

## ЛАБОРАТОРНА РОБОТА №3

В роботі досліджуються такі дані :

1. Середня величина субсидій
2. Кількість резидентів
3. Середня величина субсидій на людину

### Підготовка даних

```
N_houses = 60 * 8  
N_test = N_houses/2  
real_x = DATA[, col_residents]  
real_y = DATA[, col_subsidies]  
data_sample = get_data(N_test)  
sample_x = data_sample$residents  
sample_y = data_sample$subsidies
```

### Метод відношень

```
ratio = sum(sample_y)/sum(sample_x)  
estim_ratio = ratio * mean(real_x)
```

### Метод лінійної регресії

```
model = lm(subsidies ~ residents, data = data.frame(residents = sample_x, subsidies =  
sample_y))  
lm_coef = model$coefficients  
estim_lm = lm_coef[1] + lm_coef[2] * mean(real_x)
```

Справжнє значення	Оцінка за відношенням	Оценка лінійної регресії	Вибіркове середнє
2820	2802.501	2810.046	2871.346

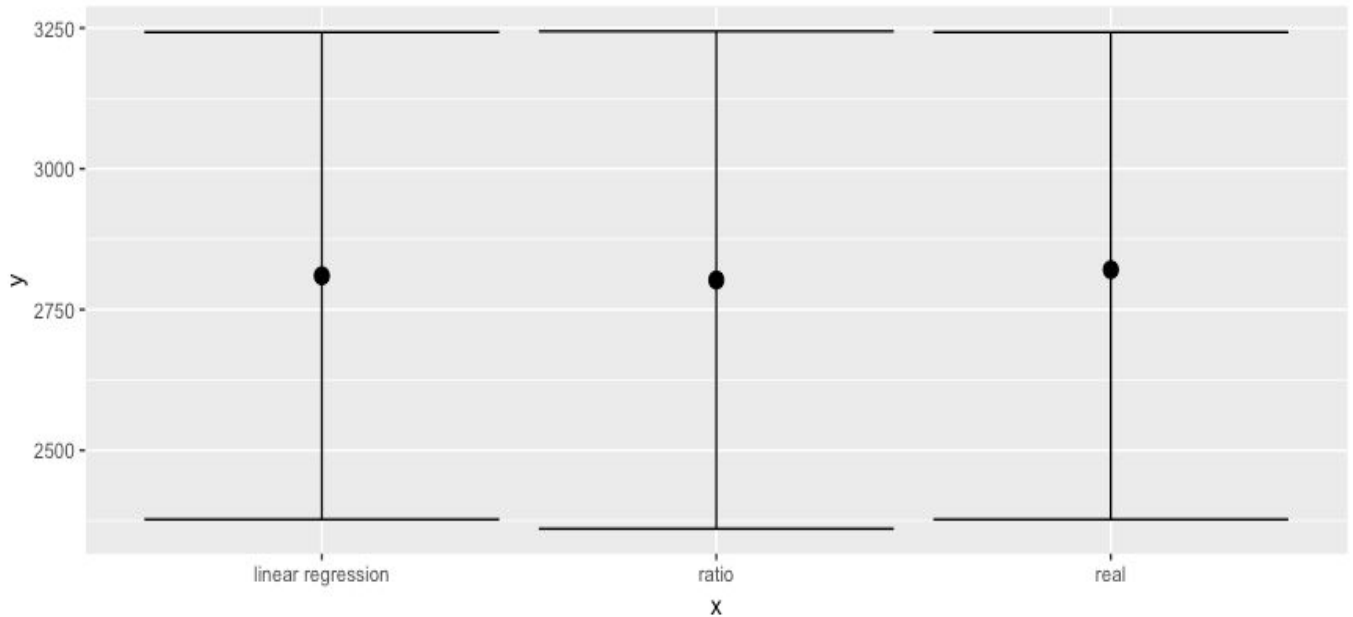
### Довірчі інтервали

```
norm_quantile = qnorm(0.975)  
D_ratio = (1 - N_test/N_houses) * sd(sample_y - estim_ratio)^2 / N_test  
D_lm = (1 - N_test/N_houses) * sd(model$residuals)^2 / N_test  
left_ratio = estim_ratio - sqrt(D_ratio) * norm_quantile  
right_ratio = estim_ratio + sqrt(D_ratio) * norm_quantile  
left_lm = estim_lm - sqrt(D_lm) * norm_quantile  
right_lm = estim_lm + sqrt(D_lm) * norm_quantile
```

	left	predicted value	right
estimation by ratio	2360.78	2802.501	3244.221
linear regression	2377.38	2810.046	3242.712

## Візуалізація результатів

```
library(ggplot2)
d = data.frame(x = c("real", "ratio", "linear regression"), y = c(mean(real_y), estim_ratio,
estim_lm),
              ylo = c(max(left_ratio, left_lm), left_ratio, left_lm),
              yhi = c(min(right_ratio, right_lm), right_ratio, right_lm))
ggplot(d, aes(x = x, y = y)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = ylo, ymax = yhi))
```



```
plot(sample_x, sample_y, pch = 16, col = sample_y/5000 + 1, ylab = "sample subs",
xlab = "sample residents", main = "Dependency")
abline(0, ratio)
```

