

28.2: Comparing Two Means

A more common question that often arises in statistics is whether there is a difference between the means of two different groups. Let's say that we would like to know whether regular marijuana smokers watch more television. We can ask this question using the NHANES dataset; let's take a sample of 200 individuals from the dataset and test whether the number of hours of television watching per day is related to regular marijuana use. The left panel of Figure 28.1 shows these data using a violin plot.

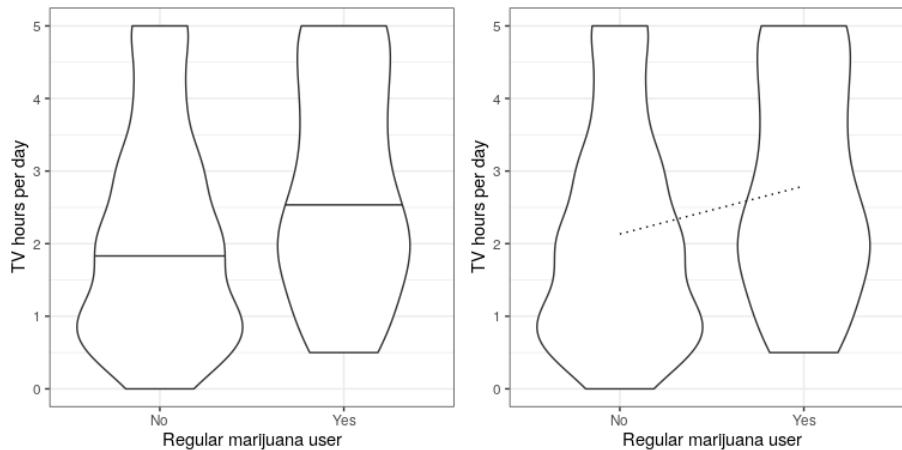


Figure 28.1: Left: Violin plot showing distributions of TV watching separated by regular marijuana use. Right: Violin plots showing data for each group, with a dotted line connecting the predicted values for each group, computed on the basis of the results of the linear model..

We can also use Student's t test to test for differences between two groups of independent observations (as we saw in an earlier chapter); we will turn later in the chapter to cases where the observations are not independent. As a reminder, the t-statistic for comparison of two independent groups is computed as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

where \bar{X}_1 and \bar{X}_2 are the means of the two groups, S_1^2 and S_2^2 are the variances for each of the groups, and n_1 and n_2 are the sizes of the two groups. Under the null hypothesis of no difference between means, this statistic is distributed according to a t distribution with $n-2$ degrees of freedom (since we have computed two parameter estimates, namely the means of the two groups). We can compute the t-test in R using the `t.test()` function. In this case, we started with the specific hypothesis that smoking marijuana is associated with greater TV watching, so we will use a one-tailed test. Since the `t.test` function orders the conditions alphabetically, the "No" group comes first, and thus we need to test the alternative hypothesis of whether the first group is less than the second ("Yes") group; for this reason, we specify 'less' as our alternative.

```
##
## Two Sample t-test
##
## data: TVHrsNum by RegularMarij
## t = -3, df = 198, p-value = 0.004
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -0.25
## sample estimates:
## mean in group No mean in group Yes
##                2.1                2.8
```

In this case we see that there is a statistically significant difference between groups, in the expected direction - regular pot smokers watch more TV.

This page titled [28.2: Comparing Two Means](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.