Assignment 4

Dmitrii Kuptsov

October 22, 2025

Task 1

Task 1.a

Task 1.a.iii

This confidence region shows that estimates of x_3 and x_4 are negatively correlated. From that we can conclude that (as is clear from dgp) we can recover the coefficient for $x_3 + x_4$ quite well, but the individual coefs have larger variance. It happens because x_4 is generated as x_3 + noise. This graph also shows, that we can reject the hypothesis $x_3 = x_4 = 0$. In comparison, we couldn't reject the hypothesis that $x_3 = 0$ or $x_4 = 0$ separately.

Task 1.a.iv

When we test significance of the regressors separately in this case we cannot reject the null, because 0 is in the confidence interval. However, when we test jointly, (0, 0) is quite far from the confidence ellipsis, which means that both regressors can hardly be insignificant at the same time.

It is especially relevant for the collinear regressors, because when the regressors have little collinearity, joined test would be somewhat similar to the separate ones. However, if the regressors are highly collinear, like in this case, it would be hard for the model to distinguish between these regressors. It would lead to high CIs. In comparison, joined test will be much more helpful, because the coef for the sum of these regressors would have smaller CIs (if there are no other strong collinearities in the model).

Task 1.b

Task 1.b.i

Estimates became much more precise for all of the coefs. (Reason: increase in the number of observation leads to the decrease of the variance.) However, for x_2 the variance is again considerably smaller than for x_3 and x_4. Therefore, collinearity problem decreased due to the larger sample, but haven't been totally solved.

Task 1.b.ii

CIs became much smaller, all of them contain the true values. Again, this happens because of the decrease in variance. We can see that CI for x_2 is few times smaller than

CIs for x_3 and x_4, with the same true coefficients. This is the effect of collinearity.

Task 1.b.iii

The confidence region has the same form as in the previous step. It is still an ellipsis due to the collinearity, but it has decreased considerably. It is still very close to the line $x_3 + x_4 = 1$, because the coefficients for sum of the regressors have much lower variance than the coefs itself again due to the collinearity.

Task 1.b.iv

$var(\hat{\beta}|X) = \sigma \cdot (XX^T)^{-1}$, so $var(\hat{\beta}_3|X) = \sigma \cdot (\sum_i x_{i3} \cdot x_{i3})^{-1}$. Therefore, $var(\hat{\beta}_3|X)$ decreases proportionally to $\frac{1}{n}$ with the number of observations. As a result, standard deviation decreases and confidence intervals became smaller.

Task 1.c

Task 1.c.i

```
> n<- 35
> set.seed(1234)
> x2<-runif(n,0,30)
> x3<-runif(n,0,30)
> x4<-x3 * (-0.5)
> y<-10+0.5*x2+0.5*x3+0.5*x4+rnorm(n,0,4)
> model <- lm(y ~ x2 + x3 + x4)
> summary(model)

Call:
lm(formula = y ~ x2 + x3 + x4)

Residuals:
   Min     1Q Median     3Q    Max 
-8.027 -3.281 -1.063  2.667 10.555 

Coefficients: (1 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.18380    2.20069   4.628 5.85e-05 ***
x2           0.46971    0.09993   4.700 4.74e-05 ***
x3           0.19849    0.10150   1.956   0.0593 .  
x4                NA         NA      NA       NA    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.701 on 32 degrees of freedom
Multiple R-squared:  0.428,	Adjusted R-squared:  0.3923 
F-statistic: 11.97 on 2 and 32 DF,  p-value: 0.0001312

> confint(model, level = 0.95)
                   2.5 %     97.5 %
(Intercept)  5.701154477 14.6664521
x2           0.266149361  0.6732690
x3          -0.008251321  0.4052339
x4                    NA         NA
>
```

Task 1.c.ii

We got an estimate only for x_3, x_4 was excluded due to the perfect collinearity. We can substitute x_4 with $-0.5 \cdot x_3$, and the true coefficient of x_3 with excluded x_4 would be 0.25 ($\beta_3 \cdot x_3 + \beta_4 \cdot x_4 = x_3 \cdot (\beta_3 - 0.5\beta_4)$). Therefore, we got one estimate for $(\beta_3 - 0.5\beta_4)$, and infinite number of estimates for each particular parameter.