BIOS 6611 Homework 6

Due Tuesday, October 22, 2019 by 11:59 pm to Canvas Assignment Basket

Provide all relevant code and output with your answers to each question summarizing your conclusion based on your results.

1. Recall the data in Homework 3 on total hospital costs per patient for either of two procedures (Standard (=1) and New (=2)) over a one-month period at one hospital. Use R to read in the **ProcedureCost.csv** data, as before, and carry out the following:

   i. We will use bootstrap sampling to examine the sampling distribution of mean costs for the "New" procedure. For the observed data:
      a. Plot the observed data (histogram, normal quantile plot, boxplot, etc.)
      b. Describe the shape of the distribution (bell-shaped, symmetric, skewed, etc.)
      c. Provide summary statistics (mean, standard deviation)

      For the bootstrap sampling distribution:

      d. Provide plots
      e. Describe the shape and spread
      f. Estimate the bootstrap mean, standard error of the mean, and bias
      g. Obtain the 95% normal percentile and the 95% bootstrap percentile confidence intervals and interpret the results. Comment on the coverage of the normal percentile confidence interval and the potential accuracy of the bootstrap percentile confidence interval.

   ii. We will use bootstrap sampling to estimate the ratio of mean costs between the two procedures: New/Standard, from the original data. Obtain a bootstrap sampling distribution of the ratio of mean costs and:
      a. Provide plots and describe the shape, mean, standard error, and bias of the bootstrap sampling distribution for the ratio of mean costs.
      b. Obtain the 95% normal percentile and the 95% bootstrap percentile confidence intervals and interpret the results. Comment on the coverage of the normal percentile confidence interval and the potential accuracy of the bootstrap percentile confidence interval.

   iii. We will now use permutation testing to estimate a p-value for the ratio of mean costs between the two procedures: New/Standard, from the original data.
      a. Create a permutation distribution. Provide a histogram of the permutation distribution with a vertical line at the observed average ratio of mean costs. Also calculate the permutation test p-value for a two-sided case of the ratio of mean costs and provide an interpretation.
      b. **(Extra credit.)** In 1.ii we used bootstrap sampling and in 1.iii we are using permutation sampling. While the bootstrap provides us an estimate of variability and confidence intervals, the permutation test provides a p-value. Would it make sense to combine these results for a single summary statement (e.g., "The average ratio of mean costs is Y (95% bootstrap percentile interval: (X,Z)) with a p-value of P from a permutation test.")? Why or why not?

2. Continuing with our total hospital costs per patient for either of two procedures data using the **ProcedureCost.csv** data, in Homework 3 we filled in the following table:

		Cost	
		Non-Zero	Zero
Procedure Group	2	65	15
	1	72	48

Assume this was actually a cross-sectional study where the "exposure" is the type of procedure, and the "disease" is if you had a cost or not. For this data we are interested in exploring if the different procedures results in different likelihood of having a cost. Answer the following questions:

i. Using an R function or by hand, calculate the risk difference, risk ratio, and odds ratio and their corresponding 95% CI for the "disease" outcome of non-zero or zero cost. Provide an interpretation of each measure and note if the measures agree on their statistical significance based on the confidence intervals. [*Hint: enter the table above as a matrix to use epi.2by2 from the epiR package so that it treats the "disease" outcome appropriately for an outcome of cost (positive for outcome) versus no cost (negative for outcome).*]

ii. Conduct the following tests of association in R or by hand based on the 2x2 table: chi-squared test without continuity correction, chi-squared test with continuity correction, Fisher's exact test, and McNemar's test with continuity correction. From these test results:

a. Fill in the following table (round values to 3 decimal places, if p-values are less than 0.001 use p<0.001):

Approach	**Test Statistic**	**p-value**
Chi-squared test *without* continuity correction	X^2=	p=
Chi-squared test *with* continuity correction	X^2=	p=
Fisher's exact test	NA	p=
McNemar's test with continuity correction	X^2=	p=

b. R (or your own hand) won't necessary stop you from conducting an inappropriate test. Given our context, are any of these tests inappropriate? Why?

c. Select the most appropriate of the tests conducted. Interpret its p-value in the context of our problem for zero versus non-zero costs by procedure.