In [61]:

```
#Сгенерируйте в Rstudio свои индивидуальные данные, используя приведенную ниже схему:
install.packages("MASS")
library(MASS)
options(digits = 4)
set.seed(70)
sigma<-matrix(c(1,0.8,0.4,-0.6,0.8,1,0.7,-0.4,0.4,0.7,1,-0.1,-0.6, -0.4, -0.1, 1), nrow

mean <- c(7 * 15, 7 ^ 2, 7 + 20, 60 - 7)
mydata <- mvrnorm(300, mean, sigma)
mydata <- as.data.frame(mydata)
names(mydata) <- c("y","x1","x2","x3")
#Проверьте, полученный набор данных
head(mydata, n = 5)
#Далее можете сохранить свой набор данных

```

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)
```

A data.frame: 5 × 4

	y	x1	x2	x3
	<dbl>	<dbl>	<dbl>	<dbl>
1	106.5	50.45	27.87	51.98
2	104.8	48.17	26.02	51.81
3	104.2	48.31	25.64	53.77
4	106.0	49.19	26.44	52.14
5	106.2	49.72	27.97	52.99

In [62]:

```
write.table(mydata, file = "Esakov.txt", sep = "\t")
```

In [63]:

```
install.packages("lmtest")
library("lmtest")
```

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)
```

Выборочная парная:

$$r_{xy} = \frac{\widehat{Cov}(x,y)}{S_x S_y} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}}.$$

In [64]:

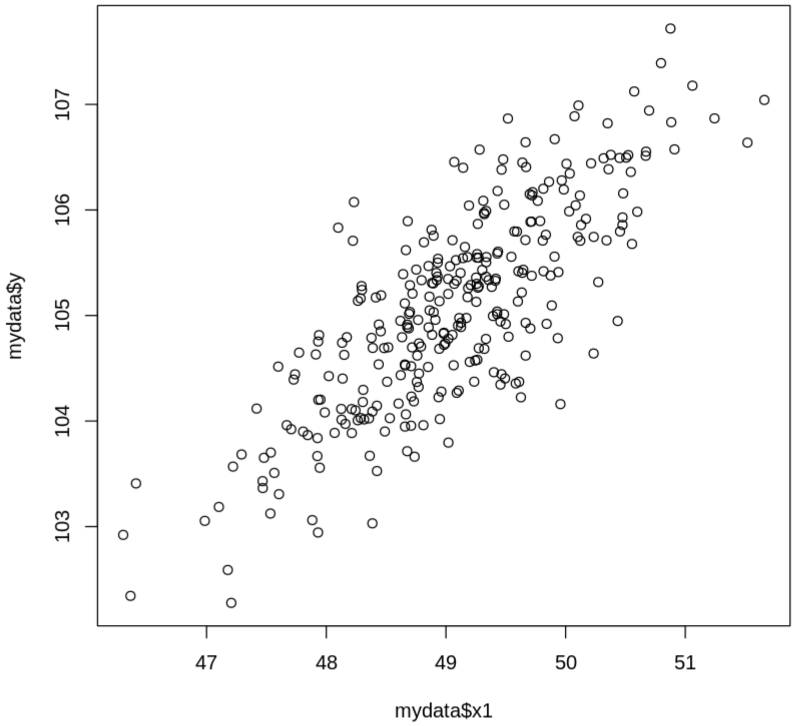
```
cor_x1_y = cor(mydata$y, mydata$x1)
cor_y_x2 = cor(mydata$y, mydata$x2)
cor_y_x3 = cor(mydata$y, mydata$x3)
cor_x1_x2 = cor(mydata$x1, mydata$x2)
cor_x1_x3 = cor(mydata$x1, mydata$x3)
cor_x3_x2 = cor(mydata$x3, mydata$x2)

print(c('Cor x1 and y', cor_x1_y))
print(c('Cor x1 and y', cor_y_x2))
print(c('Cor x1 and y', cor_y_x3))
print(c('Cor x1 and y', cor_x1_x2))
print(c('Cor x1 and y', cor_x1_x3))
print(c('Cor x1 and y', cor_x3_x2))
```

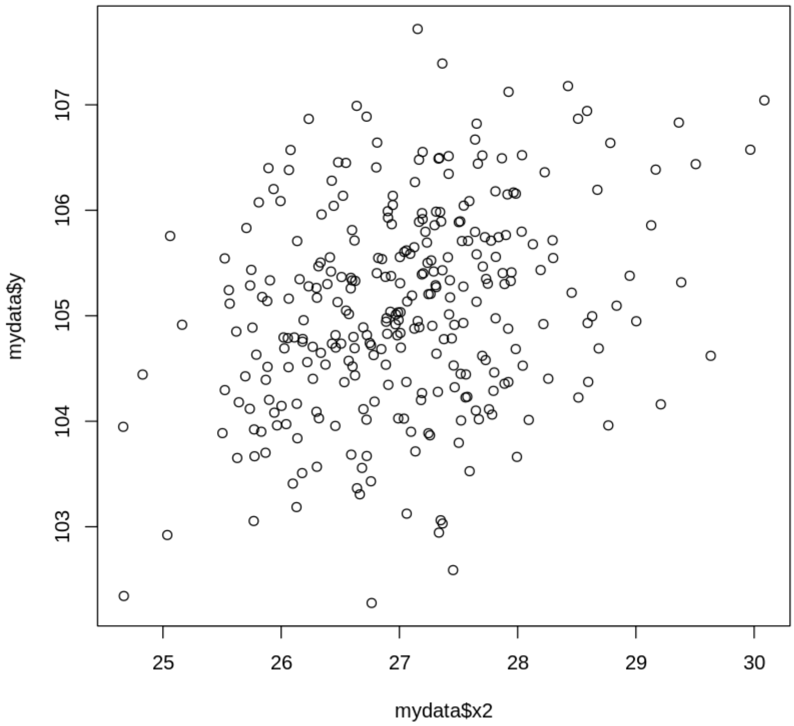
```
[1] "Cor x1 and y"      "0.794323059454161"
[1] "Cor x1 and y"      "0.289133294383633"
[1] "Cor x1 and y"       "-0.622354186746941"
[1] "Cor x1 and y"      "0.613568009949477"
[1] "Cor x1 and y"       "-0.455750485719631"
[1] "Cor x1 and y"        "-0.0948200745315119"
```

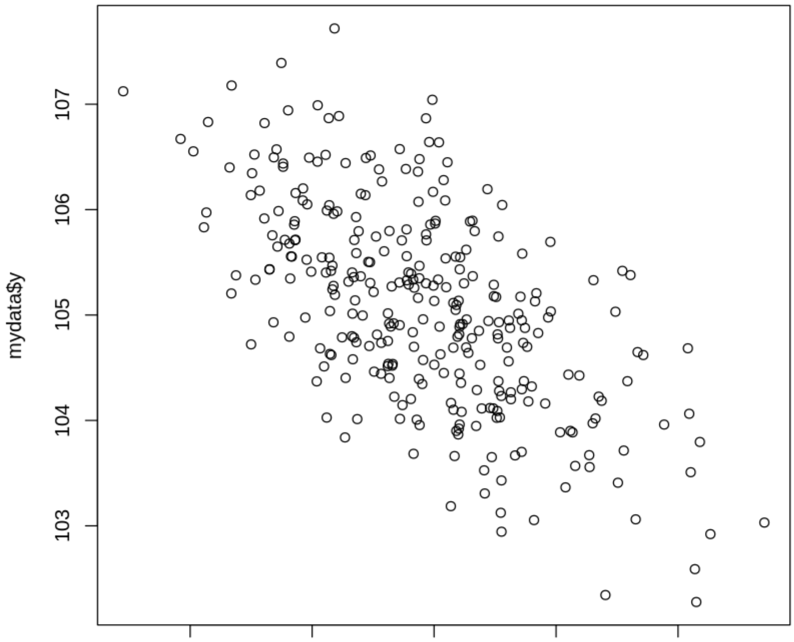
Наибольшая корреляция присутствует между y и x1

In [65]:

```
plot(mydata$x1, mydata$y)
plot(mydata$x2, mydata$y)
plot(mydata$x3, mydata$y)
```







Коэффициент детерминации:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}.$$

Скорректированный:

$$R_{adj}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k-1}.$$

In [66]:

```
#2
n = 300
m = 3
lmm = lm(mydata)
multiple = summary(lmm)
multiple
```

```
Call:
lm(formula = mydata)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2898 -0.2805 -0.0077  0.3186  1.4178

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  84.0773     3.1949   26.32  < 2e-16 ***
x1            0.8811     0.0462   19.07  < 2e-16 ***
x2           -0.2470     0.0398   -6.20  1.9e-09 ***
x3           -0.2939     0.0365   -8.06  1.9e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.486 on 296 degrees of freedom
Multiple R-squared:  0.749,     Adjusted R-squared:  0.747
F-statistic:  295 on 3 and 296 DF,  p-value: <2e-16
```

$$\hat{y_i} = \underset{(3.1949)}{84.0773} + \underset{(0.0462)}{0.8811}x_{1i} - \underset{(0.0389)}{0.247}x_{2i} - \underset{(0.0365)}{0.2939}x_{3i}$$

$$R^2 = 0.749 \; R^2_{adj} = 0.747 \quad F = 295$$

$$Se = 0.486 \quad A = 0.364$$

In [67]:

```
lmp = lm(y~x1,data = mydata)
pair = summary(lmp)
pair
```

```
Call:
lm(formula = y ~ x1, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.6780 -0.3764  0.0173  0.3925  1.7175 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   62.945      1.867    33.7   <2e-16 ***
x1             0.859      0.038    22.6   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.588 on 298 degrees of freedom
Multiple R-squared:  0.631,	Adjusted R-squared:  0.63 
F-statistic:  509 on 1 and 298 DF,  p-value: <2e-16
```

$$\widehat{y_i} = \underset{(1.876)}{62.945} + \underset{(0.038)}{0.859}x_{1i}$$

$$R^2 = 0.631\ R_{adj}^2 = 0.63 \quad F = 509$$

$$Se = 0.588 \quad A = 0.445$$

R^2 имеет недостаточно хорошее значение, значит значение y средне обусловлено факторами модели.

Близость значений R^2 и R_adj^2 обусловлено достаточным колчиеством наблюдений в соотношении с количеством факторов

Ошибка аппроксимации

$$A = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(y_i - \hat{y_i})}{y_i}\right| * 100\%$$

In [68]:

```
#3
#ошибка е
e = mydata$y - predict(lmm)
e1 = mydata$y - predict(lmp)

sum(e)
sum(e1)

#средняя относительная ошибка аппроксимации
a = sum(abs(e / mydata$y)) / length(mydata$y) * 100
a1 = sum(abs(e1 / mydata$y)) / length(mydata$y) * 100

a
a1
```

-1.52056145452661e-12

2.18847162614111e-12

0.364280908658051

0.445197644671828

Ошибка аппроксимации в процентах очень мала что свидетельствует о точности прогноза модели

t-тест Стьюдента и гипотезы

$$H0:\ \beta_j = 0;$$
$$H1:\ \beta_j \neq 0.$$

Формула t-значения

$$t_{\beta_j} = \frac{\hat{\beta}_j}{\hat{\sigma}_{\beta_j}},$$
$$t\text{табл}(\alpha;\ n - k - 1)$$

Ф-критерий Фишера и гипотезы

$$H0:\ \beta_1 = \ldots = \beta_k = 0,$$
$$H1:\ \beta_1^2 + \ldots + \beta_k^2 > 0.$$

Формула занчения критерия Фишера

$$F_{\text{набл}} = \frac{R^2/k}{(1-R^2)\big/(n-k-1)},$$
$$F_{\text{табл}}(\alpha; k; n-k-1),$$

In [69]:

```
#4
print(c('t-test table value', qt(0.975, 296)))

print(c('F-test table value(mult)', qf(0.95, 3, 296)))
print(c('F-test table value(pair)', qf(0.95, 1, 298)))
```

```
[1] "t-test table value" "1.96801072755025"
[1] "F-test table value(mult)" "2.63510620032357"
[1] "F-test table value(pair)" "3.87285281146316"
```

Все значения t-test для параметров моделей больше по модулю табличного значения, значит гипотеза H0 отвергается и парметры значимы

Значения Fкритерия Фишера для модели больше табличного для заданного кол-ва наблюдений и параметров, следовательно гипотеза H0 отвергается и модель значима.

$$\bar{Э}_j = y'(\bar{x}_j) \cdot \frac{\bar{x}_j}{\bar{y}}$$

(для линейной модели: $\bar{Э}_j = \hat{\beta}_j \cdot \frac{\bar{x}_j}{\bar{y}}$).

$$\tilde{\beta}_j = \hat{\beta}_j \frac{S_{x_{ij}}}{S_{y_i}}.$$

$$\Delta_j = r_{y_i, x_{ij}} \frac{\hat{\beta}_j}{R^2}.$$

In [70]:

```
#5
elastic = lmm$coefficients['x1'] * mean(mydata$x1)  /mean(mydata$y)
elastic

betta = lmm$coefficients['x1'] * sd(mydata$x1) / sd(mydata$y)
betta

delta = cor_x1_y * lmm$coefficients['x1'] / multiple$r.squared
delta

lmm$coefficients['x2'] * mean(mydata$x2) / mean(mydata$y)
lmm$coefficients['x3'] * mean(mydata$x3) / mean(mydata$y)
lmm$coefficients['x2'] * sd(mydata$x2) / sd(mydata$y)
lmm$coefficients['x3'] * sd(mydata$x3) / sd(mydata$y)
cor_y_x2 * lmm$coefficients['x2'] / multiple$r.squared
cor_y_x3 * lmm$coefficients['x3'] / multiple$r.squared
```

x1: 0.411424982820251

x1: 0.81511490824679

x1: 0.93430361004555

x2: -0.0635863396937481

x3: -0.148026663427373

x2: -0.2369122335324

x3: -0.273329207037122

x2: -0.0953430202337788

x3: 0.244206177083396

In [71]:

```
elastic1 = lmp$coefficients['x1'] * mean(mydata$x1) / mean(mydata$y)
elastic1

betta1 = lmp$coefficients['x1'] * sd(mydata$x1) / sd(mydata$y)
betta1

delta1 = cor_x1_y * lmp$coefficients['x1'] / multiple$r.squared
delta1
```

x1: 0.400930405987264

x1: 0.79432305945416

x1: 0.910471510804158

$$F_{\text{набл}} = \frac{RSS_R - RSS_{UR}}{RSS_{UR}} \cdot \frac{n-m}{q} = \frac{R_{UR}^2 - R_R^2}{1 - R_{UR}^2} \cdot \frac{n-m}{q}$$

In [72]:

```
#6
Rur = multiple$r.squared #множественной
Rr = pair$r.squared #парной
q = 2 #количество удаляемых факторов

f = ((Rur - Rr) * (n - m)) / ((1 - Rur) * q)
f
ftabl = qf(0.95, 1, 18)
if (f > ftabl){
  print('H0 отвергается, длинная круче')
}else{
  print('H0 принимается, короткая круче')
}
```

69.9063089053288

```
[1] "H0 отвергается, длинная круче"
```

In [73]:

```
#7
library("lmtest")
resettest(lmm, power = 2:3, data = mydata)
```

```
	RESET test

data:  lmm
RESET = 0.14, df1 = 2, df2 = 294, p-value = 0.9
```

Коэффициент детерминации (квази-R^2).

После построения линеаризированной модели необходимо перейти к исходной нелинейной и рассчитать на ее основе $\hat{y}_i$, e_i и

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

$$A = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\%.$$

In [74]:

```
#8 полином второй степени
lm2 = lm(y~poly(x2, 2), data = mydata)
sm2 = summary(lm2)
sm2

#ошибка е
e2 = mydata$y - predict(lm2)
sum(e2)

#средняя относительная ошибка аппроксимации
a2 = sum(abs(e2 / mydata$y)) / length(mydata$y) * 100
a2

#квази-R2
r22 = 1 - sum(e2 ^ 2)/(sum((mydata$y - mean(mydata$y)) ^ 2))
r22
```

```
Call:
lm(formula = y ~ poly(x2, 2), data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7227 -0.6626  0.0106  0.6375  2.5979 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  105.0719     0.0535 1962.67  < 2e-16 ***
poly(x2, 2)1   4.8283     0.9273    5.21  3.6e-07 ***
poly(x2, 2)2  -0.4324     0.9273   -0.47     0.64    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.927 on 297 degrees of freedom
Multiple R-squared:  0.0843,	Adjusted R-squared:  0.0781 
F-statistic: 13.7 on 2 and 297 DF,  p-value: 2.1e-06
```

9.70601377048297e-12

0.703205018237429

0.0842686068805899

In [75]:

```
#показательная
lmp2 = lm(log10(y)~x2, data = mydata)
smp2 = summary(lmp2)
smp2

#ошибка e
ep2 = mydata$y - 10 ^ predict(lmp2)
sum(ep2)

#средняя относительная ошибка аппроксимации
ap2 = sum(abs(ep2 / mydata$y)) / length(mydata$y) * 100
ap2

#квази-R2
r2p2 = 1 - sum(ep2 ^ 2) / (sum((mydata$y - mean(mydata$y)) ^ 2))
r2p2
```

```
Call:
lm(formula = log10(y) ~ x2, data = mydata)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.011338 -0.002724  0.000058  0.002551  0.010691 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.987778   0.006471  307.18  < 2e-16 ***
x2          0.001246   0.000239    5.21  3.5e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.00383 on 298 degrees of freedom
Multiple R-squared:  0.0835,	Adjusted R-squared:  0.0804 
F-statistic: 27.1 on 1 and 298 DF,  p-value: 3.54e-07
```

1.21684424325898

0.703347611199877

0.0835500386410406

In [76]:

```
#гиперболическая
X2 = 1 / mydata$x2
#X2
lmg2 = lm(y~X2, data = mydata)
smg2 = summary(lmg2)
smg2

#ошибка e
eg2 = mydata$y - predict(lmg2)
sum(eg2)

#средняя относительная ошибка аппроксимации
ag2 = sum(abs(eg2 / mydata$y)) / length(mydata$y) * 100
ag2

#квази-R2
r2g2 = 1 - sum(eg2 ^ 2) / (sum((mydata$y - mean(mydata$y)) ^ 2))
r2g2
```

```
Call:
lm(formula = y ~ X2, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7173 -0.6597  0.0118  0.6294  2.6050 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   113.33       1.58   71.88  < 2e-16 ***
X2           -222.97      42.57   -5.24  3.1e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.926 on 298 degrees of freedom
Multiple R-squared:  0.0843,	Adjusted R-squared:  0.0812 
F-statistic: 27.4 on 1 and 298 DF,  p-value: 3.07e-07
```

9.16600129130529e-12

0.703192152409533

0.0843059024211946

In [77]:

```
#9
lms1 = lm(log10(y)~log10(x1)*log10(x2), data = mydata)
sms1 = summary(lms1)
sms1
```

```
Call:
lm(formula = log10(y) ~ log10(x1) * log10(x2), data = mydata)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.006515 -0.001358 -0.000018  0.001574  0.006052 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)            2.509      2.163    1.16     0.25
log10(x1)             -0.216      1.278   -0.17     0.87
log10(x2)             -0.930      1.512   -0.62     0.54
log10(x1):log10(x2)    0.499      0.893    0.56     0.58

Residual standard error: 0.00222 on 296 degrees of freedom
Multiple R-squared:  0.694,	Adjusted R-squared:  0.691 
F-statistic:  224 on 3 and 296 DF,  p-value: <2e-16
```

In [78]:

```
lmp1 = lm(log10(y)~x2^2*x3^3, data = mydata)
smp1 = summary(lmp1)
smp1
```

```
Call:
lm(formula = log10(y) ~ x2^2 * x3^3, data = mydata)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.007482 -0.002086  0.000065  0.002141  0.008753 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.355390   0.281888    8.36  2.6e-15 ***
x2          -0.007096   0.010385   -0.68     0.49    
x3          -0.006817   0.005321   -1.28     0.20    
x2:x3        0.000153   0.000196    0.78     0.44    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003 on 296 degrees of freedom
Multiple R-squared:  0.442,	Adjusted R-squared:  0.437 
F-statistic: 78.3 on 3 and 296 DF,  p-value: <2e-16
```

In [79]:

```
X1 = 1 / mydata$x1
X2 = 1 / mydata$x2
lmg2 = lm(y~X1+X2, data = mydata)
smg2 = summary(lmg2)
smg2
```

```
Call:
lm(formula = y ~ X1 + X2, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.5666 -0.3423  0.0095  0.3832  1.4692 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   148.32       1.71    86.8   <2e-16 ***
X1          -2557.95     105.54   -24.2   <2e-16 ***
X2            240.60      31.25     7.7    2e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.537 on 297 degrees of freedom
Multiple R-squared:  0.692,	Adjusted R-squared:  0.69 
F-statistic:  334 on 2 and 297 DF,  p-value: <2e-16
```