Lab06\_Massey

Matt Massey

6/24/2021

1. **Which nonparametric test applies?**

The Wilcoxon Rank-Sum test would be appropriate in this case because we want to determine if geographic location has an effect on birth weights. In other words, we are testing if there is a difference between birth weights from two samples from two independent populations - Albany Medical Center Hospital and Olean General Hospital.

1. **Identify the requirements for the appropriate nonparametric test and determine if they are satisfied.**

There are three requirements for the Wilcoxon Rank-Sum test for two independent samples:

1. *The samples are independent*. Both samples are independent as one birth weight has no relationship to another birth weight. This requirement is satisfied.
2. *The data were collected from simple random samples.* The edited data for this lab has been taken from a larger data set of 400 births recorded at four hospitals. We have no information on how the original data was collected, nor how the edited data set was modified. I can only assume the data represent random samples and this requirement is satisfied.
3. *Each sample has more than 10 data points.* The edited data set contains two samples, each having n=100 individual values. This requirement is satisfied.
4. **State the null and alternative hypothesis in words.**

For this example, the null hypothesis, H0 is that both samples come from populations with equal medians. The alternative hypothesis H1 is that both samples come from populations with different medians.

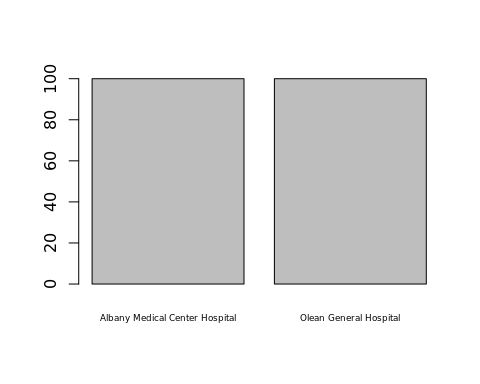
1. **Apply the test, using the code below. Make a conclusion regarding the hypothesis and defend it using the p-value.**

I would fail to reject the null hypothesis. There is sufficient evidence to suggest that both samples come from populations with equal medians and there is no significant difference in birth weights at the two locations. The calculated P-value is 0.94, which is very high well above the common confidence levels of 90%, 95%, or 99%.

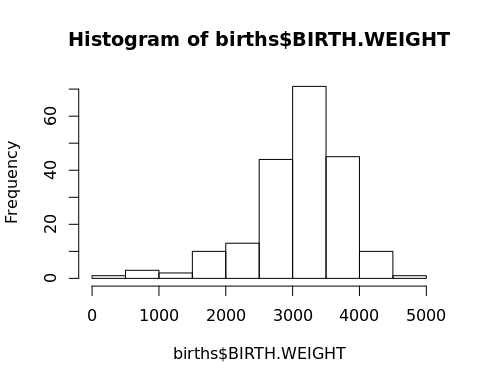
1. **Would this nonparametric test or a two-sample t-test be more appropriate in this case? Explain.**

The nonparametric Wilcoxon rank-sum test has a slightly lower efficiency (0.95) than the two-sample t test, and the requirements are satisfied for the two-sample t-test (simple random samples, independent samples, number of values in both samples is greater than 30 OR the data are approximately normally distributed, and we do not know the population standard deviations). On the other hand, the Wilcoxon rank-sum test is often preferred because of the high efficiency and easier calculations, and the dataset is slightly skewed to the left. I don’t think either test is more or less appropriate than the other in this case.

# Import your data set   
births <- read.table(file = "Lab6\_Birthsedited.txt", sep = "\t", header = TRUE)  
  
# Plot your two variables of interest - What are you looking for?  
# cex.names is the scale of the x-axis labels   
plot(births$FACILITY, cex.names = 0.55)



hist(births$BIRTH.WEIGHT)



# Wilcoxon Rank-Sum Test - What are the requirements for this test?  
wilcox.test(births$BIRTH.WEIGHT ~ births$FACILITY)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: births$BIRTH.WEIGHT by births$FACILITY  
## W = 5030, p-value = 0.9424  
## alternative hypothesis: true location shift is not equal to 0