THEORETICAL PORTION

PROBLEM #1.

P(X=1) & {P(X=2) | P(X=3) | P(X=4)} & {P(X=5) | P(X=6)}
E(.5yr) & {E(1yr.) | E(1yr.) | E(1yr.)} & {E(1.5yr) | E(1.5yr)}

for machine 1 - λ = .5 years	X~Exp(.5 yr)	$P(X<x)= 1-e^{-2x}$	$\lambda=2$
for machine 2,3,4 - λ = 1 years	X~Exp(1 yr)	$P(X<x)= 1-e^{-x}$	$\lambda=1$
for machine 5,6 - λ = 1.5 years	X~Exp(1.5 yr)	$P(X<x)= 1-e^{-(2x/3)}$	$\lambda=2/3$

When x = 2 years, the probability that the following machine fails is (by Exp):

Machine 1:	0.9817	Success is: 1-0.9817 = 0.0183
Machine 2,3,4:	0.8647	Success is: 1-0.8647 = 0.1353 (per machine)
Machine 5,6:	0.7364	Success is: 1-0.7364 = 0.2636 (per machine)

Using the binomial distribution for the “grouped” machines, 1 minus the probability that none of the machines work (per “group”) is: [the binomial coefficient in both cases = 1]

At least one machine from 2,3,4 works is: $1- (1 * 0.1353^0 * (0.8647)^3) = .3535$
At least one machine from 5,6 works is: $1- (1 * 0.2636^0 * (0.7364)^2) = .4577$

Thus the chances that the system works, with at least the 1st machine working, one machine from 2,3,4, and 1 machine from 5,6 working is: .0183 * .3535 * .4577 = 0.002961
=> **0.30% probability that the system will function uninterrupted for 2 years**.

PROBLEM #2.

a. $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ and then the variance would be $Var(\bar{X}) = E[(\bar{X}-\mu)^2]$ for σ^2, then divide by the total number of samples to obtain the variance of the entire sample.

b. $E(\bar{X}^2) = Var(\bar{X}) + E(\bar{X})^2$ = $\boldsymbol{\sigma^2/n + \mu^2}$

PROBLEM #3.

a. $E(X+Y) = E(X) + E(Y) = \mu + \theta$
   $Var(X+Y) = Var(X) + Var(Y) + 2CoVar(X,Y) = \sigma^2 + \gamma + 0$ (since X and Y are independent)
b. X+Y is normally distributed since the variables are also normally distributed.
c. 3X+20Y+8 is normally distributed with:
   $3X+20Y+8 = 3(E(X)) + 20(E(Y)) + 8 = 3\mu + 20\theta + 8$ (mean)
   $3^2(Var(X)) + 20^2(Var(Y)) = 9\sigma^2 + 400\gamma$ (variance)

PROBLEM #4.

$$E(X^3) = \int_a^b x^3 \frac{1}{b-a} dx = \frac{x^4}{4(b-a)}\Bigg|_a^b = \frac{b^4}{4b-4a} - \frac{a^4}{4b-4a}$$

PROBLEM #5.

a. Assuming normal distribution, the equation for a 98% CI is:
   For cumulative probability - $t_{.01}$
   degrees of freedom = n-1 = 21
   t-value for $t_{.01, 21} = 2.518$ (using t-table from http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf)

(continuation of Problem #5.a.)

The equation is then: **$\bar{x} \pm 2.518 * (s / (\sqrt{22}))$**

b. If the histogram appears to be exponentially distributed (skewed left), an improvement to reflect a more normally distributed histogram is to increase the sample size (n). Ideally increase n above 30.
Another method to improve the histogram is to use the mean from the sample data instead of the population mean. —> replace $\lambda = 1/\mu$ with $\lambda = 1/\bar{x}$ so it re-centers the data on the mean of the sample instead of the population
A third method to improve the skewed histogram is to resample the sample data (bootstrapping). It may have been that the sample data before was skewed due to different dependent variables such as time since recovery, weather (I've heard barometric pressure can effect artificial body components, i.e. metals), activity levels, or even medication. As the sample data is resampled, even with an exponentially distributed histogram, the data will then reflect a normal distribution the more bootstrap resamples are completed.

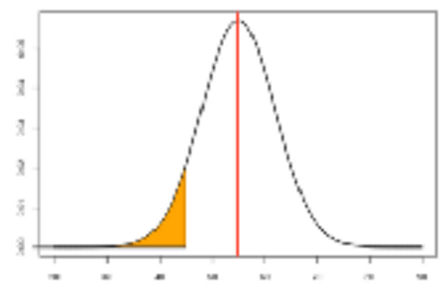
PROBLEM #6.

$\Phi(0.385) = 0.65$, $\Phi(0.428) = 0.67$, $\Phi(0.71) = 0.762$, $\Phi(1.43) = 0.924$, $\Phi(1.92) = 0.973$
$\Phi^{-1}(0.972) = 1.917$, $\Phi^{-1}(0.740) = 0.643$, $\Phi^{-1}(0.95) = 1.645$
$z_{0.0017} = 2.923$, $z_{0.016} = 2.14$, $z_{0.027} = 1.93$

a. $P(X \le 45)$

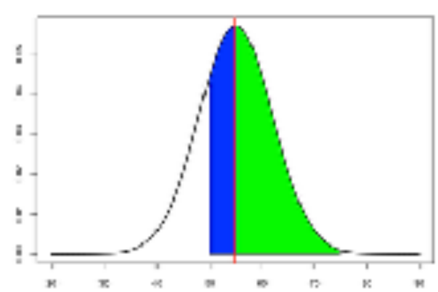
$= P\left(\frac{x-\mu}{\sigma} \le \frac{45-55}{7}\right)$
$= P(Z \le -1.4286)$
$= 0.5 - P(0 \le Z \le 1.4286)$
$= 0.5 - (0.924 - 0.5)$ [from values above]
$= 0.5 - .424$
$= 0.076$ {also could have done $1 - \Phi(1.43)$)

The probability that the paint will dry in 45 minutes or less is **7.6%**



b. $P(50 < X < 70)$

$= P\left(\frac{50-55}{7} \le Z \le \frac{70-55}{7}\right)$
$= P(-0.7143 \le Z \le 2.1429)$
$= P(-0.7143 \le Z \le 0) + (0 \le Z \le 2.1429)$
= Using $\Phi(0.71) - 0.5 = 0.762 - 0.5 = 0.262$ (for the left side)
= Using $z_{0.016} = 2.14$, $[(1-(2*0.016))/2] = 0.484$ (for the right side)
= Summed together => **0.746** (computed in R gives 0.760)



c. $P(X < x_1) = 35\%$?

$$P\left(Z < \frac{x_1 - 55}{7}\right) = 0.35$$

$Z_1 = \frac{x_1 - 55}{7} = .385$ (from $\Phi(0.385) = 0.65$)

Solving for $x_1 = 57.695$ (this is on the right side of the mean, noticed from 0.65 percentile above. So we can simply invert this to show $57.695 - 55 = 2.695$ and $55 - 2.695 =$ **52.305**
Thus, **52.3 minutes** represents the 35th percentile in drying paint.

PROBLEM #7.

a. The mean of X_1 is 0 while the mean of X_2 is also 0.
b. The variance of X_1 is 0.1 while the variance of X_2 is also 0.1.
c. The correlation between X_1 and X_2 does not exist since both variables previously are independent of one another.
(w is scalar, w+E(e1) [=0] , e1 and e2 are independent since x1 and x2 are independent)

PROBLEM #8.

a. The value of c that is needed so that $f(x)$ is a pdf, is 2. $\int_0^2 \frac{x^3}{4}dx = \frac{x^4}{16}\Big|_0^2 = \frac{16}{16} - 0 = 1$
b. $F(x) = \int_0^x \frac{t^3}{4}dt = \frac{t^4}{16}\Big|_0^x = \frac{x^4}{16}$
c. $F(x) = \frac{x^4}{16}$ $U = \frac{t^4}{16}$ $16U = x^4$ $x = \sqrt[4]{16U}$

COMPUTATIONAL PORTION

Code and graphs included with each part, not indexed.

PROBLEM #1 - COMPUTATIONAL

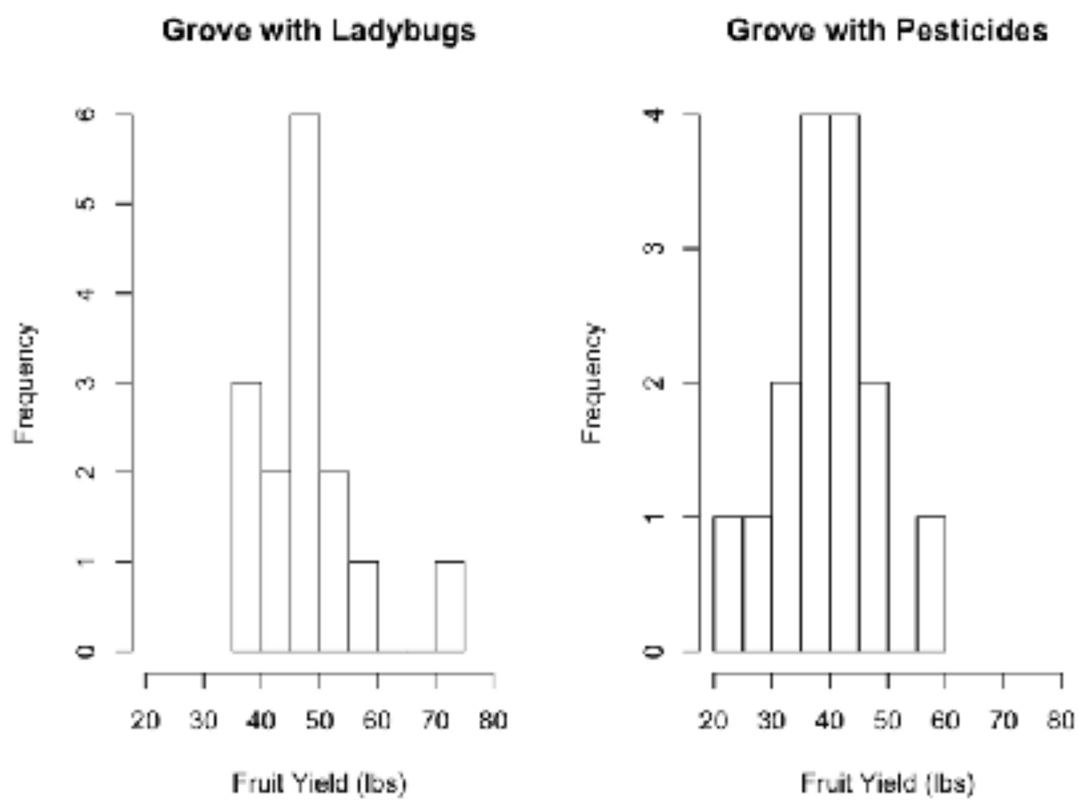
Read in of data prior to work - code: `groveData <- read.table("TreeData.txt", header=T)`

#1, part *a.*

code:

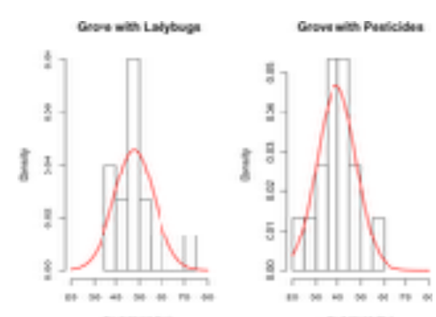
```
par(mfrow=c(1,2))
hist(groveData[,1], main ="Grove with Ladybugs", xlab = "Fruit Yield (lbs)", xlim=c(20,80))
hist(groveData[,2], main ="Grove with Pesticides", xlab = "Fruit Yield (lbs)", xlim=c(20,80))
```

output:



#1, part *b.*

The ladybugs histogram closely resembles a uniform distribution with a slight right skew, though the 1 value between 70-75 throws it off slightly. The pesticides histogram more closely resembles a normal distribution with it's frequencies though it's not perfectly a resemblance. I believe with more sampling, they would more closely resemble a normal distribution. The density histograms with the normal distribution is reflected on the right.



#1, part *c.*

code:

```
mean(groveData[,1]); mean(groveData[,2])  #1-ladybugs #2-pesticides
```

output:

```
47.77666 #ladybugs
39.27545 #pesticides
```

#1, part *d.*

code:

```
alpha = .05
x1=groveData[,1]
x2=groveData[,2]
CIL = c(mean(x1)-qnorm(1-alpha/2)*sd(x1)/sqrt(length(x1)), mean(x1)+qnorm(1-alpha/2)*sd(x1)/sqrt(length(x1)))
CIP = c(mean(x2)-qnorm(1-alpha/2)*sd(x2)/sqrt(length(x2)), mean(x2)+qnorm(1-alpha/2)*sd(x2)/sqrt(length(x2)))
CIL
CIP
```

output:

```
> CIL
[1] 43.36794  52.18539     #ladybugs
> CIP
[1] 34.96814  43.58276      #pesticides
```

#1, part *e.*

Ladybugs:

With 95% confidence, the true mean of the fruit yield (in pounds) is between 43.368 and 52.185.

Pesticides:

With 95% confidence, the true mean of the fruit yield (in pounds) is between 34.968 and 43.583.

#1, part *f.*

code:

```
var(groveData[,1])/(15)      #Ladybugs
var(groveData[,2])/(15)      #Pesticides
```

output:

```
> var(groveData[,1])/(15)
        [1] 5.05976          #Ladybugs
> var(groveData[,2])/(15)
        [1] 4.82965          #Pesticides
```

#1, part g.

code:

```
CCIL=c(14*(sd(groveData[,1])^2)/26.11894805, (14*(sd(groveData[,1])^2)/5.62872610))
CCIP=c(14*(sd(groveData[,2])^2)/26.11894805, (14*(sd(groveData[,2])^2)/5.62872610))
CCIL     #variance for Ladybugs
CCIP     #variance for Pesticides
```

output:

```
40.68118   188.77266      #variance for Ladybugs
38.83106   180.18758      #variance for Pesticides
```

PROBLEM #2 - COMPUTATIONAL

#2, part a.

code:

```
means.of.expSamples=NULL
expSample=NULL

for(i in 1:500){
   if(i==1){
          expSample = rexp(n=300, 2)
   }else{
          y=mean(rexp(n=300, 2))
          means.of.expSamples[i-1] = y
   }
}

mean(expSample); sd(expSample)
mean(means.of.expSamples); sd(means.of.expSamples)
```

output:

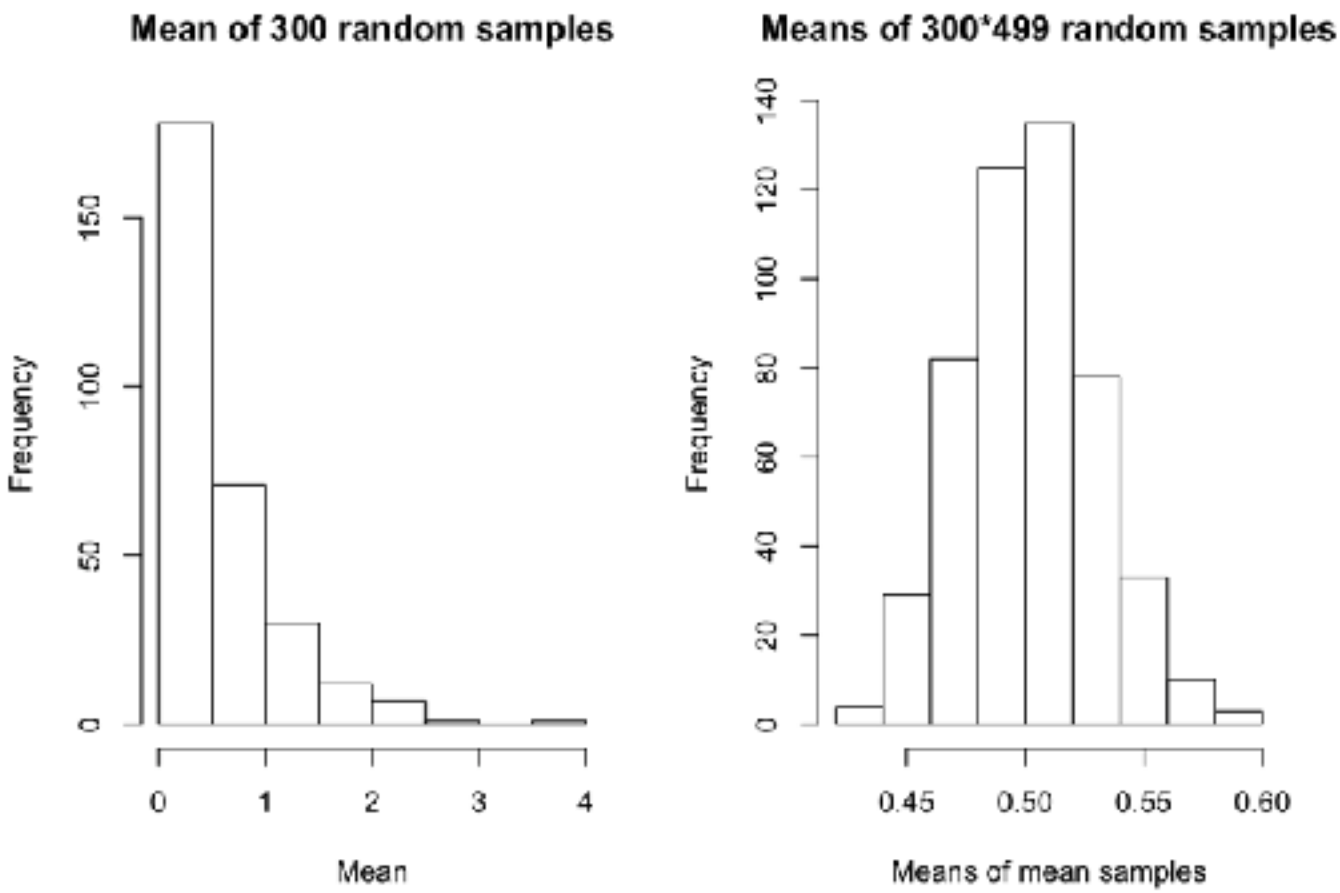
```
> mean(expSample); sd(expSample)
        [1] 0.56464       #mean
        [1] 0.5505958     #sd
> mean(means.of.expSamples); sd(means.of.expSamples)
        [1] 0.5014331     #mean
        [1] 0.02843841    #sd
```

#2, part b.

code:

```
hist(expSample,xlab="Mean",main="Mean of 300 random samples")
hist(means.of.expSamples, xlab="Means of mean samples", main="Means of 300*499
    random samples")
```

output:



As more means are created of the 300 means, it's found that the normal distribution is created to favor the mean of the exponential distribution (the first histogram).

PROBLEM #3 - COMPUTATIONAL

#3, part a.

Y=5X+2, the mean is 5(3) + 2 = 17, while the variance is $5^2(2) = 100$. Since Y comes from X (which is normally distributed), Y is also normally distributed. Y is also a linear combination of X.

#3, part b.

code:

```
pnorm(10,3,2)
pnorm(10,17,10)
```

output:

```
> pnorm(10,3,2)
    [1] 0.9997674    #P(X<10)
> pnorm(10,17,10)
    [1] 0.2419637    #P(Y<10)
```

#3, part c.

code:

```
qnorm(.67,17,10)
```

output:

```
> qnorm(.67,17,10)
[1] 21.39913     #marks the 67th percentile (for Y)
```

#3, part d.

code:

```
-.674*(2)+3
```
(used as calculator)

output:

```
> -.674*(2)+3
    [1] 1.652
```

PROBLEM #4 - COMPUTATIONAL

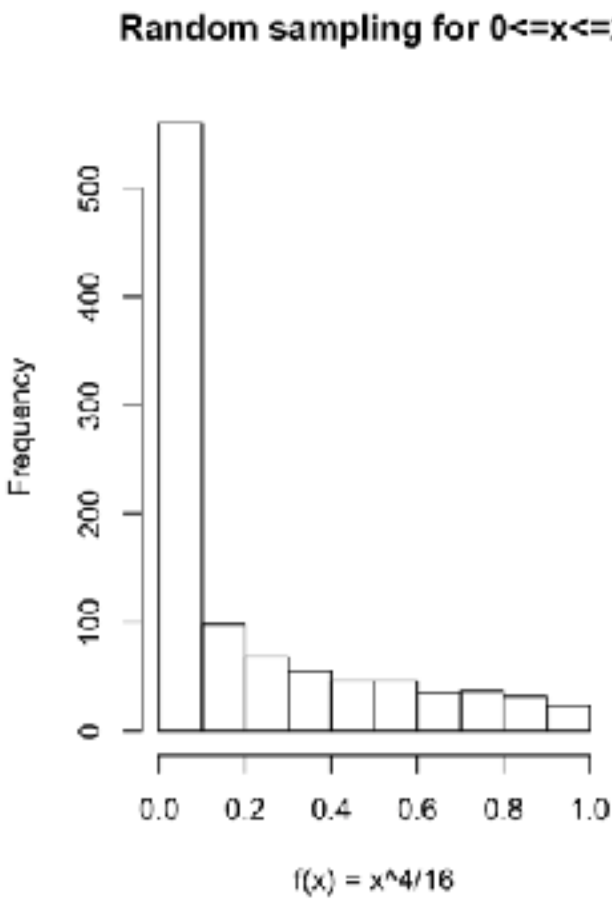
#4, part *a.*

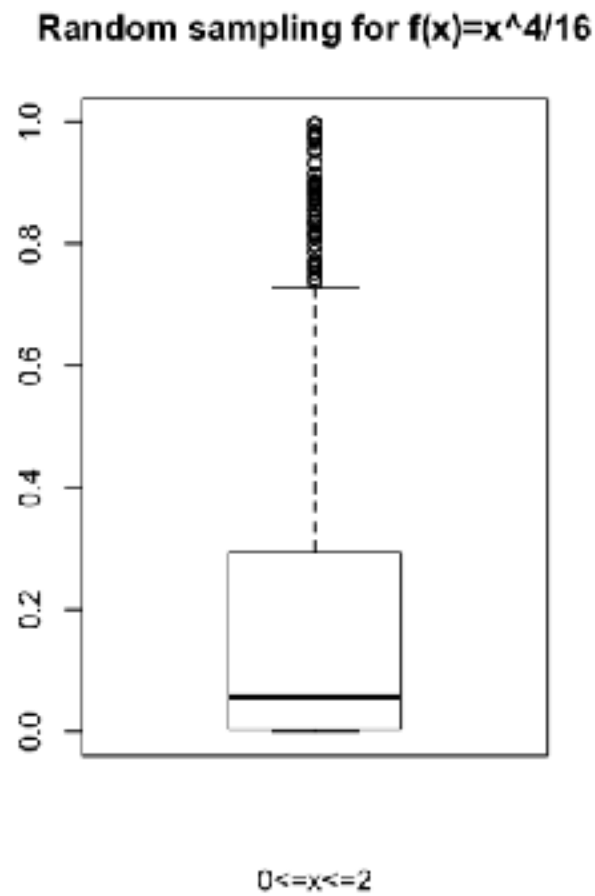
code:

```
numero8=function(u){
            z=(u^4)/16
    return(z)
}

hist(numero8(runif(1000,0,2)),main="Random sampling for 0<=x<=2", xlab="f(x) = x^4/16")
boxplot(numero8(runif(1000,0,2)),main="Random sampling for f(x)=x^4/16", xlab="0<=x<=2")
```

output:





The distribution looks to be exponential, which seems likely since the median of the sampling is 1. Thus for the equation, it would result in about 0.0625 for most of the samples (assuming the random values for x have the same probability of being generated). Even with more samples, this would remain to appear as a exponential distribution (continuous).