Stat 461 HW1

Roshan Poduval

rpoduval@wisc.edu

1) $\underline{\mathbb{E}[X]} = \sum x\, p_X(x) = 1 \cdot p_X(1) + 0 \cdot p_X(0) = 1 \cdot p + 0(1-p) = p$ ✓

$\underline{Var[X]} = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

$\mathbb{E}[X^2] = \sum x^2 p_X(x) = 1 \cdot p_X(1) + 0\, p_X(0) = p$

$Var[X] = p - p^2 = p(1-p)$ ✓

$$\underline{Skew[X]} = \frac{\mathbb{E}[X^3] - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

$\mathbb{E}[X^3] = \sum x^3 p_X(x) = 1\, p_X(1) + 0\, p_X(0) = p$

$$Skew[X] = \frac{p - 3p^2(1-p) - p^3}{\sqrt{p(1-p)}^3} = \frac{p(1-p^2) - 3p^2(1-p)}{p^{3/2}(1-p)^{3/2}} = \frac{p(1-p)(1+p) - 3p^2(1-p)}{p^{3/2}(1-p)^{3/2}} = \frac{p(1-p)(1+p-3p)}{p^{3/2}(1-p)^{3/2}} = \frac{1-2p}{\sqrt{p(1-p)}}$$ ✓

$$\underline{Kurtosis[X]} = \frac{\mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4}$$

$\mathbb{E}[X^4] = \cdots = p$

$$Kurt[X] = \frac{p - 4p^2 + 6p^3(1-p) + 3p^4}{p^2(1-p)^2} = \frac{p - 4p^2 + 6p^3 - 3p^4}{p^2(1-p^2)} = \frac{p(1-p)(3p^2-3p+1)}{p^2(1-p)^2} = \frac{1-3p+3p^2}{p(1-p)}$$ ✓

2) $$skew[X] = \frac{\mathbb{E}[X^3] - 3\mu\mathbb{E}[X^2] + 2\mu^3}{\sigma^3} = \frac{\mathbb{E}[X^3] - 3\mu\mathbb{E}[X^2] + 3\mu^3 - 3\mu^3 + 2\mu^3}{\sigma^3} = \frac{\mathbb{E}[X^3] - 3\mu(\mathbb{E}[X^2] - \mu^2) - \mu^3}{\sigma^3} = \frac{\mathbb{E}[X^3] - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$ ✓

$$kurt[X] = \frac{\mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\mathbb{E}[X^2] - 3\mu^4}{\sigma^4} = \frac{\mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\mathbb{E}[X^2] - 6\mu^4 + 6\mu^4 - 3\mu^4}{\sigma^4} = \frac{\mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2[\mathbb{E}[X^2] - \mu^2] + 6\mu^4}{\sigma^4}$$

$$= \frac{\mathbb{E}[X^4] - 4\mu\mathbb{E}[X^3] + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4}$$ ✓

3) $$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

a) $$M_t(x) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $\left(z = \frac{x-\mu}{\sigma} \to x = z\sigma + \mu\right)$ $\left(dz = \frac{dx}{\sigma} \to \sigma dz = dx\right)$

$$= \int_{-\infty}^{\infty} e^{t(z\sigma+\mu)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}z^2} \cdot \sigma\, dz = e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{t\sigma z - \frac{1}{2}z^2} dz = e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2t\sigma z)} dz$$

$$= e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2 - 2t\sigma z + t^2\sigma^2 - t^2\sigma^2)} dz = e^{t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z - 2t\sigma z + t^2\sigma^2)} \cdot e^{\frac{1}{2}t^2\sigma^2}$$

$$= e^{t\mu + \frac{1}{2}t^2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z - t\sigma)^2} dz = \underline{e^{t\mu + \frac{1}{2}t^2\sigma^2}}$$ $[\therefore Z \sim N(t\sigma, 1)]$

b) $M_X'(x) = e^{t\mu + \frac{1}{2}t^2\sigma^2}(\mu + t\sigma^2)$

$\mathbb{E}[X] = M_X'(0) = e^0(\mu + 0) = \mu$ ✓

c) $\mathbb{E}[X^2] = M_X''(0) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \cdot \sigma^2 + (\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \cdot (\mu + t\sigma^2) = \sigma^2 + \mu^2$

$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$ ✓

d) $X \sim N(\mu, \sigma^2)$, then $\mu_{2n} = 1, 3, 5, \cdots, (2n-1)\sigma^{2n}$ and $\mu_{2n+1} = 0$

$$\text{skew}(x) = \frac{\mu_3^2}{\mu_2^3} = 0 \quad \checkmark$$

e) $\text{kurt}(x) = \mu_4 / \mu_2^2 = 3\sigma^4/\sigma^4 = 3 \quad \checkmark$

4) $L(\beta_0, \beta_1, \sigma_i^2, y) = \prod_{i=1}^{n} f(y_i, \beta_0, \beta_1, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$

$$l(\beta_0, \beta_1, \sigma^2) = \text{constant} - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

a) For LSE of β_0, β_1, we minimize $S = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ with respect to β_0 and β_1

$$\frac{\partial S}{\partial \beta_0} = 0 \qquad\qquad \frac{\partial S}{\partial \beta_1} = 0$$

$$2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-1) = 0 \qquad\qquad 2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-x_i) = 0$$

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \quad (1) \qquad\qquad \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \quad (2)$$

solving (1) & (2): $\hat{\beta}_1 = \dfrac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} = \dfrac{1/n \sum x_i y_i - \bar{x}\cdot\bar{y}}{1/n \sum x_i^2 - \bar{x}^2} = \underline{\dfrac{\text{cov}(x,y)}{\text{var}(x)}}$

$$\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \underline{\bar{y} - \hat{\beta}_1 \bar{x}}$$

b) $\dfrac{\partial l}{\partial \beta_0} = 0$

$$-\frac{2}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-1) = 0$$

$$n\beta_0 + \beta_1 \sum_1^n x_i = \sum_1^n y_i \quad (*)$$

$\dfrac{\partial l}{\partial \beta_1} = 0$

$$-\frac{2}{2\sigma^2}\sum_1^n (y_i - \beta_0 - \beta_1 x_i)(-x_i)$$

$$\beta_0 \sum_1^n x_i + \beta_1 \sum_1^n x_i^2 = \sum_1^n x_i y_i \quad (**)$$

$\dfrac{\partial l}{\partial \sigma^2} = 0$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0 \quad (***)$$

Since (*) & (**) are the same as (1) & (2), the MLE estimates of β_0 & β_1 will be the same as the LSE estimates.

$$\hat{\beta}_0^{MLE} = \hat{\beta}_0^{LSE} = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1^{MLE} = \hat{\beta}_0^{LSE} = \frac{\text{cov}(x,y)}{\text{var}(x)} = \frac{1/n \sum x_i y_i - \bar{x}\cdot\bar{y}}{1/n \sum x_i^2 - \bar{x}^2}$$

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

5) a/b) see included R for table and line

It appears as x ↑, y ↑. $y = 0.02733x - 4.79841$

c) see included R for anova table

The p-value for x is extremely small, so we can see that x is related to y

d) $R^2 = .8356$

Our model fits the data reasonably well. Need more predictors (and more data) to improve model. Also, the model can be improved by removing outliers.

e) see included R for diagnostic plots used

Linear Relationship between X & Y?

Residuals vs Fitted shows an approximately linear relationship; hard to tell because n is small. ✓

Normality of Residuals?

Normal Q-Q Plot shows the residuals are normal except for two outliers. ✓

Homogeneous Residual Variance?

Scale location plot shows the variance is approx. homog.
(hard to tell w/out more data)

6) a) see included R

b) mean = 2771.034

standard deviation = 278.0726

skewness = .2617603

kurtosis = 2.518914 (no excess since not > 3)

min = 2237.4

max = 3526.65

c) calculated small p-value, so the data is not normal

(expected by looking at density and normal q-q plots for simple returns)

d) mean = .0004836705

std dev. = .013087

skewness = -1.172274

kurtosis = 25.13016

min = -.1276522

max = .089683

e) The p-value for a t-test of if the mean of log returns is 0 was .2621

So, we fail to reject H_0 (mean is 0).

f) $S^* = -14.53179$ p-value is very small, so we must reject

H_0 (the log returns are not skewed)

g) $K^* = 155.7599$ (pvalue is very small)

We reject H_0 (the tails behave normal)

7) a) computed in R

b) mean = -0.0002461043

std. dev. = 0.001378411

skewness = -0.4605336

kurtosis = 5.644422 (excess kurtosis = 2.644422)

min = $-.004928916$

max = 0.004708191

c) see included R

d) a t-test for $\mu = 0$ (H_0) gives us p-value = $.02819$ ($< .05$)

So we reject H_0 (it appears the mean is not 0)

e) a joint normality test gives us a very small p-value,

so it appears the log returns are not normal

(also see log USDCNY Normal Q-Q plot)

R Notebook

Code

Problem 5

Hide

```
x = c(300, 350, 400, 400, 450, 450, 480, 480, 530, 530,  580,  580,  620,  620,  670,  700)
y = c(5.8, 4.5, 5.9, 6.2, 6.0, 7.5, 6.1, 8.6, 8.9, 8.2, 14.2, 11.9, 11.1, 11.5, 14.5, 14.8)
x_trimmed = c(350, 400, 400, 450, 450, 480, 480, 530, 530,  580,  620,  620,  670,  700)
y_trimmed = c(4.5, 5.9, 6.2, 6.0, 7.5, 6.1, 8.6, 8.9, 8.2, 11.9, 11.1, 11.5, 14.5, 14.8)

plot(x, y, main="Problem 5a/b")
xylm = lm(y ~ x)
abline(xylm)
```

Hide

```
xylm
```

```


Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
   -4.79841      0.02733
```

Hide

```
anova(xylm)
```

```
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x          1 148.930 148.930  71.163 7.346e-07 ***
Residuals 14  29.299   2.093
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

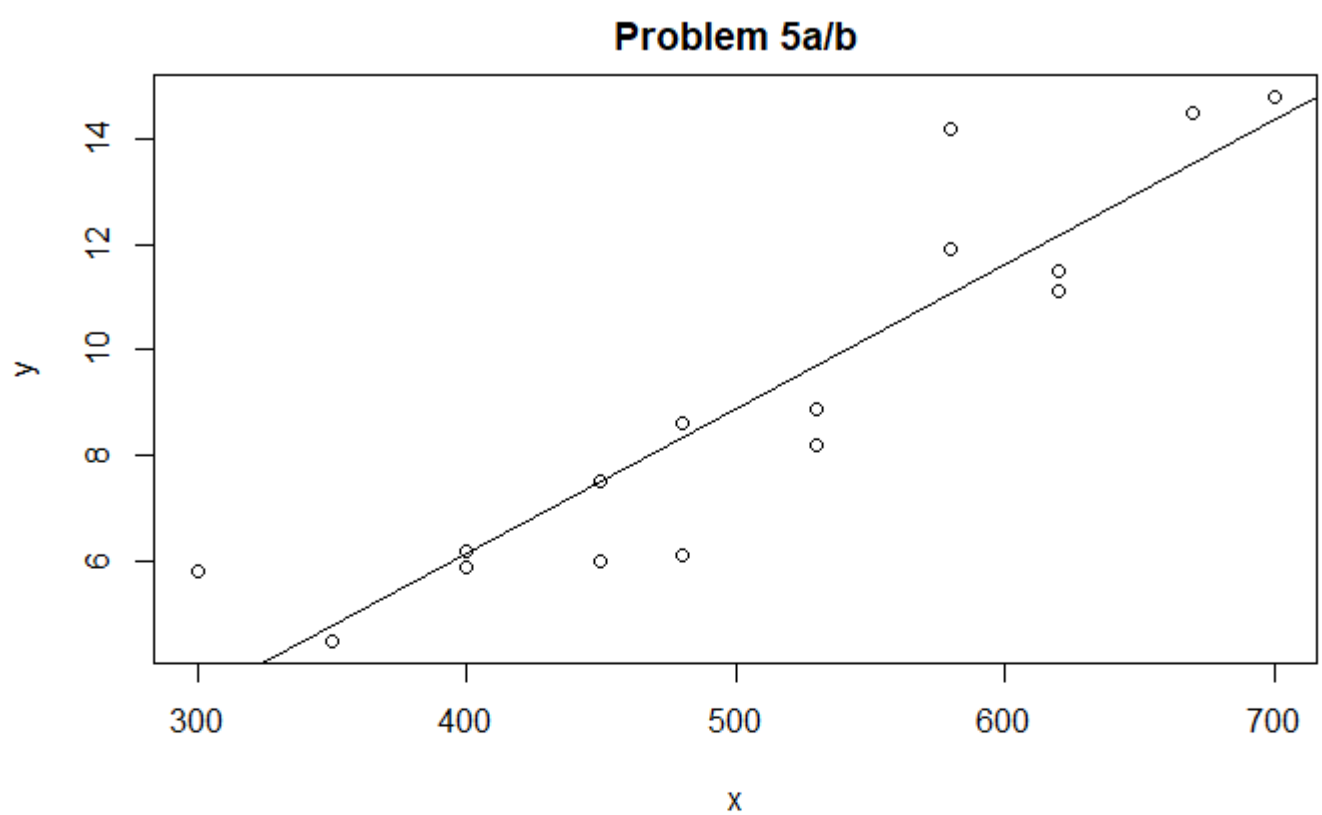
Hide

```
summary(xylm)
```

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.2205 -0.8520 -0.1173  0.5616  3.1464 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -4.79841    1.68750  -2.844    0.013 *  
x            0.02733    0.00324   8.436 7.35e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.447 on 14 degrees of freedom
Multiple R-squared:  0.8356,	Adjusted R-squared:  0.8239 
F-statistic: 71.16 on 1 and 14 DF,  p-value: 7.346e-07
```

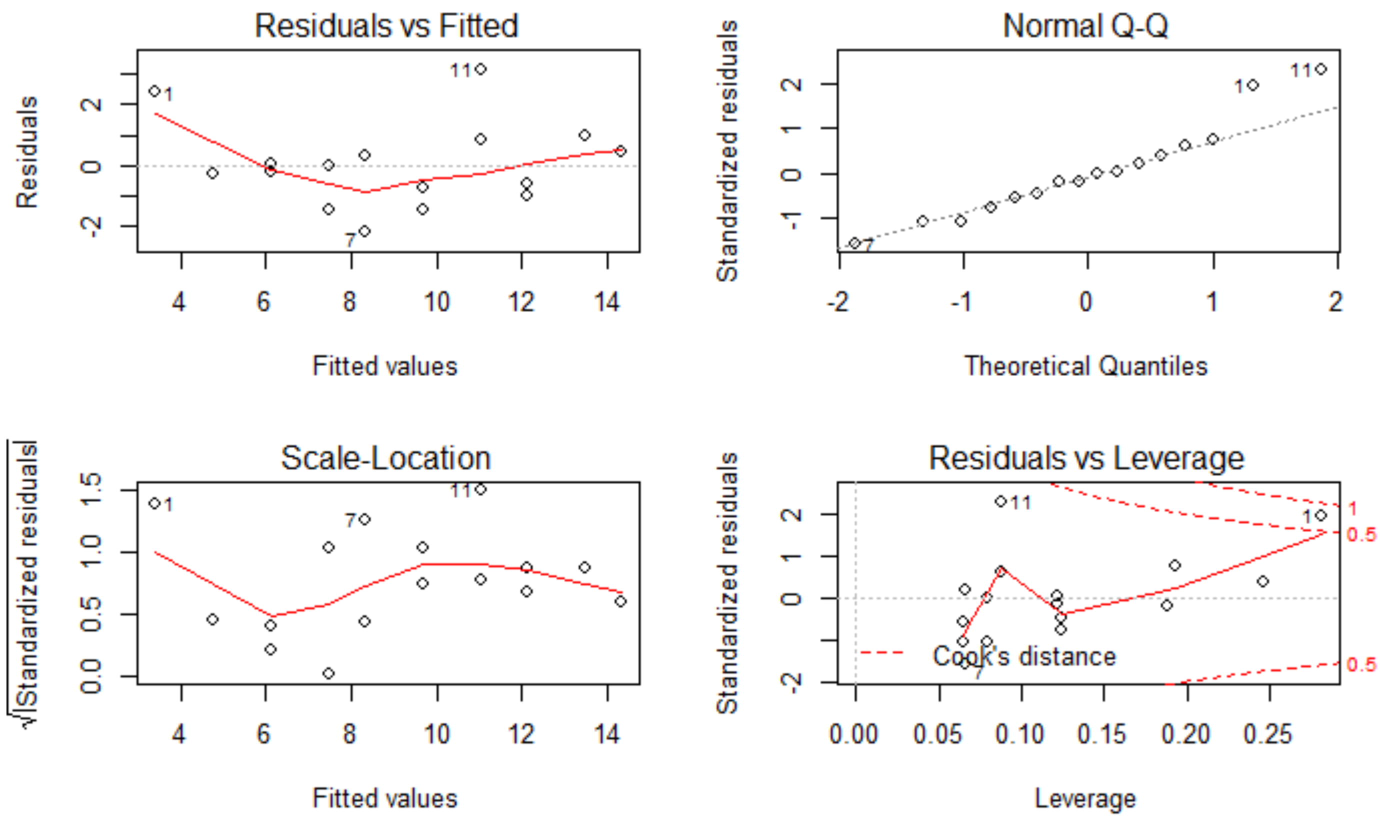
Hide

```
par(mfrow = c(2,2))
```



Hide

```
plot(xylm)
```



It appears that as x increases, so does y. The line of best fit found by simple linear regression seems to agree. Intercept = -4.79841 Slope = 0.02733 y = 0.02733x - 4.79841

Problem 6

Hide

```
require(quantmod)
require(moments)
require(symmetry)
require(ICS)
sp500 = new.env()
getSymbols("^GSPC",from="2017-1-1", to="2020-9-2", env=sp500, src="yahoo")
GSPC = sp500$GSPC$GSPC.Adjusted
chartSeries(GSPC$GSPC.Adjusted, theme="white")
mean(GSPC)
sd(GSPC)
GSPC_skew = (sum((GSPC - mean(GSPC))^3)/nrow(GSPC))/((sum((GSPC - mean(GSPC))^2)/nrow(GSPC))^(3/
2))
GSPC_skew
GSPC_kurt = (sum((GSPC - mean(GSPC))^4)/nrow(GSPC))/((sum((GSPC - mean(GSPC))^2)/nrow(GSPC))^(2
))
GSPC_kurt
min(GSPC)
max(GSPC)
GSPC_density = density(GSPC)
plot(GSPC_density, main="Simple return density plot")
qqnorm(GSPC, main="Simple return Normal Q-Q Plot")
fBasics::normalTest(as.vector(GSPC), method="jb")

logGSPC = diff(log(GSPC$GSPC.Adjusted))[-1,]
mean(logGSPC)
sd(logGSPC)
logGSPC_skew = (sum((logGSPC - mean(logGSPC))^3)/nrow(logGSPC))/((sum((logGSPC - mean(logGSPC))^
2)/nrow(logGSPC))^(3/2))
logGSPC_skew
logGSPC_kurt = (sum((logGSPC - mean(logGSPC))^4)/nrow(logGSPC))/((sum((logGSPC - mean(logGSPC))^
2)/nrow(logGSPC))^(2))
logGSPC_kurt
min(logGSPC)
max(logGSPC)
t.test(as.vector(logGSPC), mu=0)
S = skewness(logGSPC)/sqrt(6/length(logGSPC))
2*pnorm(-abs(S))
K = kurtosis(logGSPC)/sqrt(24/length(logGSPC))
2*pnorm(-abs(K))
```

Problem 7

Hide

```
getFX("USD/CNY", from="2020-4-01", to="2020-9-2")
```

```
[1] "USD/CNY"
```

Hide

```
logUSDCNY = diff(log(USDCNY$USD.CNY))[-1,]
mean(logUSDCNY)
```

```
[1] -0.0002461043
```

Hide

```
sd(logUSDCNY)
```

```
[1] 0.001378411
```

Hide

```
logUSDCNY_skew = (sum((logUSDCNY - mean(logUSDCNY))^3)/nrow(logUSDCNY))/((sum((logUSDCNY - mean
(logUSDCNY))^2)/nrow(logUSDCNY))^(3/2))
logUSDCNY_skew
```

```
[1] -0.4605336
```

Hide

```
logUSDCNY_kurt = (sum((logUSDCNY - mean(logUSDCNY))^4)/nrow(logUSDCNY))/((sum((logUSDCNY - mean
(logUSDCNY))^2)/nrow(logUSDCNY))^(2))
logUSDCNY_kurt
```

```
[1] 5.644422
```

Hide

```
min(logUSDCNY)
```

```
[1] -0.004928916
```

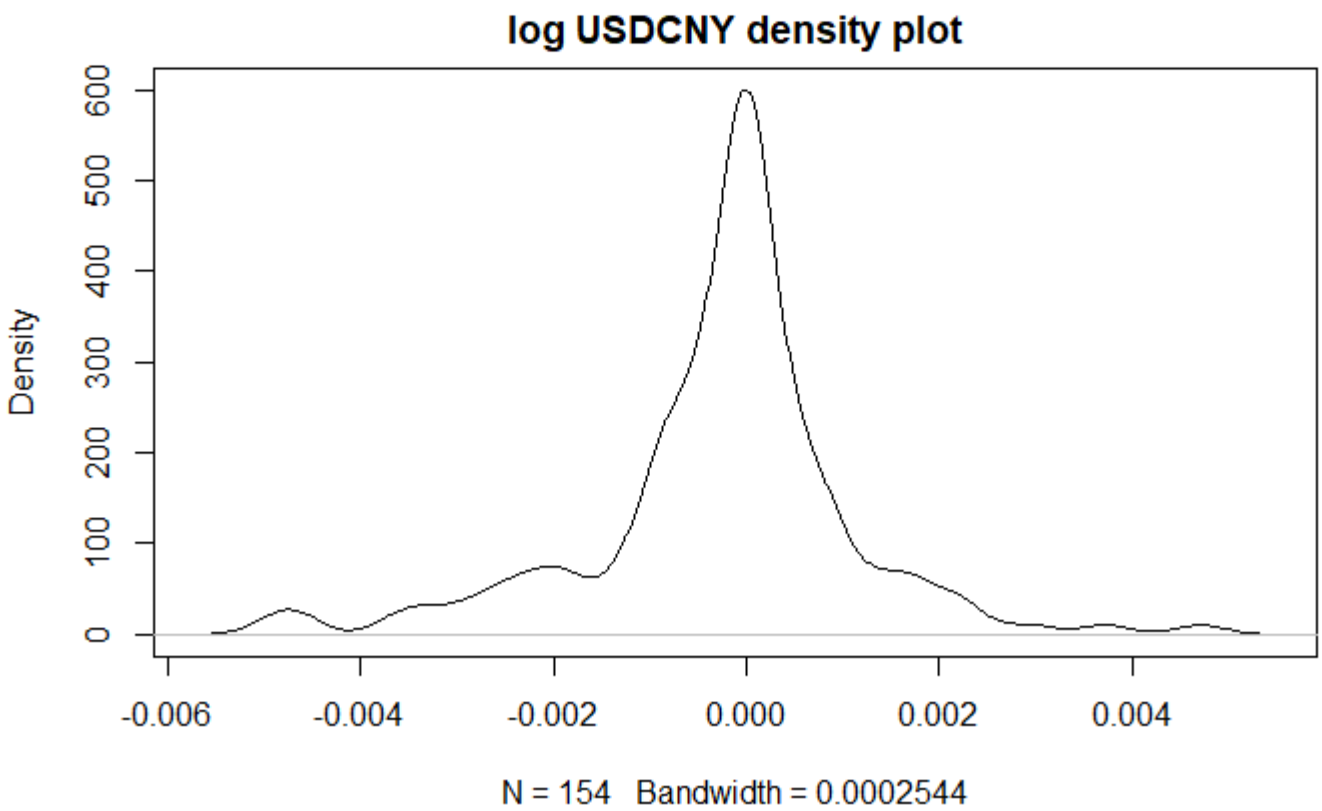
Hide

```
max(logUSDCNY)
```

```
[1] 0.004708191
```

Hide

```
logUSDCNY_density = density(logUSDCNY)
plot(logUSDCNY_density, main="log USDCNY density plot")
```



Hide

```
t.test(as.vector(logUSDCNY), mu=0)
```

```
    One Sample t-test

data:  as.vector(logUSDCNY)
t = -2.2156, df = 153, p-value = 0.02819
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -4.655441e-04 -2.666455e-05
sample estimates:
    mean of x
-0.0002461043
```

Hide

```
fBasics::normalTest(as.vector(logUSDCNY), method="jb")
```

```
Title:
 Jarque - Bera Normalality Test

Test Results:
  STATISTIC:
    X-squared: 50.3152
  P VALUE:
    Asymptotic p Value: 1.186e-11

Description:
 Sun Sep 20 19:11:28 2020 by user: rosha
```

Hide

```
qqnorm(logUSDCNY, main="log USDCNY Normal Q-Q Plot")
```