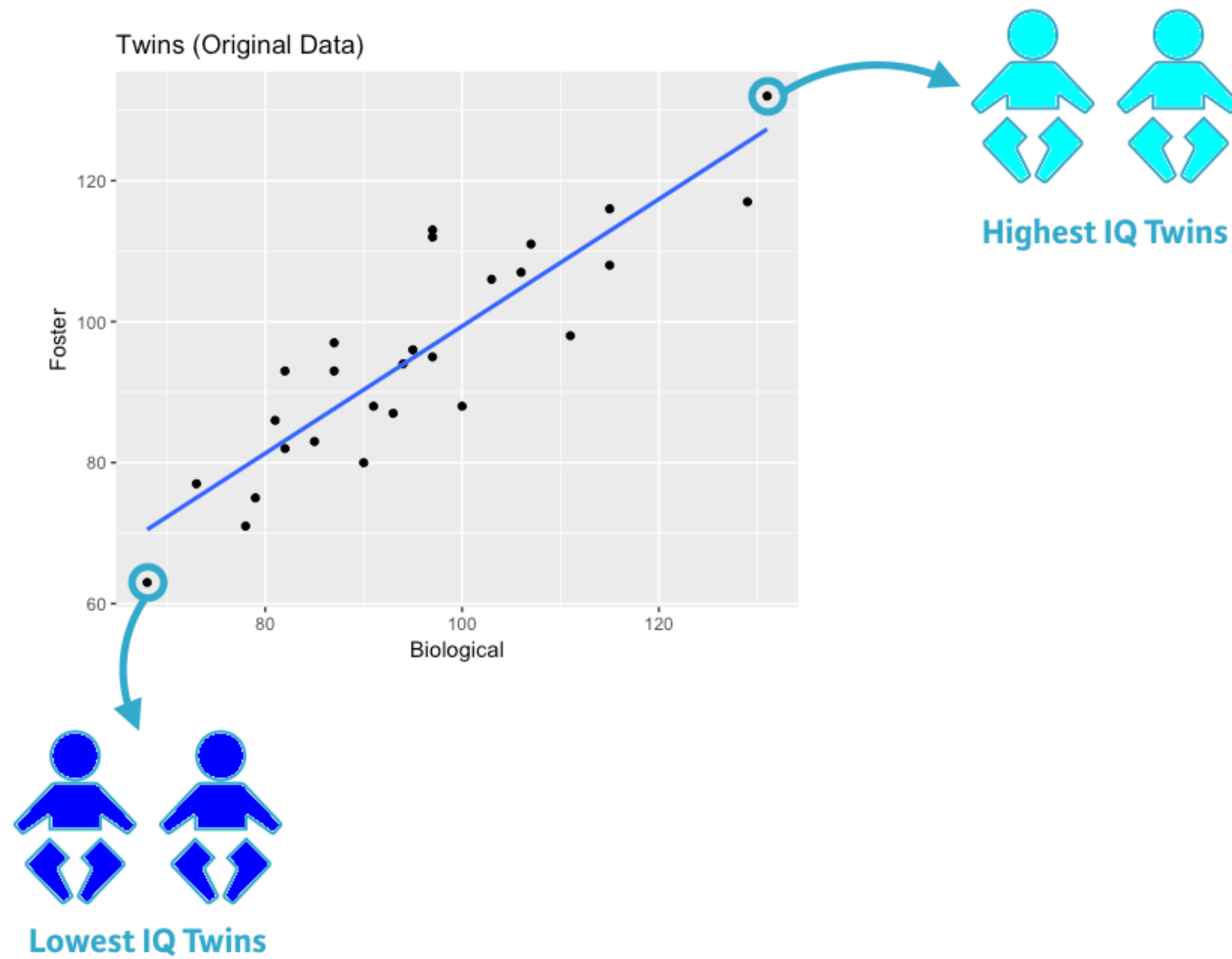INFERENCE FOR LINEAR REGRESSION

Simulation-based Inference

Jo Hardin

Professor, Pomona College

Twins (Original Data)

Twins (Original Data)

Highest IQ Twins

Lowest IQ Twins

Twin data

Foster		Biological	
	80		90
	108		115
	116		115
	93		83

Permuted twin data

Foster	Biological
108	90
93	115
116	115
80	83

Permuted data (1) plotted

Original data
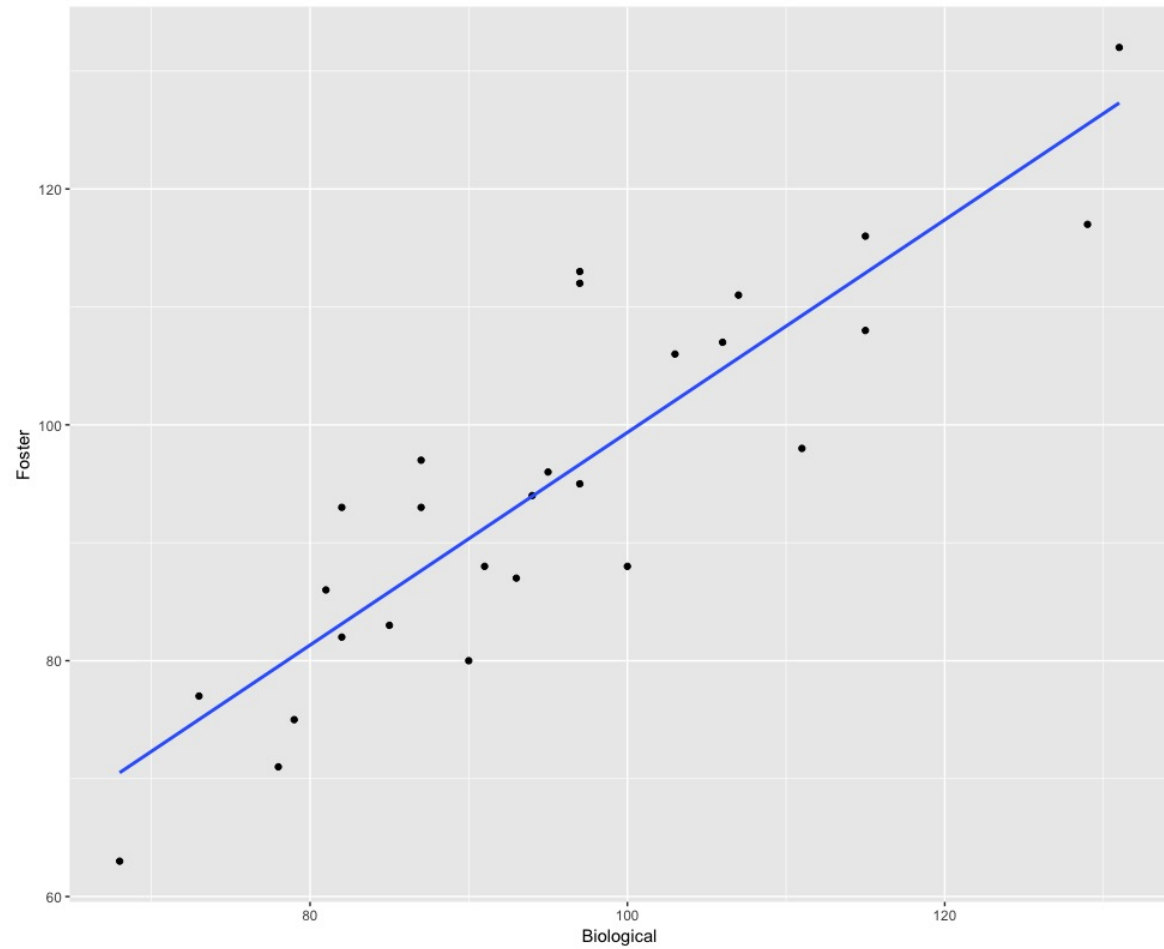
Permuted data (1)

Permuted data (2) plotted

Original data

Permuted data (2)

Permuted data (1) and (2)

Permuted data (1)

Permuted data (2)
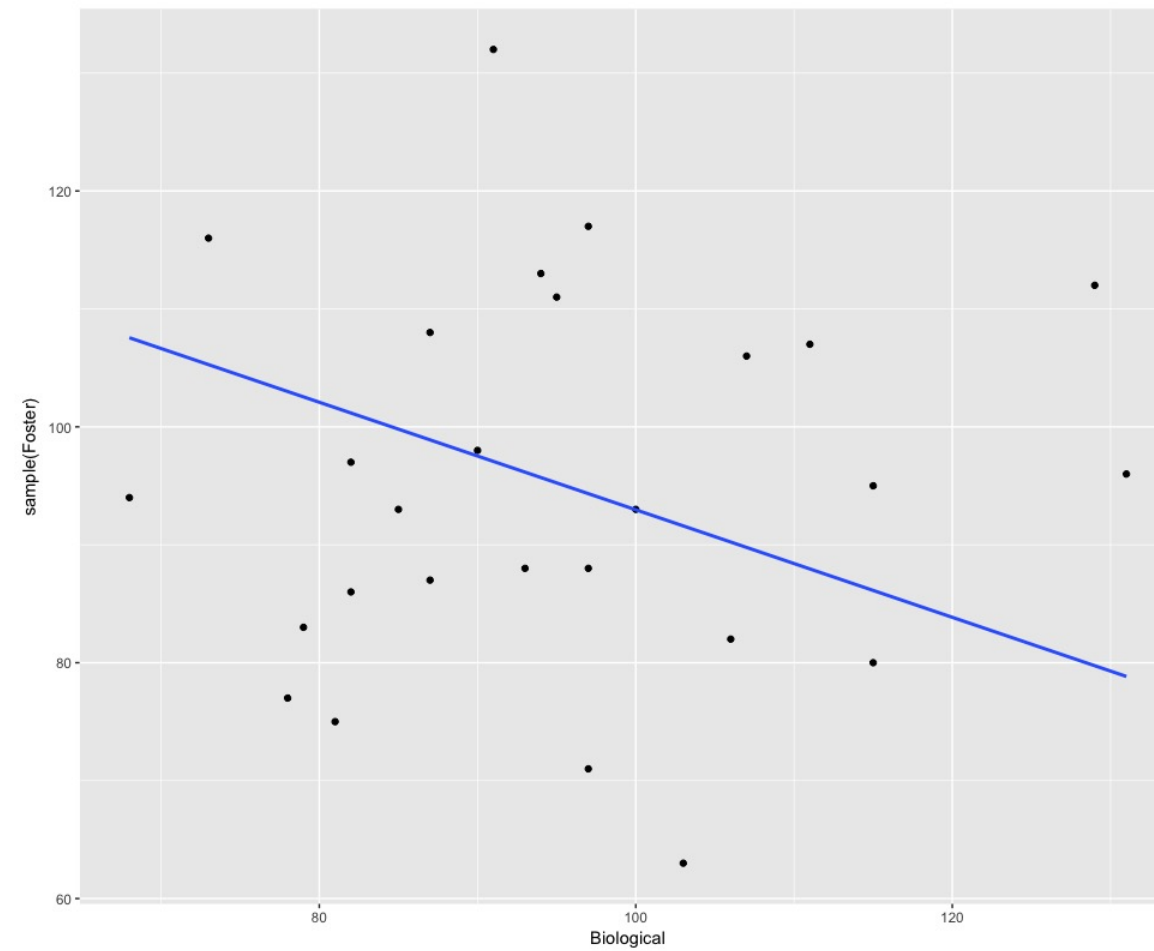
Linear model on permuted data

```
twins %>%
    specify(Foster ~ Biological) %>%
    hypothesize(null = "independence") %>%
    generate(reps = 10, type = "permute") %>%
    calculate(stat = "slope")
# A tibble: 10 x 2
#    replicate            stat
#         <int>          <dbl>
#  1          1   0.0007709302
#  2          2  -0.0353592305
#  3          3  -0.0278627974
#  4          4  -0.0072547982
#  5          5  -0.1252761541
#  6          6  -0.1669869287
#  7          7  -0.2610519170
#  8          8  -0.0157695494
#  9          9   0.0581361900
# 10         10   0.1598471947
```
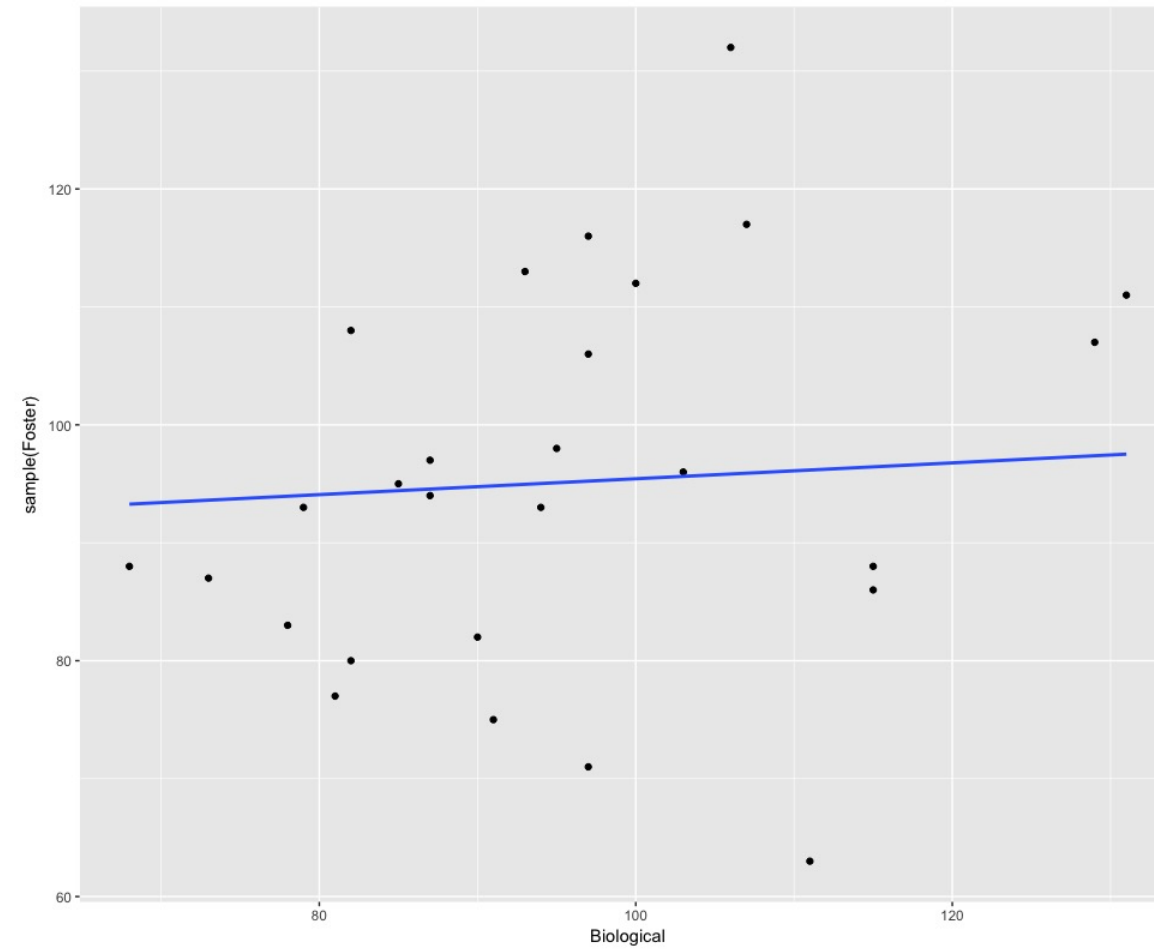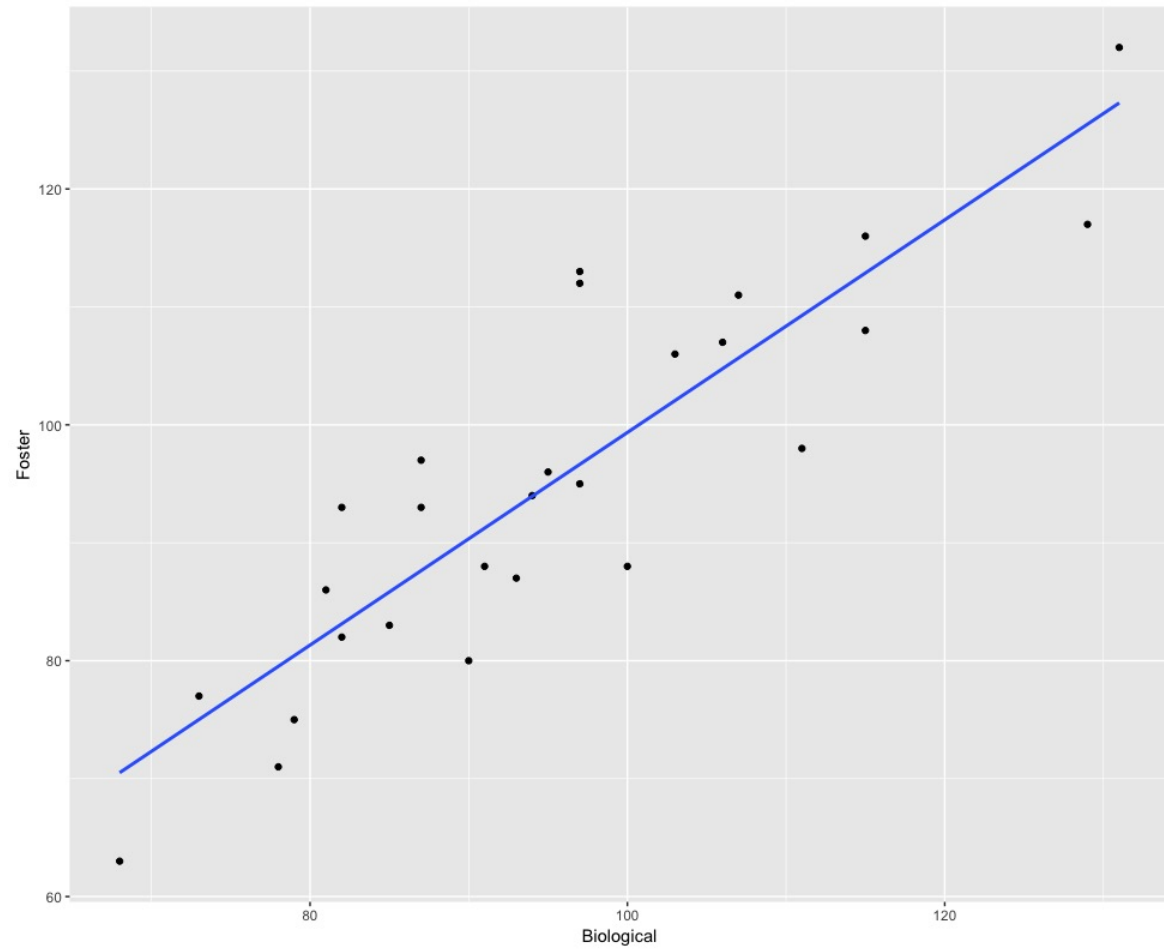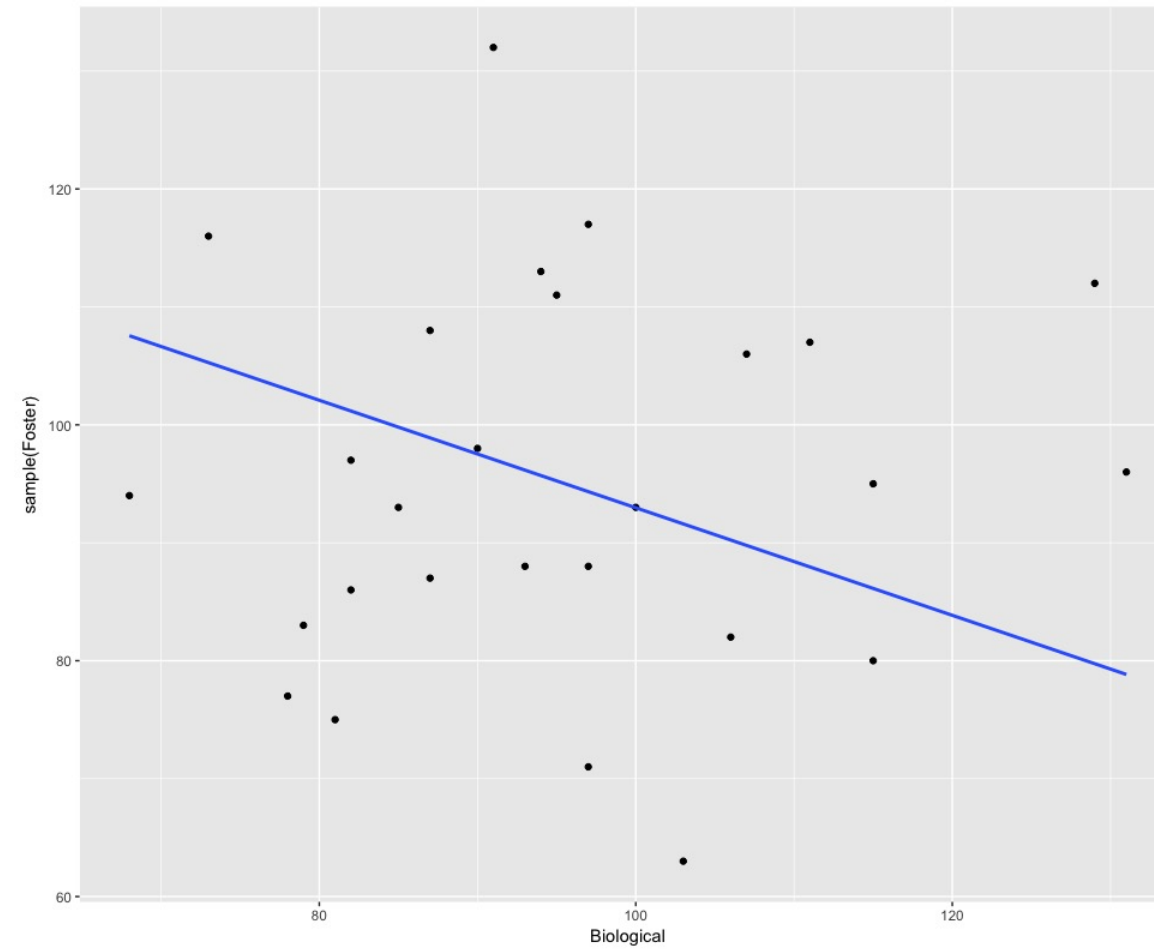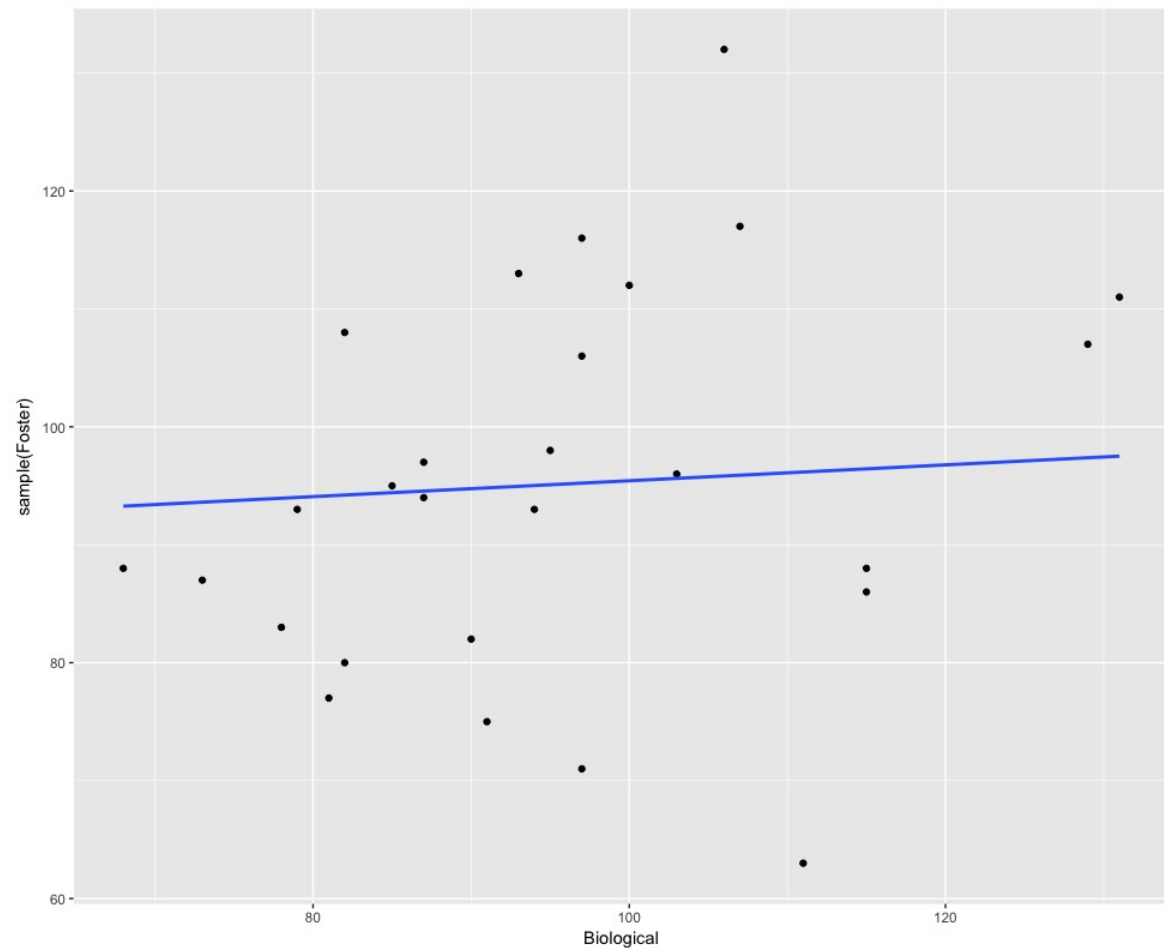
Many permuted slopes

```r
perm_slope <- twins %>%
    specify(Foster ~ Biological) %>%
    hypothesize(
        null = "independence"
        ) %>%
    generate(reps = 1000,
            type = "permute") %>%
    calculate(stat = "slope")

ggplot(data=perm_slope,
        aes(x=stat)) +
    geom_histogram() +
    xlim(-1,1)
```
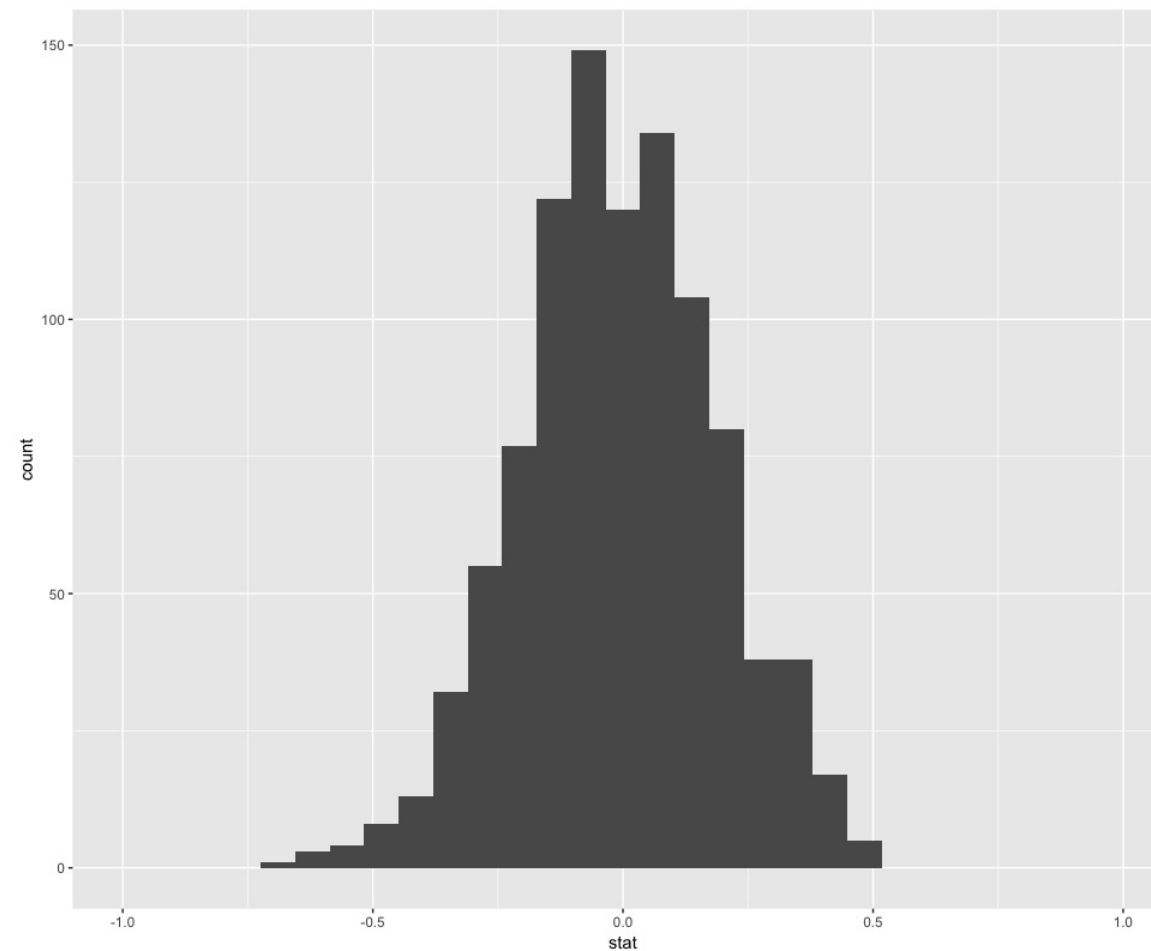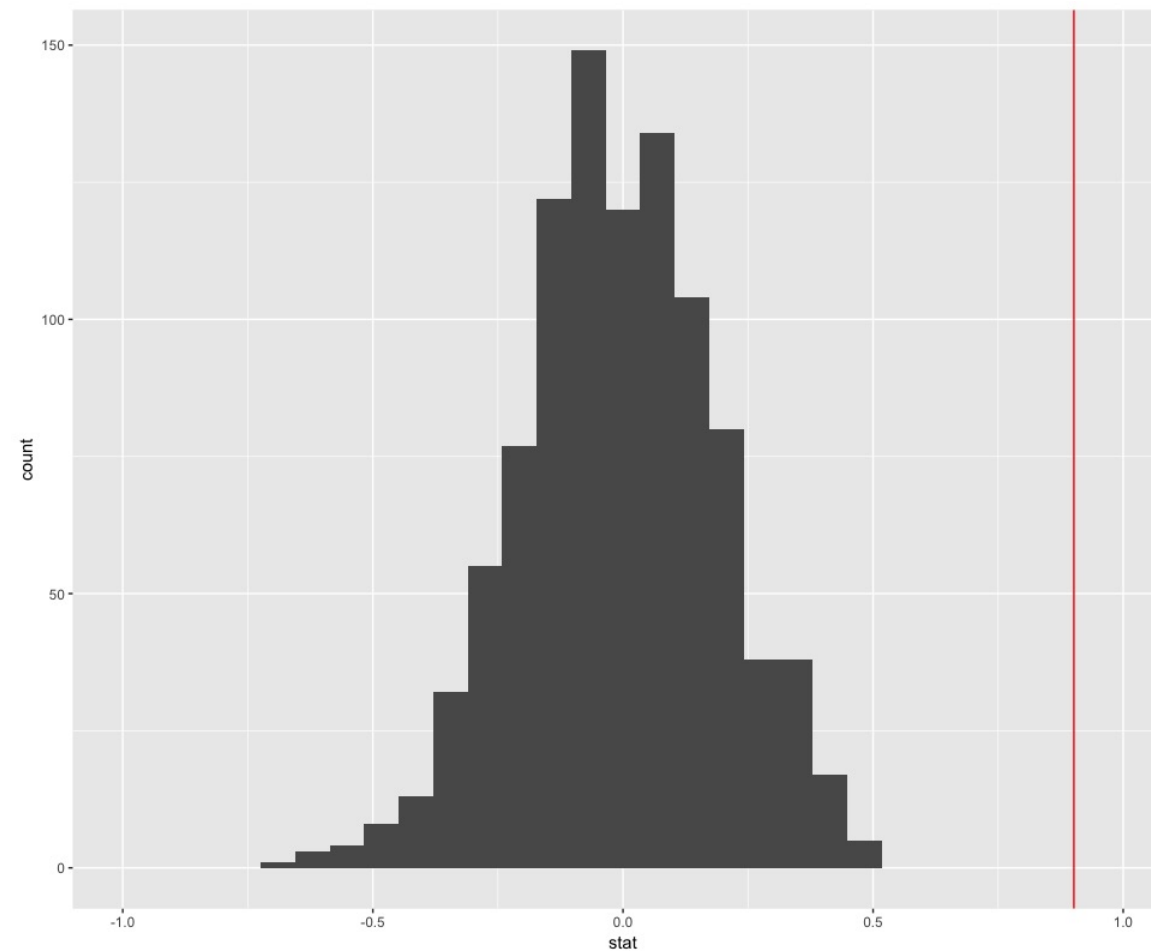
Permuted slopes with observed slope in red

```
obs_slope <- lm(Foster ~ Biological,
                 data = twins) %>%
    tidy() %>%
    filter(term == "Biological") %>%
    select(estimate) %>%
    pull()

obs_slope
# [1] 0.901436

ggplot(data = perm_slope,
       aes(x = stat)) +
    geom_histogram() +
    geom_vline(xintercept = obs_slope,
               color = "red")
    + xlim(-1,1)
```

INFERENCE FOR LINEAR REGRESSION

Let's practice!

INFERENCE FOR LINEAR REGRESSION

Simulation-based CI for slope

Jo Hardin

Professor, Pomona College

Original Sample

Foster	Biological
80	90
108	115
116	115
93	83

Bootstrapped Sample

Foster	Biological
93	83
108	115
108	115
93	83

Permutation vs. bootstrap variability

Slopes from permuted data
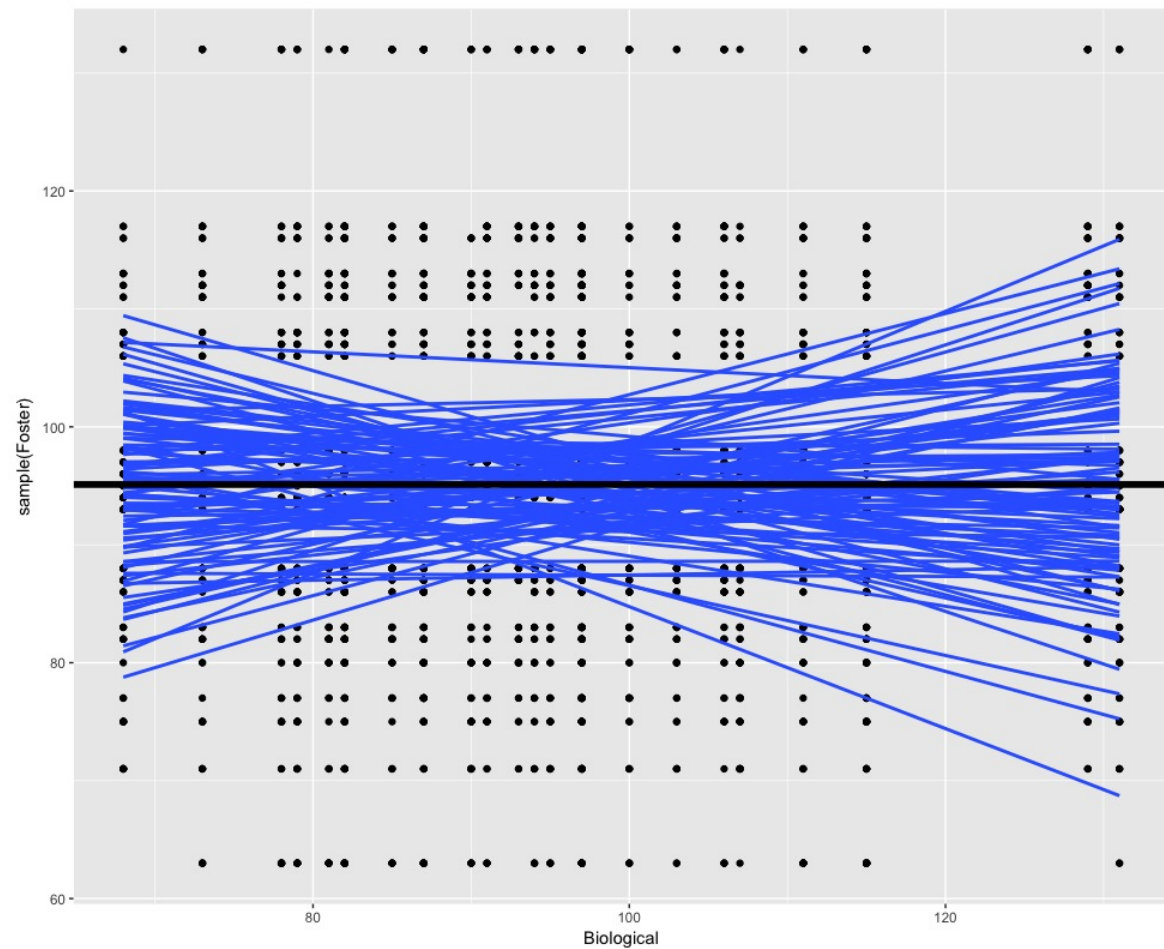
Slopes from bootstrapped data

Permutation vs. boostrap code

Permutation:
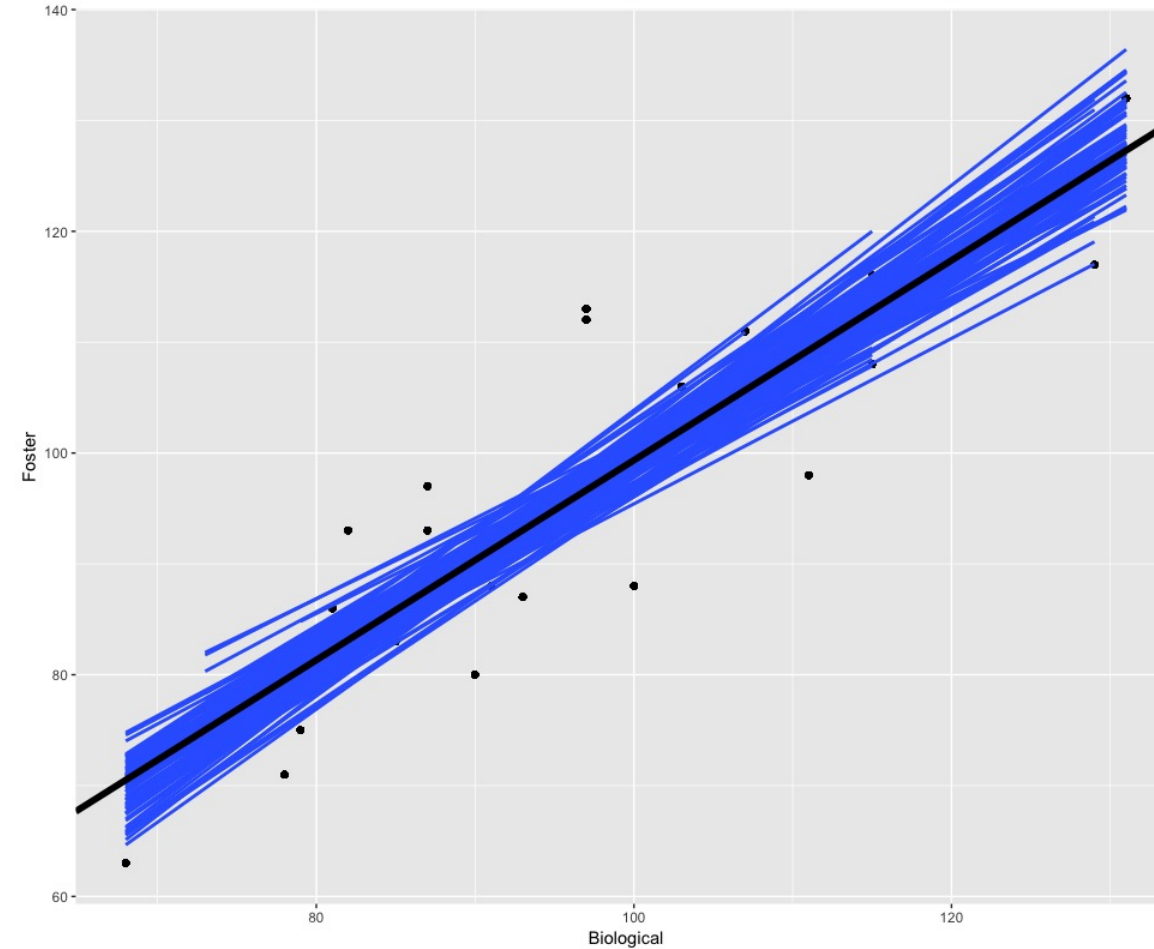
```
twins %>%
  specify(Foster ~ Biological) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 100, type = "permute") %>%
  calculate(stat = "slope")
```
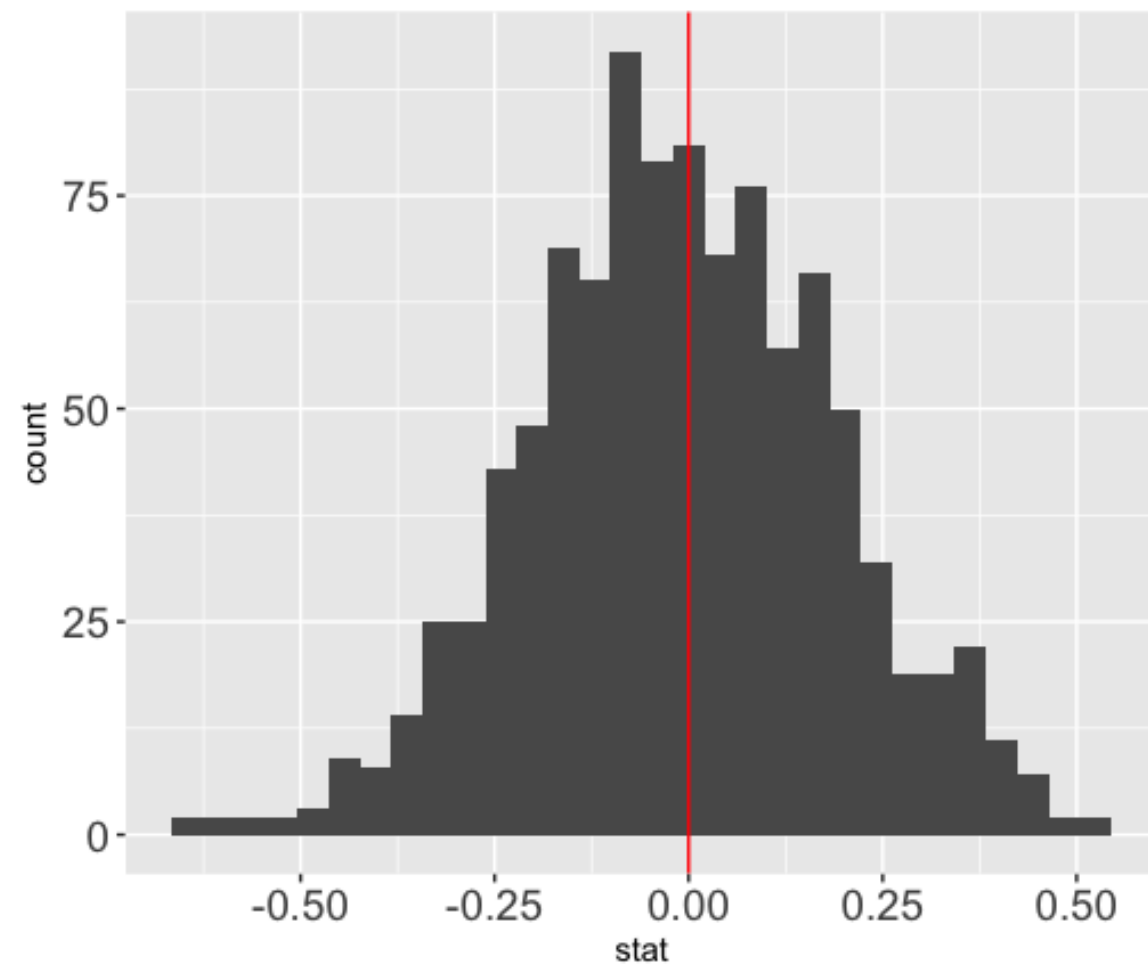
Bootstrap:

```
twins %>%
  specify(Foster ~ Biological) %>%
  generate(reps = 100, type = "bootstrap") %>%
  calculate(stat = "slope")
```
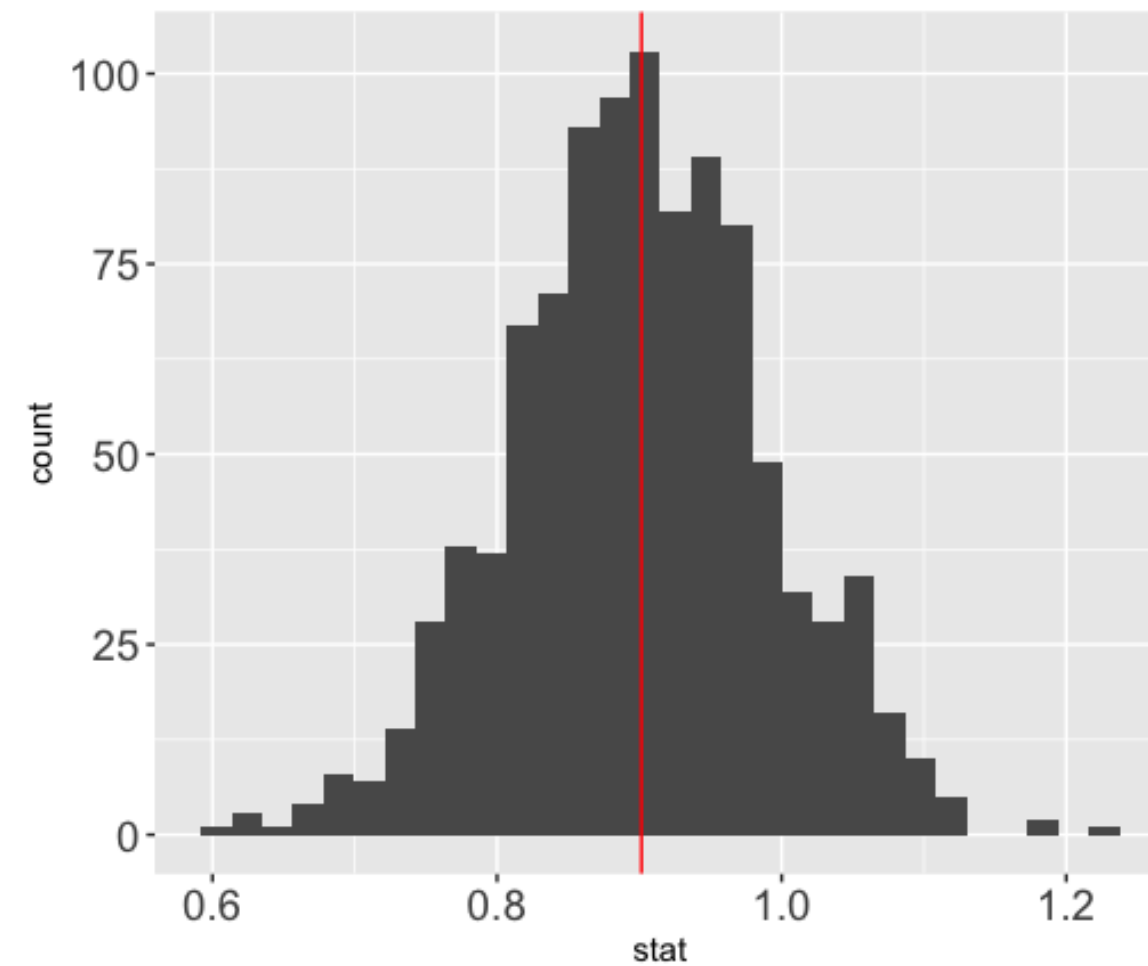
Sampling distribution: randomization vs. bootstrap

Slopes from permuted data

Slopes from bootstrapped data

INFERENCE FOR LINEAR REGRESSION

Let's practice!