Stat 131A Spring 2012

Lifetimes of Guinea Pigs

Due Tuesday, May 1st, at 11:5

1. You may work in groups of two or three. Please email me by Thursday, April 19, if you would like to change your group. If I don’t hear from you, I will assume that the groups will stay the same as for Project 2.

2. The data are available on the main projects page. The data set is called Guinea Pigs.

3. The project details:

   [Bjerkedal et al. 1960] performed a study on the lifetimes of guinea pigs after being infected with different doses of virulent tubercle bacilli. Bjerkedal formed 6 groups of guinea pigs: a control group with 107 guinea pigs and 5 treatment groups containing 72 guinea pigs each. Guinea pigs in each treatment group were inoculated with tubercle bacilli. The dosage of tubercle bacilli was different for each treatment.

   The guinea pigs were observed for two years. For each guinea pig that died, the age of the guinea pig (in days) at the time of death was recorded. Not all guinea pigs died within the two years of observation.

   The data set gpigs.csv contains data from the lowest-dosage treatment groups (trt = 1) and the control group (trt = 0).

   (a) Compare the treatment group and control group numerically and graphically (include the 5 number summary, mean, SD, boxplot, and histogram for both treatment and control groups). Do you notice any differences between the groups?

   (b) Do the treatment and control groups look approximately normal? Check with a QQ plot.

   (c) For the guinea pigs that died, is the average age for the treatment group significantly different than that for the control group? Perform a hypothesis test to check.

   (d) Make a 95% confidence interval for the true difference of the mean of the control group and the treatment group.

Bootstrap procedures can be very helpful in finding the sampling distribution of statistics:

   (a) From the control group, take 1000 samples of size 65 with replacement. For each sample, find the mean. Please DO NOT turn in these numbers.

   (b) From the treatment group, take 1000 samples of size 60 with replacement. For each sample, find the 1st quartile, the median, the mean, the standard deviation and the IQR. Please DO NOT turn in these numbers.

   (c) For the treatment group, draw a histogram of the means obtained from the 1000 samples. This is an approximation of the sampling distribution of the sample mean of the control group; no assumptions on approximate normality have been made. Describe the shape of the histogram. Does the histogram look normal? Make a QQ plot and comment.

   (d) For the treatment group, draw histograms for the 1st quartile, the median, the standard deviation, and the IQR obtained from the 1000 samples.

   Do all of the histograms look normal? Make QQ plots for each statistic and comment.

   (e) Suppose that the guinea pigs in the treatment group are a SRS of all guinea pigs (they’re not, but let’s suppose they are). For the treatment group, sort the means of the 1000 samples in increasing order and select the 26th and the 975th largest means. Note that 950 of the 1000
means are between these numbers. This is a 95% bootstrap confidence interval for the mean lifetime of all guinea pigs that die within two years and are inoculated with the given dosage of tubercle bacilli. Write down this confidence interval.

(f) For the treatment group, give 95% bootstrap confidence intervals for the 1st quartile, the median, the standard deviation, and the IQR. We do not have nice mathematical formulas for these confidence intervals (we only have the nice formulas for the means and proportions). However, we can still make approximate confidence intervals for these choices of statistic using the bootstrap method.

(g) Obtain 1000 differences of sample means by taking the 1000 sample means for the control group subtracted by the 1000 sample means for the treatment group. Draw a histogram of these differences of sample means. Comment on the shape of the histogram. Do these differences look normal? Make a QQ plot and comment.

(h) Obtain a 95% bootstrap confidence interval for the true difference of the mean of the control group and the treatment group. Compare this confidence interval to the interval in 4).

References