HW 5

Problem 1

(1)

```
> shapiro.test(mtcars$mpg) # p-value > 0.05; mpg is normally distributed

        Shapiro-Wilk normality test

data:  mtcars$mpg
W = 0.94756, p-value = 0.1229
```

Since p-value of Shapiro test is greater than 0.05, we fail to reject the null hypothesis that mpg is normally distributed. Namely, we can say mpg is normally distributed.

(2)

Before using t-test, we should proceed var.test.

```
> var.test(mpg~am_factor, data=mtcars)$p.value # 0.0669052 -> equal variance
[1] 0.06690592
```

Since p-value is greater 0.05, so mpg with am=0 and mpg with am=1 have the same variance.

```
> two <- t.test(mpg~am_factor, data=mtcars,
+               alternative="two.sided", var.equal=TRUE)
> two

        Two Sample t-test

data:  mpg by am_factor
t = -4.1061, df = 30, p-value = 0.000285
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.84837  -3.64151
sample estimates:
mean in group 0 mean in group 1
       17.14737        24.39231

> less <- t.test(mpg~am_factor, data=mtcars,
+                alternative="less", var.equal=TRUE)
> less

        Two Sample t-test

data:  mpg by am_factor
t = -4.1061, df = 30, p-value = 0.0001425
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -4.250255
sample estimates:
mean in group 0 mean in group 1
       17.14737        24.39231
```

First, I use t.test with alternative “two.sided” to get rough view on the difference of two variables (mpg with am=0 and mpg with am=1). Through t-test, the p-value 0.000285 indicates that difference of two variables are significantly different from 0. (i.e. mean of each variables are different). Since the statistic of second variable is larger than that of first variable, so we use one-tail test here with alternative “less”. As seen in above, the p-value is much smaller than 0.05, hence I will accept the alternative hypothesis that difference of two variables are smaller than 0.

Problem 2

```
> # Problem 2
> data <- NOxEmissions$LNOx
> m <- mean(data); s <- sd(data);
> interval <- qnorm(0.975)/sqrt(length(data))
> left <- m - interval
> right <- m + interval
> cat("The 95% confidence interval is [",left,", ",right,"]")
The 95% confidence interval is [ 4.356897 ,  4.400484 ]
```