

## BIOS 6611 Homework 7

Due Monday, October 29, 2018 by midnight to Canvas Assignment Basket

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)
    - 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.

2. Suppose we have separately analyzed the effects of 10 single nucleotide polymorphisms (SNPs; [https://en.wikipedia.org/wiki/Single-nucleotide\\_polymorphism](https://en.wikipedia.org/wiki/Single-nucleotide_polymorphism)) comparing people with type I diabetes vs. controls. The p-values from these separate analyses are given below.

Effects of 10 SNPs on Type I Diabetes			
SNP	p-value	SNP	p-value
1	0.040	6	0.620
2	0.100	7	0.001
3	0.400	8	0.010
4	0.550	9	0.800
5	0.340	10	0.005

Use the FDR method to correct for multiple testing using an FDR = 0.05. After correction, which SNPs show statistically significant effects?

3. Twenty-two young asthmatic volunteers were studied to assess the short-term effects of sulfur dioxide (SO<sub>2</sub>) exposure under various conditions. The baseline data in the table (Table 12.30 from Rosner) were presented regarding the relationship of bronchial reactivity to SO<sub>2</sub> (cm H<sub>2</sub>O/s) stratified by lung function (as defined by forced expiratory volume / forced vital capacity [FEV<sub>1</sub>/FVC]) at screening.

Lung-Function Group		
Group A FEV <sub>1</sub> /FVC ≤ 74%	Group B FEV <sub>1</sub> /FVC 75-84%	Group C FEV <sub>1</sub> /FVC ≥ 85%
20.8	7.5	9.2
4.1	7.5	2.0
30.0	11.9	2.5
24.7	4.5	6.1
13.8	3.1	7.5
	8.0	
	4.7	
	28.1	
	10.3	
	10.0	
	5.1	
	2.2	

Using SAS or R:

- Assume that the variances across the groups are equal and test the hypothesis that there is an overall mean difference in bronchial reactivity among the three lung-function groups.
- If justified, compare the means of each pair of groups using the Tukey HSD method and summarize the results. Otherwise note why it isn't justified.
- EXTRA CREDIT:** Carry out part (i) assuming that the variances across the groups are not equal. If justified, describe a way to compare the means of each pair of groups, but do not carry out any further analysis.