Doggy Demographics:
Discerning Dimensions of Depth, Dye, and Dartiness

Collin Nolte

Introduction

My first paragraph will be an introduction to my report, including broad-stroke information. In for this particular study, there are primarily two research questions that I have. The first asks if the size and color of dogs are independent, while the second asks if dogs differ in how fast they are, based on their size.

Next, I will describe the data I have and how it was collected. For example, for this study I visited several parks, befriending various dog until I had collected data on 400 individual observations. For each dog, I recorded the dog's breed, its size, color, and the highest speed obtained in a one hundred meter dash, measured in miles per hour. As one of my primary research questions relates to size and color, I have included summary information below. I don't have to include formatted tables, but if I wanted to, the Rmd file associated with this document shows how to create this one.

Table 1: Table with size and color

	black	brown	white	yellow
large	100	0	25	75
medium	50	25	25	0
small	50	25	0	25

As the relationship between size and speed makes up our second research question, I have presented a plot demonstrating this association below.

1
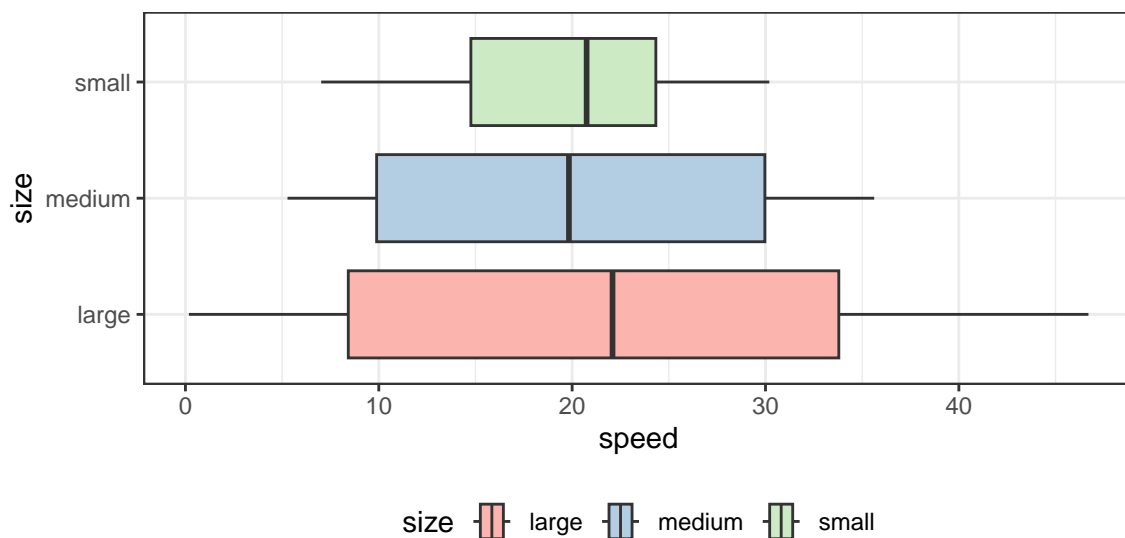
Methods

My first research question addresses the issue of independence between the size of a dog and its color. As these are both quantitative variables, I will be conducting a χ^2 test of independence at a significance level of $\alpha = 0.05$. I should note that if I try and copy α or χ^2 directly into my document, I will *not* be able to get it to knit. Instead, I should look at the Rmd file to see how these symbols are created.

My second research question asks if there is any difference in speed between the different sizes of dogs. As this question involves both a categorical variable and a quantitative variable, I have a number of options available. First, I could do a simple ANOVA to test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3.$$

Additionally, I could could use linear regression to see if there is an association or not. Because I am a man of rather eloquent taste, I have decided on using ANOVA, again testing at the $\alpha = 0.05$ level.

Results

Generally speaking, we should avoid including raw R output in our final documents as it often looks janky, however we can make an exception for the output from statistical tests. That being said, you should limit the total number of tests you do to 1-3.

In this case, our first hypothesis is that of independence between the size and color of the dog. Using the χ^2 test of independence, we find the following result:

```
##
##  Pearson's Chi-squared test
##
## data:  tab
## X-squared = 112, df = 6, p-value <0.0000000000000002
```

Here we derive a χ^2 test statistic of $\chi^2 = 112$ with 6 degrees of freedom, giving us a p-value of $p < 0.00002$. As $p < \alpha$, we handedly reject the null hypothesis that the speed and the color of a dog are independent.

Next we consider an ANOVA model to investigate whether or not there are differences in group means in speed between dogs of different sizes. The results of this test are as follows:

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## size          2    498     249     1.9   0.15
## Residuals   397  51900     131
```

Our ANOVA results in an F statistic of $F = 1.9$ with 2 and 397 degrees of freedom, giving us a p-value of $p = 0.15$. As $p > \alpha$, we fail to reject our hypothesis that dogs of each size have the same average speed.

Discussion

Based on our statistical tests, we have concluded that there does appear to be an association between the size of a dog and its color, though surprisingly there does not appear to be any association between a dogs size and its speed; both of these outcomes are contrary to what we expected to be true prior to beginning our study. Although there is no statistical test to confirm this, these results appear to suggest that perhaps this data has been artificially generated, lending support to the theory that we do indeed live in a simulation.