EXAMPLE: COMPARING PROPORTIONS

Data. The data from two independent samples is summarized in the following table:

Students	Drink Coffee	Sample Size
High School	4	26
College	15	34

Data analysis.
$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$

Using `prop.test` to compare proportions with 5% level of significance.

p-value: 0.03654
Point estimate for the difference of proportions: -0.2873303

Conclusion. There is sufficient evidence to support the claim that the proportion of college students who report drinking coffee regularly is greater than the proportion of high school students who report the same.

Discussion. The value of $n\hat{p}(1-\hat{p})$ is rather small for both samples, so larger sample sizes would have been preferred.

EXAMPLE: COMPARING MEANS

Data. Two independent samples of trees were taken: 26 evergreen and 21 deciduous trees, and their girths were mesured in cm. The data is provided in a separate file: **sample-data-two-means.csv**

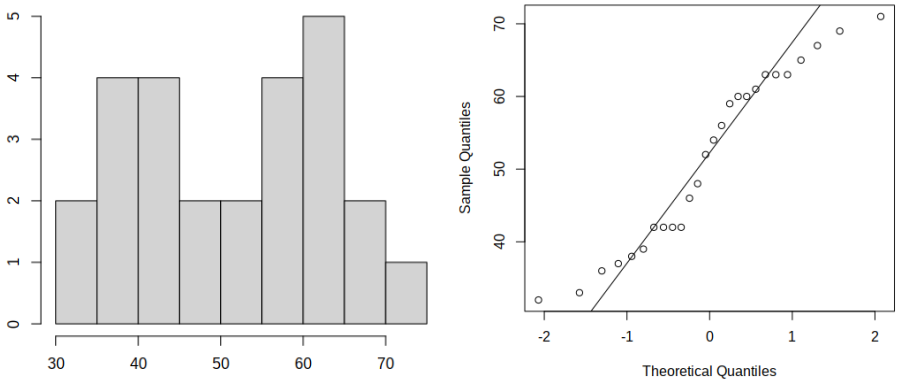
Data analysis.
$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

Using `t.test` to compare means with 1% level of significance.

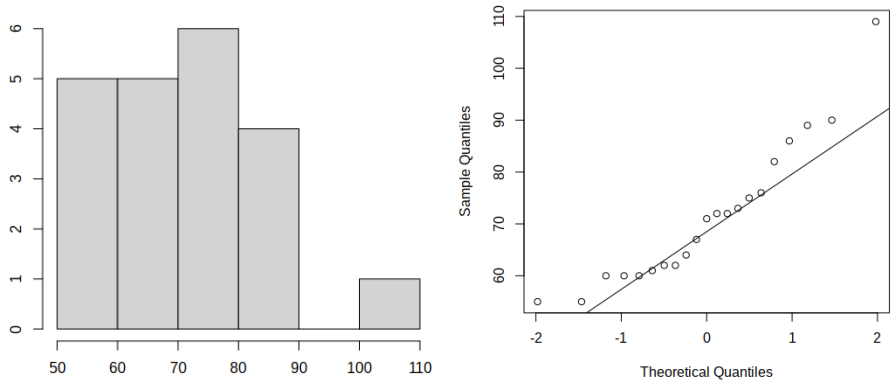
p-value: 5.677×10^{-6}
Point estimate for the difference of means: -19.93773

Conclusion. There is sufficient evidence to support the claim that the mean girth of an evergreen tree is less than the mean girth of a deciduous tree.

Discussion. Evergreen sample distribution:



Deciduous sample distribution:



The distributions do not look very normal, so larger samples would have been preferred.

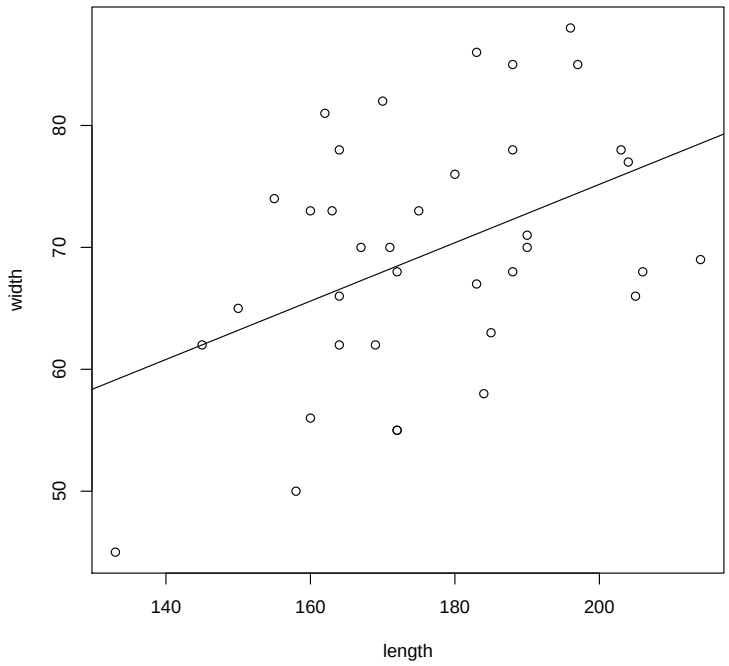
EXAMPLE: TEST FOR CORRELATION

Data. Alhambra Triangle street list:

R St, S St, Serra Way, T St, U St, 30th, 31st, 34th, 35th, Stockton Blvd

For our cluster sample, we picked all of the above streets except for 34th and Stockton Blvd. The raw sample data can be found in a separate file: **sample-data-correlation.csv**

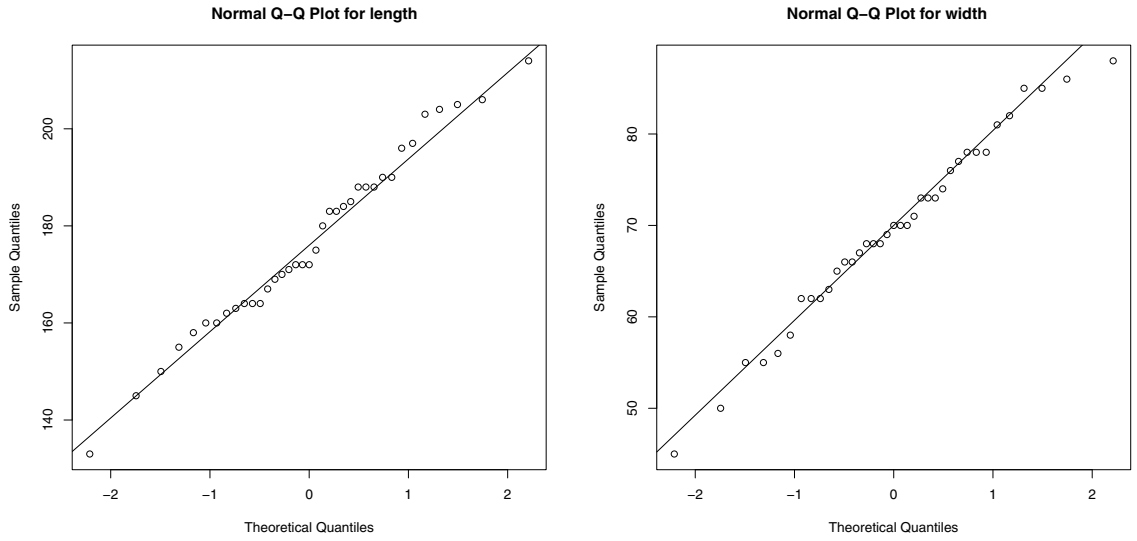
Data analysis.
$H_0 : \rho = 0$
$H_1 : \rho \neq 0$
This is a two-tail test with $\alpha = 0.01$
Sample size: $n = 37$
Pearson's Correlation Coefficient: $r = 0.4360384$



The sampling distribution is t with 35 degrees of freedom.
Test statistic: $t_0 = 2.866494$
Critical values of the sampling distribution: ± 2.723806
p-value is 0.006980393

Conclusion. We can reject H_0 and conclude that there is sufficient evidence to support the claim that the length and the width of a car are linearly correlated.

Discussion. The following QQ-plots summarize the strength of normality in the sample data:



(a) Though the sample size is relatively small, the confidence in our results is improved by the approximately normal distribution of the measured variables, seen in the plots above.

(b) Just as we presumed before commencing the study, there seems to be a positive association. A future study may well take a larger sample and seek the evidence of a strictly positive trend.