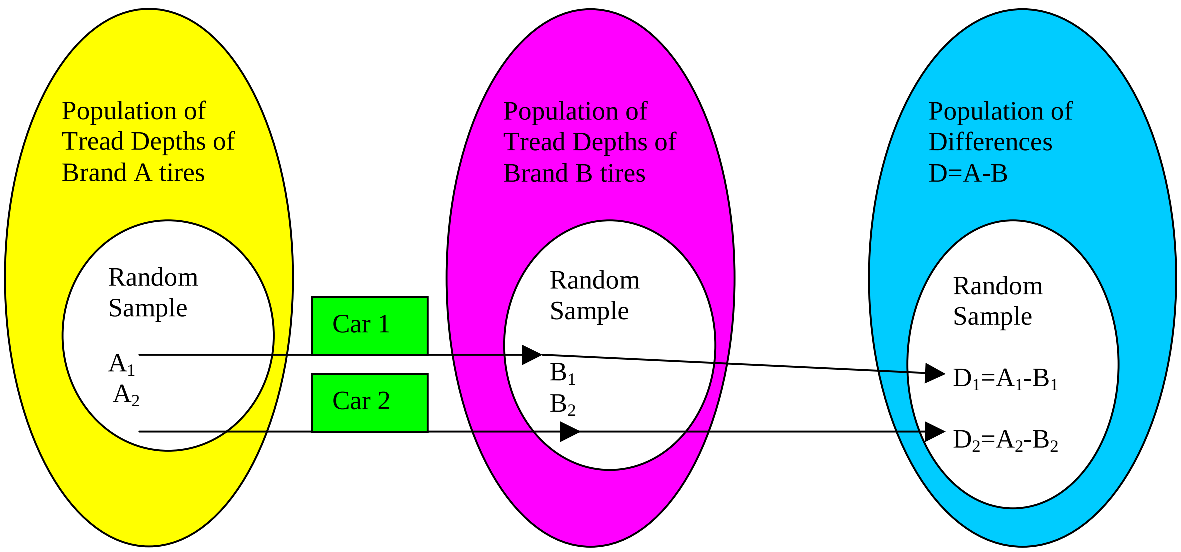
7. The Paired t-Test and a Confidence Interval

We demonstrated before how matched pairs data in our case study can be restated as single-sample data by taking differences within each pair. Thus, we can treat the 20 differences as a random sample of 20 independent observations from the population of the differences.



Hence, in order to make the inferences about the data, we will be using the statistical tools for single populations. In our case study, the population is the population of all differences within matched pairs.

Denote by μ_A and μ_B the population mean tread depth of all brand A and brand B tires, respectively. Then the difference between two population means μ_A and μ_B is equivalent to the mean of the paired differences. In other words, $\mu_D = \mu_A - \mu_B$, where D=A-B. Thus, when an inference is to be made about the difference of two means and paired data design is being used, the inference will in fact be about the mean of the paired differences. The mean of the sample paired differences will be used as the point estimate for these inferences.

We can use the sample of 20 differences to make inferences about the mean of the population of differences, μ_D -which is equal to the difference $\mu_A-\mu_B$. Thus, our test becomes

$$H_0: \mu_D=0 \ (\mu_A-\mu_B=0) \text{ versus } H_a: \mu_D>0 \ (\mu_A-\mu_B>0).$$

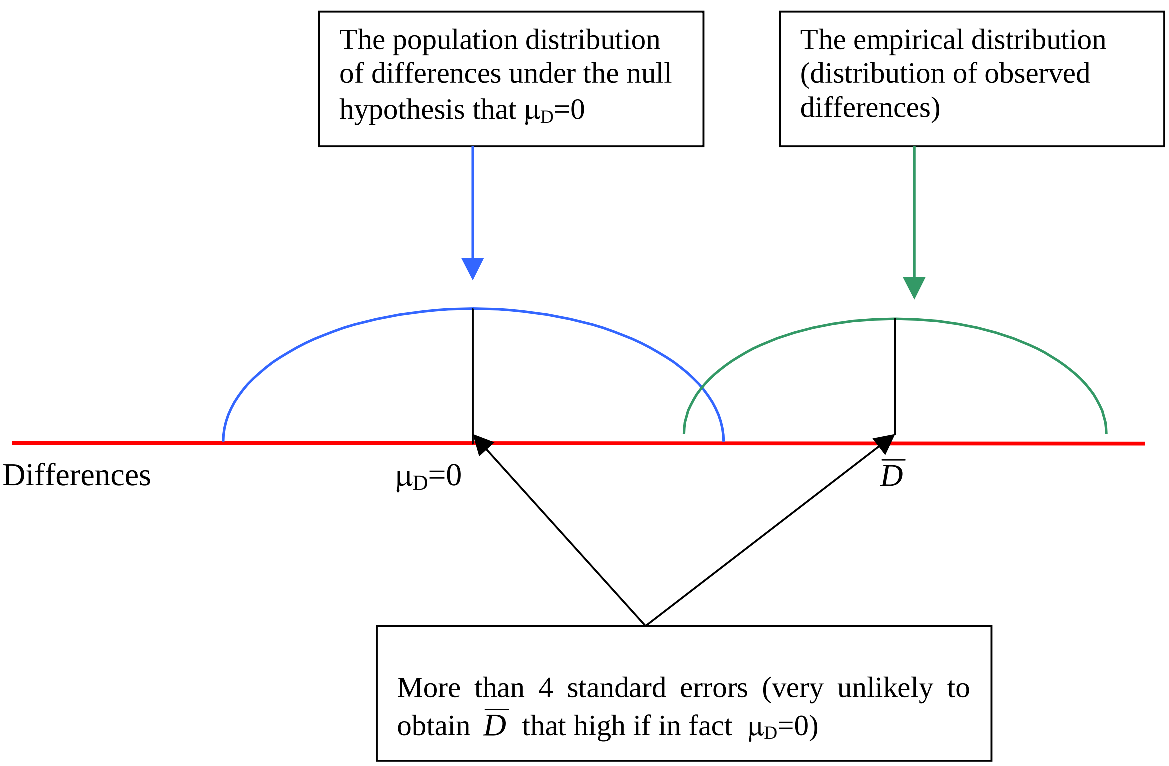
In other words, the null hypothesis states that there is no difference between the average tread depth of the brand A and brand B tires. The alternative hypothesis states that the average tread depth has increased and therefore the durability has improved.

The test statistic is a one-sample t, since we are now analyzing a single sample of differences:

$$t = \frac{\overline{D} - \mu_D}{s_D / \sqrt{n}}, \text{ where } \text{df} = \text{n-1}.$$

Here $\overline{D}$ is the sample mean of the differences, μ_D is the hypothesized value of the population mean, s_D is the sample mean of the differences, and n is the sample size.

Before we will apply the test, let us try to guess the outcome of the test based on the preliminary analysis of the data. If there was no difference in the tread depth, the mean of D would be 0, where D is the population of differences. Then the sample mean $\overline{D}$ would be close to zero. Moreover, under the assumption of normality of D, in a random sample of 20 pairs, we expect that about half differences would be negative, about half would be positive. In our case, 16 of 20 differences are positive and only 4 are negative. Moreover, the sample mean of 0.2765 is large relative to the magnitude of the standard error of 0.0611. Notice that the sample mean of 0.2765 is very, very unlikely if in fact the population mean were zero. Indeed, the distance between the sample mean and the hypothesized mean 0 (in standard errors) is more than 4. Therefore, we believe that there is strong evidence against the null hypothesis that the population mean μ_D=0.



The Paired-Samples T Test in SPSS compares the means of two variables. It computes the differences between values of the two variables for each case and tests whether the average differs from zero. The output of the test is modified here slightly compared to its SPSS version to make it easier to read:

PAIRED SAMPLES STATISTICS				
VARIABLES	**NUMBER**	**MEAN**	**STD. DEVIATION**	**STD. ERROR MEAN**
BRAND A	20	6.4370	0.622	0.139
BRAND B	20	6.1605	0.616	0.138

VARIABLE	**PAIRED DIFFERENCES**			**INFERENCES**				
	Mean	Std. Deviation	Std. Error Mean	**95% CI**		**TEST**		
				Lower	Upper	t	Df	Sign
D= A-B	.2765	.2734	.0611	.1486	.4044	4.52	19	.000

The above output confirms the results of our analysis carried out before performing the test. The value of the test statistic t=4.52 expresses the distance between the hypothesized value of the population mean $\mu_D=0$ and the observed sample mean $\overline{D}$. The p-value of the two-sided test is zero, thus the p-value of the one-sided test considered here is zero, too. Thus the data provide strong evidence that the durability has increased after the new technology was implemented. However, we cannot claim that the better durability is due to the new technology. The 95% confidence interval for the change in the tread depth is (0.1486, 0.4044) in 1/32 of an inch.

REMARK:

In our analysis above, we have made inferences about a single population, the population of all differences within matched pairs. It is inappropriate to use the two-sample t-test because the assumption of independent samples is invalid. The experiment was carried out in the way that makes the observations within pairs dependent.