# Inference for simple linear regression

## Randomization test for the slope

Prof. Dr. Jan Kirenz

# Case study:
# Sandwich store

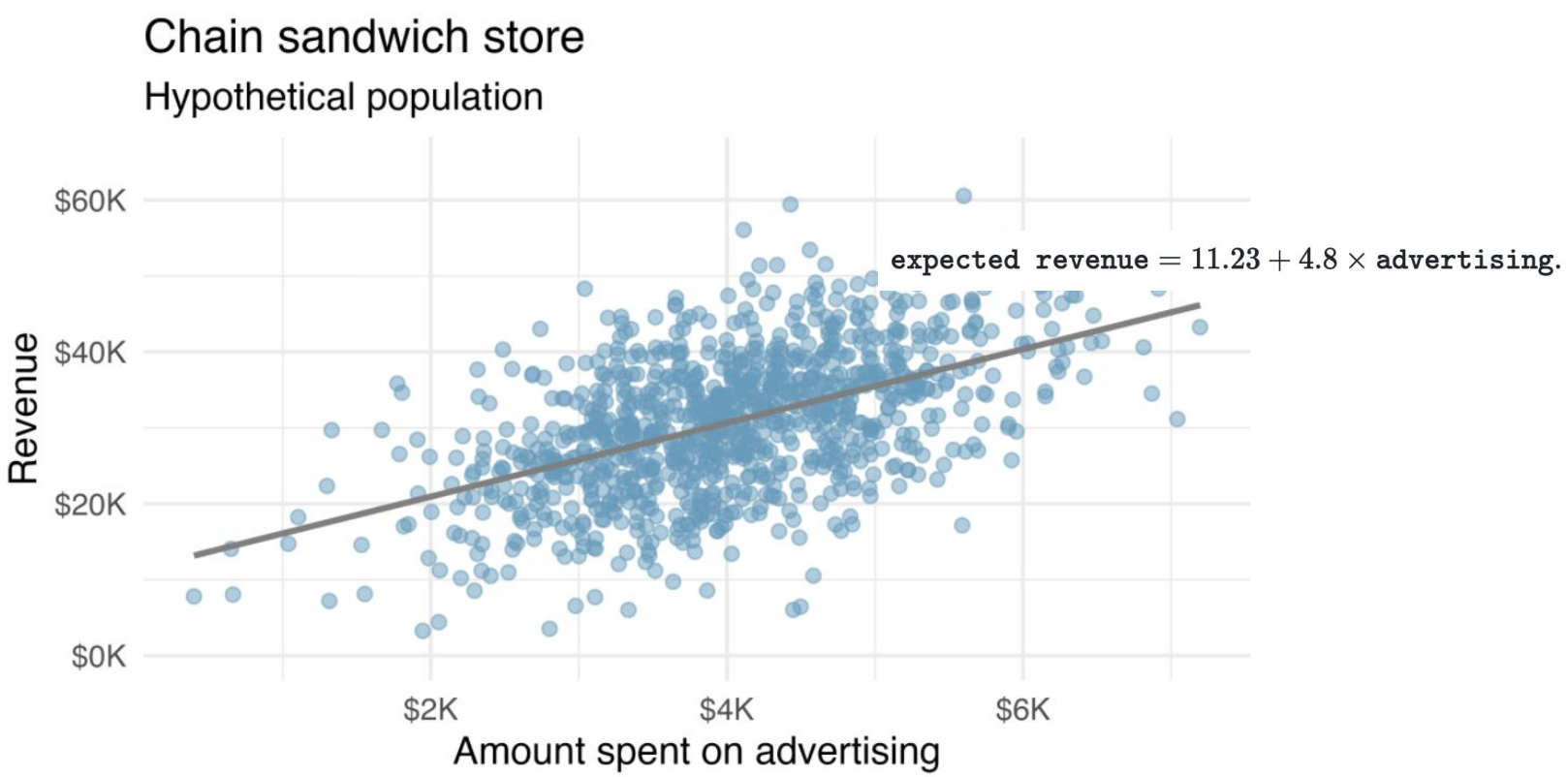



Figure 24.1: Revenue as a linear model of advertising dollars for a population of sandwich stores, in thousands of dollars.

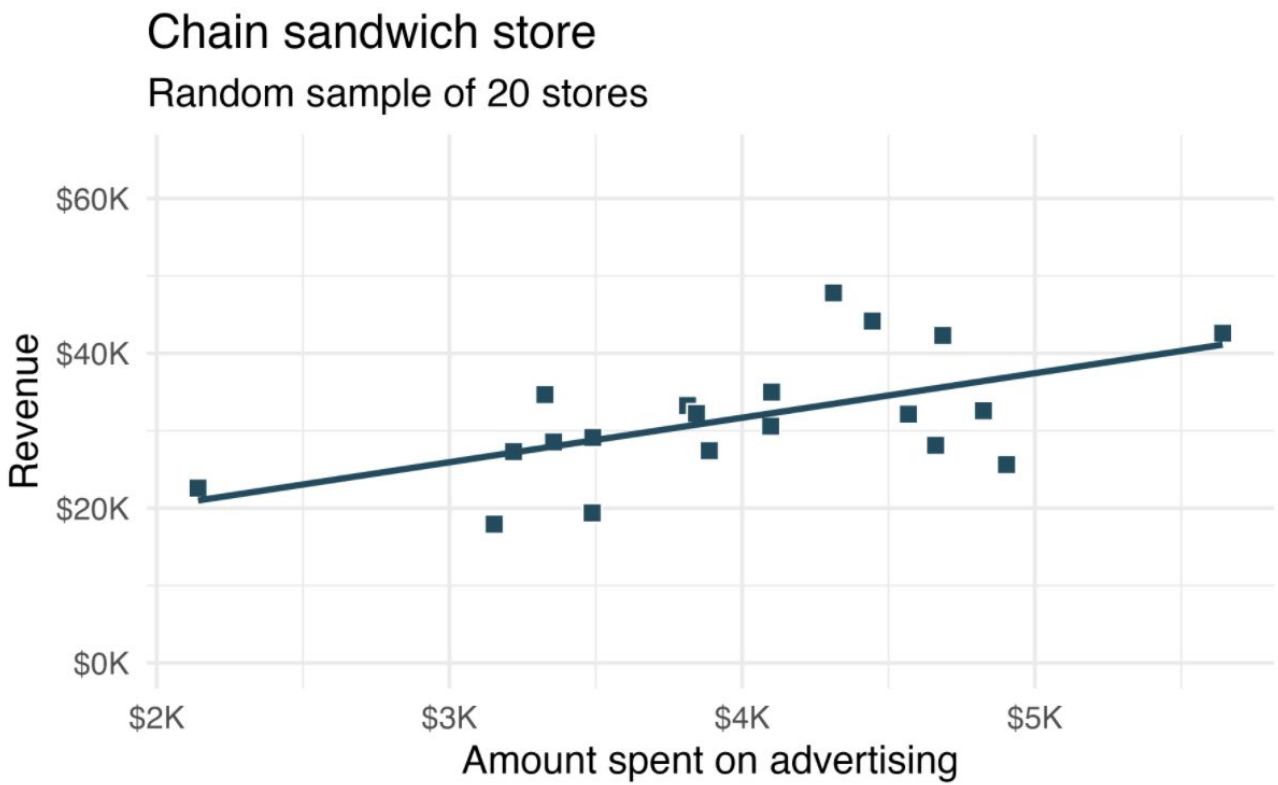



Figure 24.2: A random sample of 20 stores from the entire population. A linear trend between advertising and revenue continues to be observed.

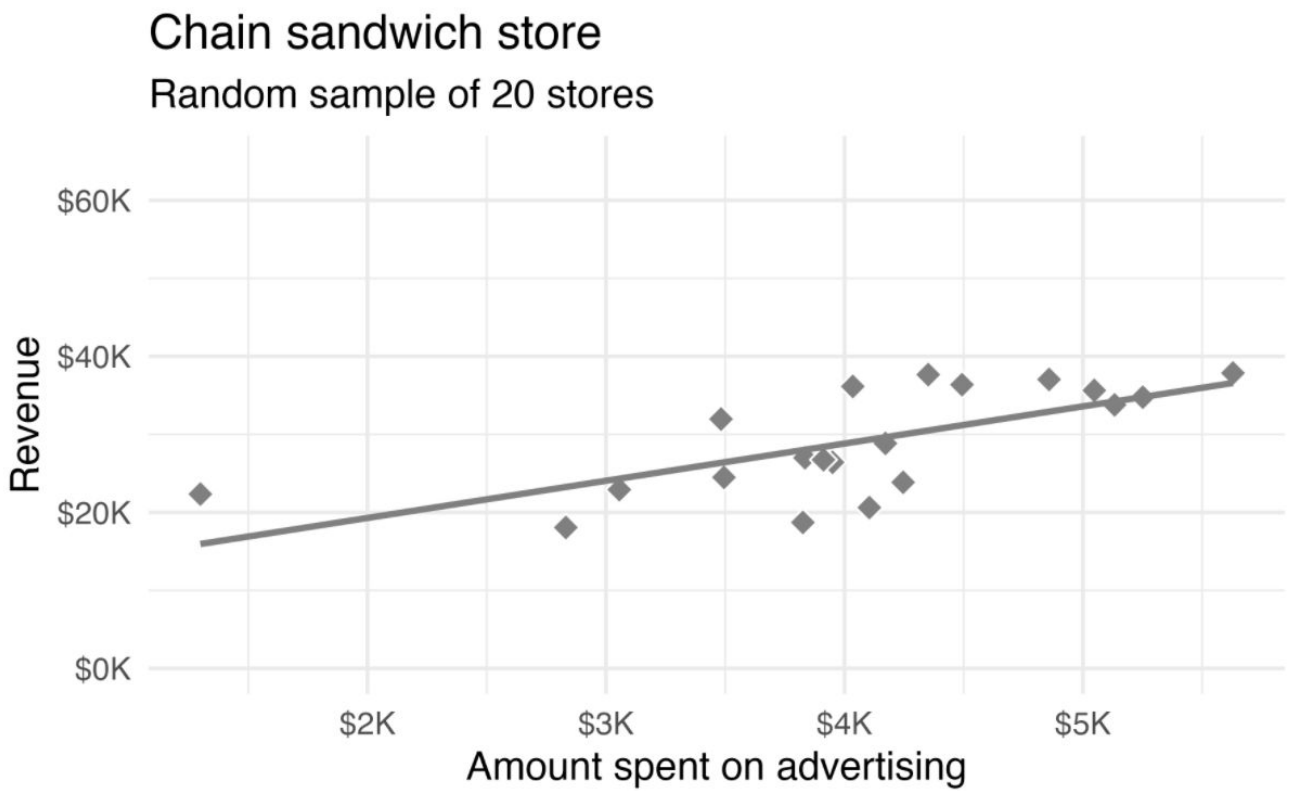
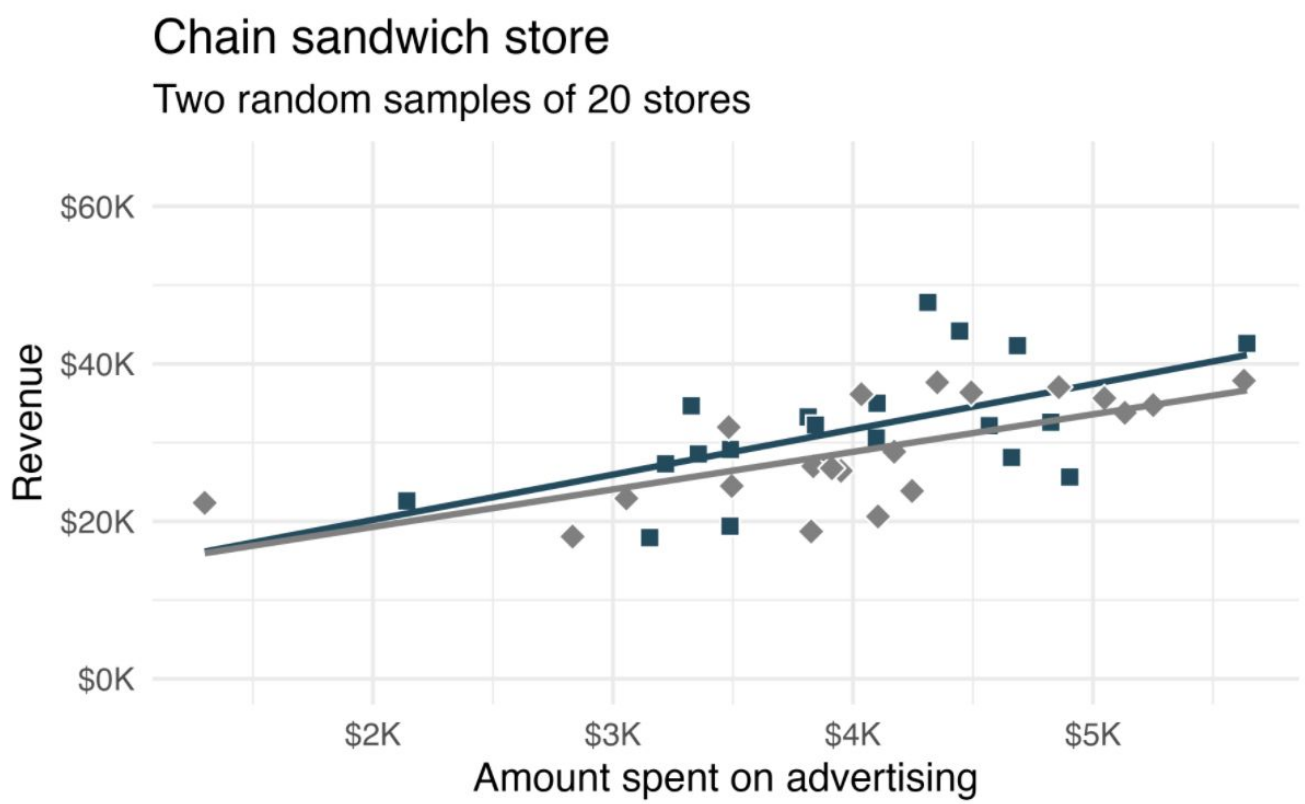



Figure 24.3: A different random sample of 20 stores from the entire population. Again, a linear trend between advertising and revenue is observed.




there is **variability** in the regression line from sample to sample.

Figure 24.4: The linear models from the two different random samples are quite similar, but they are not the same line.

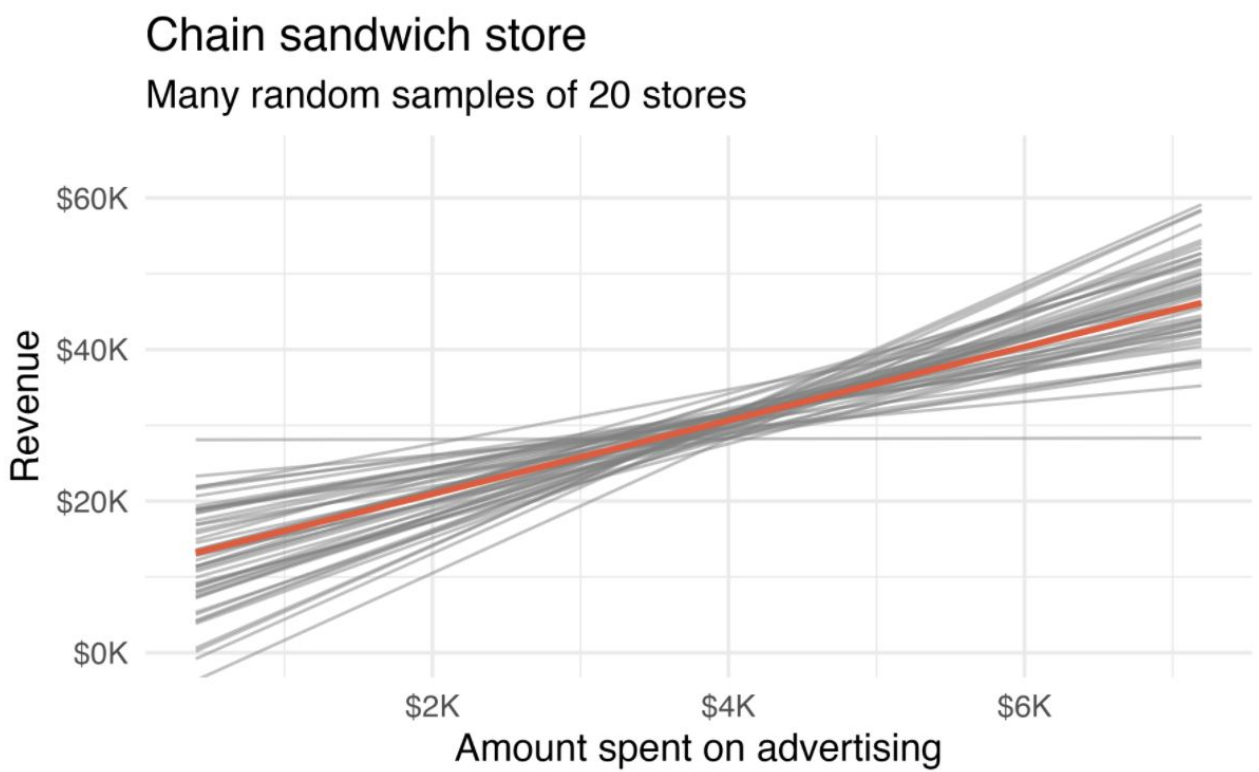



Figure 24.5: If repeated samples of size 20 are taken from the entire population, each linear model will be slightly different. The red line provides the linear fit to the entire population.

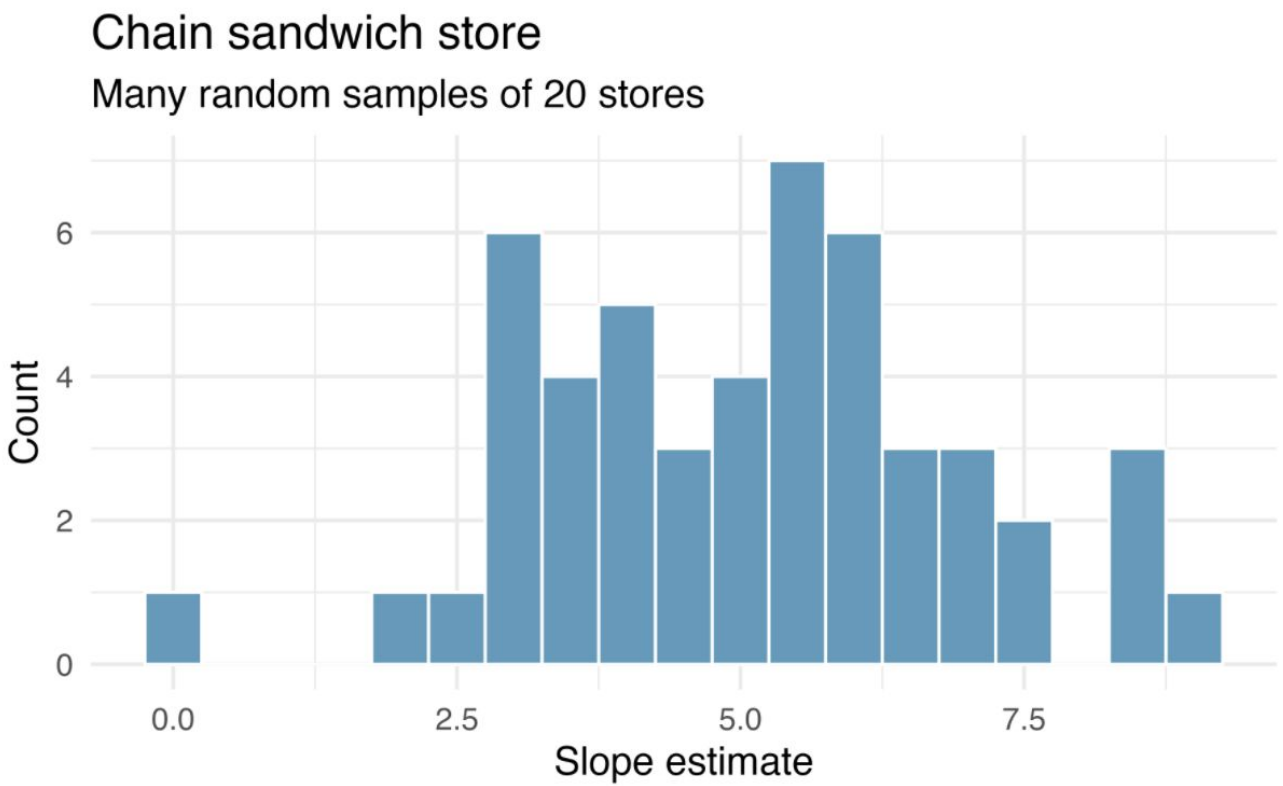



Figure 24.6: Variability of slope estimates taken from many different samples of stores, each of size 20.

# Randomization

test for the slope

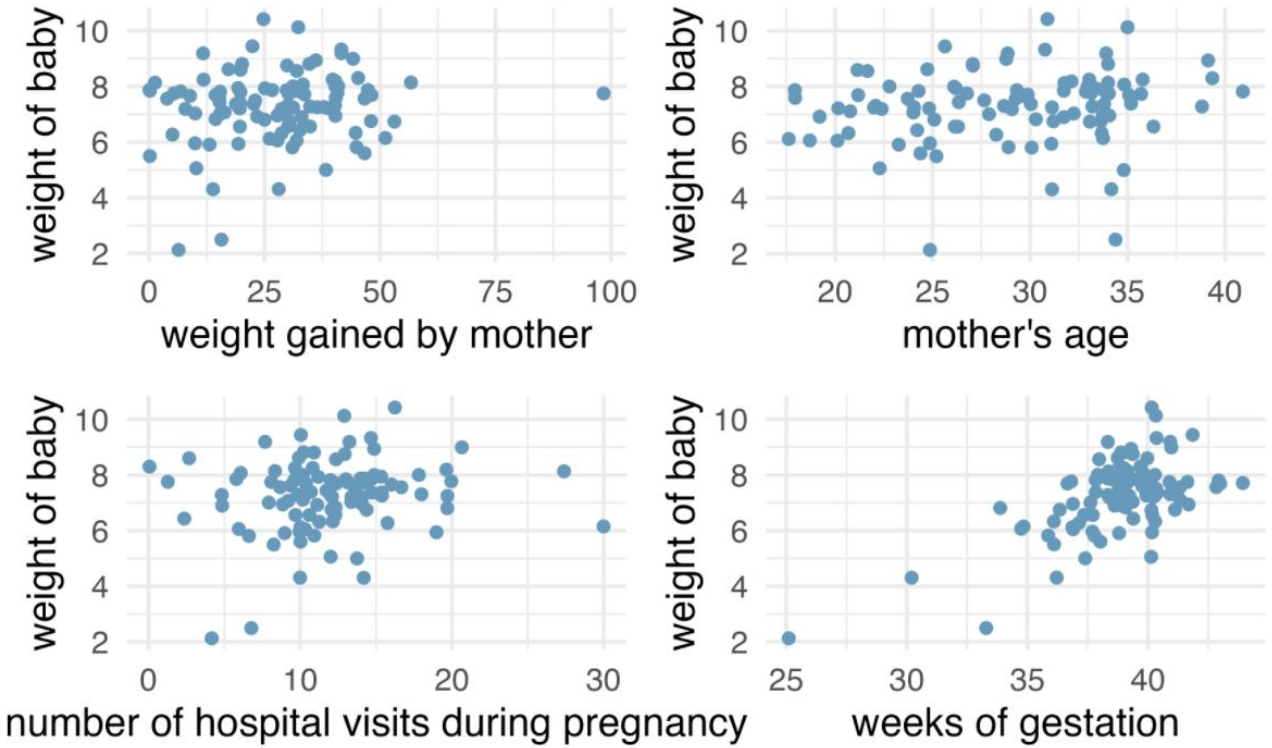



Figure 24.7: Weight of baby at birth (in lbs) as plotted by four other birth variables (mother's weight gain, mother's age, number of hospital visits, and weeks gestation).

# Hypotheses

$H_0$: $\beta_1=0$

There is no linear relationship between weight and weeks.

$H_A$: $\beta_1\neq0$

There is some linear relationship between weight and weeks

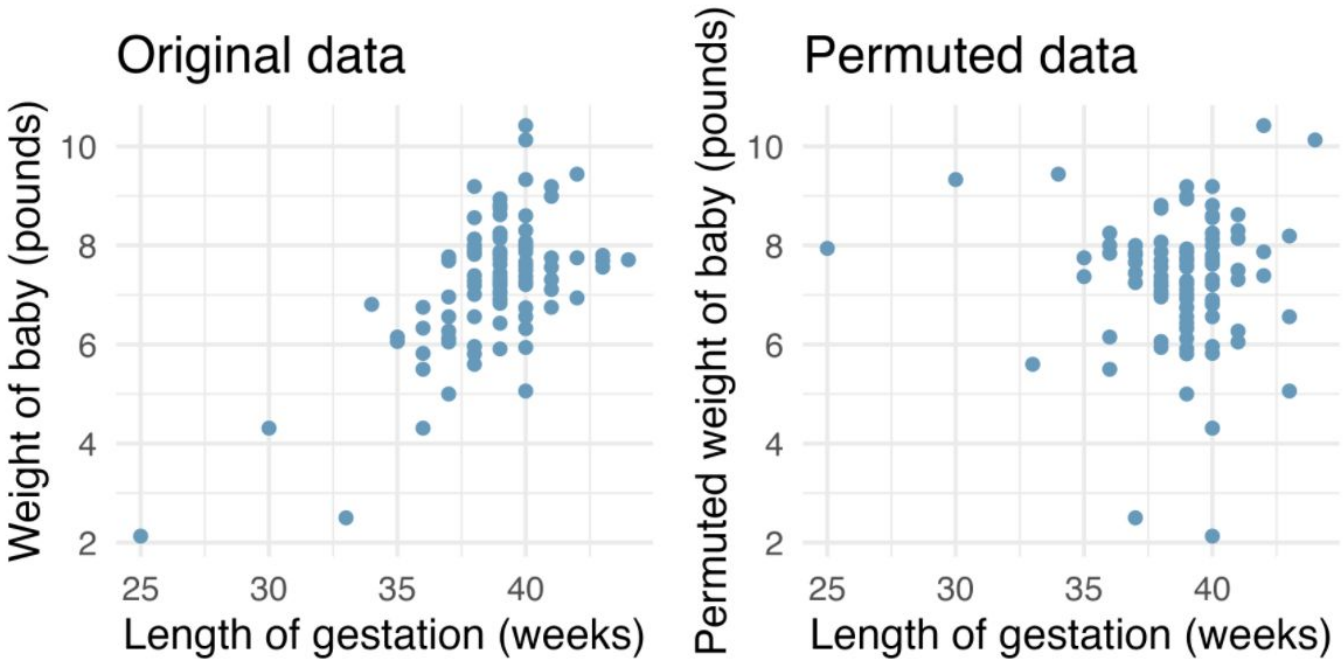



Figure 24.8: Original (left) and permuted (right) data. The permutation removes the linear relationship between `weight` and `weeks`. Repeated permutations allow for quantifying the variability in the slope under the condition that there is no linear relationship (i.e., that the null hypothesis is true).

# Observed data

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | -5.72 | 1.61 | -3.54 | 6e-04 |
| weeks | 0.34 | 0.04 | 8.07 | <0.0001 |

Table 24.1: The least squares estimates of the intercept and slope are given in the estimate column. The observed slope is 0.335.

# Variability of the statistic

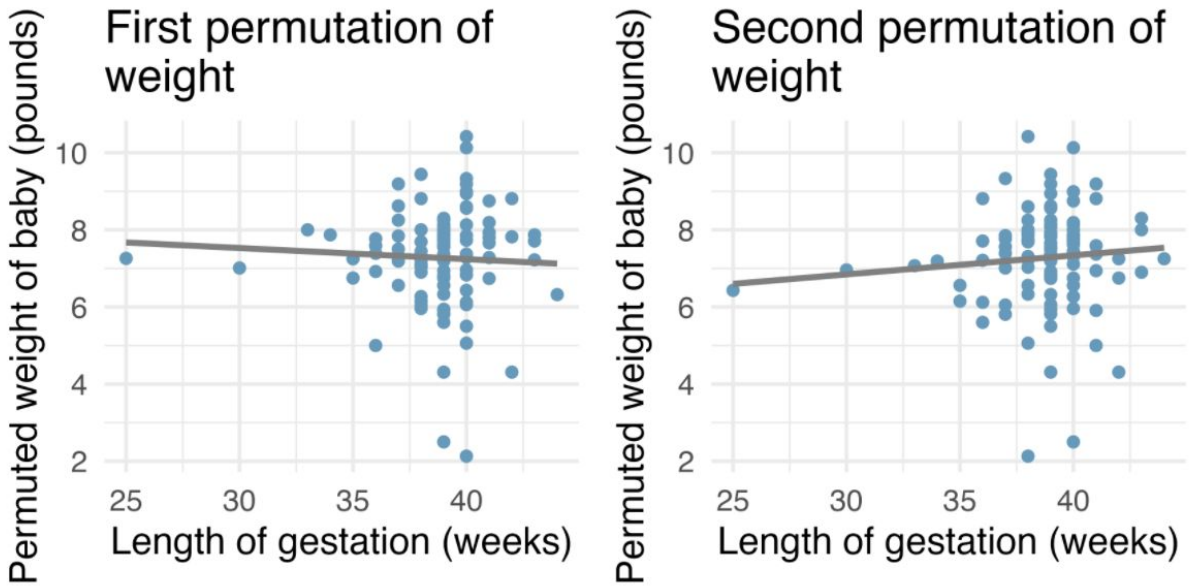



Figure 24.9: Two different permutations of the `weight` variable with slightly different least squares regression lines.

# Observed statistic vs. null statistics

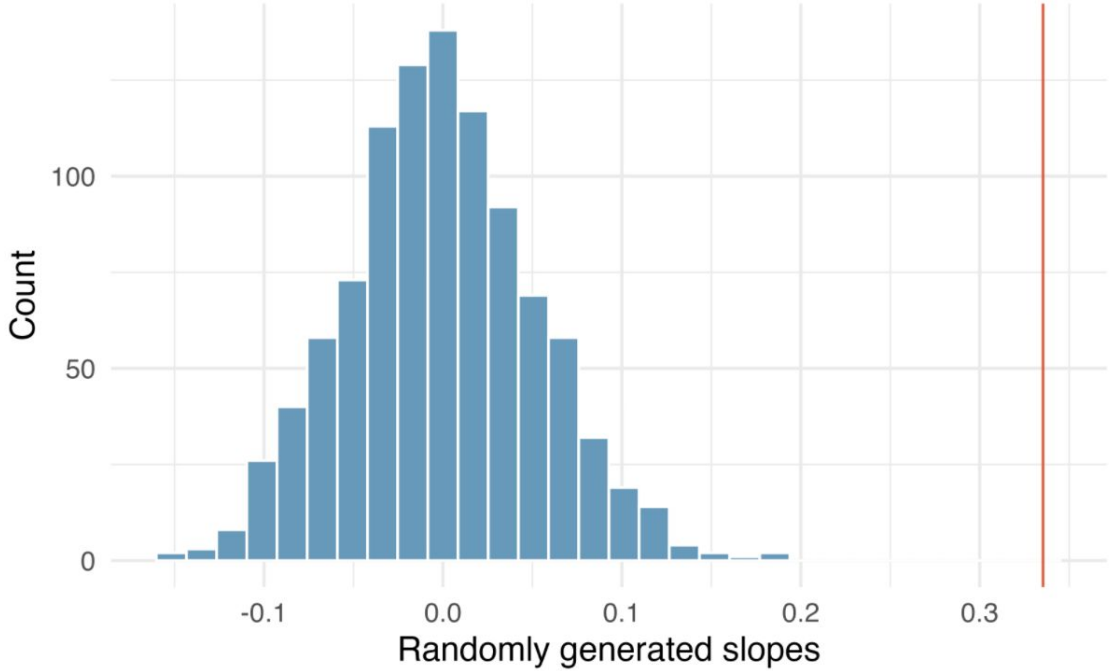
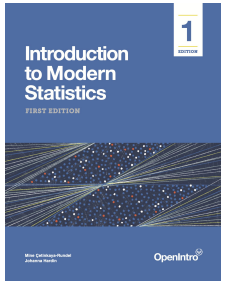



Figure 24.10: Histogram of slopes given different permutations of the `weight` variable. The vertical red line is at the observed value of the slope, 0.335.

# Resources




The content of this presentation is mainly based on the excellent book “Introduction to Modern Statistics” by Mine Çetinkaya-Rundel and Johanna Hardin (2021).

The online version of the book can be accessed for free:

https://openintro-ims.netlify.app/index.html