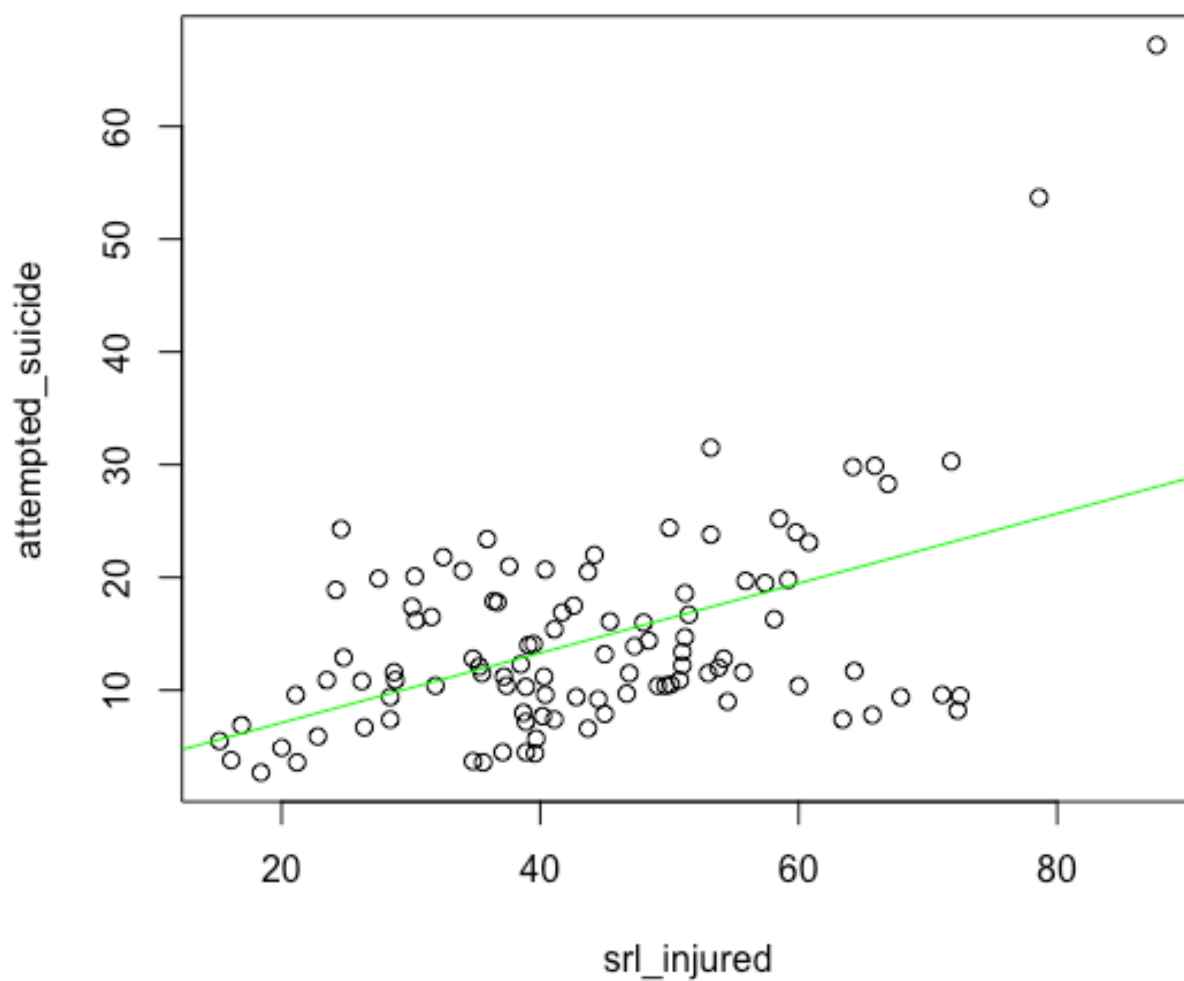
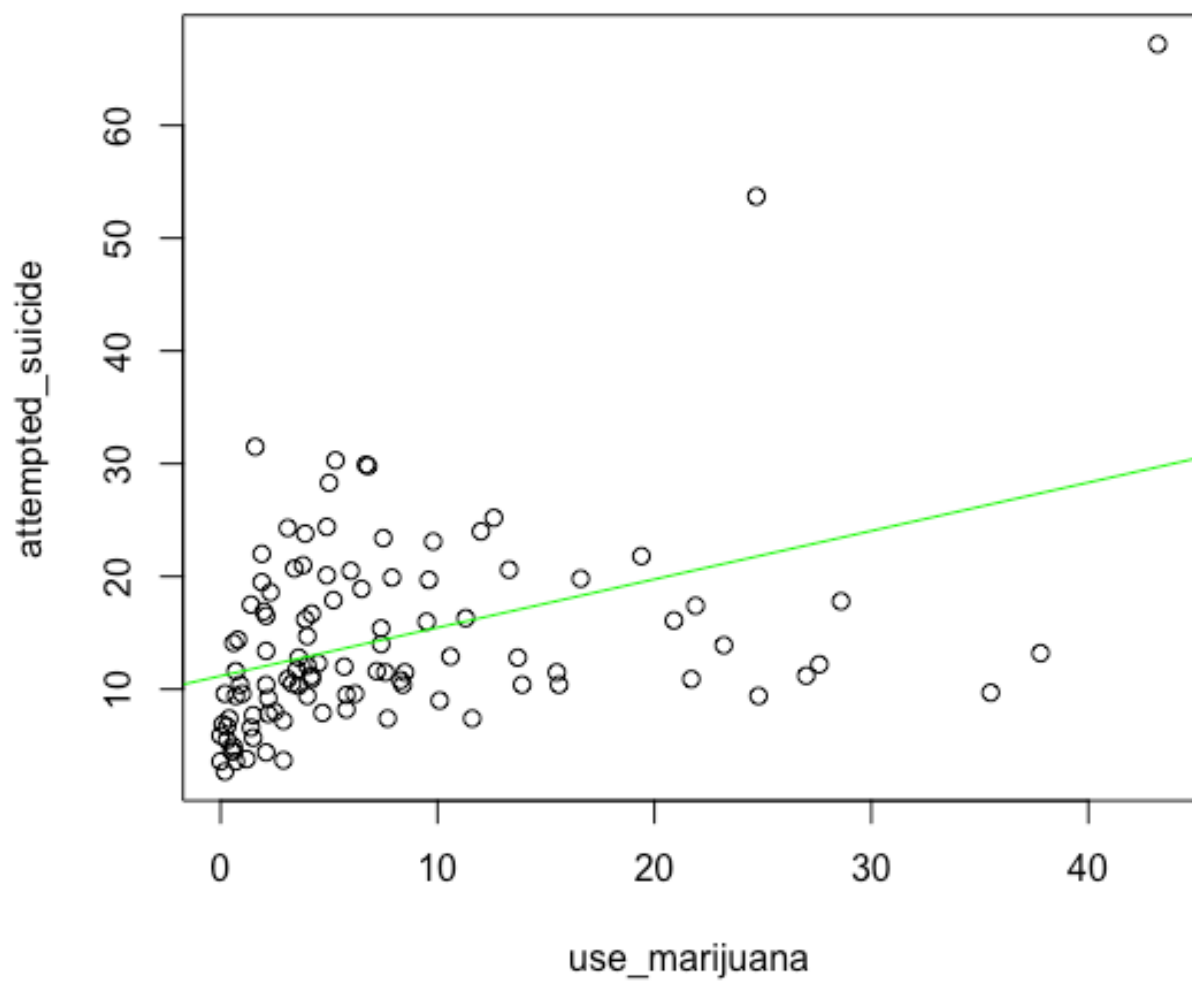
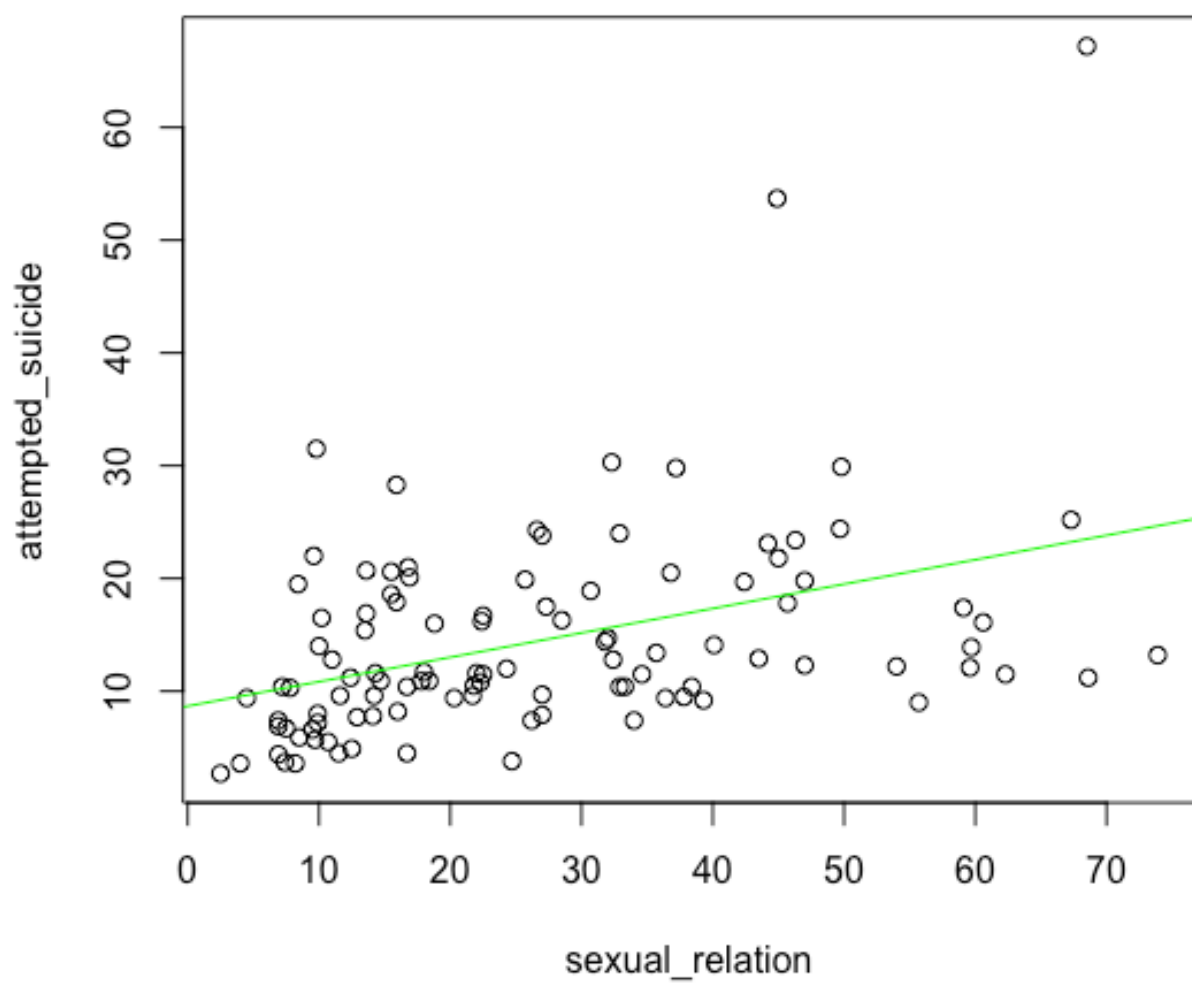


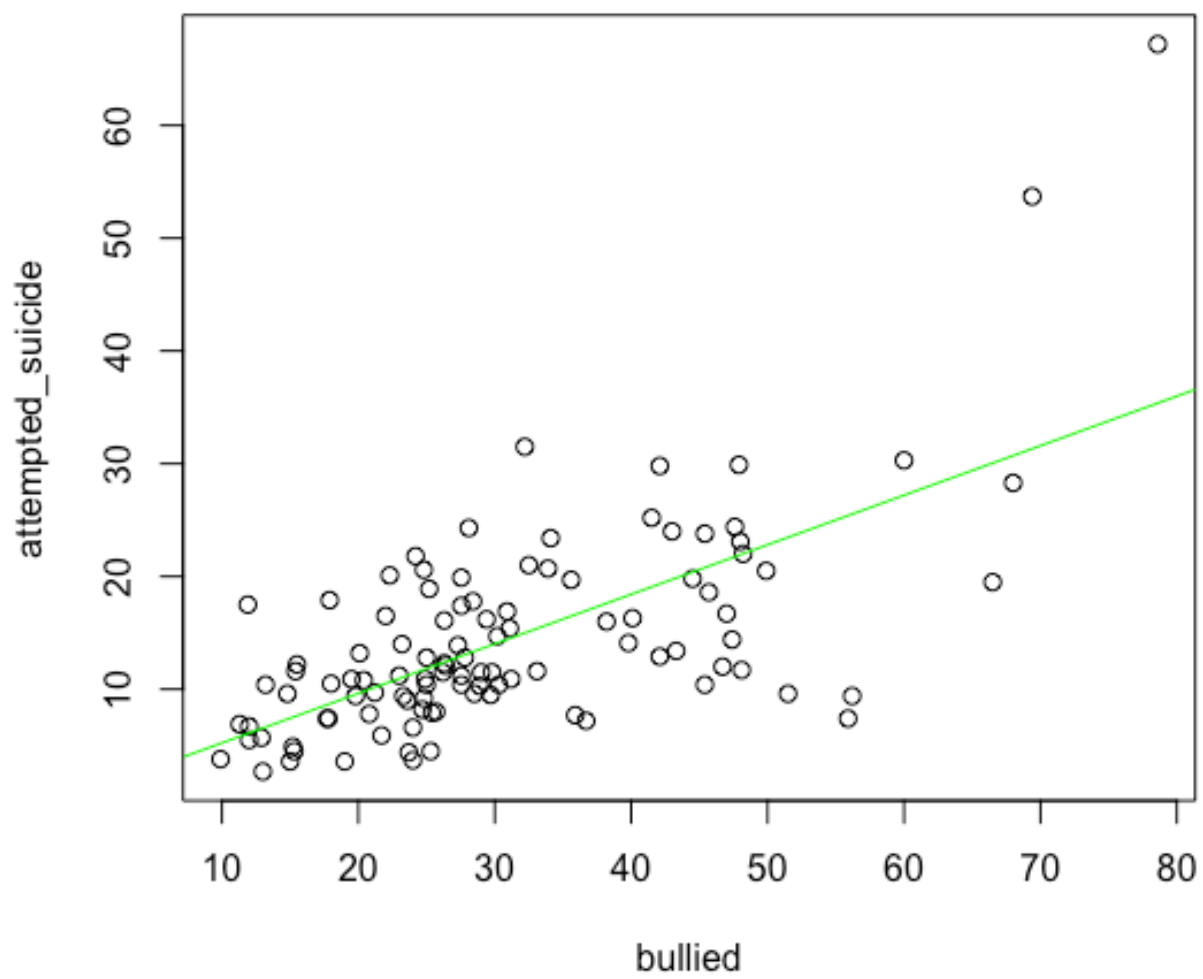
Для факторів x_1 , x_2 , x_3 , x_4 , x_5 та y побудувати в R модель mi : $\text{lm}(y \sim x_i)$ та виконати наступні завдання (A)-(L):

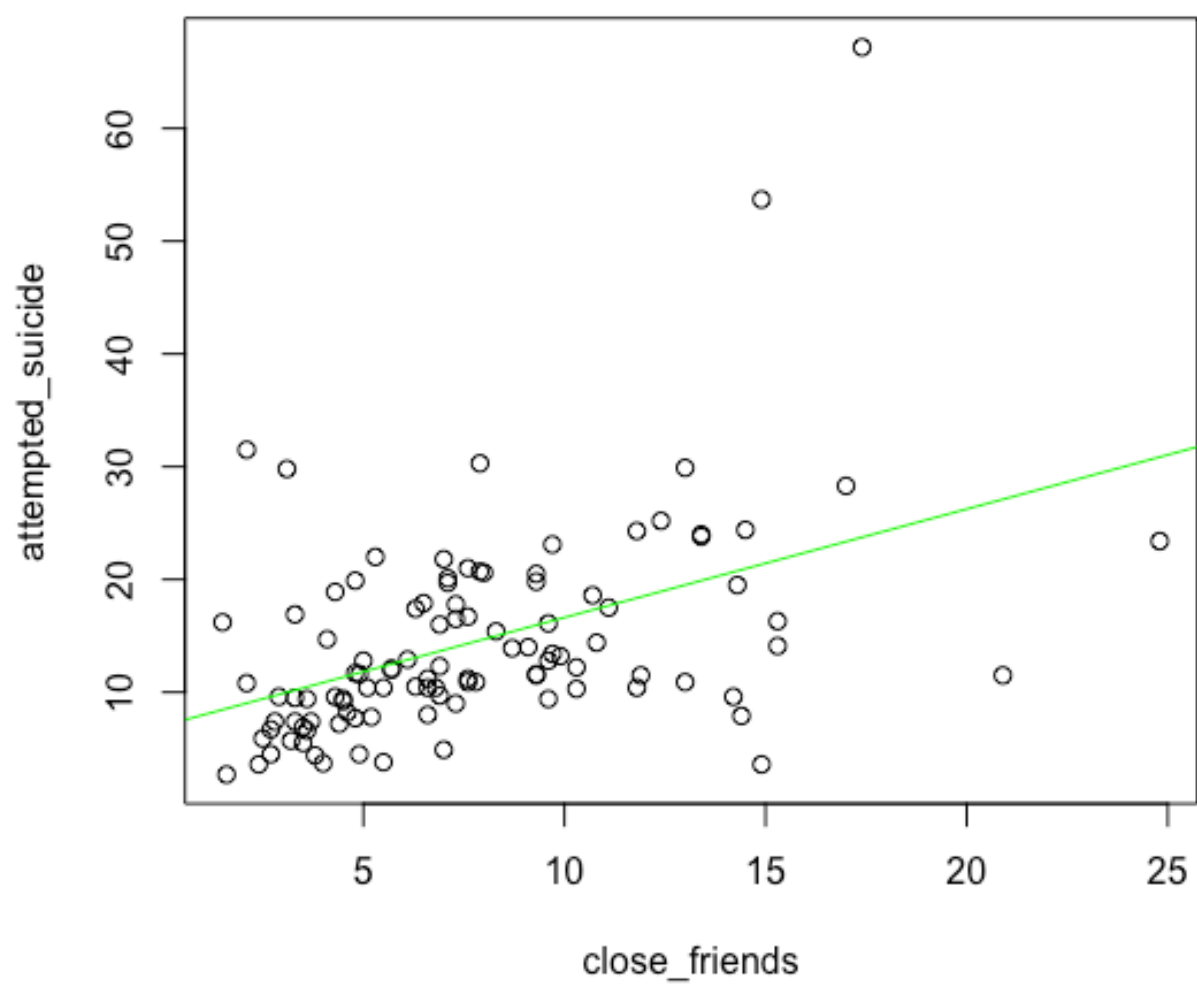
Побудувати діаграму розсіювання $\text{plot}(*\$x_i, *\$y)$ та накласти регресійну лінію









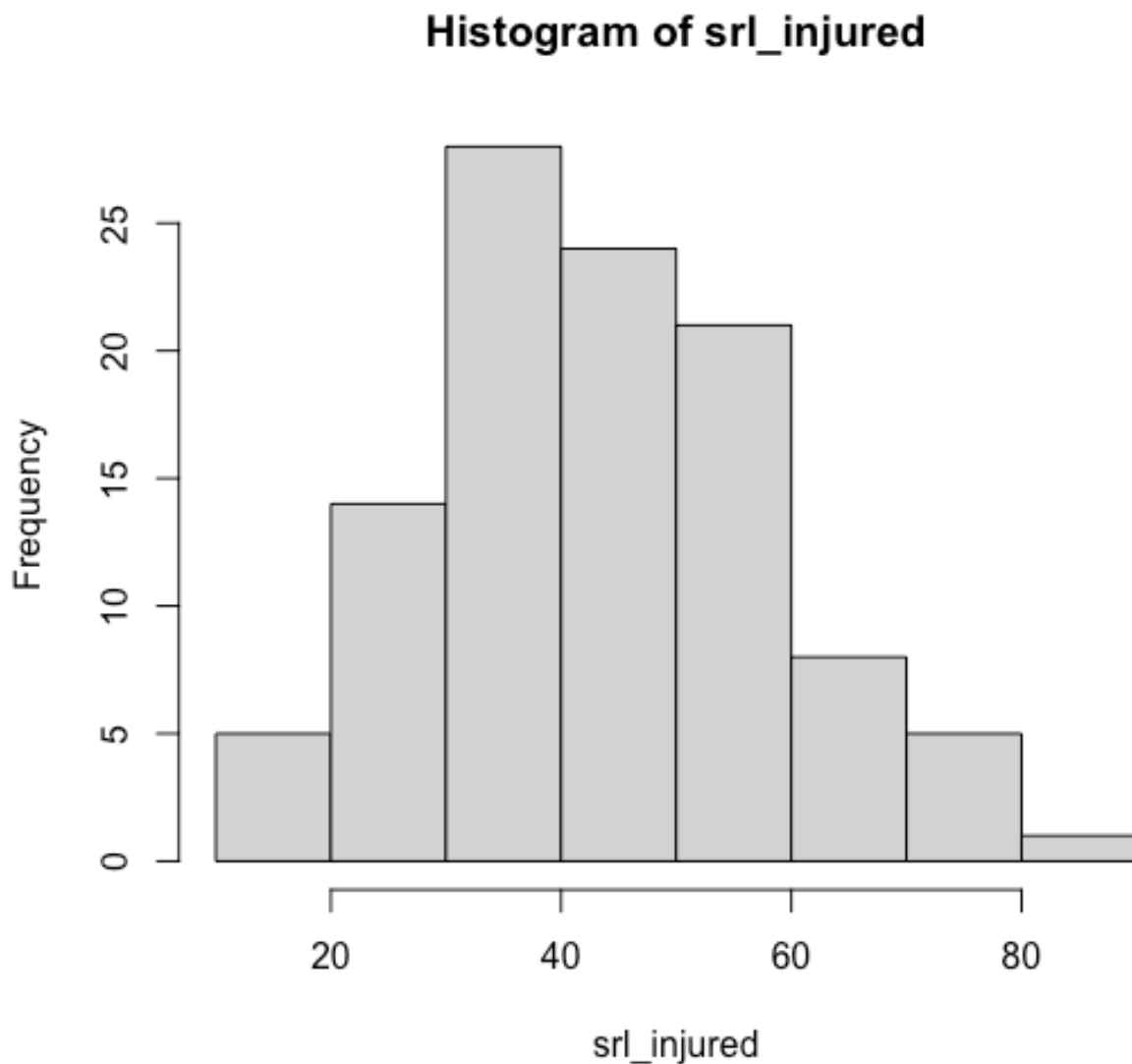


Перевірити значення $*\$r.squared$ та зробити висновки:

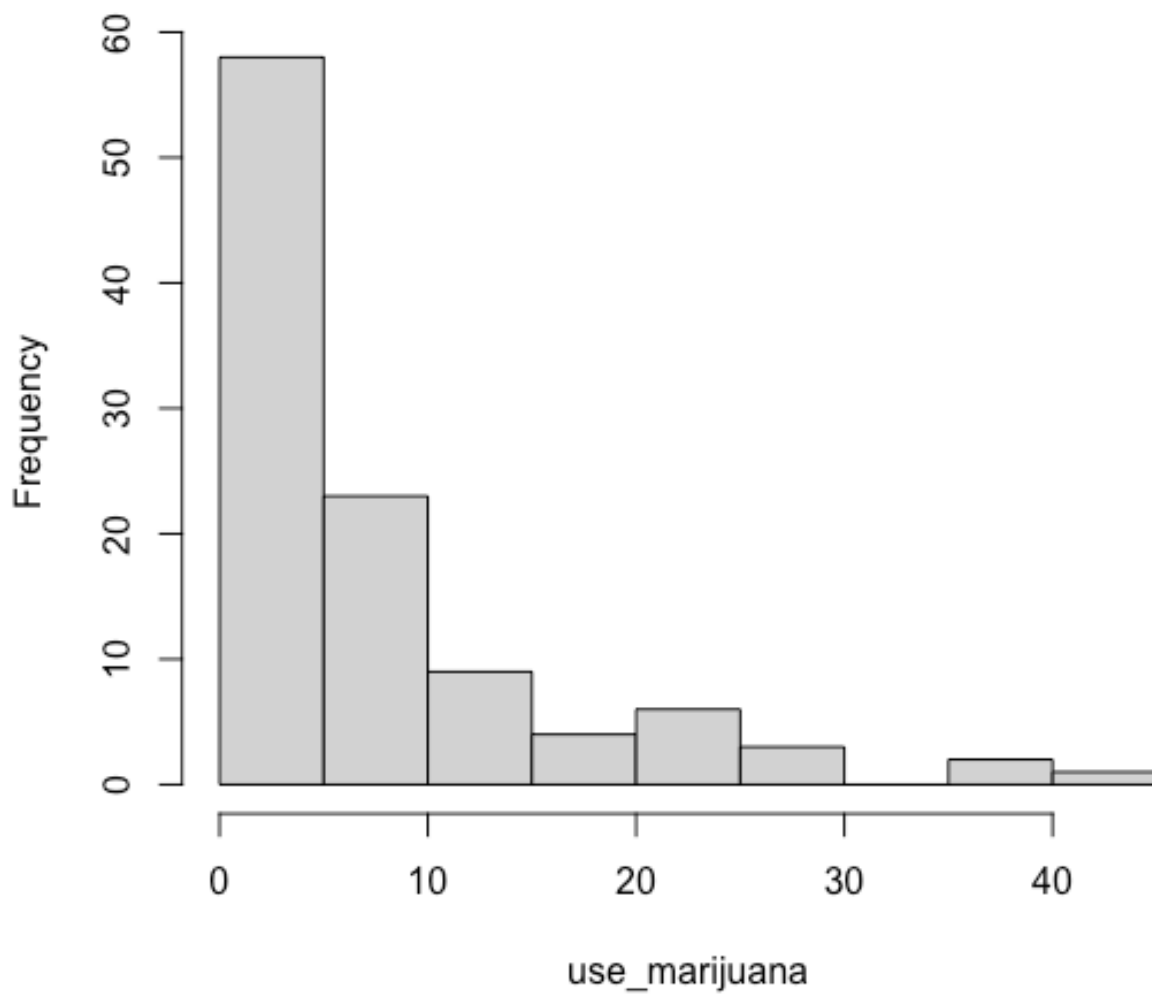
injured - 0.2436819,
use_marijuana - 0.1624589
sexual_relation - 0.1646748
bullied - 0.4355967
close_friends - 0.20477

Перевірити $\text{sum}(*\$residuals^2)$ та зробити висновки:

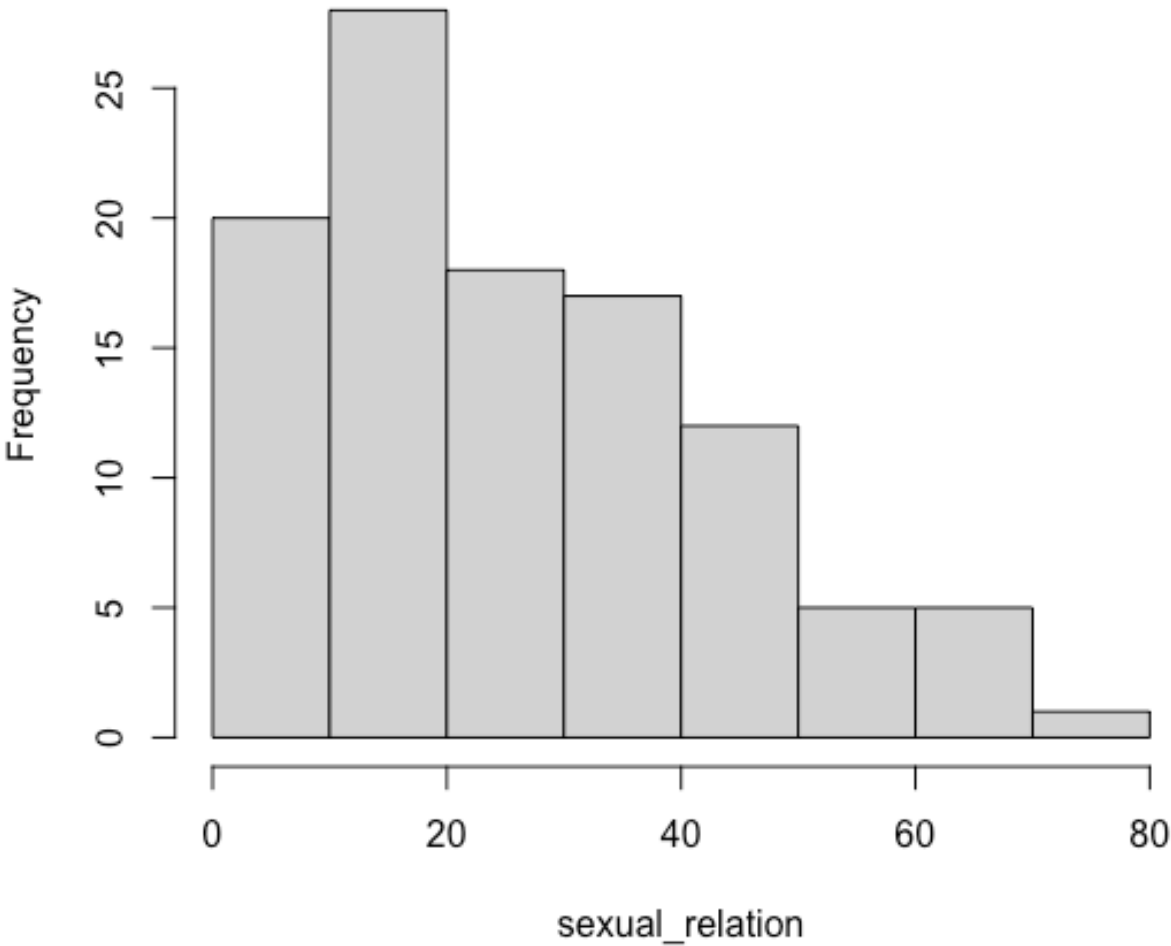
injured - 6829.556,
use_marijuana - 7563
sexual_relation - 7542.99
bullied - 5096.565
close_friends - 7180.931
Обчислити $\text{var}(*\$xi)$:
injured - 219.2893,
use_marijuana - 75.9257
sexual_relation - 302.8059
bullied - 194.0265
close_friends - 19.05544
Обчислити $\text{var}(*\$y)$:
suicidal_behaviour - 86.00004
Побудувати $\text{hist}(*\$xi)$:



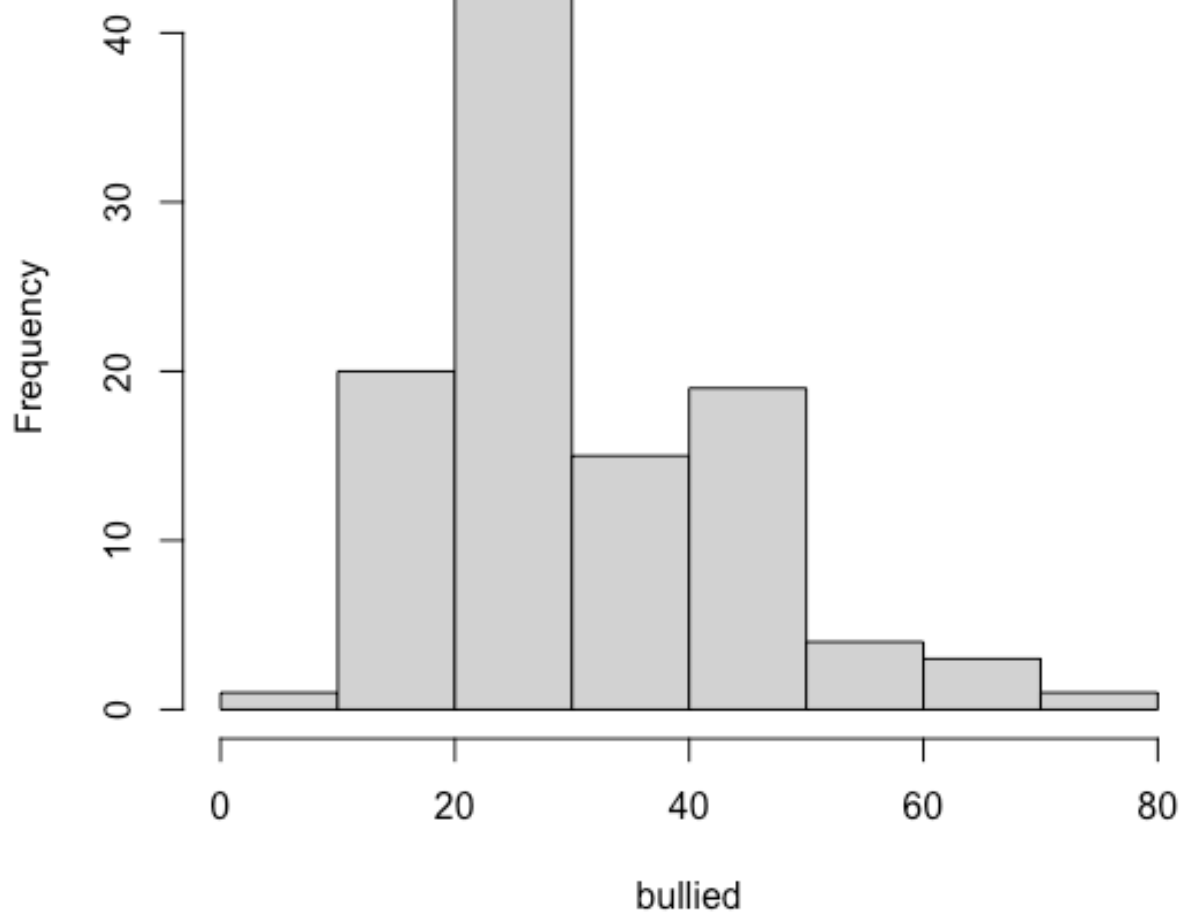
Histogram of use_marijuana



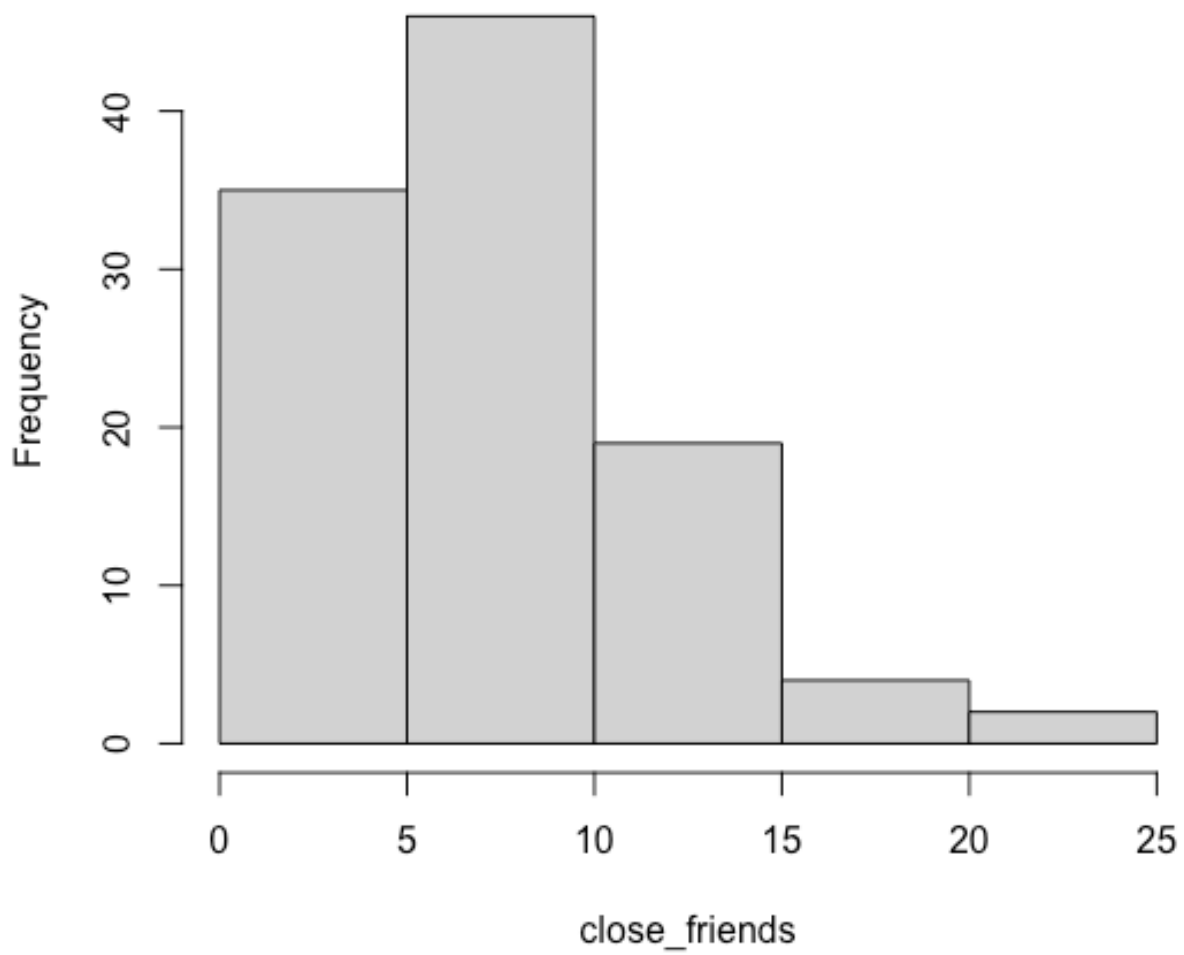
Histogram of sexual_relation



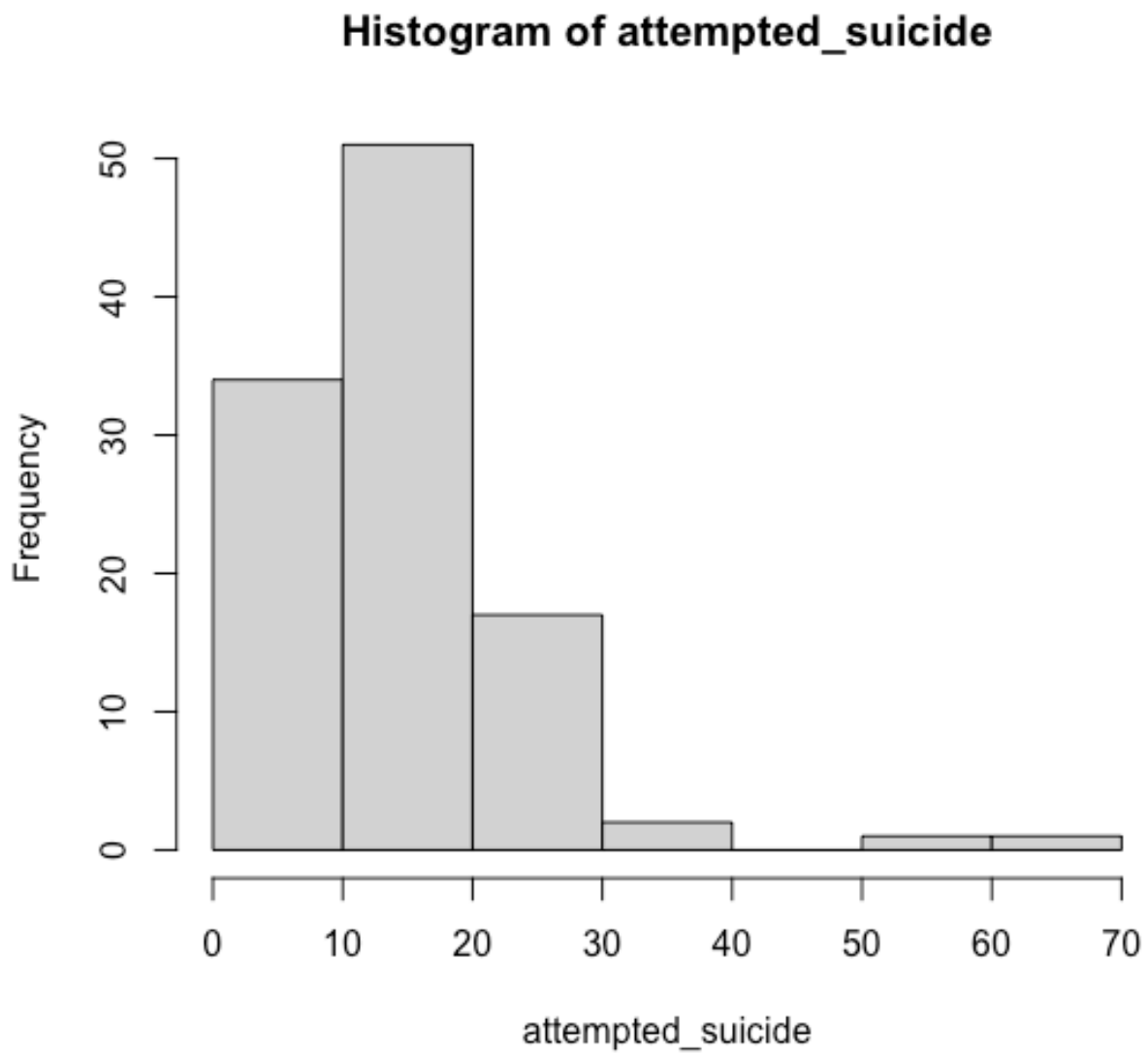
Histogram of bullied



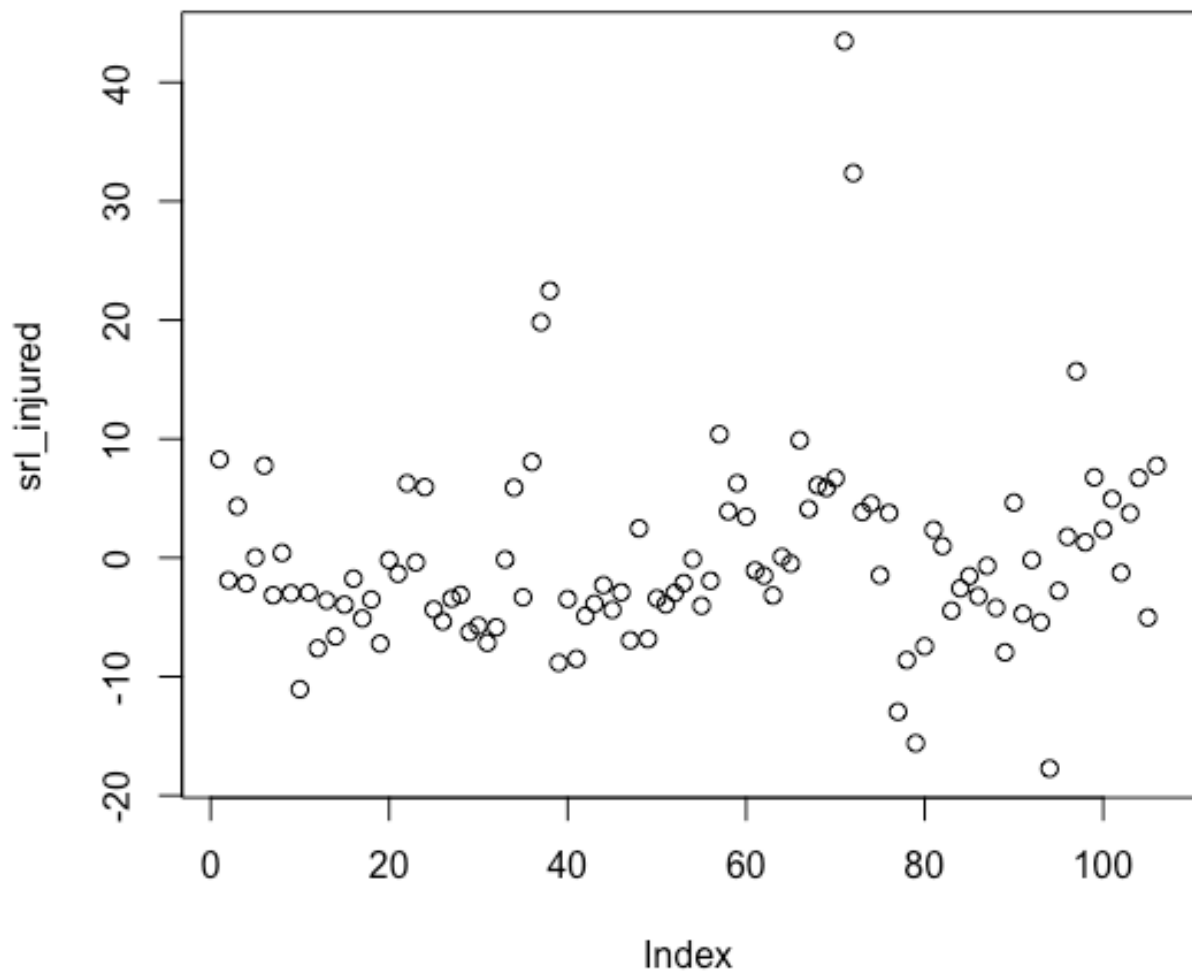
Histogram of close_friends

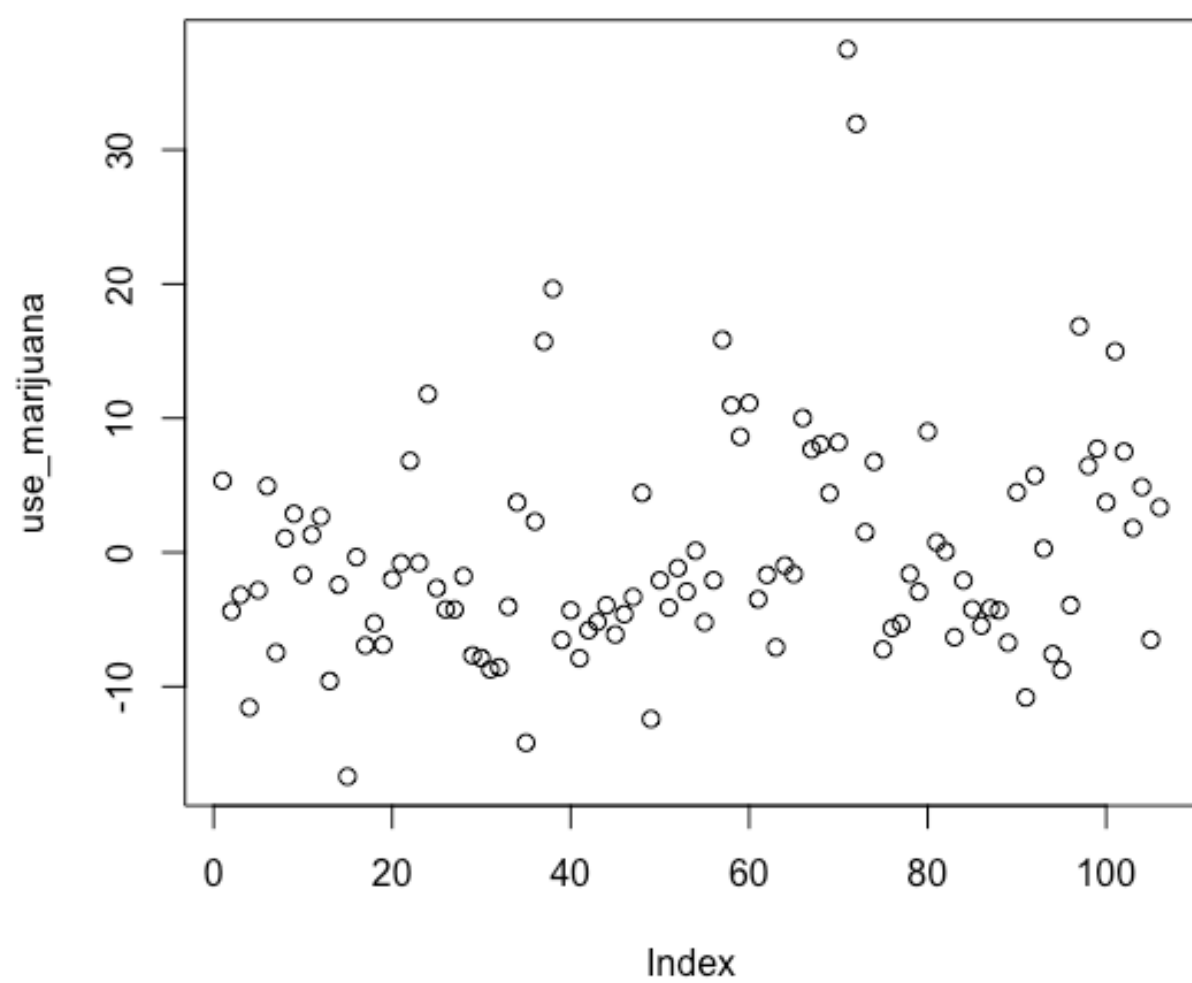


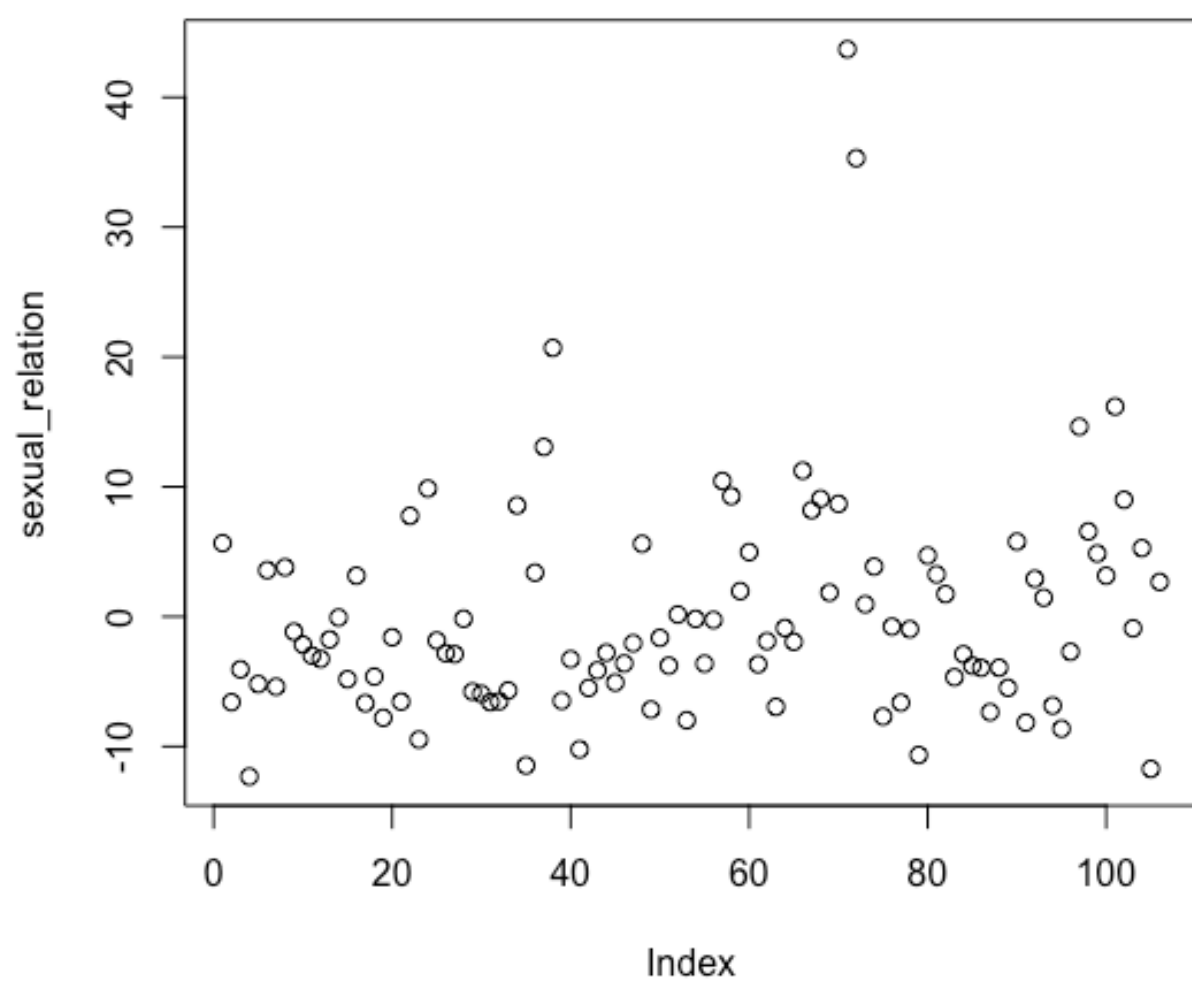
(G) Побудувати hist(*\$y)

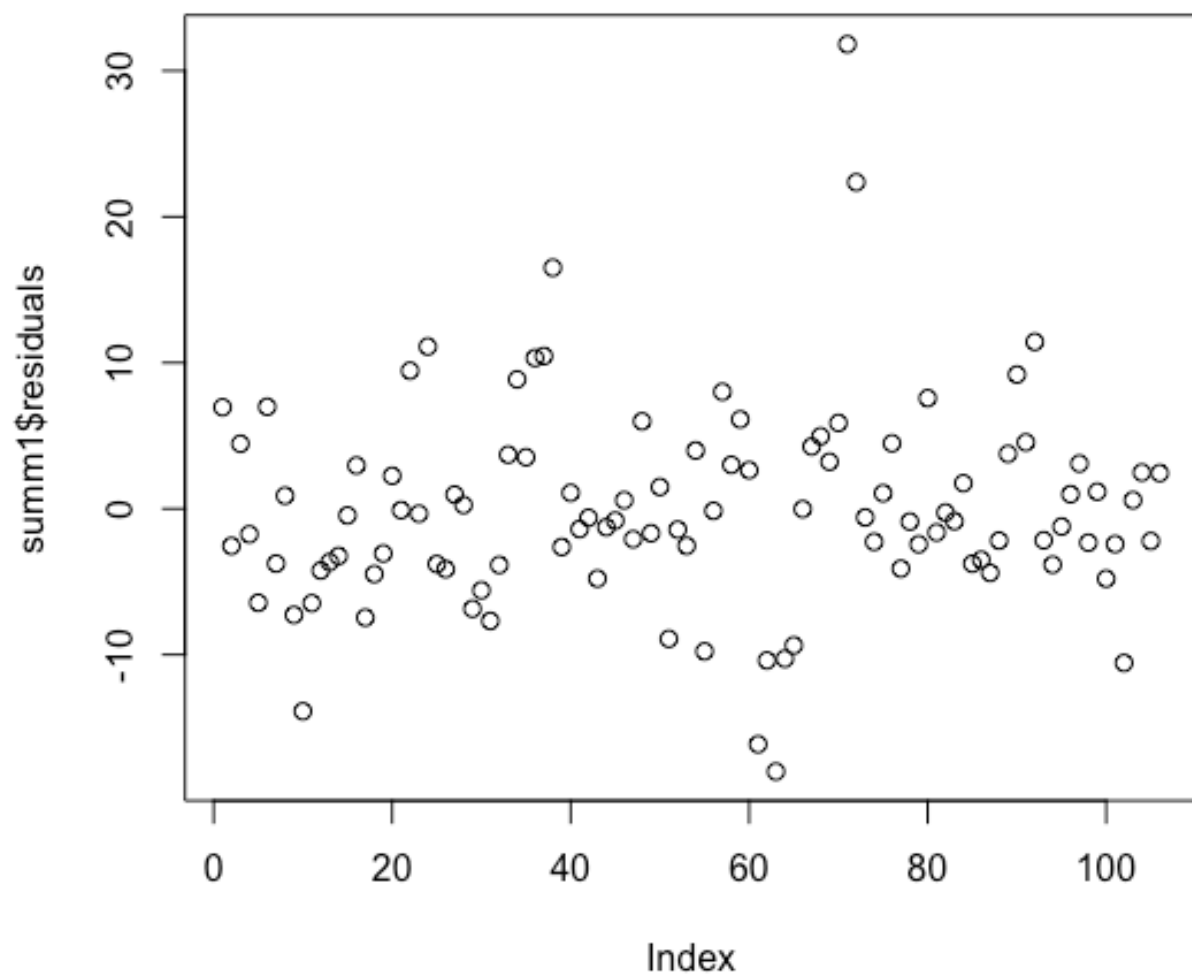


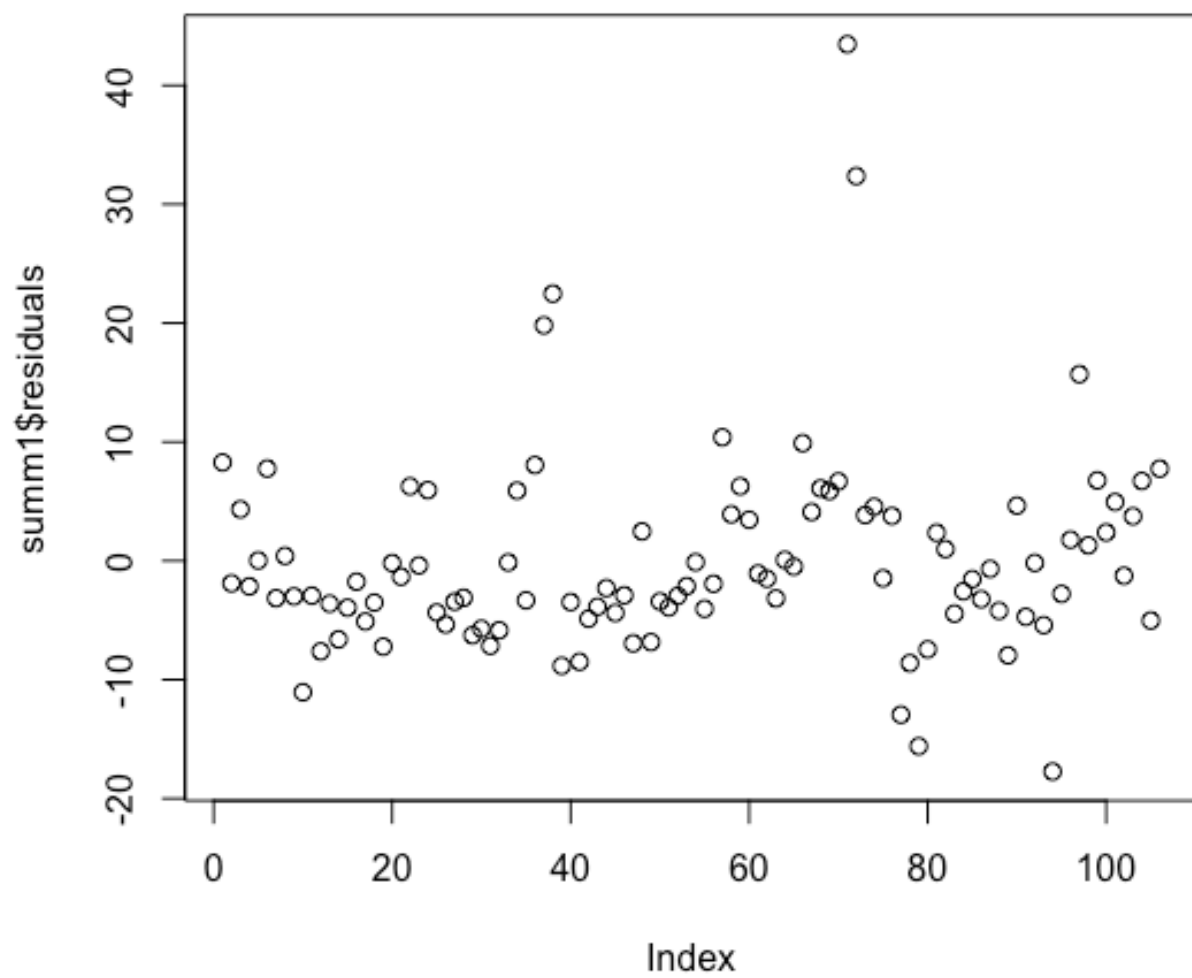
(H) Побудувати `plot(*$residuals)` зробити припущення чи відповідає $N(0; 1)$;











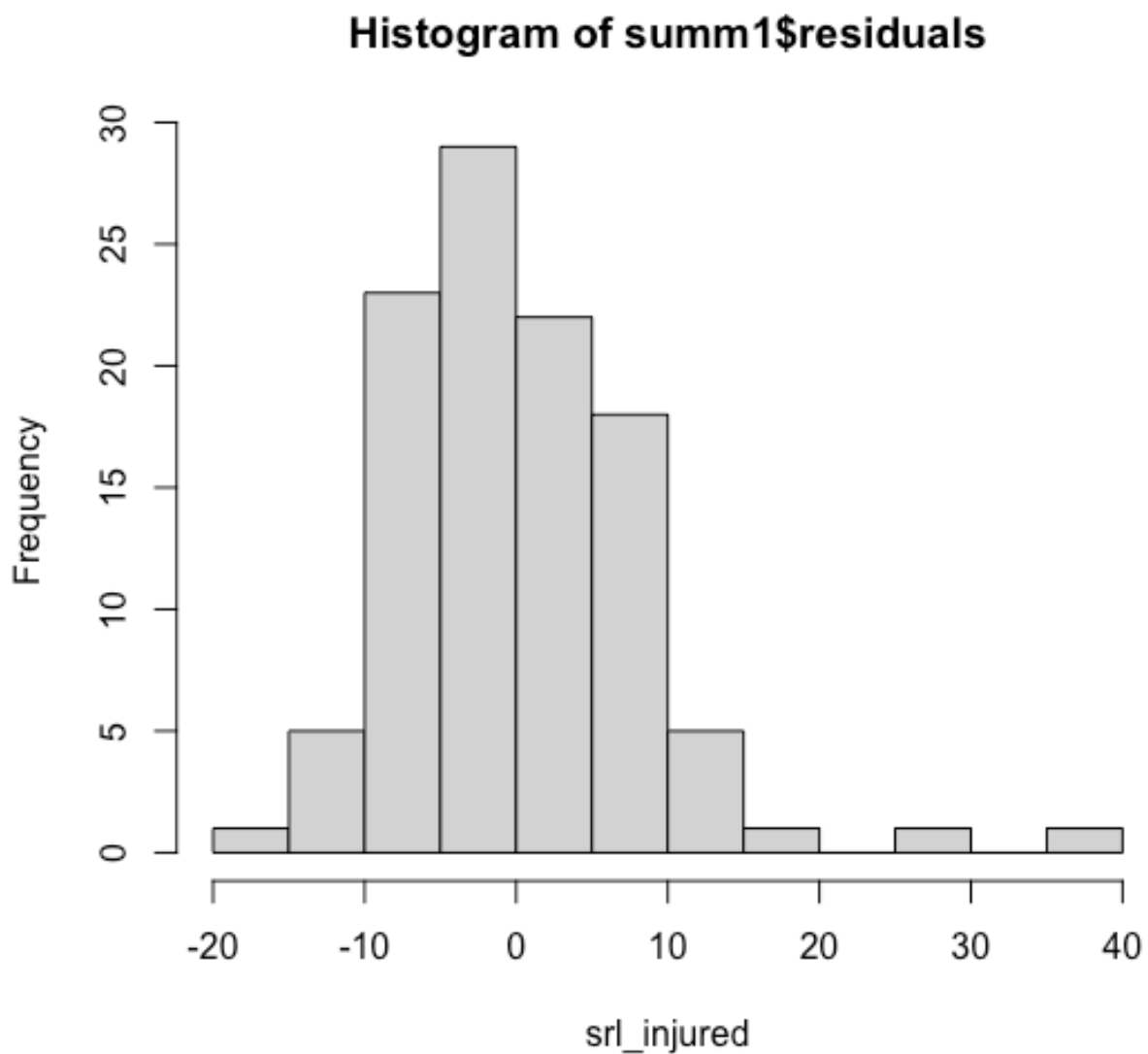
(I) Перевірити `mean(*$residuals)`;

injured - 2.01097e-16,
use_marijuana - 5.655853e-16
sexual_relation - 6.242386e-16
bullied - 3.05835e-16
close_friends - (-6.368072e-16)

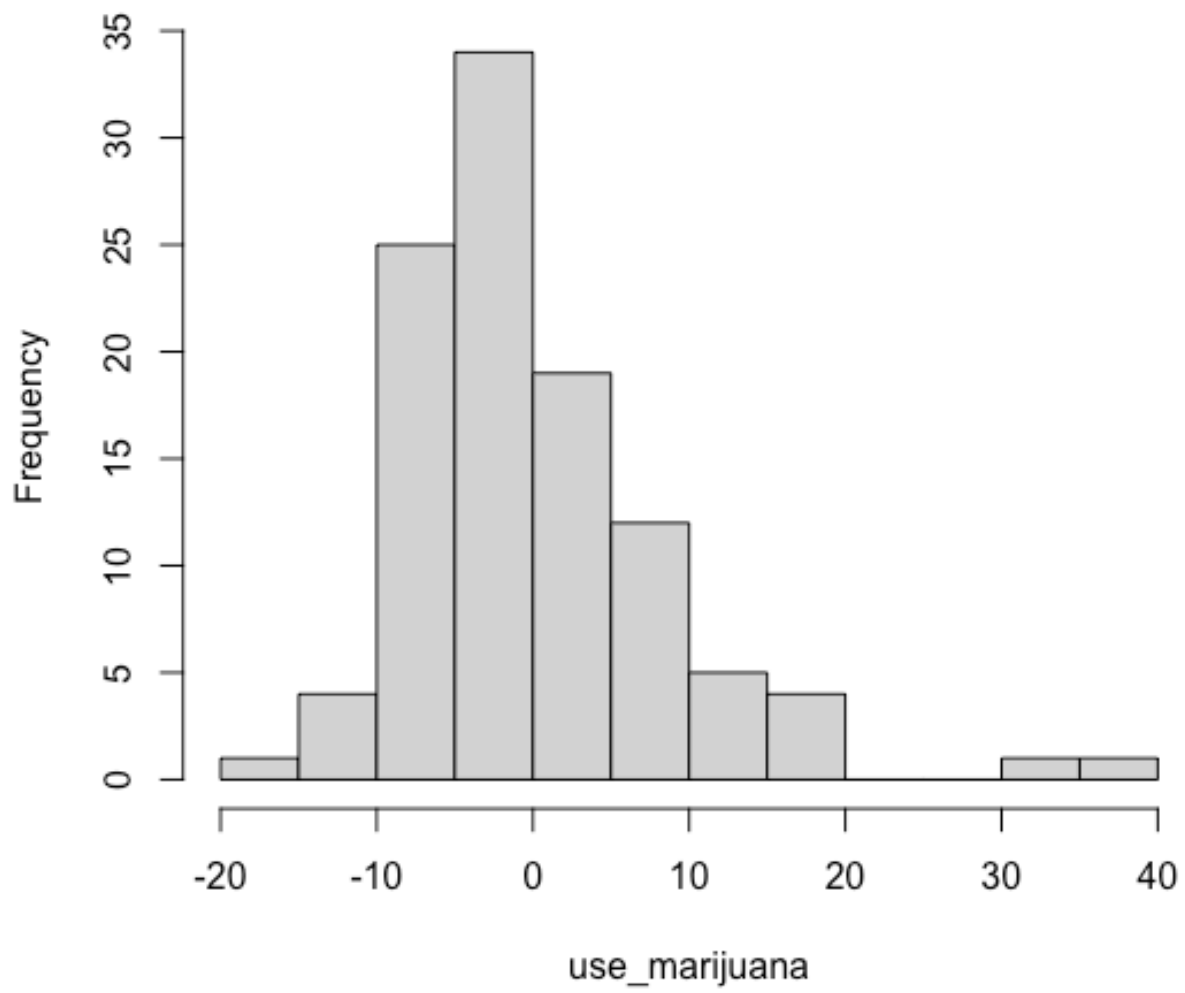
(J) Обчислити `var(*$residuals)`;

injured - 65.04339,
use_marijuana - 72.02857
sexual_relation - 71.838
bullied - 48.53871
close_friends - 68.38981

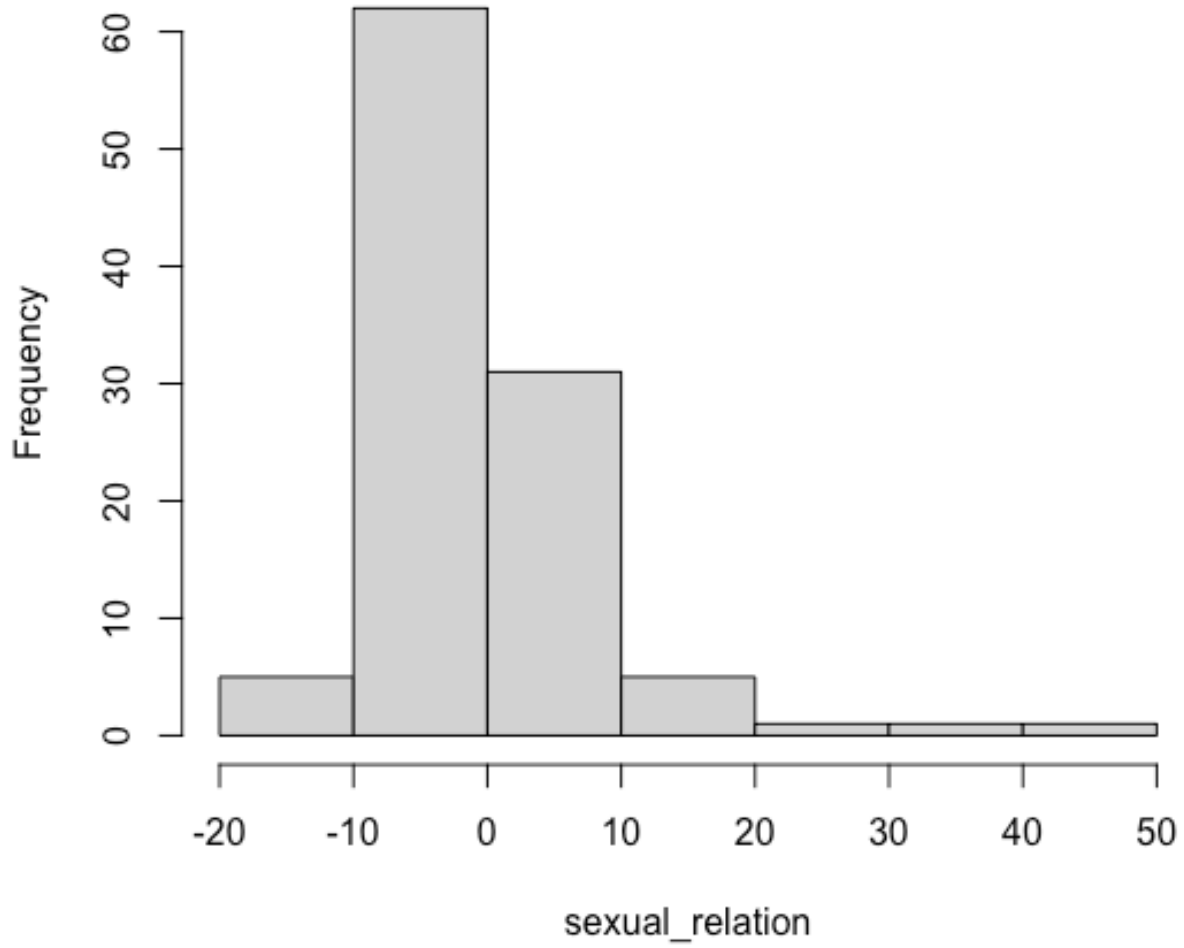
(K) Побудувати `hist(*$residuals)` та перевірити чи відповідає $N(0; 1)$:



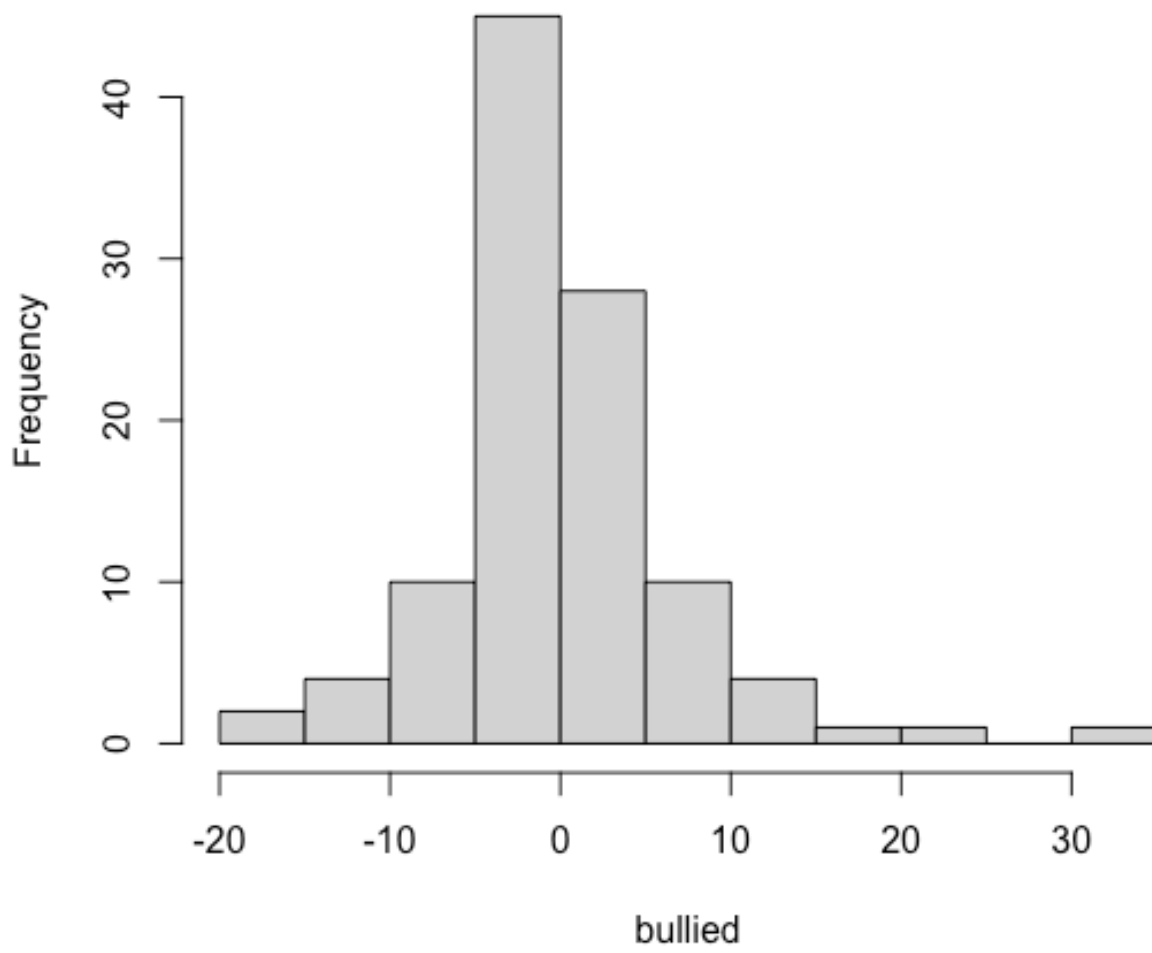
Histogram of summ1\$residuals



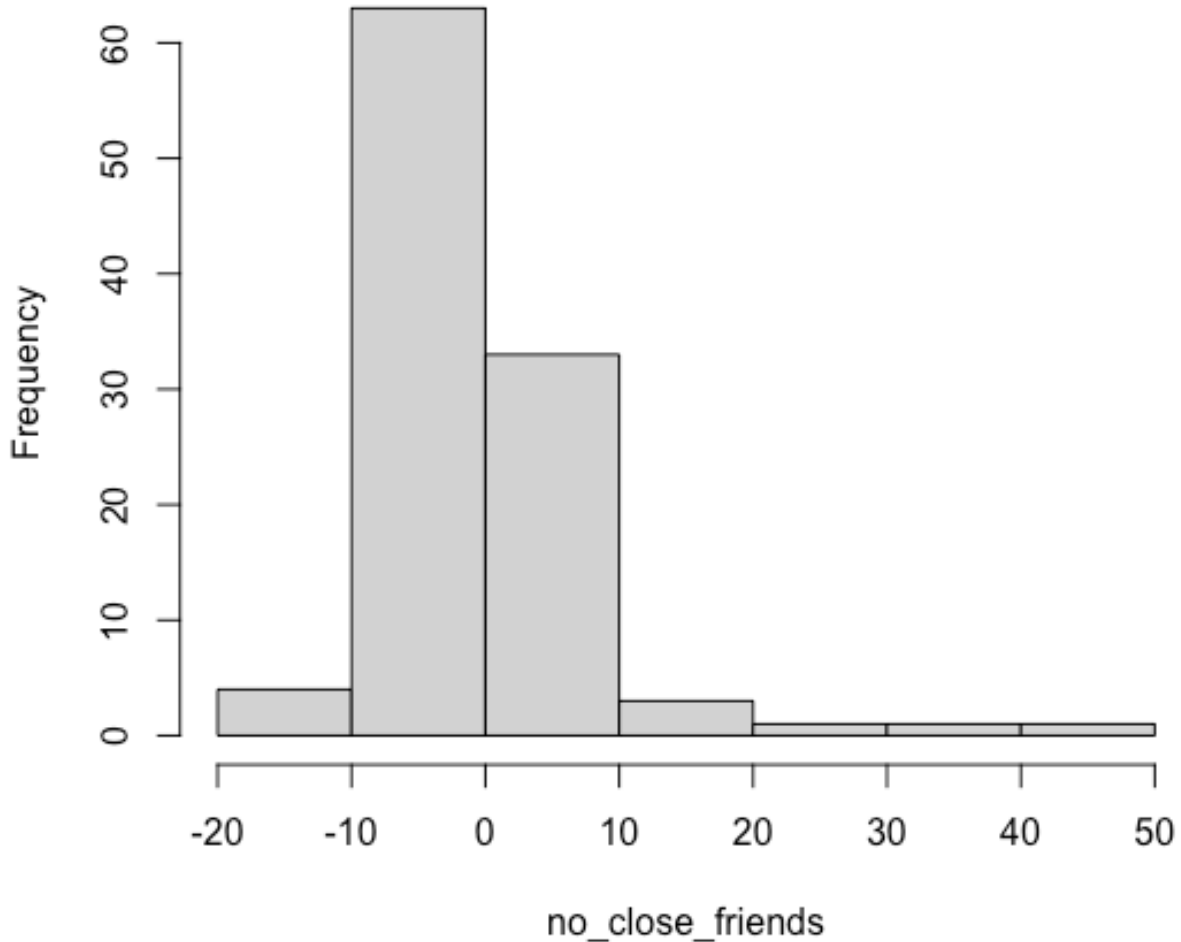
Histogram of summ1\$residuals



Histogram of summ1\$residuals



Histogram of summ1\$residuals



(L) Зробити висновок за 3-ма припущеннями для МНК.

- a. Перевірити припущення 1 (лінійний зв'язок між незалежною змінною та залежною "y")
 - srl_injured - це майже лінійний зв'язок, якщо прибрати пару викидів
 - use_marijuana - це не лінійний зв'язок.
 - sexual_relation - це не лінійний зв'язок, схоже більше на графік квадратного корня(вітка параболи).
 - bullied - це не лінійний зв'язок, схоже більше на якусь криву, що починає загинатись
 - close_friends - мені здається пряма лінія не погано підходить під цей графік, але все таки він не лінійний.
- b. Перевірити припущення 2 (гомоскедастичність)
 - srl_injured - якщо прибрати викиди, то це гомоскедастичність
 - use_marijuana - гетероскедастичність
 - sexual_relation - гетероскедастичність
 - bullied - гетероскедастичність
 - close_friends - гетероскедастичність

с. Перевірити припущення 3 (випадкова похибка " $u \in N(0; 1)$ " і u та y незалежні)
srl_injured -це досить схоже на нормальний розподіл, але все таки гістограма не симетрична.
use_marijuana -це не є нормальний розподіл.
sexual_relation - це не є нормальний розподіл
bullied - це не є нормальний розподіл
close_friends - це не є нормальний розподіл

Завдання 2: Аналіз множинною регресії

Побудувати лінійну модель (m1) за не менше ніж 5-ма параметрами

```
srl_injured = suicidal_behaviours$Got_Seriously_injured  
use_marijuana = suicidal_behaviours$Use_Marijuana  
sexual_relation = suicidal_behaviours$Had_sexual_relation  
bullied = suicidal_behaviours$Bullied  
close_friends = suicidal_behaviours$No_close_friends  
attempted_suicide = suicidal_behaviours$Attempted_suicide
```

```
mult.mod <- lm(attempted_suicide ~ srl_injured + use_marijuana + sexual_relation  
+ bullied + close_friends, data = suicidal_behaviours)
```

Визначити з summary() чому дорівнює RSE та порахувати вручну, а також перевірити чи вони співпадають.

```
summary(mult.mod)
```

RSE = 6.392 on 100 degrees of freedom

```
n <- nrow(suicidal_behaviours)  
mod_summary <- summary(mult.mod)  
SSR <- sum(mod_summary$residuals^2)
```

```
RSE <- sqrt(SSR / (n-2))
```

RSE

6.268126

Значення похибки відрізняються на 0.1

(C) Створити модель (m2) в якій на 1-н параметр менше;

Спробуємо побудувати ще одну множинну регресію, тільки на цей раз з 4 параметрами

```
mult.mod1 <- lm(attempted_suicide ~ srl_injured + use_marijuana +  
sexual_relation + close_friends, data = suicidal_behaviours)
```

```
# Порівняти моделі (m1) та (m2) за допомогою функцій summary() та  
car::compareCoefs(m1, m2) на предмет:  $R^2$ , RSE,  $SE(\beta_i)$ .  
# Зробити висновок, яка модель краща.
```

```
summary(mult.mod1)
```

```
car::compareCoefs(mult.mod, mult.mod1)
```

```
# В другій моделі вилучили фактор Bullied. Бачимо, що ефективність досить  
погіршилась,  
# стандартна похибка збільшилась, Adjusted R Squared зменшився, тому  
перша модель краще
```

```
# Визначити  $t_{кр}$  для моделі (m1):
```

```
# Визначити ступені вільності для (m1);  
106 - 5 - 1 = 100
```

```
#Перевірити t-статистику для кожного з 5-ти коефіцієнтів моделі (m1);  
#Сформулювати гіпотези  $H_0$  та  $H_1$ ;
```

```
#Вказати значення t-статистики (t-value) для відповідного коефіцієнта;
```

	t value
(Intercept)	-0.720
srl_injured	-0.385
use_marijuana	2.756 **
sexual_relation	0.515
bullied	5.985 ***
close_friends	1.754 .

```
#Значення p-значення ( $Pr(>|t|)$ ) для відповідного коефіцієнта:
```

	$Pr(> t)$
(Intercept)	0.47313
srl_injured	0.70113
use_marijuana	0.00695 **
sexual_relation	0.60753
bullied	3.4e-08 ***
close_friends	0.08246 .

```
#Вказати яка гіпотеза виконується;
```

```
(Intercept) - нульова виконується  
srl_injured - нульова виконується  
use_marijuana - не нульова виконується  
sexual_relation - нульова виконується  
bullied - не нульова виконується
```

close_friends - нульова виконується

#Зробити графічне представлення:

#Вказати довірчі інтервали для коефіцієнтів з рівнем надійності 95%, 90% та 99%:

#Виконати масштабування (центрування) моделі (m^2) та перевірити чи співпадають коефіцієнти β_1 , β_2 , β_3 , β_4 ;

Завдання 3: F-статистика

(A) Обчислити SST, SSR, SSE;

визначити компоненти

n <- nrow(suicidal_behaviours) # кількість спостережень (рядків)

k <- 5

y_mean <- mean(suicidal_behaviours\$Attempted_suicide) # mean для середнього test-scores

SSR <- sum((fitted(mult.mod) - Csuicidal_behaviours\$Attempted_suicide)^2) #
сума квадратів залишків

SST <- sum((suicidal_behaviours\$Attempted_suicide - y_mean)^2) # загальна
сума квадратів

SSE <- sum((fitted(mult.mod) - y_mean)^2) # Пояснена сума квадратів

SER <- sqrt(1/(n-k-1) * SSR) # standard error of the regression

Rsq <- 1 - (SSR / SST) # R^2

SSE/SST

adj_Rsq <- 1 - (n-1)/(n-k-1) * SSR/SST # adj. R^2

cof_F <- (SSE/k)/(SSR/(n-k-1)) # (Rsq/k)/((1-Rsq)/(n-k-1))

друк статистики в консоль

c("SER" = SER, "R2" = Rsq, "Adj.R2" = adj_Rsq)

#SER R2 Adj.R2

#6.3922595 0.5474976 0.5248725

cof_F

24.19866

summary(mult.mod)