Time Series

Homework 2

Helen Kafka, Sven Bergmann

02/28/24

Time series analysis Homework 2

1. $Y_1 = \mu + e_1 - e_0$
$Y_2 = \mu + e_2 - e_1$
$Y_3 = \mu + e_3 - e_2$
$\vdots$
$Y_n = \mu + e_n - e_{n-1}$

$$\bar{Y}_1 = \frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n \mu + e_i - e_{i-1}$$

$$= \frac{1}{n}\left(n\mu + \sum_{i=1}^n e_i - e_{i-1}\right)$$

$$\overset{\text{telescopic sum}}{=} \frac{1}{n}(n\mu + e_n - e_0)$$

$$\mathrm{Var}(\bar{Y}_1) = \mathrm{Var}\left(\frac{1}{n}(n\mu + e_n - e_0)\right)$$

$$= \frac{1}{n^2}\mathrm{Var}(\underbrace{n\mu}_{\text{constant}} + e_n - e_0)$$

$$= \frac{1}{n^2}\mathrm{Var}(e_n - e_0)$$

$$\overset{e_i \perp\!\!\!\perp e_j \text{ for } i\neq j}{=} \frac{1}{n^2}(\mathrm{Var}(e_n) + \mathrm{Var}(e_0))$$

$$= \frac{2\sigma_e^2}{n^2}$$

$Y_1 = \mu + e_1$
$Y_2 = \mu + e_2$
$\vdots$
$Y_n = \mu + e_n$

$$\bar{Y}_2 = \frac{1}{n}\sum_{i=1}^n Y_i = \frac{1}{n}\sum_{i=1}^n \mu + e_i = \frac{1}{n}\left(n\mu + \sum_{i=1}^n e_i\right)$$

$$\mathrm{Var}(\bar{Y}_2) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \mu + e_i\right)$$

$$= \frac{1}{n^2}\mathrm{Var}(\underbrace{n\mu}_{=\text{const.}} + e_1 + \dots + e_n)$$

$$= \frac{1}{n^2}\mathrm{Var}(e_1 + \dots + e_n)$$

$$\overset{e_i \perp\!\!\!\perp e_j \text{ for } i\neq j}{=} \frac{1}{n^2}(\mathrm{Var}(e_1) + \dots + \mathrm{Var}(e_n))$$

$$= \frac{n\sigma_e^2}{n^2} = \frac{\sigma_e^2}{n}$$

$$\Rightarrow \mathrm{Var}(\bar{Y}_1) = \frac{2\sigma_e^2}{n^2} \geq \frac{\sigma_e^2}{n} = \mathrm{Var}(\bar{Y}_2) \quad \text{for } n = 1,2$$

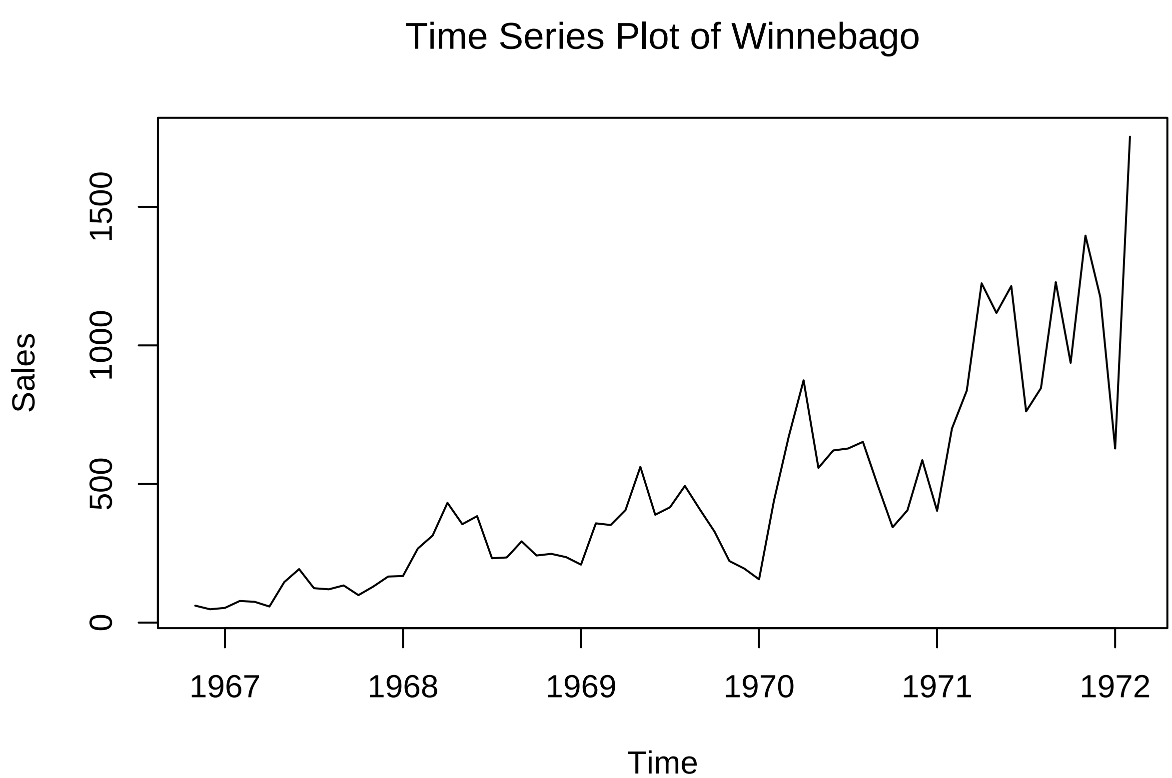
$$\mathrm{Var}(\bar{Y}_1) = \frac{2\sigma_e^2}{n^2} < \frac{\sigma_e^2}{n} = \mathrm{Var}(\bar{Y}_2) \quad \text{for } n \geq 3$$

Problem 2

a)

```r
library(TSA)
library(tseries)
```

```r
data(winnebago)
plot(winnebago, xlab = expression("Time"), ylab = expression("Sales"),
    main = expression("Time Series Plot of Winnebago"))
```



Interpretation:

- The time series shows an overall upwards trend between the years 1967 and 1972.
- Between the years 1970 and 1972 the increase is at its highest.

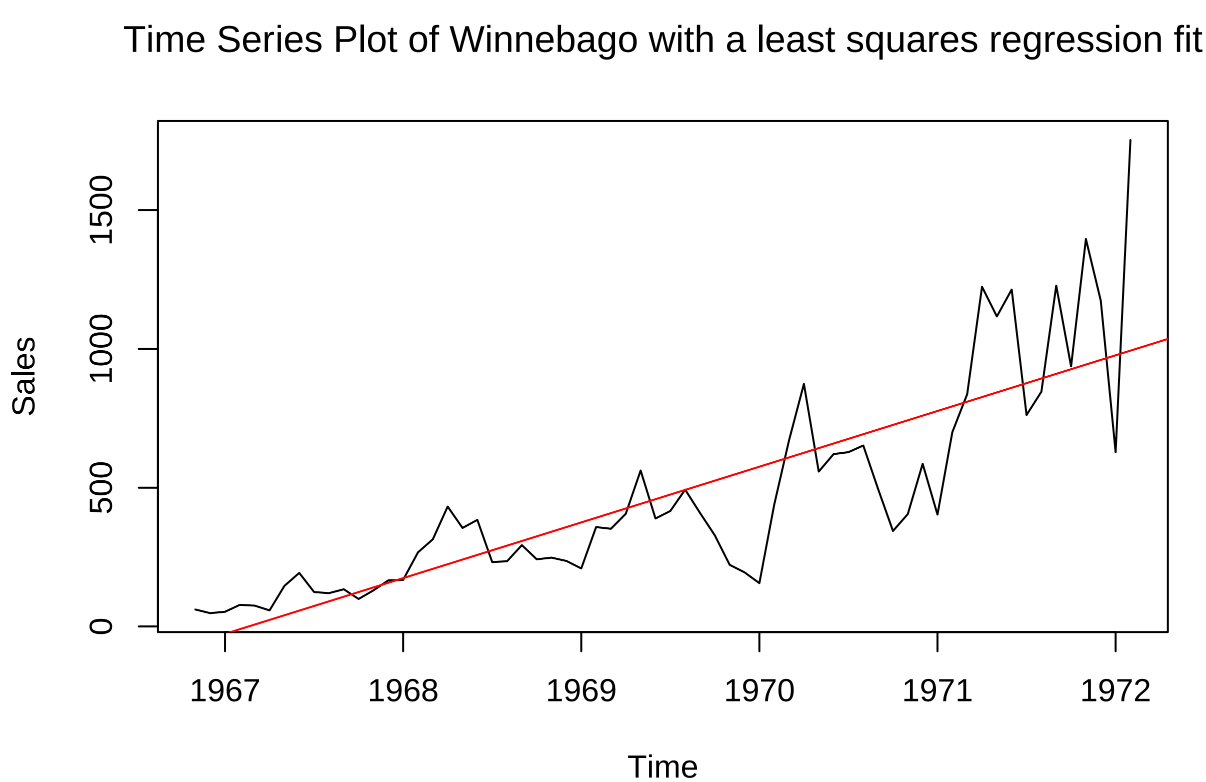
b)

```r
lm_model_a <- lm(winnebago ~ time(winnebago))
summary(lm_model_a)
```

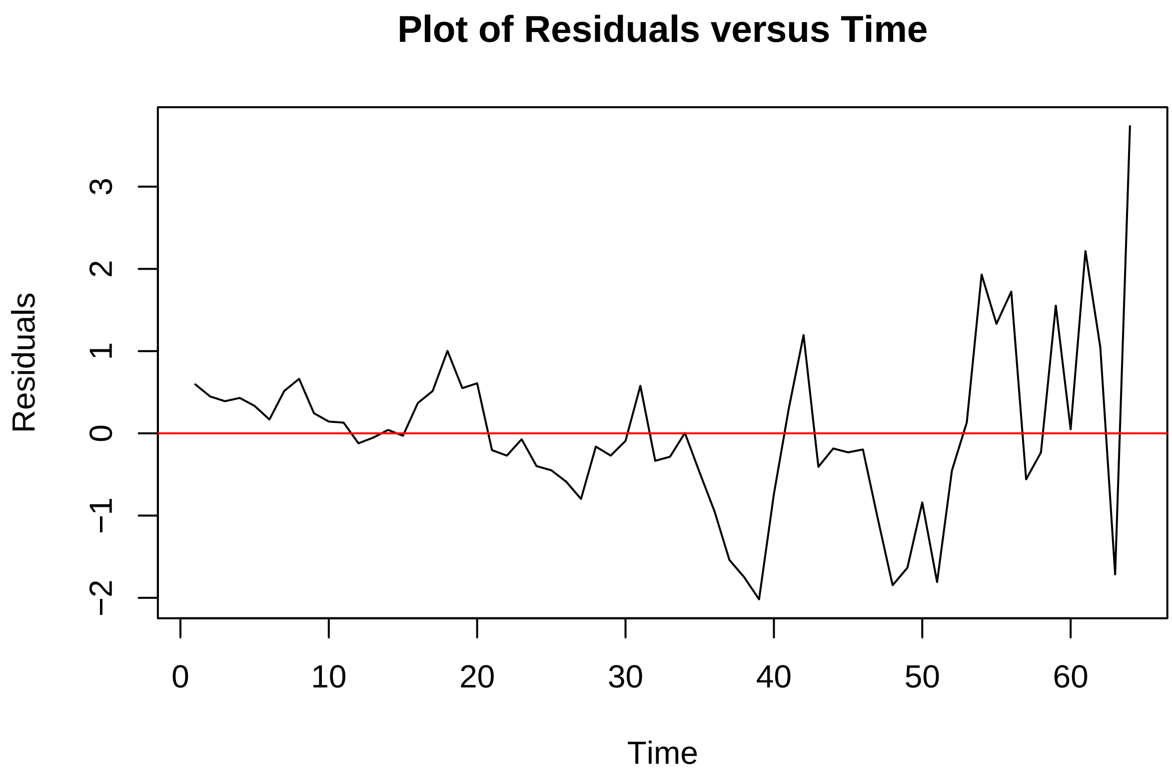
```
##
## Call:
## lm(formula = winnebago ~ time(winnebago))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -419.58  -93.13  -12.78   94.96  759.21
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -394885.68   33539.77  -11.77   <2e-16 ***
## time(winnebago)     200.74      17.03   11.79   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.7 on 62 degrees of freedom
## Multiple R-squared:  0.6915, Adjusted R-squared:  0.6865
## F-statistic: 138.9 on 1 and 62 DF,  p-value: < 2.2e-16
```

- We expect the wages to increase by \$200.74 per year
- 69.15% of the variance can be explained by the predictor variables
- the linear trend is significant because the p-value $2.2 \cdot 10^{-16}$ is smaller than 10^{-12}

```
plot(winnebago, xlab = expression("Time"), ylab = expression("Sales"),
    main = expression("Time Series Plot of Winnebago with a least squares regression fit"))
abline(lm_model_a, col = "red")
```



```r
lm_model_a_std_res <- as.ts(rstandard(lm_model_a))
plot(lm_model_a_std_res, xlab = expression("Time"), ylab = expression("Residuals"),
    main = "Plot of Residuals versus Time")
abline(h = 0, col = "red")
```



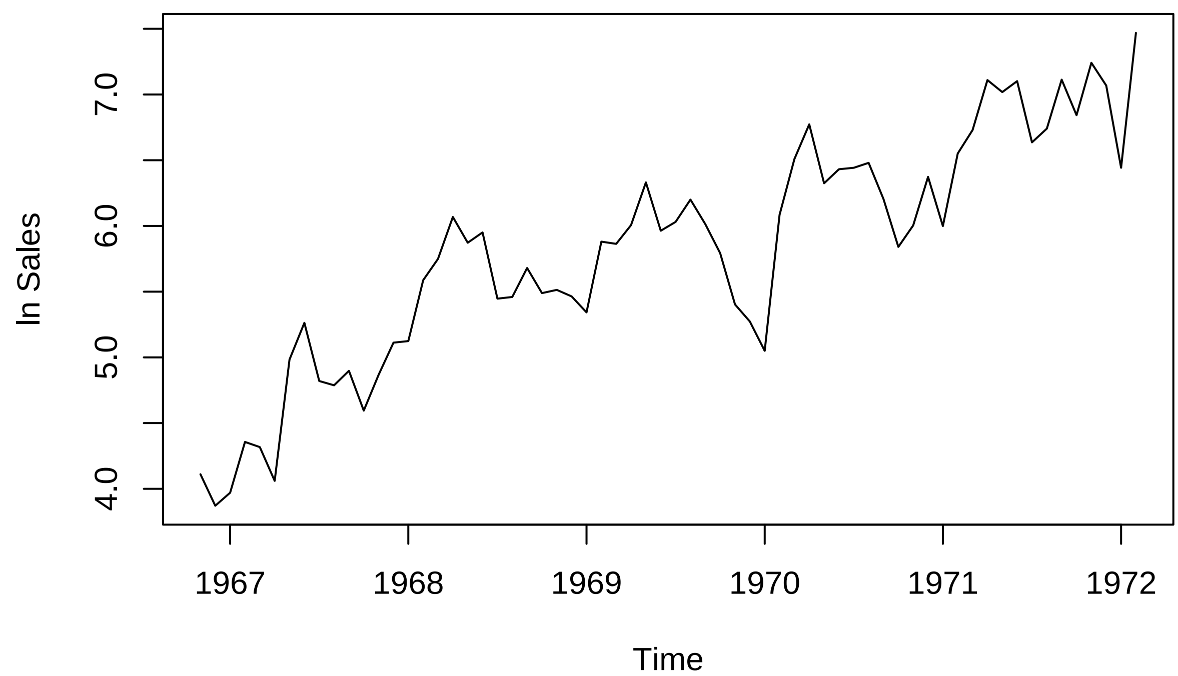
Interpretation:

- The residuals plot shows somewhat random movement around zero.
- More uneven spread between ca. 35 to 65 in comparison to 0 to 35.
- There may be a “seasonal” cyclical trend.

c)

```r
ln_winnebago <- log(winnebago)
plot(ln_winnebago, xlab = expression("Time"), ylab = expression("ln Sales"),
    main = expression("Time Series Plot of Winnebago"))
```

Time Series Plot of Winnebago



Interpretation:

- The time series plot of the transformed values shows a linear upward trend (which corresponds with the untransformed data)
- There is one downward spike around the year 1970
- The "seasonal" trend seems more pronounced

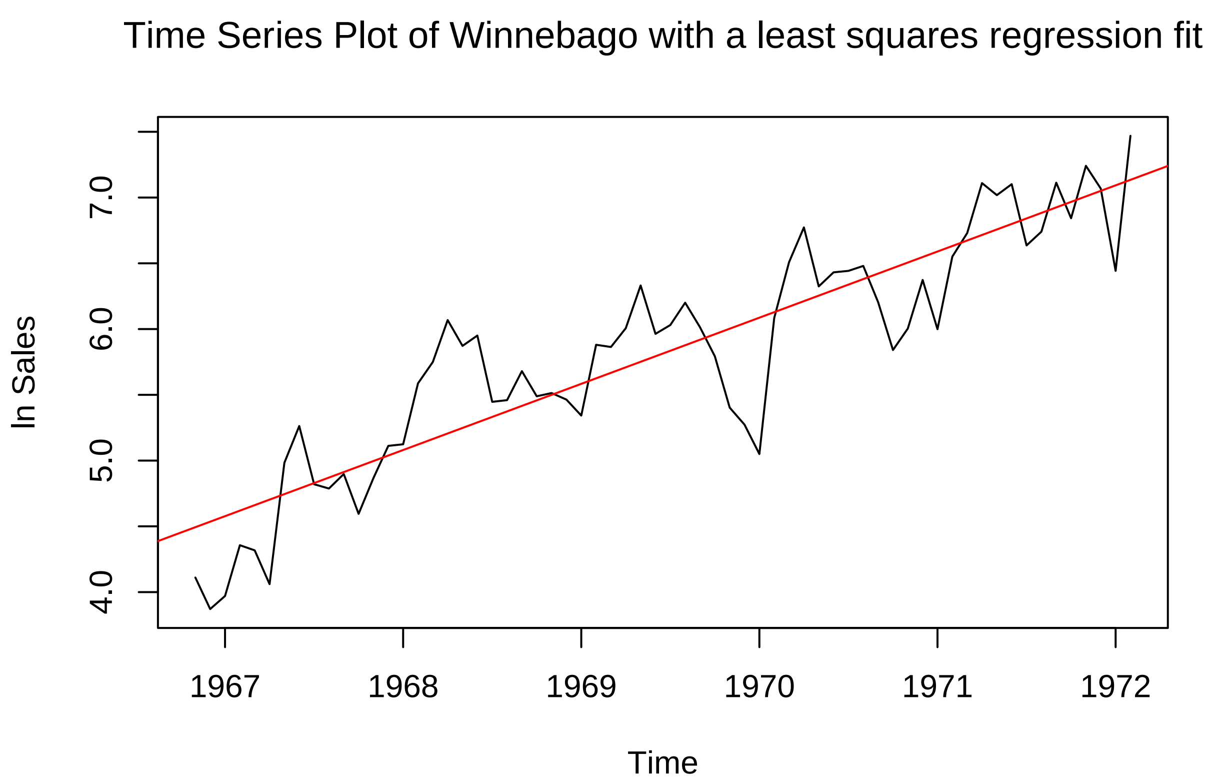
d)

```r
lm_model_d <- lm(ln_winnebago ~ time(ln_winnebago))
summary(lm_model_d)
```

```
##
## Call:
## lm(formula = ln_winnebago ~ time(ln_winnebago))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.03669 -0.20823  0.04995  0.25662  0.86223
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -984.93878   62.99472  -15.63   <2e-16 ***
## time(ln_winnebago)    0.50306    0.03199   15.73   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3939 on 62 degrees of freedom
## Multiple R-squared:  0.7996, Adjusted R-squared:  0.7964
## F-statistic: 247.4 on 1 and 62 DF,  p-value: < 2.2e-16
```

```
plot(ln_winnebago, xlab = expression("Time"), ylab = expression("ln Sales"),
    main = expression("Time Series Plot of Winnebago with a least squares regression fit"))
abline(lm_model_d, col = "red")
```




```
lm_model_d_std_res <- as.ts(rstandard(lm_model_d))
plot(lm_model_d_std_res, xlab = expression("Time"), ylab = expression("Residuals"),
    main = "ln Plot of Residuals versus Time")
abline(h = 0, col = "red")
```

Interpretation:

- the ln-transformed residuals plot shows random movement around 0
- There seems to be an overall cyclical trend

e)

```
months <- season(winnebago)
lm_model_e <- lm(ln_winnebago ~ months + time(winnebago))
summary(lm_model_e)
```

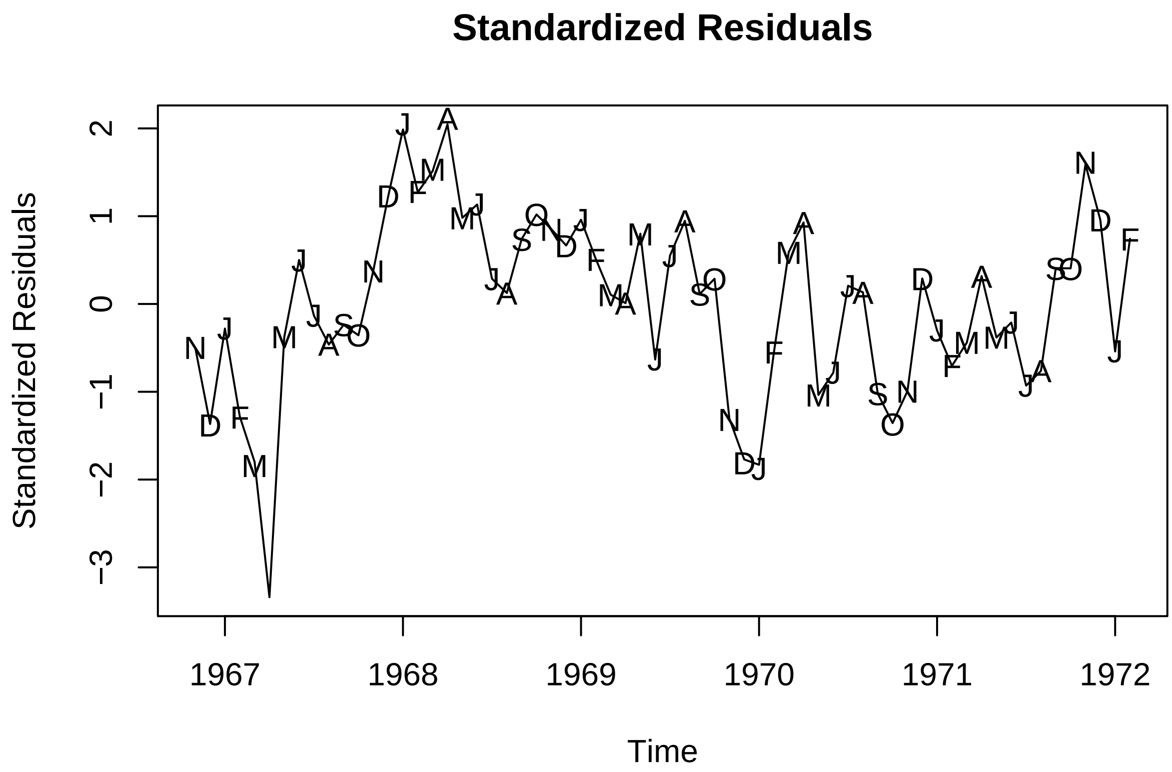
```
##
## Call:
## lm(formula = ln_winnebago ~ months + time(winnebago))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92501 -0.16328  0.03344  0.20757  0.57388
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -997.33061   50.63995 -19.695  < 2e-16 ***
## monthsFebruary     0.62445    0.18182   3.434 0.001188 **
## monthsMarch        0.68220    0.19088   3.574 0.000779 ***
```

```
## monthsApril        0.80959    0.19079   4.243 9.30e-05 ***
## monthsMay          0.86953    0.19073   4.559 3.25e-05 ***
## monthsJune         0.86309    0.19070   4.526 3.63e-05 ***
## monthsJuly         0.55392    0.19069   2.905 0.005420 **
## monthsAugust       0.56989    0.19070   2.988 0.004305 **
## monthsSeptember    0.57572    0.19073   3.018 0.003960 **
## monthsOctober      0.26349    0.19079   1.381 0.173300
## monthsNovember     0.28682    0.18186   1.577 0.120946
## monthsDecember     0.24802    0.18182   1.364 0.178532
## time(winnebago)    0.50909    0.02571  19.800  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3149 on 51 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.8699
## F-statistic: 36.09 on 12 and 51 DF,  p-value: < 2.2e-16
```

- 89.46% of the variance can be explained by the predictor variables
- the linear trend is significant because the p-value $2.2 \cdot 10^{-16}$ is smaller than 10^{-12}

f)

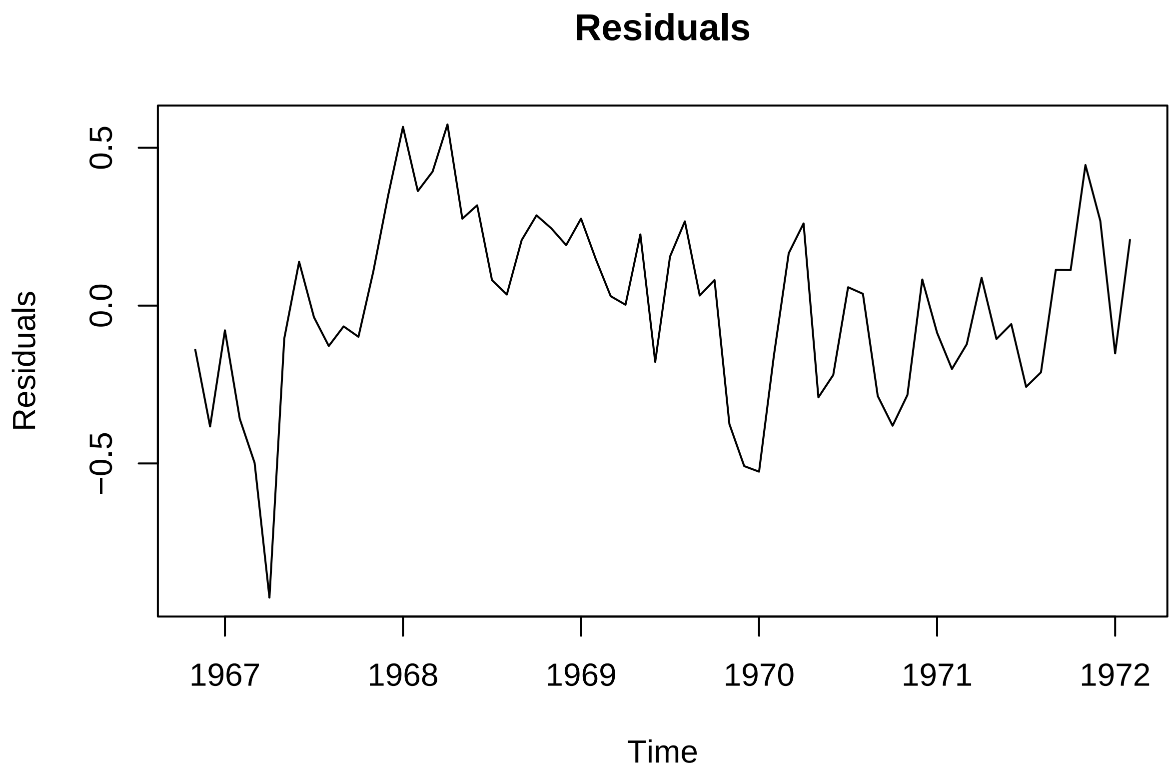
```
lm_model_e_std_res <- as.ts(rstandard(lm_model_e))
plot(lm_model_e_std_res, x = as.vector(time(winnebago)), xlab = "Time",
    ylab = "Standardized Residuals", main = "Standardized Residuals",
    type = "l")
points(y = rstudent(lm_model_e), x = as.vector(time(winnebago)),
    pch = as.vector(season(winnebago)))
```



- looking at the residuals, we can revert our assumption of a seasonal trend
- we do not see an obvious pattern of the months at the highs and lows

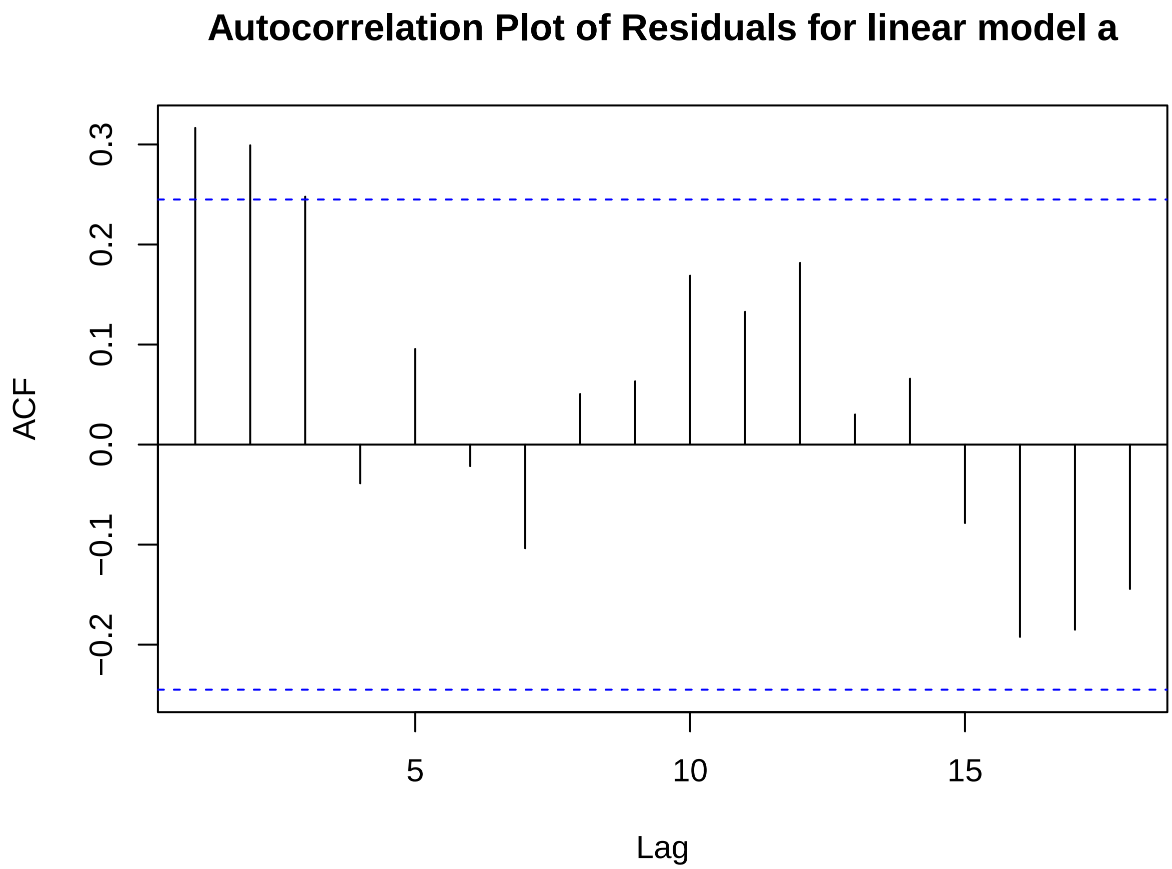
g)

```
lm_model_e_res <- residuals(lm_model_e)
plot(lm_model_e_res, x = as.vector(time(winnebago)), xlab = "Time",
    ylab = "Residuals", main = "Residuals", type = "l")
```



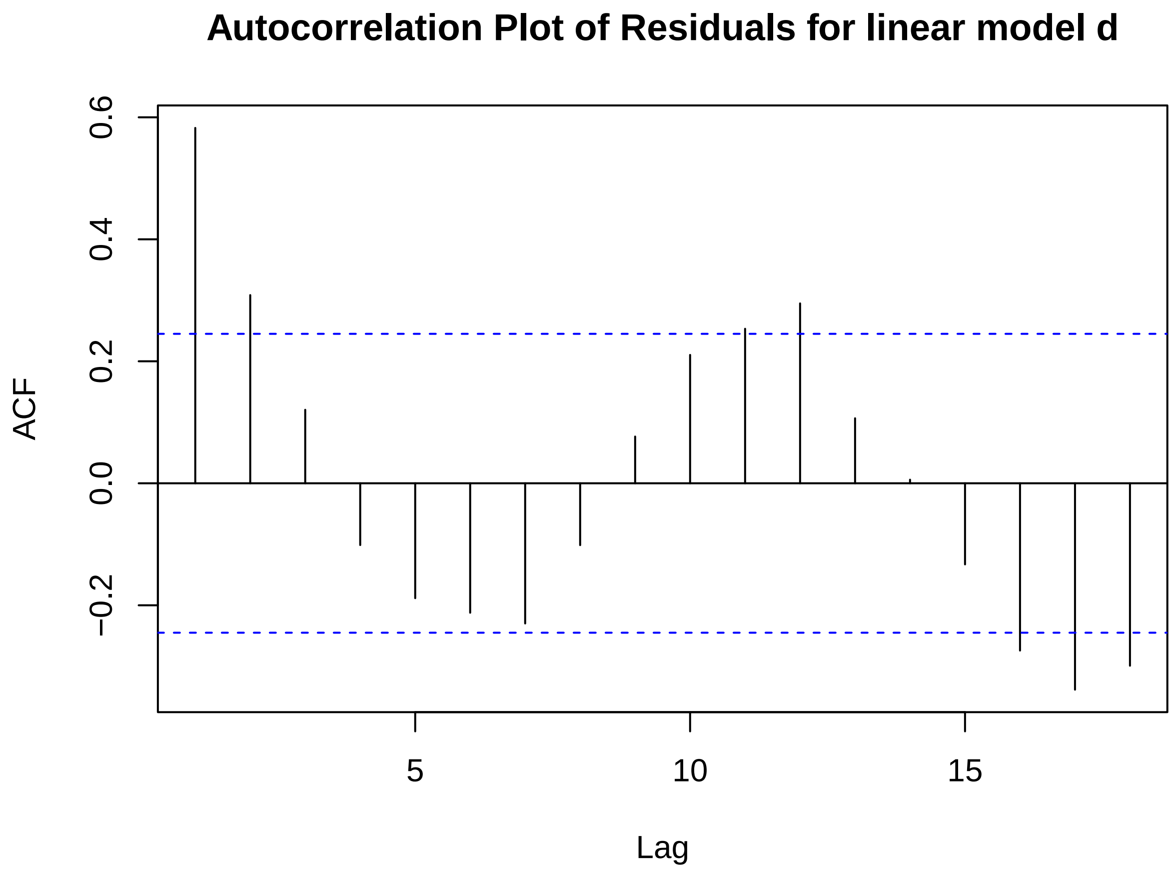
h)

```
acf(lm_model_a_std_res, main = "Autocorrelation Plot of Residuals for linear model a")
```



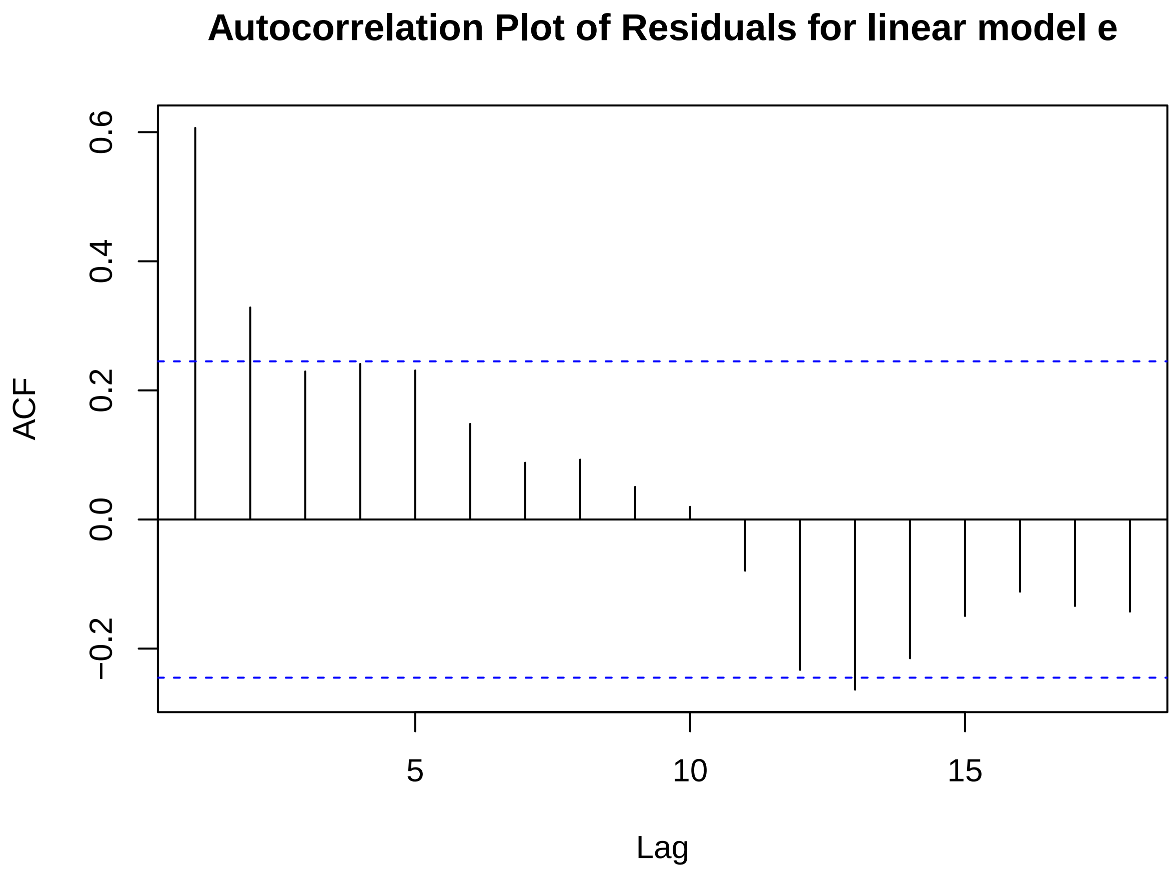
- significant autocorrelation at lags 1, 2, 3
- somewhat periodic behaviour starting at lag 7
- small magnitude with ca. .3

```
acf(lm_model_d_std_res, main = "Autocorrelation Plot of Residuals for linear model d")
```



- significant autocorrelation at lags 1, 2, 11, 12, 16, 17, 18
- periodic behaviour

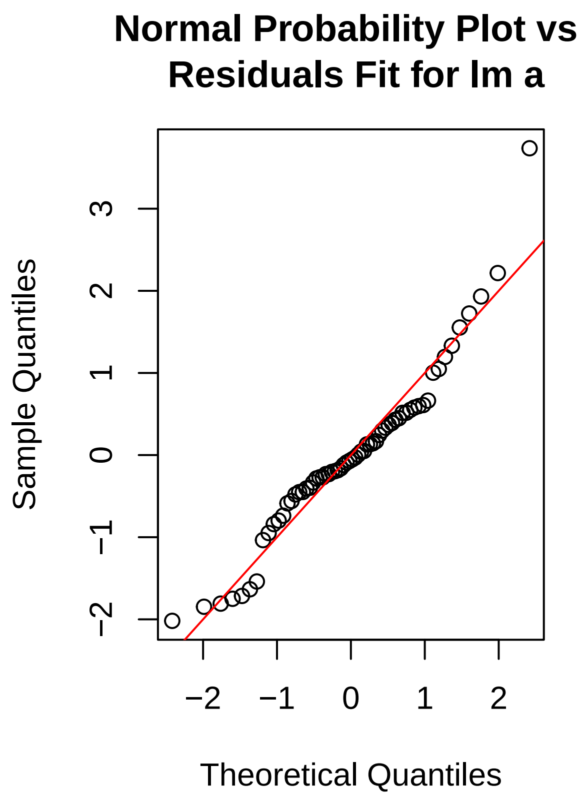
```
acf(lm_model_e_std_res, main = "Autocorrelation Plot of Residuals for linear model e")
```

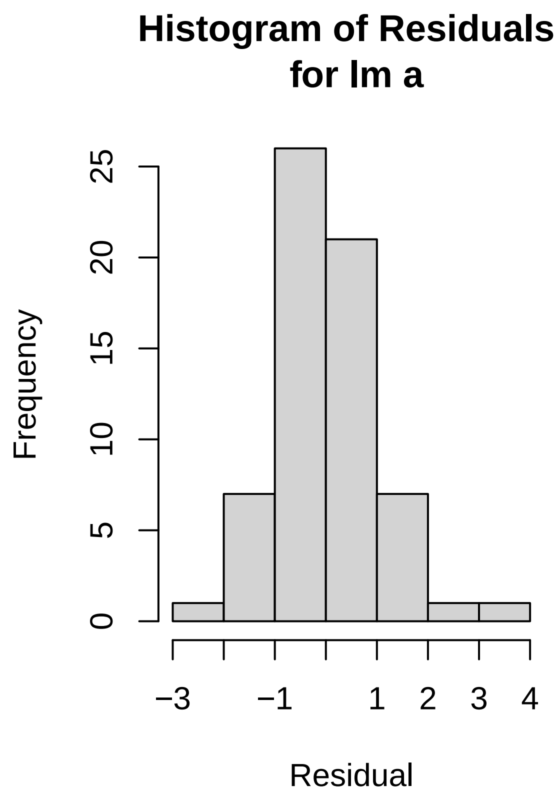


- significant autocorrelation at lags 1, 2, 13
- cyclic trend
- magnitude with about .6

i)

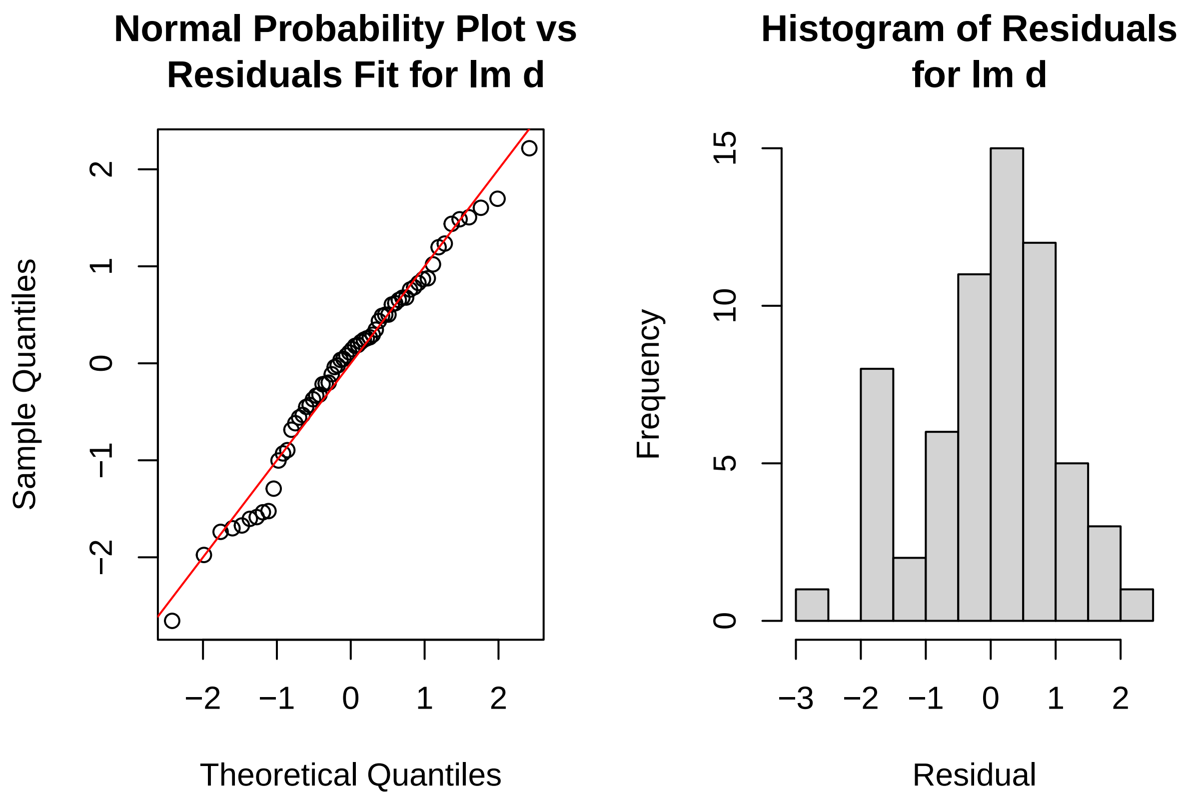
```
par(mfrow = c(1, 2))
qqnorm(lm_model_a_std_res, main = "Normal Probability Plot vs \n Residuals Fit for lm a")
abline(a = 0, b = 1, col = "red")
hist(lm_model_a_std_res, xlab = "Residual", main = "Histogram of Residuals \n for lm a")
```





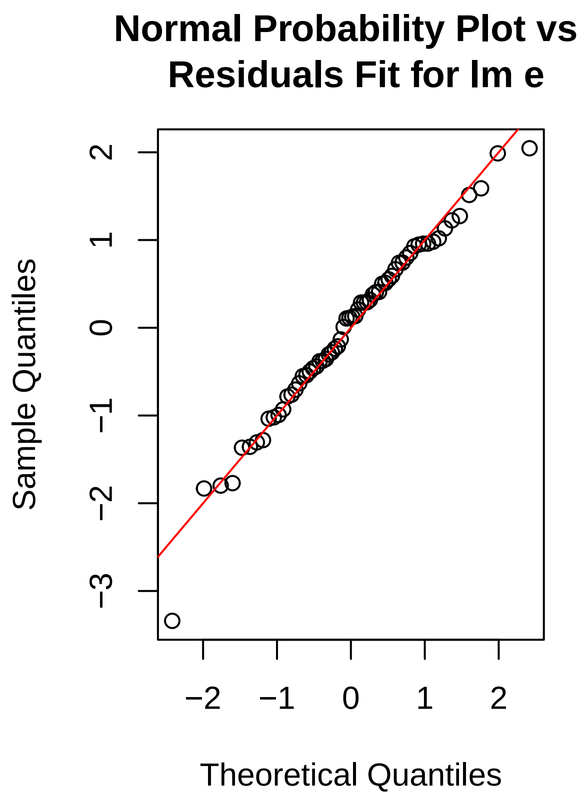
- The quantiles of the linear model a are a little bit off everywhere but somewhat equally spreaded around the line and we have one outlier at about 4.
- Data looks somewhat normally distributed based on the histogram.

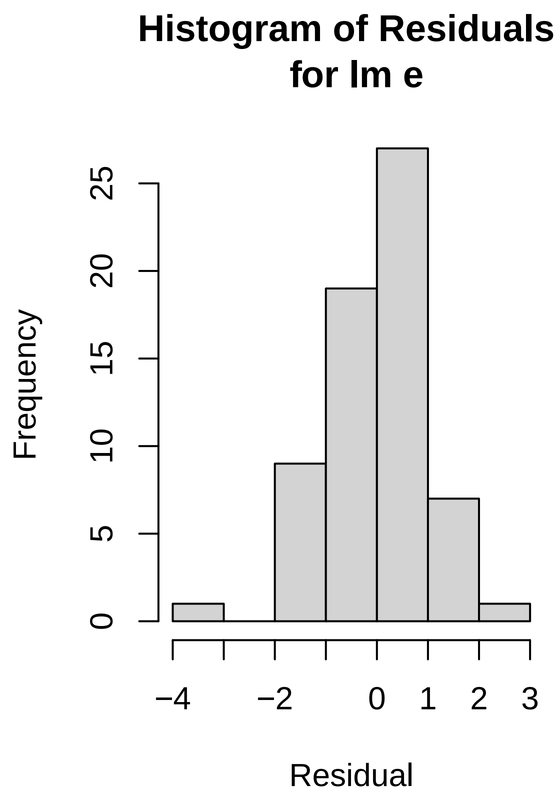
```r
par(mfrow = c(1, 2))
qqnorm(lm_model_d_std_res, main = "Normal Probability Plot vs \n Residuals Fit for lm d")
abline(a = 0, b = 1, col = "red")
hist(lm_model_d_std_res, xlab = "Residual", main = "Histogram of Residuals \n for lm d")
```



- Again, we have outliers at -3 and 2 but not as bad as in linear model a.
- Looking at the histogram, we have normal distributed data if we do not take the values at -2 into account.

```
par(mfrow = c(1, 2))
qqnorm(lm_model_e_std_res, main = "Normal Probability Plot vs \n Residuals Fit for lm e")
abline(a = 0, b = 1, col = "red")
hist(lm_model_e_std_res, xlab = "Residual", main = "Histogram of Residuals \n for lm e")
```





- The quantiles look great, nearly all of them are very close to the line.
- The histogram shows a nearly normal distribution but the values from -1 to 0 and 0 to 1 are not that symmetric as they should be.

3. $\rho_k = \varphi^k$

a) show $Var(\bar{Y}) = \frac{\gamma_0}{n}\left(\frac{1+\varphi}{1-\varphi} - \frac{2\varphi}{n}\frac{1-\varphi^n}{(1-\varphi)^2}\right)$

use 3.2.3 $Var(\bar{Y}) = \frac{\gamma_0}{n}\left(1 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho_k\right)$ ①

$$\sum_{k=0}^{n}\varphi^k = \frac{1-\varphi^{n+1}}{1-\varphi} \quad ②$$

$$\sum_{k=0}^{n} k\varphi^{k-1} = \frac{d}{d\varphi}\sum_{k=0}^{n}\varphi^k \quad ③$$

$$Var(\bar{Y}) \overset{①}{=} \frac{\gamma_0}{n}\left(1 + 2\sum_{k=1}^{n-1}\left(1-\frac{k}{n}\right)\rho_k\right)$$

$$= \frac{\gamma_0}{n}\left(1 + 2\underbrace{\sum_{k=1}^{n-1}\underbrace{\rho_k}_{\varphi^k}}_{=-1+\sum_{k=0}^{n-1}\varphi^k} - \frac{2}{n}\underbrace{\sum_{k=1}^{n-1}k\underbrace{\rho_k}_{\varphi^k}}_{=\sum_{k=0}^{n-1}k\varphi^k = \varphi\sum_{k=0}^{n-1}k\varphi^{k-1}}\right)$$

$$= \frac{\gamma_0}{n}\left(1 - 2 + 2\sum_{k=0}^{n-1}\varphi^k - \frac{2\varphi}{n}\sum_{k=0}^{n-1}k\varphi^{k-1}\right)$$

$$\overset{②,③}{=} \frac{\gamma_0}{n}\left(-1 + 2\frac{1-\varphi^n}{1-\varphi} - \frac{2\varphi}{n}\frac{d}{d\varphi}\left(\sum_{k=0}^{n-1}\varphi^k\right)\right)$$

$$\overset{②}{=} \frac{\gamma_0}{n}\left(-1 + 2\frac{1-\varphi^n}{1-\varphi} - \frac{2\varphi}{n}\underbrace{\frac{d}{d\varphi}\frac{1-\varphi^n}{1-\varphi}}\right)$$

$$= \frac{(1-\varphi)(-n\varphi^{n-1}) + 1-\varphi^n}{(1-\varphi)^2}$$

$$= \frac{n\varphi^n - n\varphi^{n-1}}{(1-\varphi)^2} + \frac{1-\varphi^n}{(1-\varphi)^2}$$

$$= \frac{\gamma_0}{n}\left(-1 + 2\frac{1-\varphi^n}{1-\varphi} - \frac{2\varphi}{n}\cdot\frac{n\varphi^n - n\varphi^{n-1}}{(1-\varphi)^2} - \frac{2\varphi}{n}\frac{1-\varphi^n}{(1-\varphi)^2}\right)$$

$$= \frac{\gamma_0}{n}\left(-1 + \underbrace{\frac{(2-2\varphi^n)(1-\varphi)}{(1-\varphi)^2} - \frac{2\varphi^{n+1} - 2\varphi^n}{(1-\varphi)^2}} - \frac{2\varphi}{n}\frac{1-\varphi^n}{(1-\varphi)^2}\right)$$

$$= \frac{2 - 2\varphi - 2\varphi^n + 2\varphi^{n+1} - 2\varphi^{n+1} + 2\varphi^n}{(1-\varphi)^2}$$

$$= \frac{2(1-\varphi)}{(1-\varphi)^2} = \frac{2}{1-\varphi}$$

$$= \frac{\gamma_0}{n}\left(\frac{2}{1-\varphi} - \frac{1-\varphi}{1-\varphi} - \frac{2\varphi}{n}\frac{1-\varphi^n}{(1-\varphi)^2}\right)$$

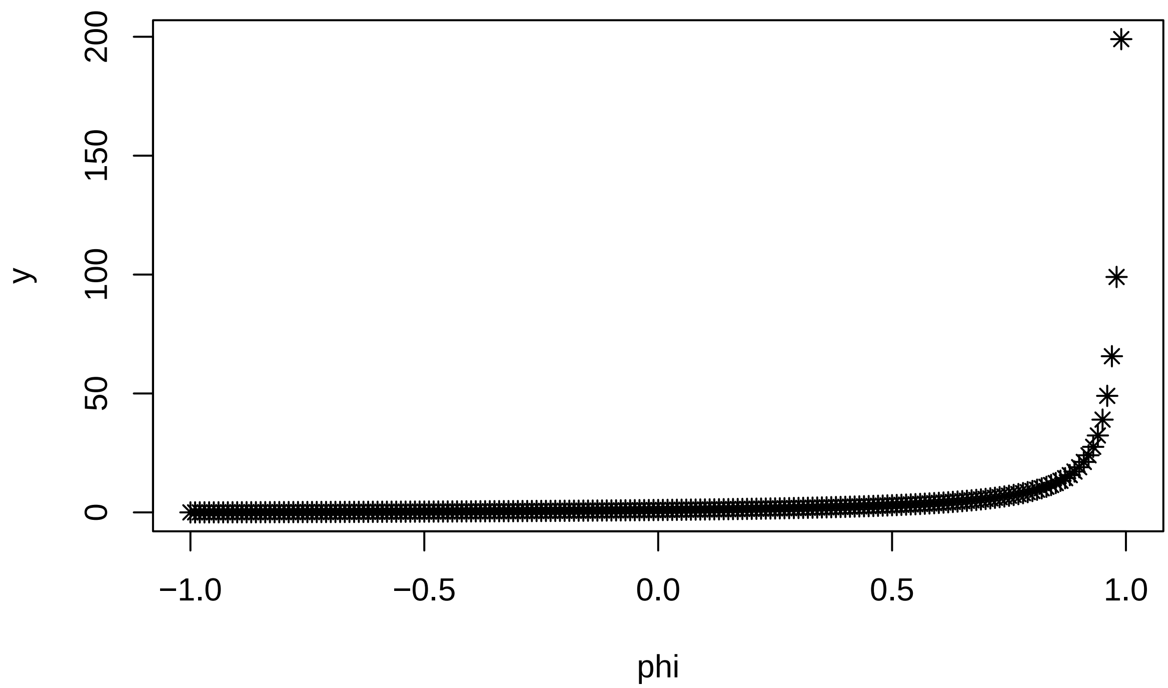
$$= \frac{\gamma_0}{n}\left(\frac{1+\varphi}{1-\varphi} - \frac{2\varphi}{n}\frac{1-\varphi^n}{(1-\varphi)^2}\right)$$

b) $\operatorname{Var}(\bar{Y}) = \frac{\gamma_0}{n} \left(\frac{1+\varphi}{1-\varphi} - \underbrace{\underbrace{\frac{2\varphi}{n}}_{\xrightarrow{n\to\infty} 0} \cdot \underbrace{\frac{1-\varphi^n}{(1-\varphi)^2}}_{\xrightarrow{n\to\infty} \frac{1}{(1-\varphi)^2}}}_{\xrightarrow{n\to\infty} 0} \right)$ $\qquad \varphi^n \xrightarrow{n\to\infty} 0$ because $|\varphi| < 1$

$\Rightarrow$ for n large enough: $\operatorname{Var}(\bar{Y}) \approx \frac{\gamma_0}{n} \frac{1+\varphi}{1-\varphi}$

c)

```r
phi <- seq(from = -1, to = 1, by = 0.01)
y <- (1 + phi)/(1 - phi)
plot(phi, y, pch = 8)
```



The closer ϕ is to -1, the closer the variance is to 0, thus the precision increases. The closer ϕ is to 1, the more the variance increases, and thus the precision decreases.