

Examining Serum Levels of Zinc and Copper in Individuals with Chronic Tinnitus

Reference

This is an original study conducted by myself, Valerie Ingalls, BA, as well as Dr. Ishan Bhatt, PhD, member and PI, respectively, of the Audiogenomics Lab, part of the Department of Communication Sciences and Disorders at the University of Iowa. Data for the study are taken from the National Health and Nutrition Examination Survey (NHANES), which has its data published publicly available at their website.

Introduction

Tinnitus, described as the phantom perception of sound and commonly experienced as ringing in the ears, affects an estimated 10% of US adults each year. While noise exposure is known to cause tinnitus, the exact biological origins of the disorder yet remain unclear. One leading theory is that tinnitus is connected to damage within the cochlea, the primary structure of the inner ear. Zinc and copper are known to prevent free radical damage in the cochlea over time by neutralizing the superoxide radical. Because of this protective role, it is theorized that these molecules may have a protective effect against tinnitus and that their deficiency may result in tinnitus susceptibility. Previous studies examining this hypothesis have shown some promising results, but have largely been underpowered and lacking statistical significance. Therefore, it is necessary to conduct a large-scale investigation to determine whether there are actually differences in serum levels of these molecules among tinnitus patients as compared to the no-tinnitus population.

Figure

As this is an original study, I do not have a figure to replicate. Additionally, the previously mentioned studies did not use the type of visualization that I am planning to use, because their sample sizes were too small for a histogram to be an effective visualization. As such, I have pulled a somewhat similar example and will describe the differences that I plan for my final version.

As you can see from this example, the goal is to create a plot with two histogram distributions on the same axis. On the x-axis will be level of serum copper (or zinc), while the y-axis will be density (preferable to a count y-axis because generalized sample size is in the millions). One of the distributions will be for the no chronic tinnitus group, while the other distribution will be for individuals with chronic tinnitus. I will also add a vertical line at the clinical cutoff value

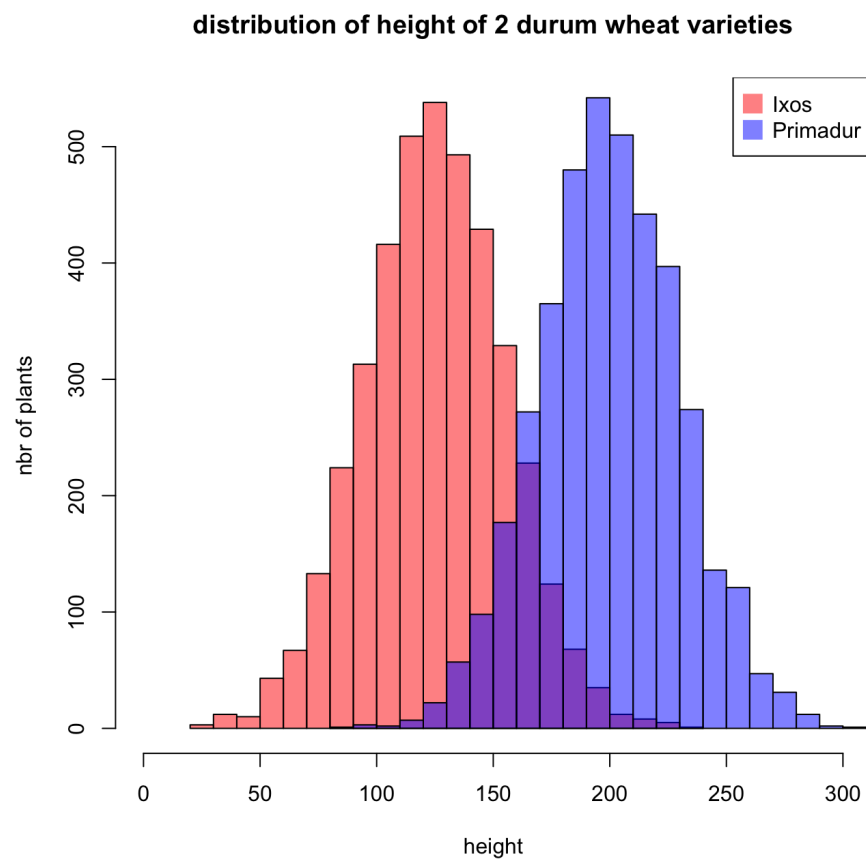


Figure 1: An example histogram

for deficiency in those metals. In this way, I will visualize the difference in zinc and/or copper in those with chronic tinnitus as opposed to those without chronic tinnitus, while simultaneously showing the proportion of individuals in each population who have a clinically significant deficiency in that metal.

Materials and Methods

This is a large-sample epidemiological study using data from the NHANES, a nationally representation health survey that makes use of both questionnaires and physical examination to gather a robust set of health data. The NHANES also utilizes a complex stratified survey design in order to increase its representativeness. To summarize that survey design:

1. Subpopulations of particular public health interest are identified (eg non-white Hispanic individuals)
2. The total population is broken up into strata based on location, often but not always by county
3. Strata are assigned a weight that increases their likelihood of being randomly selected for sampling based on the proportion of their population that includes the targeted subpopulation(s)
4. Weighted randomized selection of strata for the survey
5. Individuals from within those strata are actually surveyed

Weights and strata information are then included in the NHANES data so that it is possible to accurately extrapolate results to the population level. These data are publicly available on the NHANES website, but do require work in order to be effectively usable in R. Data cleaning involves the following rough steps (there's a bit more nuance in reality):

1. Data for each survey cycle are contained in numerous .xpt files. Need to import each of these into R
2. Merge the imported data frames into one df for each cycle
3. Two cycles are being used for this study, so those cycle data frames are merged into one large one
4. Given values for answers of "Unsure" or "Refused to Answer" need to be converted to *NA*
5. Some variables need to be refactored, eg. linear age to turned into age group by decade

In order to make use of the complex survey design in our analysis, we need to use the **survey** package. This package contains functions for logistic regression, t tests, graphing, and many other things. With that package, we can implement the survey design, and perform our statistical analyses (t test and logistic regression) and construct our figures (histograms) in such a manner as to be generalizable to the entire US population.