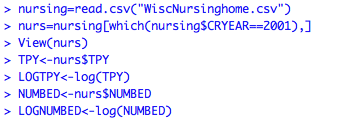
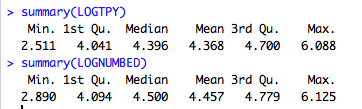
Homework 2

Yubing Li

00808366

**1. Nursing home Utilization:**

**(a)**



For natural log of total patients years in 2001, with “patients” as units:

The minimum is 2.511.

The 1st quantile is 4.041.

The median is 4.396.

The 3rd quantile is 4.700.

The maximum is 6.088.

The mean is 4.368.

For natural log of the number of beds in 2001, with “beds” as units:

The minimum is 2.890.

The 1st quantile is 4.094.

The median is 4.500.

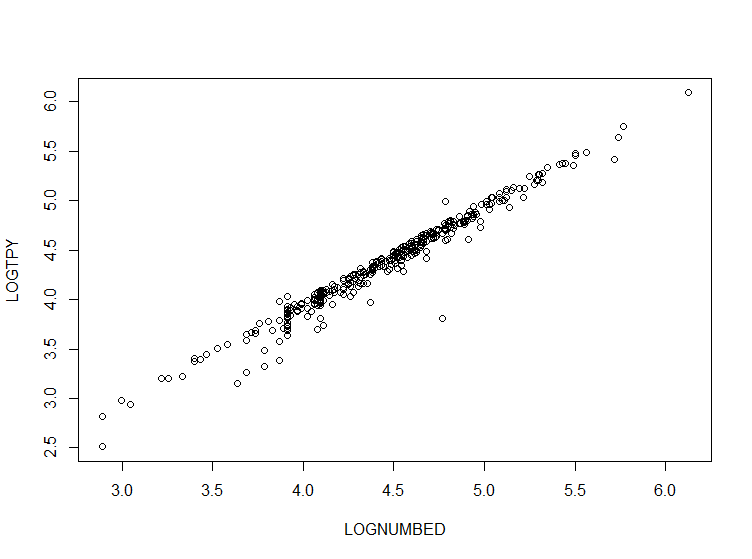
The 3rd quantile is 4.779.

The maximum is 6.125.

The mean is 4.457.

Scatter Plot:

In the scatter plot, the LOGNUMBEDs are the explanatory variables and LOGTPYs are the response variables.



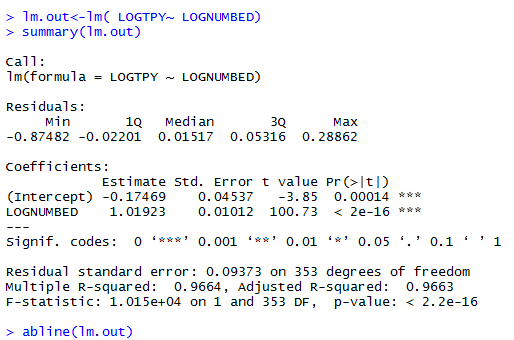


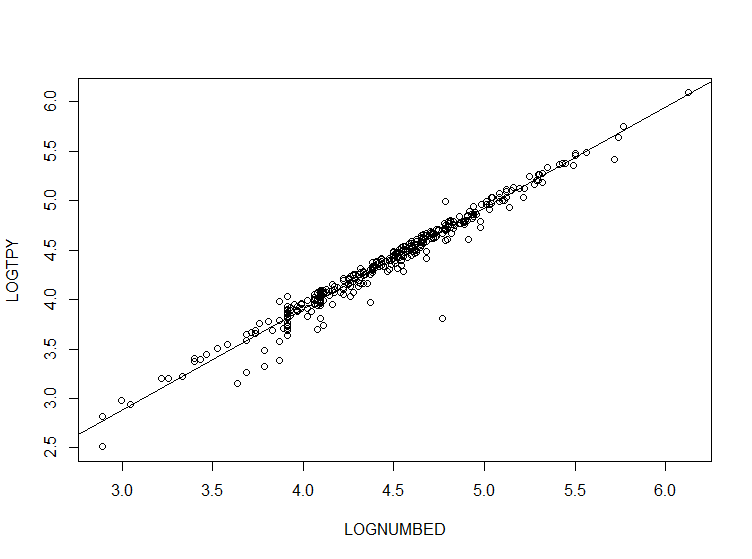
Correlation coefficient for these two variables is:

Correlation: r = 0.98305 R2 = r2 = 0.983052 = 0.9664

The r value is very close to 1, which indicates a strong linear association between LOGTPY and LOGNUMBED in 2001. It also indicates that 96.64% of the variability in LOGTPY is accounted for by variation in LOGNUMBED. The scatter plot also suggests that the two variables have strong positive linear association. In general, the association between LOGTPY and LOGNUMBED is strong, positive linear and there’s no unusual features.

**(b) Basic linear model:**





The coefficient of determination, R2 is

R2 = 0.9664

which indicates that 96.64% of the variability in the LOGNUMBED is accounted for by variation in LOGTPY.

The least squares regression line in terms of LOGNUMBED and LOGTPY is

= -0.17469 + 1.01923 (LOGNUMBED)

The regression coefficient for LOGNUMBED is 1.01923

which indicates that for every unit to LOGNUMBED increase, LOGTPY will increase by 1.01923 units.

The corresponding t-statistic is 100.73.

**(c) Hypothesis testing:**

(i) b0 = 0 vs b1 ≠ 0:



Test stat. = (1.01923-0)/0.01012 = 100.7144

p-value = <2e-16

Since p-value is very small (<2e-16 < α=0.05), we reject the null hypothesis b0 = 0 and conclude that there is sufficient evidence that the slop is different from 0.

(ii) b0 = 1 vs b1 ≠ 1:



Test stat. = (1.01923-1)/0.01012 = 1.9002

p-value = 0.05822

Since p-value is large (0.05822 > α=0.05), we fail to reject the null hypothesis b0 = 1 and conclude that there is insufficient evidence that the slop is different from 1.

(iii) b0 = 1 vs b1 >1:



Test stat. = (1.01923-1)/0.01012 = 1.9002

p-value = 0.02911

Since p-value is small (0.02911 < α=0.05), we reject the null hypothesis b0 =1 comparing to the alternative b1 >1.

(iv) b0 = 1 vs b1 < 1:



Test stat. = (1.01923-1)/0.01012 = 1.9002

p-value = 0.97089

Since p-value is very large (0.97089 > α=0.05), we fail to reject the null hypothesis b0 =1 comparing to the alternative b1 <1.

**(d) Confidence interval:**

(i) For marginal change in LOGNUMBED of 2, the point est. of expected change in LOGTPY is

change in = 2×1.01923 = 2.03846

(ii) 95% CI

CI for b1

t353, 1-0.025 = 1.966707

1.0192 ± t353, 1-0.025 (0.01012) = [0.9993, 1.0391]

CI for change of LOGTPY: [0.9993×2, 1.0391×2] = [1.9986, 2.0782]

(iii) 99% CI

t353, 1-0.005 = 2.589828

1.0192 ± t353, 1-0.005 (0.01012) = [0.9930, 1.0454]

99% CI is [0.9930×2, 1.0454×2] = [1.9860, 2.0908]

**(e) x\* = 100:**

(i) Predicated value is = -0.17469 + 1.01923 (ln100) = 4.5190

(ii) Standard error = 0.09373×√1+1/355+ (4.5190-4.457)2/(354\*0.2424) = 0.0938

(iii) 95% PI

4.5190 ± (1.966707) × (0.0938)

95% PI is [4.3344, 4.7034]

(iv) point prediction of TPY is e4.5190 = 91.7438

PI is [e4.3344 , e4.7034] = [76.2792, 110.3216]

(v) 90% PI

4.5190± (1.6492) × (0.0938)

90% PI is [4.3643, 4.6737]

PI of TPY is [e4.3643, e4.6737] = [78.5944, 107.0933]

**2. Initial Public Offerings**

**(a) Return & Revenue:**

(i) The least squares regression:

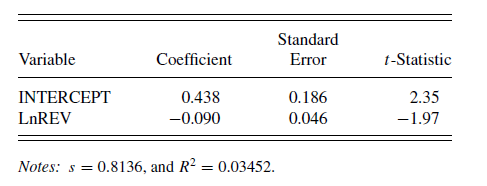
b1 = r×[Std(RETURN)/Std(REVENUE)] = -0.0175×(0.824/261.881) = -0.0000551

b0 = – b1 () = 0.106 – (-0.0000551) ×134.487 = 0.11341

In general, the LSR regression is = 0.11341 - 0.0000551× (REVENUE)

(ii) The fitted return is = 0.11341 - 0.0000551× (95.55) = 0.10815 with unit of return.

**(b) Return & LnRevenue:**



(i) New regression model is

= 0.438 - 0.090× (LnREVENUE)

The fitted return is = 0.438 - 0.090× (95.55) =-8.1615, is different from the one in a(ii).

(ii) According to the linear model, R2 = 3.42% of the variability in the RETURN is accounted for by variation in LnREVENUE.

Hypothesis testing:

H0: b1 = 0 vs.

We test Hypo: H1: b1 ≠ 0