

Statistics 345: Homework 6

Due on 2014-11-11

Li Li

Please complete the problems on a separate sheet of paper with your name at the top. Make sure to show your work and/or provide an explanation for each problem. Partial credit will be given when merited. The total credit is 7.

Problem 1

An article in the Journal of Physiology studied gene expression as a function of resistance exercise. Expression data (measure of gene activity) from one gene are shown in the following. One group of rats was exercised for six hours while the other received no exercise. Remember to use the R codes I put on the slide 15.

6 Hours of Exercise : 425, 223, 389, 139, 213, 324, 209, 287, 244, 408, 158, 436

No Exercise : 485, 160, 478, 249, 236, 253, 407, 335

- (a) Compute the mean of the expression data for the exercise group and non-exercise group. (0.5 point)
- (b) Compute the standard deviation of the expression data for the exercise group and non-exercise group. Which data has more variability? (0.5 point)
- (c) Compute the median of the expression data for the exercise group and non-exercise group. (0.5 point)
- (d) Construct histogram of the expression data for the exercise group and non-exercise group. (0.5 point)
- (e) Construct the box-plots of the expression data for the exercise group and non-exercise group. Be sure to label where are the median, Q_1 and Q_3 . (0.5 point)
- (f) Which group do you think has higher gene activity? Why? (0.5 point)

Problem 2

- (a) Suppose X_1, \dots, X_{10} follow normal distribution with mean 2 and variance 1.5^2 . Let $\bar{X}_1 = \sum_{i=1}^{10} X_i/10$. What would be the mean, variance of \bar{X}_1 ? (1 points)
- (b) Suppose X_1, \dots, X_{20} follow normal distribution with mean 2 and variance 1.5^2 . Let $\bar{X}_2 = \sum_{i=1}^{20} X_i/20$. What would be the mean, variance of \bar{X}_2 ? (1 points)

In the following, we use simulations to verify the above theoretical results. Use the following R code to take 100 samples out of normal distribution with mean 2 and standard deviation 1.5; suppose the sample size for each sample is 10. The samples are stored in the matrix "mysample1" and the mean for each sample is stored in the vector "xbar1".

```

n=100
samplesize=10
mysd=1.5
mysample1=matrix(0,ncol=samplesize,nrow=n)
xbar1=NULL
for(i in 1:n){
mysample1[i,]=rnorm(samplesize, mean=2, sd=mysd)
xbar1[i]=mean(mysample1[i,])
}

```

- (a) What is the mean of data vector \bar{x} ? What is the difference between mean of data vector \bar{x} and the theoretical mean of \bar{X}_1 ? (0.5 point)
- (b) What is the variance of data vector \bar{x} ? What is the difference between variance of data vector \bar{x} and the theoretical variance of \bar{X}_1 (0.5 point)

Now we increase the sample size to 20. Use the following R code to take 100 samples out of normal distribution with mean 2 and standard deviation 1.5. The samples are stored in the matrix "mysample2" and the mean for each sample is stored in the vector "xbar2".

```

n=100
samplesize=20
mysd=1.5
mysample2=matrix(0,ncol=samplesize,nrow=n)
xbar2=NULL
for(i in 1:n){
mysample2[i,]=rnorm(samplesize, mean=2, sd=mysd)
xbar2[i]=mean(mysample2[i,])
}

```

- (a) What is the variance of data vector \bar{x} ? What is the difference between variance of data vector \bar{x} and the theoretical variance of \bar{X}_2 (0.5 point)
- (b) What do you observe on the variances of \bar{x} and \bar{x} ? Which is smaller? (0.5point)