BMEG 802 – Advanced Biomedical Experimental Design and Analysis

Assignment 1

Joshua G. A. Cashaback, PhD

Question 2

```r
A <- c(4,7,2,4,8,6,5,1,0,9,10,3, 8,9,5,2,7,3,2,5,7)
summary(A)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.000   5.000   5.095   7.000  10.000
```

```r
range(A)
```

```
## [1]  0 10
```

```r
sd(A) # sample
```

```
## [1] 2.879319
```

```r
sd(A)^2 # sample variance
```

```
## [1] 8.290476
```

```r
std <- function(x) sd(x)/sqrt(length(x)) # make a function
std(A) # sample standard error
```

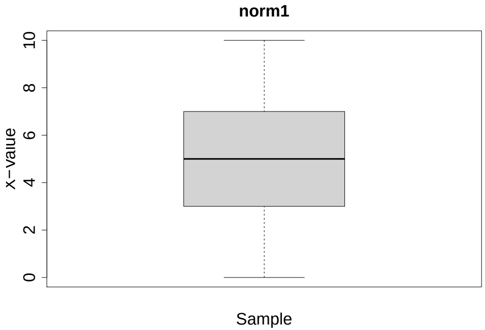
```
## [1] 0.6283189
```

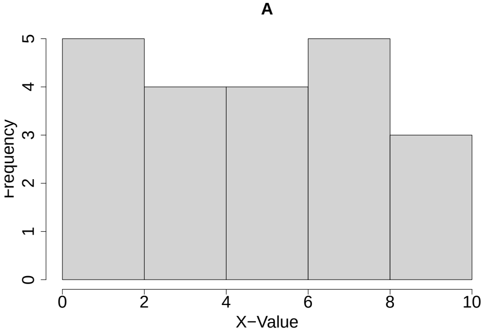
```r
sd(A) / mean(A) * 100
```

```
## [1] 56.50999
```

Question 3

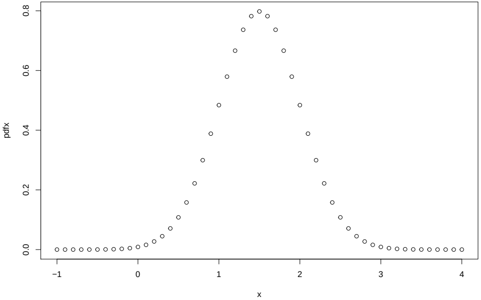
```
boxplot(A, main="norm1", xlab="Sample",
     ylab="X-Value", cex.lab=2.0, cex.axis=2.0, cex.main=2.0)
hist(A, main="A", xlab="X-Value", ylab="Frequency",
     cex.lab=2.0, cex.axis=2.0, cex.main=2.0)
```





Question 3

```r
x <- seq(from = -1.0, to = 4, by = 0.1)
mu = 1.5
sigma = 0.5
pdfx = (1/(sigma * sqrt(2 * pi)))*exp(-(1/2) * ((x - mu)/sigma)^2)
plot(x,pdfx)
```



Question 3

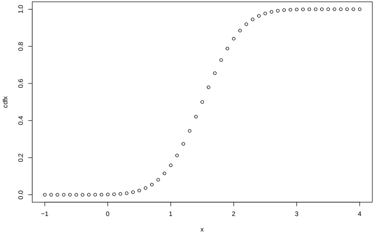
install.packages("NORMT3")

```r
library(NORMT3)
```

```
## NORMT3: Evaluates erf, erfc, Faddeeva functions and Gaussian/T sum dens
```

```
## Copyright: Guy Nason 2005-2012
```

```r
cdfx = 1/2*(1 + erf((x - mu)/(sigma*sqrt(2))))
plot(x,cdfx)
```



Question 3

```
# probability of x being less than 1.5
1/2*(1 + erf((1.5 - mu)/(sigma * sqrt(2))))
```

```
## [1] 0.5+0i
```

```
# probability of x being greater than 1
1 - 1/2*(1 + erf((1.0 - mu)/(sigma * sqrt(2))))
```

```
## [1] 0.8413447+0i
```

```
# probability of x being between 1.1 and 1.6
1/2*(1 + erf((1.6 - mu)/(sigma * sqrt(2)))) - 1/2*(1 + erf((1.1 - mu)/(sig
```

```
## [1] 0.3674043+0i
```

Question 4

Sample from $\mathcal{N}(10.0, 2.5^2)$ and estimate sample mean and SE

```
sample10 = rnorm(10,10,2.5)  # n = 10
mean(sample10) # mean
```

```
## [1] 10.39617
```

```
sd(sample10) / length(sample10)^(1/2) #SE
```

```
## [1] 0.8071385
```

```
(mean(sample10) - 11) / (sd(sample10) / length(sample10)^(1/2)) # t-score
```

```
## [1] -0.7481131
```

```
sample10000 = rnorm(10000,10,2.5) # n = 10000
mean(sample10000)
```

```
## [1] 10.00374
```

```
sd(sample10000) / length(sample10000)^(1/2)
```

```
## [1] 0.02487661
```

```
(mean(sample10000) - 11) / (sd(sample10000) / length(sample10000)^(1/2)) # t-score
```

```
## [1] -40.04791
```

```
# standard error decreases, we are more confident in our estimate of the mean.
# the t score increases as we become more confident with our mean
# IMPORTANCE: we can increase our sample size in an experiment!
```

Question 5

Lets sample 20 values from group 1 ($\mathcal{N}(10.0, 2.5^2)$) and 20 values for group 2 ($\mathcal{N}(12.0, 2.5^2)$).

```
sample20a = rnorm(20,10,2.5)
sample20b = rnorm(20,12,2.5)
mean(sample20b) - mean(sample20a) # mean difference
```

```
## [1] 2.558228
```

```
(sd(sample20b) ^ 2 / length(sample20b) + sd(sample20a) ^ 2 / length(sample20a))^(
```

```
## [1] 0.7265502
```

```
(mean(sample20b) - mean(sample20a) - 0) / (sd(sample20b) ^ 2 / length(sample20b) 
```

```
## [1] 3.521061
```

```
# the t-score is the normalized distance away that the group difference is from 0
```

Question 6

Disorder	Patients (n)	Treatment time ($\bar{x}$)	Std. Deviation (s)
Schizophrenia	18	4.7	9.3
Bipolar	10	8.8	11.5

see Primer for answer

Question 7

a. $p(A) = 1 - P(A') = 1 - 0.8 = 0.2$
b. $p(A) = 1 - 4/10 - 4/10 = 0.2$
c. $p(A \cap B) = p(B|A) \cdot p(A) = 4/10 \cdot 4/10 = 0.16$
d. $p(A \cap B) = p(B|A) \cdot p(A) = 3/9 \cdot 4/10 = 0.133$
e. $p(B|A) = p(A \cap B)/p(A) = .15/0.9 = 0.167$
f. $p(A \cup B) = p(A) + p(B) - p(B|A) \cdot p(A) = p(A) + p(B) - p(A \cap B) = 0.4 + 0.3 - 0.2 = 0.5$
g. $1 - 0.5 = 0.5$

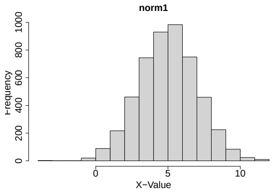
Question 8

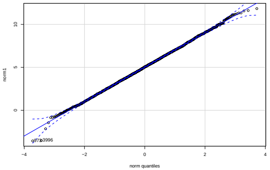
```
library("car")
```

```
## Loading required package: carData
```

```
norm1 <- rnorm(5000, mean=5, sd=2)
hist(norm1, main="norm1", xlab="X-Value", ylab="Frequency",
     cex.lab=2.0, cex.axis=2.0, cex.main=2.0)
qqPlot(norm1)
```

```
## [1]  973 3996
```





Question 8 cont'd

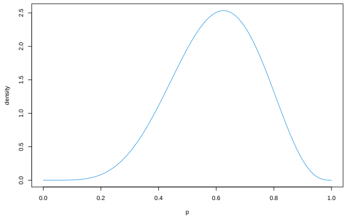
```r
library("car")
shapiro.test(norm1)
```

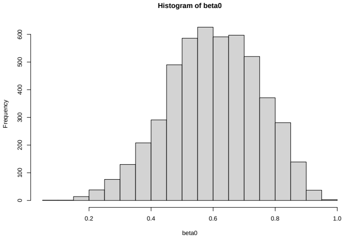
```
##
##  Shapiro-Wilk normality test
##
## data:  norm1
## W = 0.99962, p-value = 0.4607
```

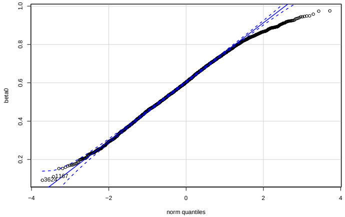
Question 9

```r
beta0 <- rbeta(5000, 6, 4)
p = seq(0,1, length=100)
plot(p, dbeta(p, 6, 4), ylab="density", type ="l", col=4)
hist(beta0)
qqPlot(beta0)
```

```
## [1] 3624 1167
```







Question 9 cont'd

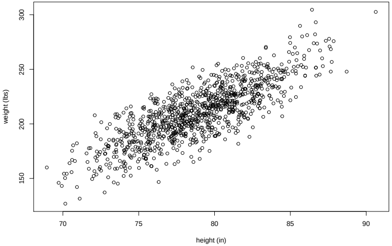
```r
shapiro.test(beta0)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  beta0
## W = 0.99432, p-value = 3.672e-13
```

Question 10 (Joint Probabilities)

install.packages("MASS")

```
library("MASS")
my_n1 <- 1000
my_mu1 <- c(78.8, 211)
my_Sigma1 <- matrix(c(3.668^2, 0.81*3.668*26.904, 0.81*3.668*26.904, 26.904
binorm <- mvrnorm(n = my_n1, mu = my_mu1, Sigma = my_Sigma1)
plot(binorm, xlab="height (in)", ylab="weight (lbs)")
```



Question 10 Cont'd (Marginal Probabilities)

$p(height) = \mathcal{N}\left(\mu_{height},\ \sigma^2_{height}\right)$ and $p(weight) = \mathcal{N}\left(\mu_{weight},\ \sigma^2_{weight}\right)$

```
mu_h = 78.8
sigma_h = 3.668
mu_w = 211
sigma_w = 26.904
# What is the probability a player is over 85 inches tall?
1 - 1/2*(1 + erf((85 - mu_h)/(sigma_h * sqrt(2))))
```

```
## [1] 0.04548582+0i
```

```
# What is the probability a player is under 190 lbs?
1/2*(1 + erf((190 - mu_w)/(sigma_w * sqrt(2))))
```

```
## [1] 0.2175327+0i
```

```
# What is the probability a player is between 200 and 220 lbs?
1/2*(1 + erf((220 - mu_w)/(sigma_w * sqrt(2)))) - 1/2*(1 + erf((200.0 - mu_w)/(sigma_w * sqrt(2))))
```

```
## [1] 0.2896867+0i
```

Question 10 Cont'd (Conditional Probabilities)

$$p(X \mid Y = y) = \mathcal{N}\left(\mu_X + \frac{\sigma_X}{\sigma_Y}\rho(y - \mu_Y),\ (1 - \rho^2)\sigma_X^2\right); X = height,\ Y = weight$$

```
# What is the probability that a player is under 75.5 inches given they are 200 lbs?
mu_h = 78.8
sigma_h = 3.668
mu_w = 211
sigma_w = 26.904
rho = 0.81

mu_c1 = mu_h + sigma_h / sigma_w * rho * (200 - mu_w)
sigma_c1 = ((1 - rho^2) * sigma_h^2)^(1/2) # SD
1/2*(1 + erf((75.5 - mu_c1)/(sigma_c1 * sqrt(2))))
```

```
## [1] 0.1661685+0i
```

and $p(Y \mid X = x) = \mathcal{N}\left(\mu_Y + \frac{\sigma_Y}{\sigma_X}\rho(x - \mu_X),\ (1 - \rho^2)\sigma_Y^2\right); X = height,\ Y = weight$

```
# What is the probability that a player is over 250 lbs given they are 86 inches?
mu_c2 = mu_w + sigma_w / sigma_h * rho * (86 - mu_h)
sigma_c2 = ((1 - rho^2) * sigma_w^2)^(1/2)
1 - 1/2*(1 + erf((250 - mu_c2)/(sigma_c2 * sqrt(2))))
```

```
## [1] 0.5945873+0i
```