# Data Analysis Four

Author:
Thomas Noelcke

Instructor:
Katie Jager

I. Part 1

A.

The random sample that I took from the data had a sample mean of 418.2667 and a sample deviation of 85.174. When compared to the population mean of 399.87 this seems a little high. This is reflected below in the histogram of the data. The center of the data appears to be just after 400. The standard deviation on the other hand is actually a bit lower than the population standard deviation of 89.8.

Random Sample of 45 Vehicles Sold in US
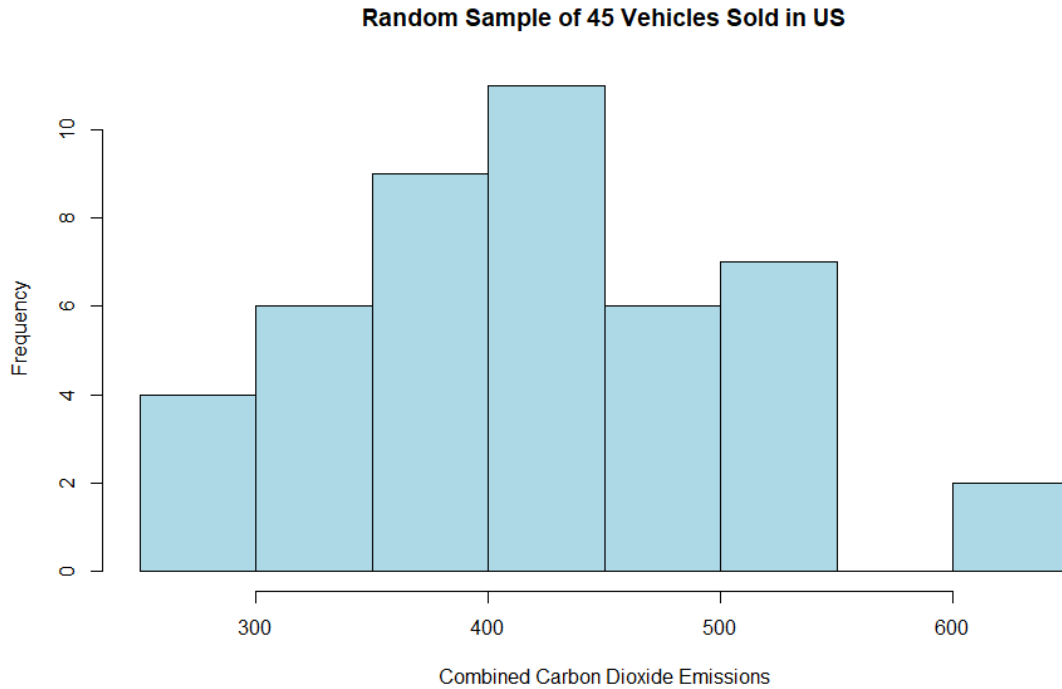
B.

Below I show a 95% confidence interval given my sample from above.

$$Interval = X_bar + -\frac{z^*\sigma}{\sqrt{n}}$$

$$= 418.267 + -\frac{1.96 * 89.8}{sqrt45} = [392.86, 443.612]$$

The interval above does contain the population mean for co2 emissions.

C.

I would expect that around 5% of the students would not find the true population mean in the their 95% confidence interval. Meaning that about 7 students answers would not contain the true population mean.

II. PART 2

A.

As mentioned above in question 1.c. only about 7 students out of the whole class are going to reject the null hypothesis. It is very likely that most of us will fail to reject the null hypothesis.

B.

Below is the solution to the hypothesis test:

$$Z_s tat = \frac{x_b ar - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$= \frac{418.2667 - 399.87}{\frac{89.8}{\sqrt{45}}} = 1.37$$

$$1 - Z(1.37) = 1 - 0.9147 = 0.08353 > 0.05$$

C.

There is not convincing evidence that the sample average does not equal the true population mean. The sample estimate of the average value for CO2 emissions is 418.267 PPM with a 95% confidence interval of 392.86 to 445.612. We have failed to reject the null hypothesis at a significance level of 0.05 with a p-value of 0.0835.

D.

The same data that doesn't contain the true population mean will also reject the null hypothesis because that sample is out side of the 95% interval and exists out side the 0.05 value we have set for Alpha.

III. PART 3

A.

According to the central limit theorem the distribution of the sample means will be normal. In theory the theoretically mean would be the same as the population mean at 399.87. The calculation for the standard deviation is shown below.

$$sd = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{89.8}{\sqrt{45}} = 13.386$$

B.

The shape of the histogram appears as I would have expected normal centered around 400. The mean of the samples was 399.937 which is very close to our expected mean. The standard deviation of the samples was also close to our expected value of 13.120. This supports the central limit theorem.
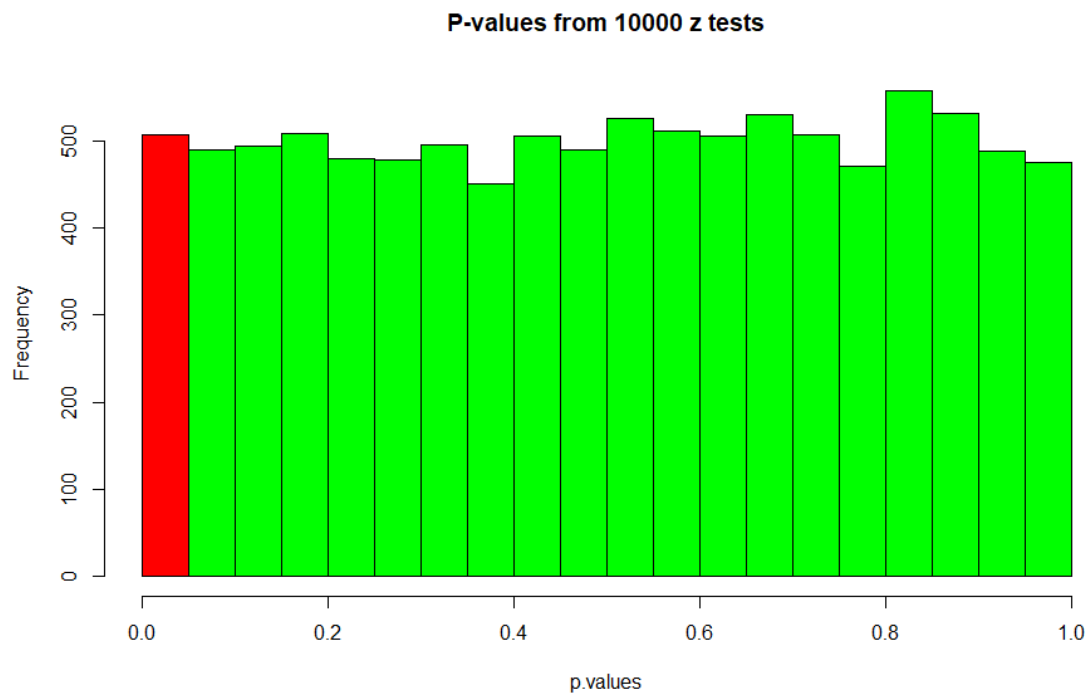
Sampling Distribution

Sample Means , n = 45
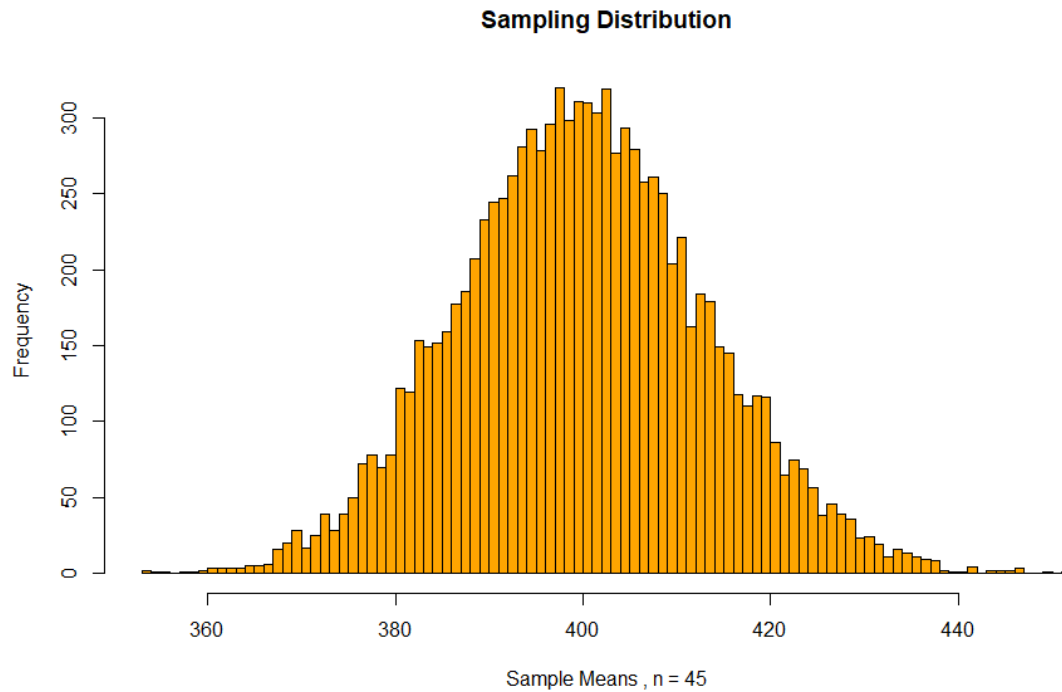
C.

The Z statics can be modeled by the standard normal distribution. They should be centered around zero and have a bell shaped curve.

Histogram of 10000 z statistics
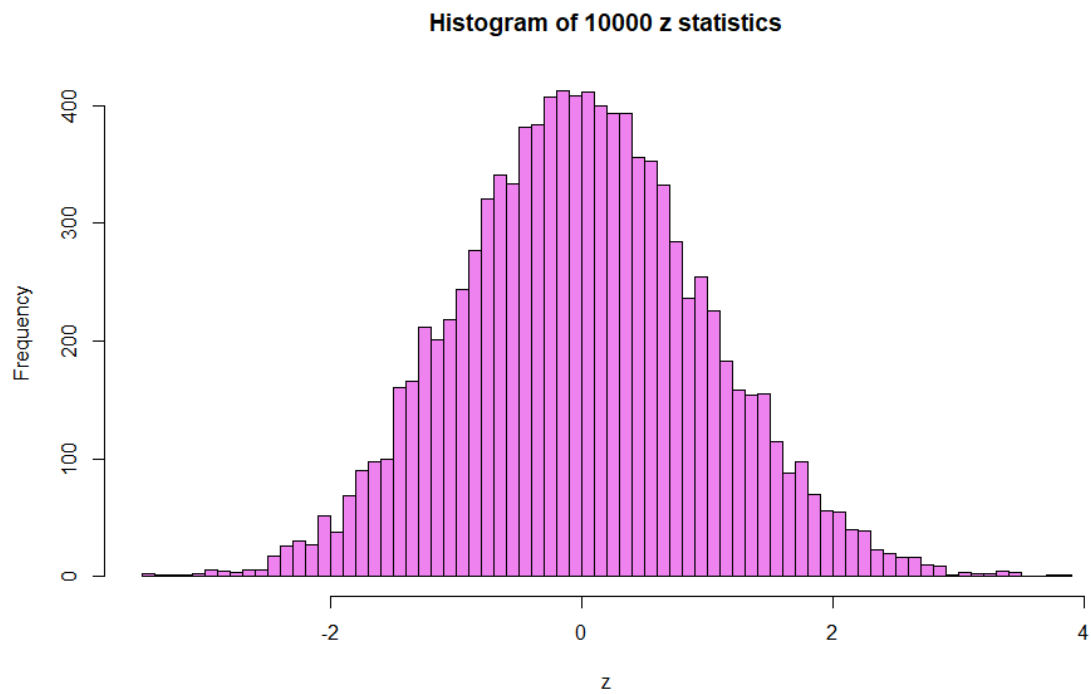
D.

    In this case it appears we reject the null hypothesis about 5 percent of the time. This seems to be inline with what we would expect with Alpha equal to 0.05. In the case where we reject the null hypothesis we are making a Type 1 error because we are rejecting the null hypothesis when it is true.

P-values from 10000 z tests

IV. BIBLIOGRAPHY