

Activity: Translating Research Questions into Hypotheses

Suppose that the process used for conducting the 1969 Vietnam War draft was truly a fair and random process.

1. Under that assumption, what would the pattern of draft numbers versus day rank values look like in a scatter plot? Sketch an example.
2. Under that assumption, what would the simple linear regression line summarizing the relationship between draft number and day rank look like? Add it to your scatter plot above.

Activity: Conducting a randomization test for the slope

Step 1: Formulate hypotheses

1. Write down the null hypothesis for the slope coefficient in this regression model.
2. What is the name of the response variable in the `draft_data` data frame?

```
# A tibble: 6 x 6
  month   day min_year max_year day_rank draft_number
  <dbl> <dbl>   <dbl>   <dbl>   <dbl>       <dbl>
1     1     1     1944    1950     1         305
2     1     2     1944    1950     2         159
3     1     3     1944    1950     3         251
4     1     4     1944    1950     4         215
5     1     5     1944    1950     5         101
6     1     6     1944    1950     6         224
```

1. What is the name of the explanatory variable?
2. Use your answers to the previous questions to complete the R code below.

```
vietnam_setup <- draft_data |>
  specify(____<response_variable>_____ ~ ____<explanatory_variable>_____)
```

Step 2: Calculate the test statistic

1. What numeric component of the regression model are we using as the test statistic? [Note: The *value* of the test statistic is shown in the output.]

```
observed_slope <- vietnam_setup |>
  calculate(stat = "____<type of test statistic>____")
```

Response: draft_number (numeric)

Explanatory: day_rank (numeric)

```
# A tibble: 1 x 1
  stat
  <dbl>
```

1 -0.226

Step 3: Generate the null distribution

1. What kind of relationship between the response and explanatory variables did you encode in your null hypothesis above?
 - a. What word do we typically use in statistics to describe such a pair of variables?
 - b. *How* are we using randomization to build the null distribution?

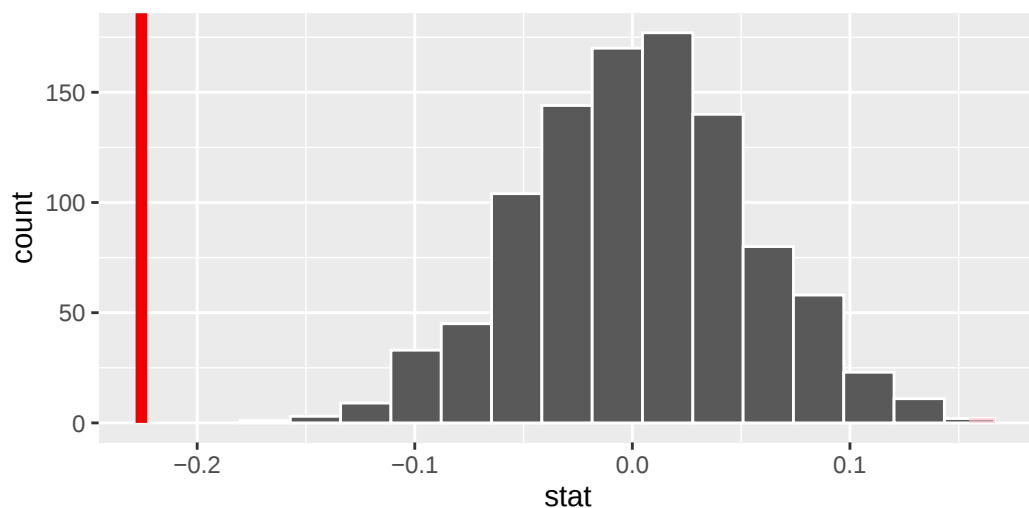
```
permutation_dist <- vietnam_setup |>
  hypothesize(null = "___<type of relationship>___") |>
  generate(reps = 1000, type = "___<type of randomization>___") |>
  calculate(stat = "___<test statistic>___")
```

Step 4: Situate the test statistic in the null distribution

1. What did we call the R object that stored the value of the test statistic?
2. On which sides of the null distribution might we expect to find unusual observations?

```
visualise(permutation_dist) +
  shade_p_value(obs_stat = ___<R object>___, direction = "___<direction>___")
permutation_dist |>
  get_p_value(obs_stat = ___<R object>___, direction = "___<direction>___")
```

Simulation-Based Null Distribution



A tibble: 1 x 1

```
p_value  
<dbl>  
1      0
```

1. Does the randomization test support the null hypothesis?