

Di Zhen (1717719)

dizhen

2020.3.20.

Set working directory:

```
setwd("D:/git/DPH112-xjtlw/week04")
```

Data

```
IGE <- c(24.87,12.90,9.87,8.74,6.88,5.90,4.85,3.53,2.25,2.14,1.94,1.29,0.94,0.91,0.55,0.30,0.14,0.11,0.05,0.041034,0.050909,0.046,0.039032,0.050909,0.042142,0.055,4.333333,0.55,0.050909,0.44615)
SKIN <- c(0.055,0.041034,0.050909,0.046,0.039032,0.050909,0.042142,0.055,4.333333,0.55,0.050909,0.44615)
mydata <- data.frame(IGE,SKIN)
str(mydata)
```

```
## 'data.frame':    23 obs. of  2 variables:
##  $ IGE : num  24.87 12.9 9.87 8.74 6.88 ...
##  $ SKIN: num  0.055 0.041 0.0509 0.046 0.039 ...
```

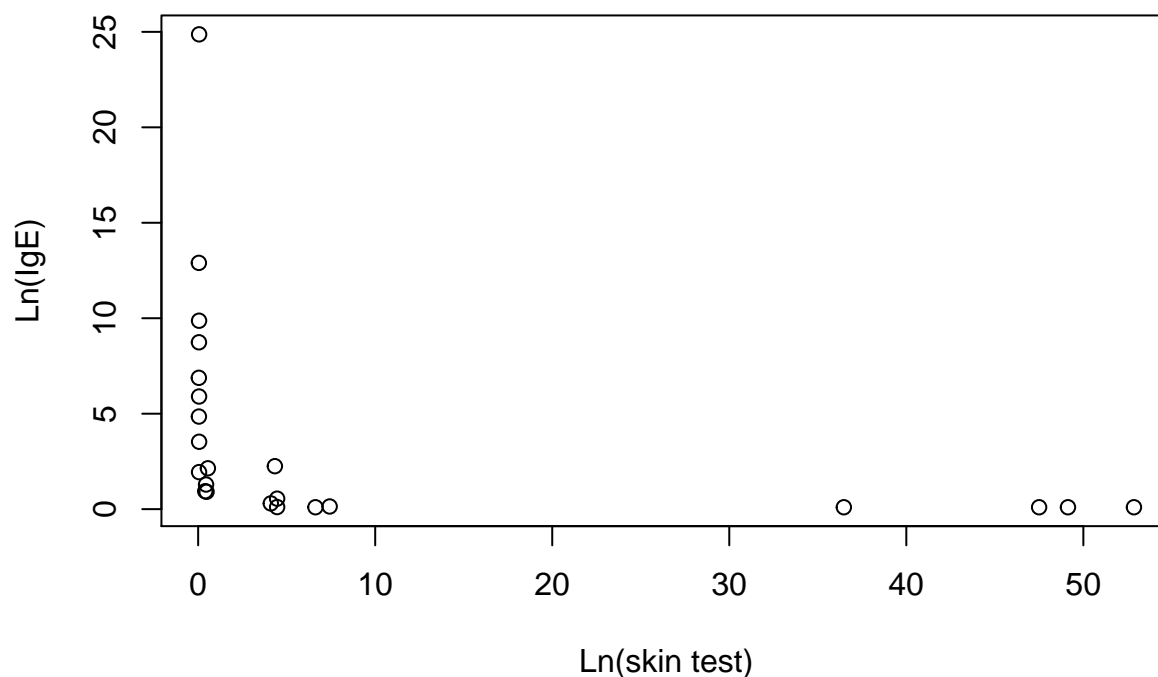
```
head(mydata)
```

```
##      IGE      SKIN
## 1 24.87 0.055000
## 2 12.90 0.041034
## 3  9.87 0.050909
## 4  8.74 0.046000
## 5  6.88 0.039032
## 6  5.90 0.050909
```

Visualize the data:

```
plot(IGE ~ SKIN, data = mydata,
     main = "Immunoglobulin E (IgE) and skin test levels\n following exposure to Lol p 5.in natural log",
     xlab = "Ln(skin test)",
     ylab = "Ln(IgE)")
```

Immunoglobulin E (IgE) and skin test levels following exposure to Lol p 5.in natural log units



Question 1

A properly-formatted scatterplot of the data:

```
if(!require(car)){install.packages("car")}
```

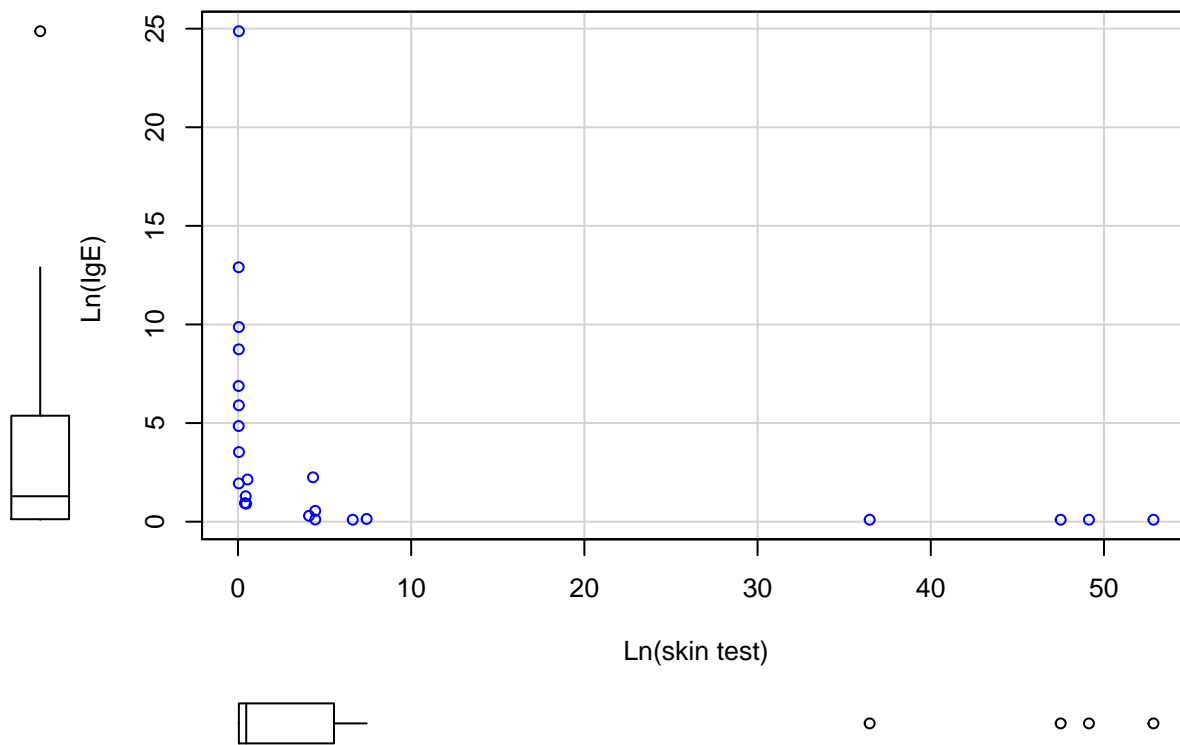
```
## Loading required package: car
```

```
## Warning: package 'car' was built under R version 3.6.3
```

```
## Loading required package: carData
```

```
library(car)
scatterplot(IGE ~ SKIN, data = mydata, smooth = FALSE, regLine=FALSE,
  main = "Immunoglobulin E (IgE) and skin test levels\n following exposure to Lol p 5.in natural log",
  ylab = "Ln(IgE) ",
  xlab = "Ln(skin test)")
```

Immunoglobulin E (IgE) and skin test levels following exposure to Lol p 5.in natural log units



Question 2

```
mydata.LM <- lm(IGE ~ SKIN,data = mydata)
summary(mydata.LM)
```

```
##
## Call:
## lm(formula = IGE ~ SKIN, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.339  -3.930  -1.436   1.129  19.904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.97247    1.33380   3.728  0.00124 **
## SKIN        -0.11725    0.06754  -1.736  0.09722 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.599 on 21 degrees of freedom
## Multiple R-squared:  0.1255, Adjusted R-squared:  0.08385
## F-statistic: 3.014 on 1 and 21 DF,  p-value: 0.09722
```

Interpretation: The estimated intercept is 4.972(p=0.001), the estimated slope is -0.117 (p=0.097).

Question 3

Test the null hypothesis of $H_0 : \beta_0 = 0$: As $\alpha = 0.1$, $p\text{-value} = 0.001$, $0.001 < 0.1$. We reject the null hypothesis and conclude that $\beta_0 \neq 0$.

Test the null hypothesis of $H_0 : \beta_1 = 0$: As $\alpha = 0.1$, $p\text{-value} = 0.097$, $0.097 < 0.1$. We reject the null hypothesis and conclude that $\beta_1 \neq 0$.

Interpretation of p-value: there is 0.1% probability that we would observe at least as extreme a result as this for the intercept estimate under the null hypothesis. There is 9.7% probability that we would observe at least as extreme a result as this for the slope estimate under the null hypothesis.

Question 4

```
confint(mydata.LM, level = 0.90)
```

```
##                5 %                95 %  
## (Intercept)  2.6773390  7.267596558  
## SKIN        -0.2334721 -0.001028099
```

The 90% confidence interval for intercept is from 2.677 to 7.268. We are 90% confident that the intercept estimate lies between the interval 2.677 and 7.268, because on repeated sampling, 90% of intervals constructed in this manner will contain the true intercept.

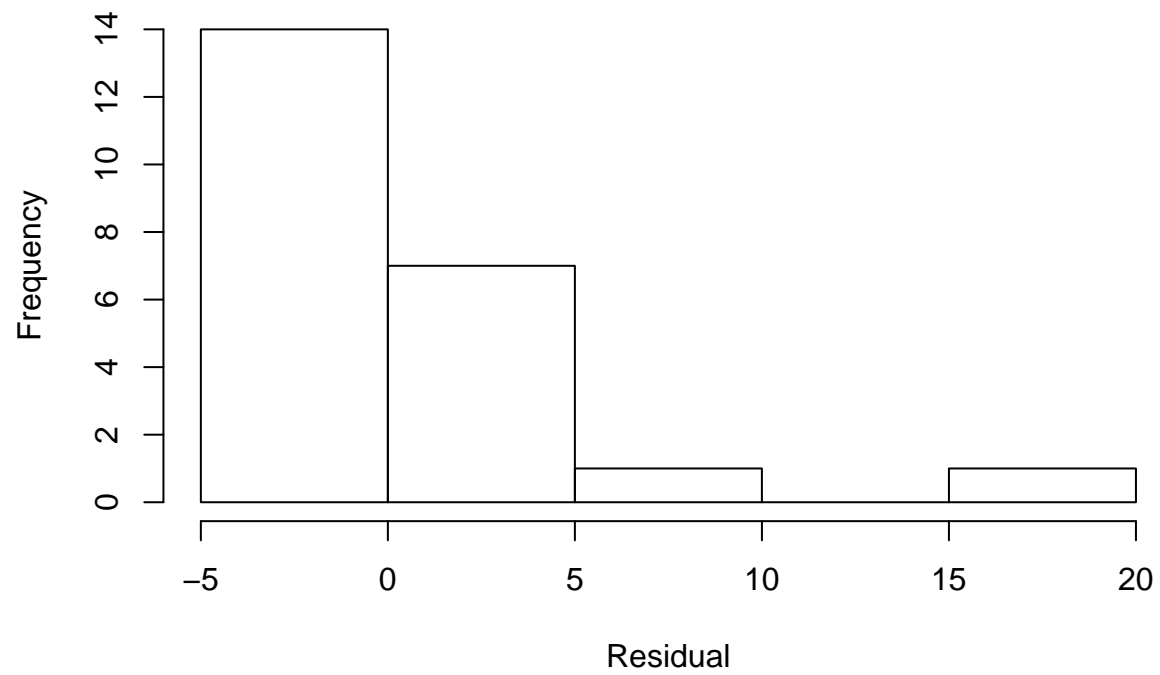
The 90% confidence interval for slope is from -0.233 to -0.001. We are 90% confident that the slope estimate lies between the interval -0.233 and -0.001, because on repeated sampling, 90% of intervals constructed in this manner will contain the true slope.

Question 5

1. Assessing normality

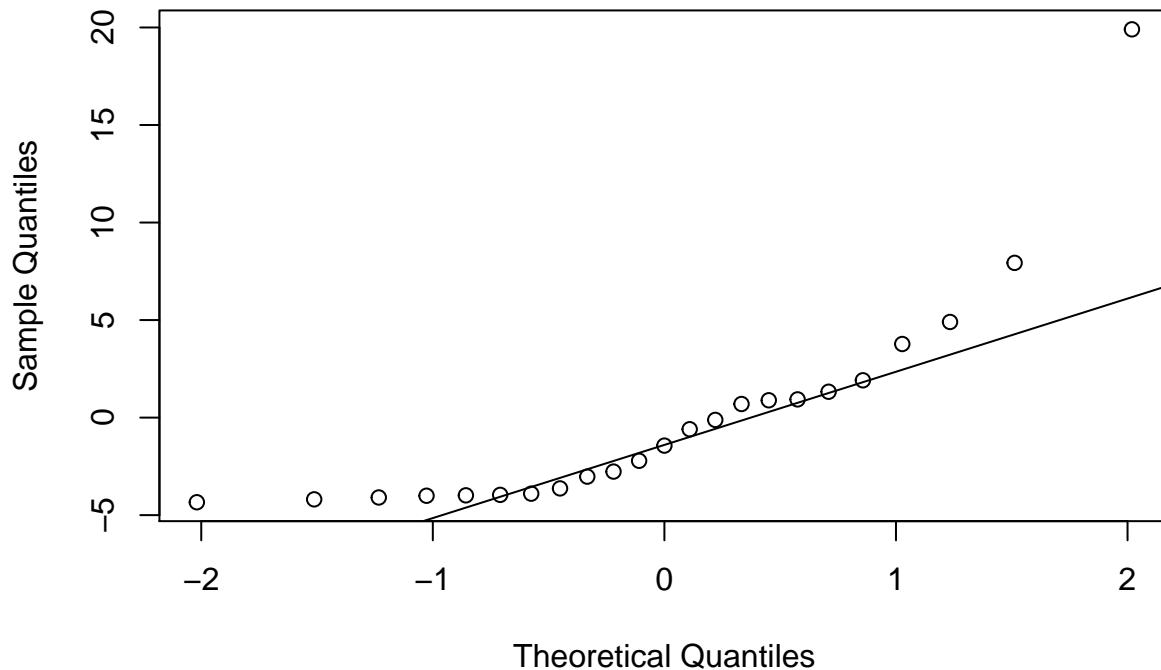
```
mydata.RESID <- resid(mydata.LM)  
hist(mydata.RESID,  
     main = "Histogram of Ordinary Residuals",  
     xlab = "Residual")
```

Histogram of Ordinary Residuals



```
qqnorm(mydata.RESID,  
       main = "Normal Quantile-Quantile Plot of Ordinary Residuals")  
qqline(mydata.RESID)
```

Normal Quantile–Quantile Plot of Ordinary Residuals



Assessment. The residuals deviate substantially from normality.

Conclusion. The assumption of normality is not met.

2. Assessing linearity

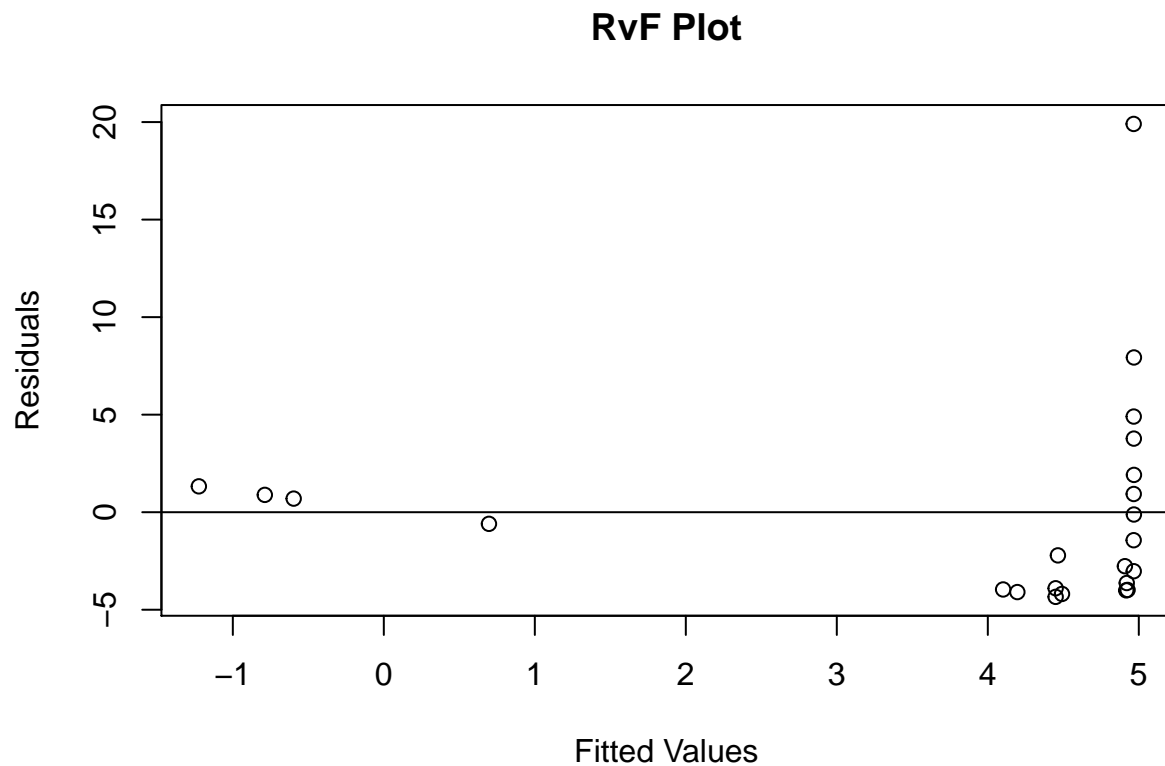
```
FITTED <- predict(mydata.LM,type = "response")
RESID <- resid(mydata.LM)
mydata2 <- cbind(mydata,FITTED, RESID)
str(mydata2)
```

```
## 'data.frame':  23 obs. of  4 variables:
## $ IGE    : num  24.87 12.9 9.87 8.74 6.88 ...
## $ SKIN   : num  0.055 0.041 0.0509 0.046 0.039 ...
## $ FITTED: num  4.97 4.97 4.97 4.97 4.97 ...
## $ RESID  : num  19.9 7.93 4.9 3.77 1.91 ...
```

```
head(mydata2)
```

```
##      IGE      SKIN  FITTED      RESID
## 1 24.87 0.055000 4.966019 19.9039810
## 2 12.90 0.041034 4.967657  7.9323435
## 3  9.87 0.050909 4.966499  4.9035013
## 4  8.74 0.046000 4.967074  3.7729257
## 5  6.88 0.039032 4.967891  1.9121087
## 6  5.90 0.050909 4.966499  0.9335013
```

```
plot(RESID ~ FITTED, data = mydata2,
     main = "RvF Plot",
     ylab = "Residuals",
     xlab = "Fitted Values")
abline(h=0)
```

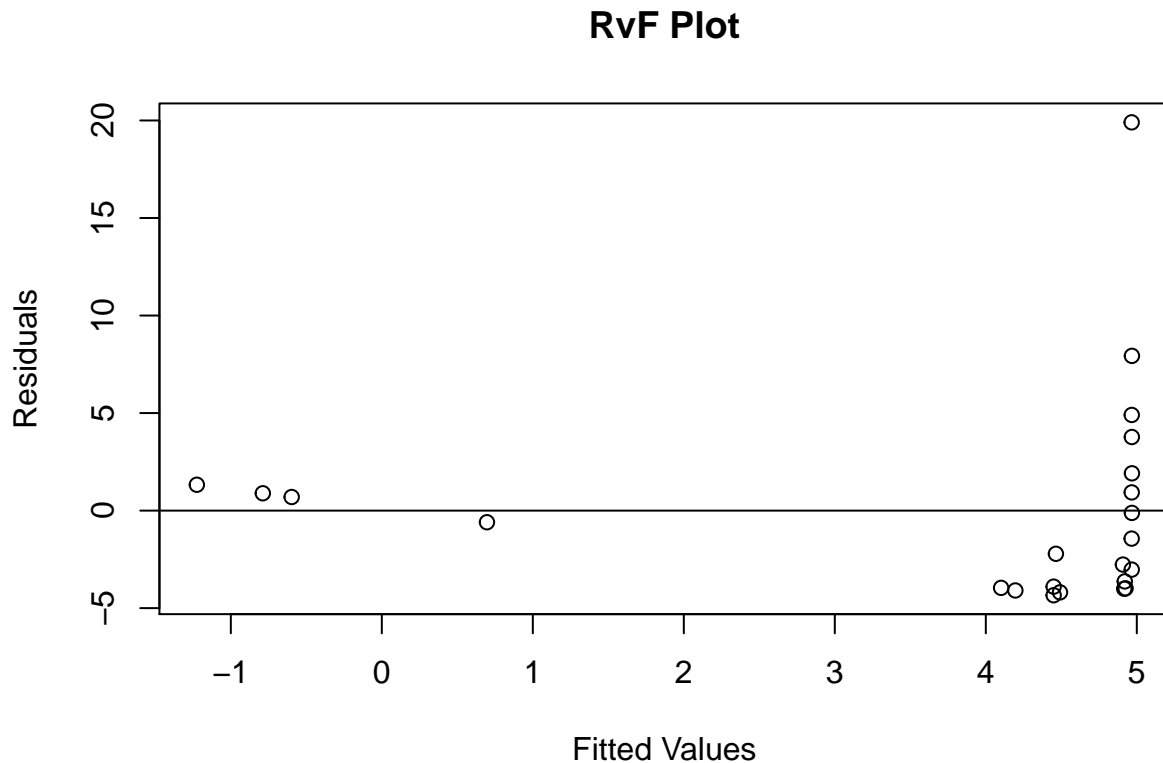


Assessment. The spread of residuals shows obvious pattern.

Conclusion. The assumption of linearity is not met.

3. Assessing homoskedasticity

```
plot(RESID ~ FITTED, data = mydata2,
     main = "RvF Plot",
     ylab = "Residuals",
     xlab = "Fitted Values")
abline(h=0)
```



Assessment. The RvF plot show evidence of some patterns

Conclusion. The assumption of homoskedasticity is not met.

4. Assessing independence

From the design of the experiment, we know that statistical independence of skin test presents.

Question 6

```
predict(mydata.LM, data.frame(SKIN = 0.1), interval = "prediction", level = 0.90)
```

```
##          fit      lwr      upr
## 1 4.960743 -4.94162 14.86311
```

IgE level on an individual in natural log units is 4.961 if the skin test level in natural log units is 0.1.90% intervals is from -4.942 to 14.863.

Question 7


```
predict(mydata.LM, data.frame(SKIN = 0.1), interval = "confidence", level = 0.90)
```

```
##          fit      lwr      upr  
## 1 4.960743 2.671212 7.250273
```

The mean IgE level in natural log units is 4.961, if the skin test level in natural log units is 0.1. 90% intervals is from 2.671 to 7.250.

THE END