```
library(ggplot2)

# problem 1 - Norm dist vs t-dist as ref dist
set.seed(1908)
?rexp
x2 <- rexp(10,rate=1)
x2

# 1.a - picture Exp dist
# x axis values
x <- seq(from=0,to=10,by=0.01)
x
# y value from PDF of Exp dist
y <- dexp(x,rate=1)
y
# plot
qplot(x,y,geom="point")
qplot(x,y,geom="area")
qplot(x,y,geom="line")

# 1.b - t-test mu=2
t.test(x2,mu=2,conf.level = 0.95, alternative = "t")
# One Sample t-test
#
# data:  x2
# t = -1.6165, df = 9, p-value = 0.1404
# alternative hypothesis: true mean is not equal to 2
# 95 percent confidence interval:
#   0.7924928 2.2010150
```

```
# sample estimates:
#   mean of x
# 1.496754

# There is insufficient evidence to support rejecting the null
# hypothesis that the population mean is equal to 2 (t-test,
# p-val = 0.14).
# We estimate the mean to be 1.497, with a 95% confidence
# interval of (0.79, 2.2).

# 1.c - z-stat
Z <- (mean(x2) - 2)/(sd(x2)/sqrt(length(x2)))
Z
# [1] -1.616477
P <- 2 * pnorm(abs(Z), mean = 0, sd = 1, lower.tail = FALSE)
P
# [1] 0.1059913

# 1.d - conceptual
# The p-value differs because the approximation to the Normal
# distribution underestimates the likelihood of large, extreme
# values. This is due to the small sample size and the
# approximation of the population standard deviation as the
# sample standard deviation.
#
# The t-distribution is more appropriate for real life
# examples, where the population sigma is unknown, so long
# as sample sizes are reasonably large.
```

```
# problem 2 - Norm approx vs exact test

?rbinom

x <- sum(rbinom(n=20,size=1,prob=0.25))

x

# 8


# 2.a - 2 tests of H0

# prop.test

A <- prop.test(x=x,n=20,p=0.25,conf.level = 0.95,correct = FALSE)

A

# 1-sample proportions test without continuity correction

#

# data:  x out of 20, null probability 0.25

# X-squared = 2.4, df = 1, p-value = 0.1213

# alternative hypothesis: true p is not equal to 0.25

# 95 percent confidence interval:

#   0.2188065 0.6134185

# sample estimates:

#   p

# 0.4

# Z 1.55


# binom.test

E <- binom.test(x=x,n=20,p=0.25,conf.level = 0.95)

E

# Exact binomial test

#
```

```
# data:  x and 20
# number of successes = 8, number of trials = 20, p-value
# = 0.1261
# alternative hypothesis: true probability of success is not equal to 0.25
# 95 percent confidence interval:
#   0.1911901 0.6394574
# sample estimates:
#   probability of success
# 0.4

# 2.b - check confidence intervals
str(A)
(0.25 > A$conf.int[1]) & (0.25 < A$conf.int[2])
(0.25 > E$conf.int[1]) & (0.25 < E$conf.int[2])

# 2.c - function approx() and exact()
approx <- function(n){
  x <- sum(rbinom(n=n,size=1,prob=0.25))
  A <- prop.test(x=x,n=n,p=0.25,conf.level = 0.95,correct = FALSE)
  return((0.25 > A$conf.int[1]) & (0.25 < A$conf.int[2]))
}

exact <- function(n){
  x <- sum(rbinom(n=n,size=1,prob=0.25))
  E <- binom.test(x=x,n=n,p=0.25,conf.level = 0.95)
  return((0.25 > E$conf.int[1]) & (0.25 < E$conf.int[2]))
}

# 2.d - replicate() 10k
```

```
?replicate

a_prop=sum(replicate(10000,approx(20)))/10000

e_prop=sum(replicate(10000,exact(20)))/10000

# a_prop = 0.9342 < e_prop = 0.964

# in this case, the result from exact is within the

# confidence interval over 96% of the time, while

# the approx result is less than 94% of the time.


# 2.e - replicate() 100

a_prop_100=sum(replicate(100,approx(20)))/100

e_prop_100=sum(replicate(100,exact(20)))/100

# with fewer replications, the results are further apart:

# exact: 0.98

# approx: 0.92
```