Math 152 - Statistical Theory - Homework 1

write your name here

Due: 9/7/2018

General notes on homework assignments (also see syllabus for policies and suggestions):

- please be neat and organized which will help me, the grader, and you (in the future) to follow your work.
- be sure to include your name, and staple pages together *prior* to turning in the assignment
- please include at least the number of the problem, or a summary of the question (which will also be helpful to you in the future to prepare for exams).
- it is strongly recommended that you start the questions as soon as you get the assignment. This will help you to start thinking how to solve them!
- for R problems, it is required to use R Markdown (or R Sweave)
- in case of questions, or if you get stuck please don't hesitate to email me (though I'm much less sympathetic to such questions if I receive emails within 24 hours of the due date for the assignment).

Homework assignments will be graded out of 5 points, which are based on a combination of accuracy and effort. Below are rough guidelines for grading.

Score & Description

5 All problems completed with detailed solutions provided and 75% or more of the problems are fully correct.

4 All problems completed with detailed solutions and 50-75% correct; OR close to all problems completed and 75%-100% correct

3 Close to all problems completed with less than 75% correct

2 More than half but fewer than all problems completed and $> 75\%$ correct

1 More than half but fewer than all problems completed and $< 75\%$ correct; OR less than half of problems completed

0 No work submitted, OR half or less than half of the problems submitted and without any detail/work shown to explain the solutions.

```
You must first install both R and R Studio onto your computer (or use the Pomona server:  https://rstudi
* http://pages.pomona.edu/~jsh04747/courses/RTutorial.pdf
* A Student's Guide to R; Horton, Pruim, Kaplan (click on "Raw" to download)
https://github.com/hardin47/LittleBooks/blob/master/StudentGuide/MOSAIC-StudentGuide.pdf
* DataCamp free R tutorial https://www.datacamp.com/courses/free-introduction-to-r
```

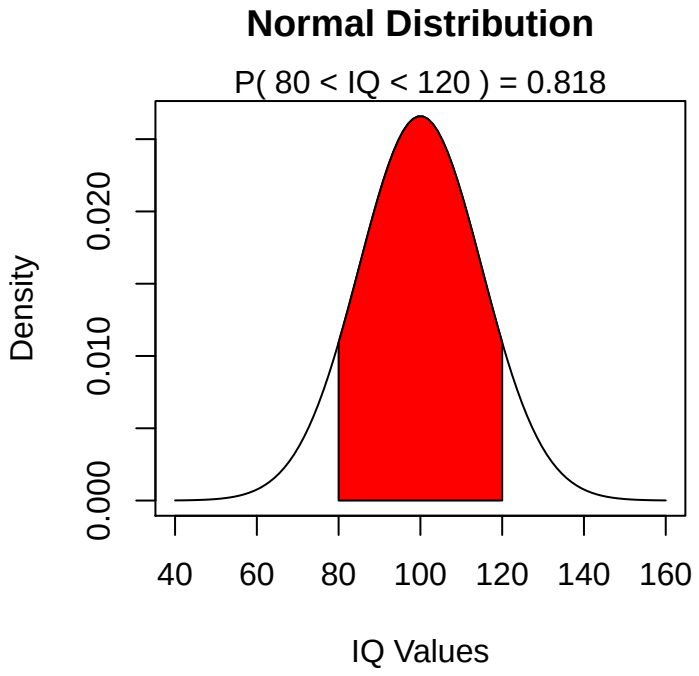
Click on the *Knit PDF* icon at the top of R Studio to run the R code and create the PDF document simultaneously. The swirl tutorial (see below) is fun:

```
install.packages("swirl")
require(swirl)
swirl()
```

Problem # 1 (sample R code provided, use help().)

Children's IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. One might, for example, be interested in the proportion of children having an IQ between 80 and 120.

```
mean=100
sd=15
lower_bound=80
upper_bound=120

x <- seq(-4,4,length=100)*sd + mean
x_pdf <- dnorm(x, mean, sd)

plot(x, x_pdf, type="n", xlab="IQ Values", ylab="Density",
  main="Normal Distribution")

i <- x >= lower_bound & x <= upper_bound
lines(x, x_pdf)
polygon(c(lower_bound,x[i],upper_bound), c(0,x_pdf[i],0), col="red")

# calculate the area under the curve
area <- pnorm(upper_bound, mean, sd) - pnorm(lower_bound, mean, sd)

# create a probability statement as a text string
result <- paste("P(",lower_bound,"< IQ <",upper_bound,") =",
   signif(area, digits=3))

# add the probability statement to the graph (at the top)
mtext(result,3)
```



a. The R function **dnorm** gives the height of the density (i.e., f(x), the pdf). What is a density?

b. The R function **pnorm** gives the area associated with the density. What is the area associated with the density?

c. Look up the R function **qnorm**. What does **qnorm** provide?

d. Look up the R function **rnorm**. What does **rnorm** provide?

Problem # 2 (simulating data)

Consider the beta distribution (google it if you are unfamiliar with it).

a. Provide the pdf for any *symmetric* beta distribution (specify a and b which should not equal to 1).

- What are the theoretical mean, variance, and standard deviation of your beta distribution? (not found using R)
- Using the R function **rbeta**, generate 100 random beta variates, and make a histogram of the random sample.
- Find the sample mean, sample variance, and sample standard deviation of your 100 random deviates.
- plot the density (pdf); try **help(rbeta)**.

b. Provide the pdf for any non-symmetric beta distribution (specify a and b).

- What are the theoretical mean, variance, and standard deviation of the beta distribution? (not found using R)
- Using the R function rbeta, generate 100 random beta variates, and make a histogram.
- Find the sample mean, sample variance, and sample standard deviation of your 100 random deviates.
- plot the density (pdf)

Problem # 3 (Bayes probabilities, no R)

Suppose that the rate of infection with TB is 1 in 1000 (about 0.1 percent = 0.001). Suppose a TB test is used which is 90% accurate – it gives a positive result for 10 percent of people who do not actually have TB but do have a reaction to the skin test. Also, 10% of the people who actually have TB fail to react to the test.

a. What's the chance that someone has TB if they test positive?

b. What's the chance that a randomly chosen person is such that they test negative and have TB?

c. There is another TB test which gives fewer false positives, but it is more expensive. Would it be better to use that one? Explain.

d. What is the prior probability of having TB?

e. What is the posterior probability of having TB (given a positive test)?