

Read [solutions to previous computing homework](#).

(Archive) Math 448 Computing Homework (Fall 2015)

Due: 8 pm, November 5.

Each of you will receive two rows of 0 or 1's. Each row has 31 observations (0 or 1's). It is assumed that the two rows correspond to two samples from two populations, and each observation follows Bernoulli distribution with probability p_1 for Population 1 or p_2 for Population 2. We will conduct a hypothesis test to tell whether $p_1 \neq p_2$. Define a new parameter $\theta = p_1 - p_2$.

If you use the following R code you can extract the two rows and save them as two vectors.

```
####Set the R working directory###
setwd("C:/")
###read in the data file, make sure your data file is under the working directory##
dat = read.csv('data_6.csv',header=FALSE)
### the "dat" you just read into R is a data frame, need to convert it into a matrix
dat <- as.matrix(dat) ##dat now is a 2x31 matrix
x1 <- dat[1,] ##take the first row of this matrix as your sample 1
x2 <- dat[2,] ##take the second row of this matrix as your sample 2
```

1. Suppose we use the test statistic introduced in class to build the test, namely $TS = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}$. What is the probability of making a Type I error if we use $|TS| > 1$ as the rejection rule? Hint: what distribution does TS approximately have under the null hypothesis? Note that you do not need to make use of the data that you have received to answer this question.
2. Suppose we use $|TS| > t$ (where $t > 0$) as the rejection rule. Determine the value of t if we want to control the probability of making a Type I error at 0.10. Note that you do not need to make use of the data that you have received to answer this question.
3. Suppose we approximate the $SE(\hat{\theta})$ by $\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$. Then calculate the observed value of the TS using the data given.
4. Make a decision/conclusion with significance level 0.10 using the TS whose denominator is approximated as in Question 3.
5. Suppose we approximate the $SE(\hat{\theta})$ by $\sqrt{\hat{p}(1-\hat{p})/n_1 + \hat{p}(1-\hat{p})/n_2}$, where \hat{p} is the pooled sample proportion, defined as $\hat{p} = (Y_1 + Y_2)/(n_1 + n_2)$. Then calculate the observed value of the TS using the data given.
6. Make a decision/conclusion with significance level 0.10 using the TS whose denominator is approximated as in Question 5.

Computing Homework 5

Due: 8 pm, October 18.

Each of you receive 20 numbers. It is known that they are from an exponential distribution with mean parameter θ . Answer the following questions.

1. Report the Method of Moment estimate of θ by matching the population first moment and sample first moment.

2. Report the Method of Moment estimate of θ by matching the population second moment and sample second moment.
3. Report the Minimum Variance Unbiased Estimate of θ .
4. Now let's consider a new parameter $\psi = \theta^2$. Report the Method of Moment estimate of ψ by matching the population first moment and sample first moment.
5. Report the Method of Moment estimate of ψ by matching the population second moment and sample second moment.
6. Report the Minimum Variance Unbiased Estimate of ψ .

In Questions 3 and 6, you need to derive the sufficient statistic for the parameter (θ or ψ), calculate its bias, and correct the bias by applying some transformations to the sufficient statistics if necessary.

Note: in practice, we discourage people from using the method in Questions 2 and 5, compared to that in Questions 1 and 4. This exercise is merely for an illustration why would this be the case.

Computing Homework 4

Due: 8 pm, September 27.

Each of you receive 255 numbers, denoted as X_1, \dots, X_{255} , all of which follow a normal distribution with an **unknown** mean and an **unknown** variance. **Please read following questions carefully. Note that not all numbers will be used!!!**

The goals include finding a point estimator and a confidence interval for μ with good accuracy.

1. Choose the first 10 observations, X_1, \dots, X_{10} , as your sample. Report an estimate of the unknown population variance. In R, the following command extracts the first 10 elements of vector X and save them as a new vector y . `y=x[1:10]` Recall Computing Homework 2 on how to calculate the sample variance. Parts (2)-(4) in that homework gave you the numerator of the sample variance formula.
2. Pretend that 10 is a large number (although it is not that large, let's pretend it that way for now.) Construct a 90% confidence interval using only the first 10 observations (the sample mean is based on the 10 observations, and the sample standard deviation is also based on the 10 observations.) Report the two-sided confidence interval: lower bound in (2a) and upper bound in (2b). For this question, ignore the materials in Section 8.8. That is, please use the formula in Section 8.6.
3. The confidence interval obtained above does not provide a lot of information about μ since it is too wide. In addition, the sample mean based on only 10 observations is also unlikely to be accurate. **In this problem, we want to find an estimator (sample mean) whose error of estimation is no greater than 0.5 with a probability of 99%.** One way to achieve this goal is to increase your sample size. Your answer in Question 1 above (based on 10 data points) gives you an estimate to the true population variance of random variables X_i 's. Now calculate the minimum sample size (i.e. number of observations) needed to achieve the desired accuracy (that is, the error of estimation has to be less than 0.5 with a probability 0.99). Round your answer up as an integer. Denote the required sample size as n . Report the total cost of collecting these n observations (remember: each observation costs $\$12$ dollars.)

4. Use the first n data points in the data set that you have received, and use these n observations to calculate an unbiased point estimate for μ . Recall the command `x[1:n]`.
5. Again, use the first n data points, provide a 90% confidence interval for μ . Note that with $n > 10$ observations in your hand (which you have paid for $\$12$ each), you can get a more accurate estimate of the population variance. Remember to use the n data points to calculate a new sample mean and a new standard error. Report the 90% two-sided confidence interval, the lower bound in (5a) and the upper bound in (5b).

To find the percentile point of a standard normal distribution, do not use the SOA normal table or Table 4 in the text book. Instead, you can use command `z=qnorm(0.995)`, `z=qnorm(0.975)`, `z=qnorm(0.95)`, etc, to find the percentile point. This request is to help us grade your answer numerically. For example, `qnorm(0.975)` gives 1.959964, which corresponds to 1.96 in the normal tables. If you use 1.96 instead of 1.959964, your answer may be mistakenly graded as incorrect. Type in `?qnorm` in R for more information on the function `qnorm()`.

Computing Homework 3

Each of you receive 200 numbers, denoted as X_1, \dots, X_n where $n=200$. It is known that $X_i \sim \text{Unif}(0, \theta)$ for θ unknown.

1. report the maximum of the 200 numbers.
2. slightly adjust the maximum of the 200 numbers so that it becomes an unbiased estimator for θ , then report the realized value based on the data.
3. report the mean of the 200 numbers.
4. slightly adjust the mean of the 200 numbers so that it becomes an unbiased estimator for θ , then report the realized value based on the data.
5. use the pivotal method to find a 95% confidence interval for θ . The pivotal quantity is based on the maximum of the 200 numbers. See Ex. 8.43. Then report the confidence lower and upper limits as the answers to (5a) and (5b) in the Google Form.

Computing Homework 2

Each of you receive 1000 numbers, denoted as X_1, \dots, X_n where $n=1000$

1. calculate the sum of the squares: $\sum_{i=1}^n X_i^2$
2. calculate $\sum_{i=1}^n (X_i - \bar{X})^2$, where \bar{X} is the sample mean.
3. calculate $\sum_{i=1}^n X_i^2 - n(\bar{X})^2$
4. use R command `var(x)` to calculate the following: $\text{var}(x) * (n - 1)$, where x is the vector of your sample (the 1000 numbers in the data file).
5. Suppose X_1, \dots, X_n follow some distribution with mean μ .
 - I. Based on your sample, give an unbiased estimate of μ
 - II. **(Bonus points)** Based on your sample, provide the standard error of the estimator.

Here are some R codes which might help you

Please install R before the beginning of the semester. In addition to R, some may find RStudio to be handy. Downloads:

- R - mirror hosted at UC Berkeley. For Windows machines, use the “base” binaries for the time being.
- R Studio - a more user friendly platform for R.

```
####Set the R working directory###
setwd("C:/")
###read in the data file, make sure your data file is under the working directory##
dat = read.csv('data_2.csv',header=FALSE)
### the "dat" you just read into R is a data frame, need to convert it into a matrix
dat <- as.matrix(dat) ##dat now is a 1000x1 matrix
x <- dat[1,] ##take the first row of this matrix as your sample
```

—At this point, you have loaded the data into your R program, whose name is “x”. You can start manipulating the data—

```
###Assign a value to a variable "n"
n <- 1000
###Find the length of a vector "x"
n <- length(x)
#### find the average of all elements in vector "x" and assign this value to "xbar" #####
xbar <- mean(x)
#### find the summation of all data points in vector "x" and assign this value to "Sx"
Sx <- sum(x)
#### define a new vector "y" whose elements are square of all elements in "x"
y <- x^2
#### subtract a number "z" from a vector x and define this new vector as "xsZ"
z <- 10
xsZ <- x-z
####Compute the sample variance of data points in vector "x"
var(x)
####multiply "*"
3*2
```

—You should read this book for more detailed explanations An Introduction to R (Chapters 2.1,2.2,2.3 are

most relevant to this homework)

From:

<http://www2.math.binghamton.edu/> - **Binghamton University Department of Mathematical Sciences**

Permanent link:

http://www2.math.binghamton.edu/p/people/qiao/teach/448/448_cp

Last update: **2016/01/24 23:45**

