Мета

Дослідити вплив різних характеристик на рівень гемоглобіну людей не молодше 20 років.

Початкові моделі

Побудуємо лінійні моделі:

Дані: всі спостереження для яких age Є [20; 60], weight Є [60, 75].

Залежна змінна: hemoglobin

Модель 1:

      Незалежна змінна: age
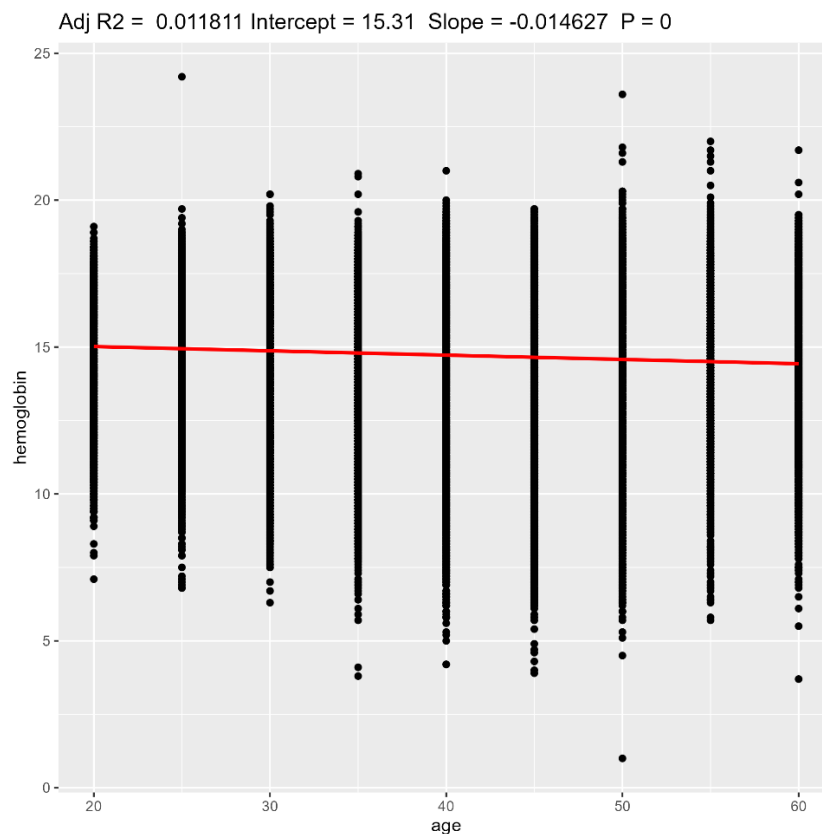
Модель 2:

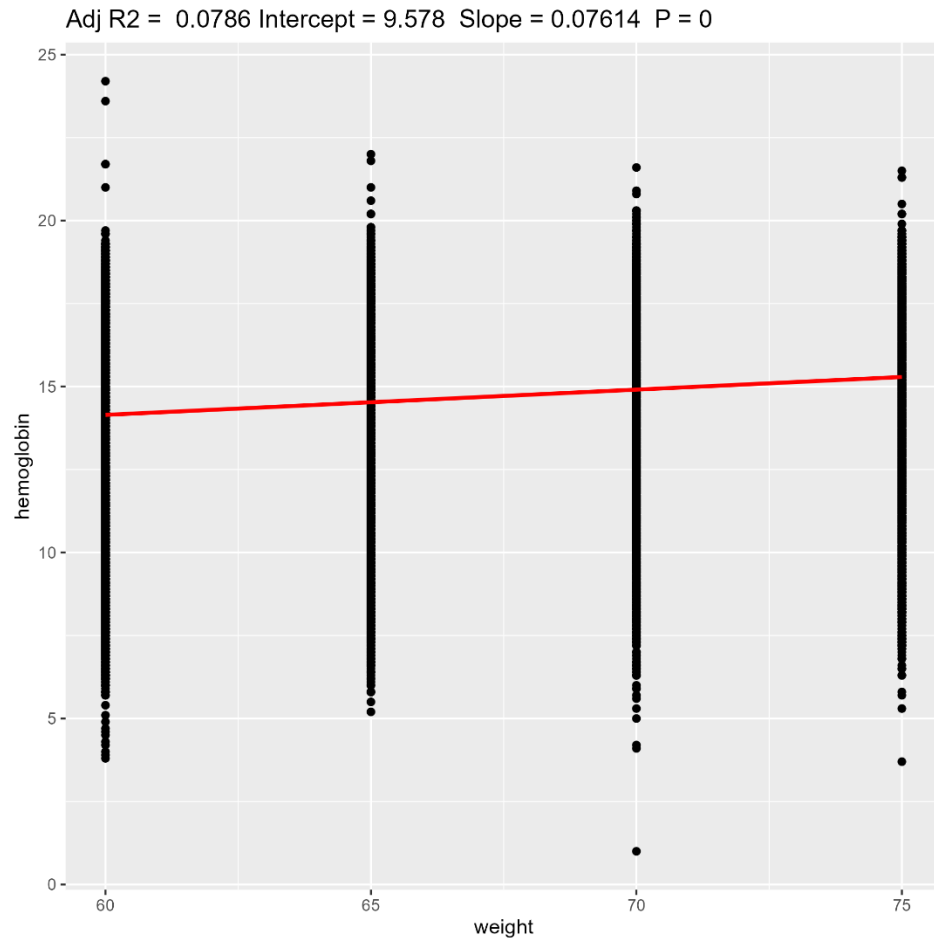   Незалежна змінна: weight

```
data_age_20_60_weight_60_75 <- data %>% filter(age >= 20, age <= 60, weight >= 60, weight <= 75)
lm_in_age_out_hemoglobin <- lm(hemoglobin ~ age, data = data_age_20_60_weight_60_75)
plot_lm_in_age_out_hemoglobin <- ggplotRegression(lm_in_age_out_hemoglobin)
ggsave("./output/Lab3_Regression/plot_lm_in_age_out_hemoglobin.png",
        plot = plot_lm_in_age_out_hemoglobin)
lm_in_weight_out_hemoglobin <- lm(hemoglobin ~ weight, data = data_age_20_60_weight_60_75)
plot_lm_in_weight_out_hemoglobin <- ggplotRegression(lm_in_weight_out_hemoglobin)
ggsave("./output/Lab3_Regression/plot_lm_in_weight_out_hemoglobin.png",
        plot = plot_lm_in_weight_out_hemoglobin)
```

Обмеження на змінні age та weight накладено з метою полегшення виявлення кореляції.

Візуалізуємо моделі:

Модель hemoglobin ~ age

Adj R2 = 0.0786 Intercept = 9.578 Slope = 0.07614 P = 0

Модель hemoglobin ~ weight

Має місце деяка кореляція.

Побудуємо лінійну модель:

Дані: весь очищений датасет (991284 спостережень)

Залежна змінна: hemoglobin

Незалежні змінні: age, weight

```
model <- lm(hemoglobin ~ age + weight, data = data)
model_hc0 <- coeftest(model, vcov. = vcovHC(model, type = "HC0"))
stargazer(model, out = "./output/Lab3_Regression/alldata_age_weight.html", type = "html",
          label = "table:evals-reg",
          dep.var.labels = c("hemoglobin"),
          dep.var.caption = "",
          se = list(model_hc0[, 2]),
          no.space = TRUE,
          omit.stat = c("rsq", "f", "ser"))
```

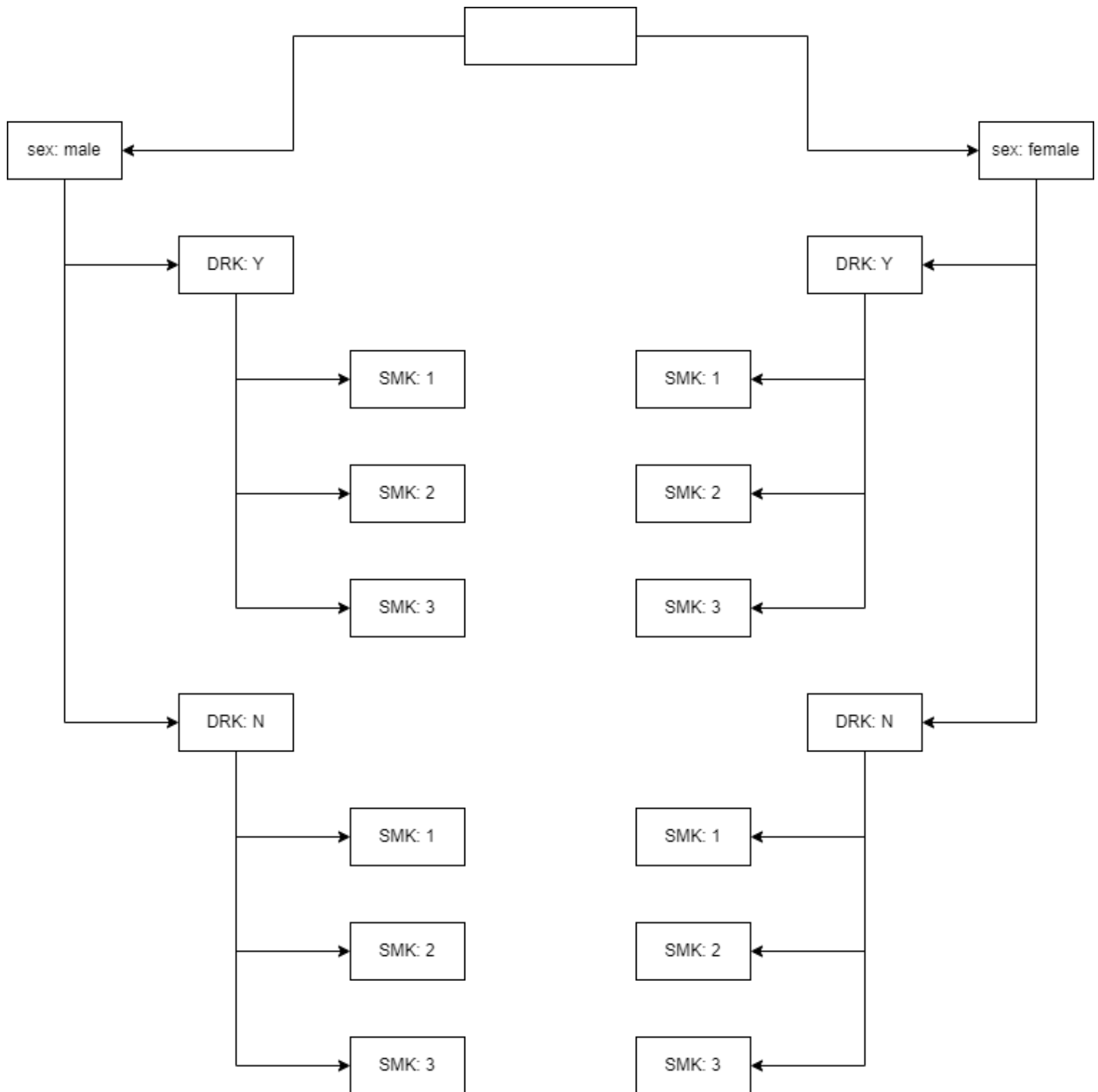|	hemoglobin
age	-0.009***
	(0.0001)
weight	0.061***
	(0.0001)
Constant	10.767***
	(0.009)
Observations	991,284
Adjusted R^2	0.255

All data; hemoglobin ~ age, weight

Спостерігаємо велике значення Constant, що свідчить про необхідність подальшого дослідження впливу інших змінних на гемоглобін.

Покращення моделей

Замість побудови та дослідження однієї великої моделі, було вирішено досліджувати 12 окремих моделей.

```
                              ┌──────────┐
              ┌───────────────┤          ├───────────────┐
              │               └──────────┘               │
         ┌────▼─────┐                               ┌─────▼──────┐
         │ sex: male│                               │ sex: female│
         └────┬─────┘                               └─────┬──────┘
              │        ┌──────────┐       ┌──────────┐    │
              ├───────►│  DRK: Y  │       │  DRK: Y  │◄────┤
              │        └────┬─────┘       └─────┬────┘    │
              │             │   ┌────────┐      │         │
              │             ├──►│ SMK: 1 │   ┌──┤         │
              │             │   └────────┘   │  │         │
              │             │   ┌────────┐   │  │         │
              │             ├──►│ SMK: 2 │   ◄──┤         │
              │             │   └────────┘   │  │         │
              │             │   ┌────────┐   │  │         │
              │             └──►│ SMK: 3 │   ◄──┘         │
              │                 └────────┘                │
         ┌────▼─────┐                               ┌─────▼────┐
         ├─────────►│  DRK: N  │       │  DRK: N  │◄─┤
         └────┬─────┘       └─────┬────┘           │
              │   ┌────────┐      │                │
              ├──►│ SMK: 1 │   ◄──┤                │
              │   └────────┘      │                │
              │   ┌────────┐      │                │
              ├──►│ SMK: 2 │   ◄──┤                │
              │   └────────┘      │                │
              │   ┌────────┐      │                │
              └──►│ SMK: 3 │   ◄──┘                │
                  └────────┘                       │
```

Дерево варіантів датасетів (1 датасет = 1 модель)

Моделі відрізняються даними, на яких відбувається побудова.

Перелік наборів даних:

1. sex: male, DRK: Y, SMK: 1

2. sex: male, DRK: Y, SMK: 2

3. sex: male, DRK: Y, SMK: 3

4. sex: male, DRK: N, SMK: 1

5. sex: male, DRK: N, SMK: 2

6. sex: male, DRK: N, SMK: 3

7. sex: female, DRK: Y, SMK: 1

8. sex: female, DRK: Y, SMK: 2

9. sex: female, DRK: Y, SMK: 3

10. sex: female, DRK: N, SMK: 1

11. sex: female, DRK: N, SMK: 2

12. sex: female, DRK: N, SMK: 3

Розбиття початкового датасету:

```
data <- data %>% filter(age >= 20)
data_male_y_1 <- data %>%
  filter(sex == "Male", DRK_YN == "Y", SMK_stat_type_cd == 1) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_male_y_2 <- data %>%
  filter(sex == "Male", DRK_YN == "Y", SMK_stat_type_cd == 2) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_male_y_3 <- data %>%
  filter(sex == "Male", DRK_YN == "Y", SMK_stat_type_cd == 3) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_male_n_1 <- data %>%
  filter(sex == "Male", DRK_YN == "N", SMK_stat_type_cd == 1) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_male_n_2 <- data %>%
  filter(sex == "Male", DRK_YN == "N", SMK_stat_type_cd == 2) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_male_n_3 <- data %>%
  filter(sex == "Male", DRK_YN == "N", SMK_stat_type_cd == 3) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_female_y_1 <- data %>%
  filter(sex == "Female", DRK_YN == "Y", SMK_stat_type_cd == 1) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_female_y_2 <- data %>%
  filter(sex == "Female", DRK_YN == "Y", SMK_stat_type_cd == 2) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_female_y_3 <- data %>%
  filter(sex == "Female", DRK_YN == "Y", SMK_stat_type_cd == 3) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_female_n_1 <- data %>%
  filter(sex == "Female", DRK_YN == "N", SMK_stat_type_cd == 1) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_female_n_2 <- data %>%
  filter(sex == "Female", DRK_YN == "N", SMK_stat_type_cd == 2) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
data_female_n_3 <- data %>%
  filter(sex == "Female", DRK_YN == "N", SMK_stat_type_cd == 3) %>%
  select(-sex, -DRK_YN, -SMK_stat_type_cd)
```

Для кожного піддатасету побудуємо модель із незалежними змінними: age, weight, height, waistline:

```
models12 <- list()
for (i in 1:length(data_12models)) {
  model <- lm(hemoglobin ~ age + weight + height + waistline, data = data_12models[[i]])
  model_hc0 <- coeftest(model, vcov. = vcovHC(model, type = "HC0"))
  models12[[i]] = model
  print(sprintf("coeftest for model %d", i))
  print(model_hc0)
}
stargazer(models12, out = "./output/Lab3_Regression/model12_age_weight.html", type = "html",
          label = "table:evals-reg",
          dep.var.labels = c("hemoglobin"),
          dep.var.caption = "",
          se = list(model_hc0[, 2]),
          no.space = TRUE,
          omit.stat = c("rsq", "f", "ser"))
```

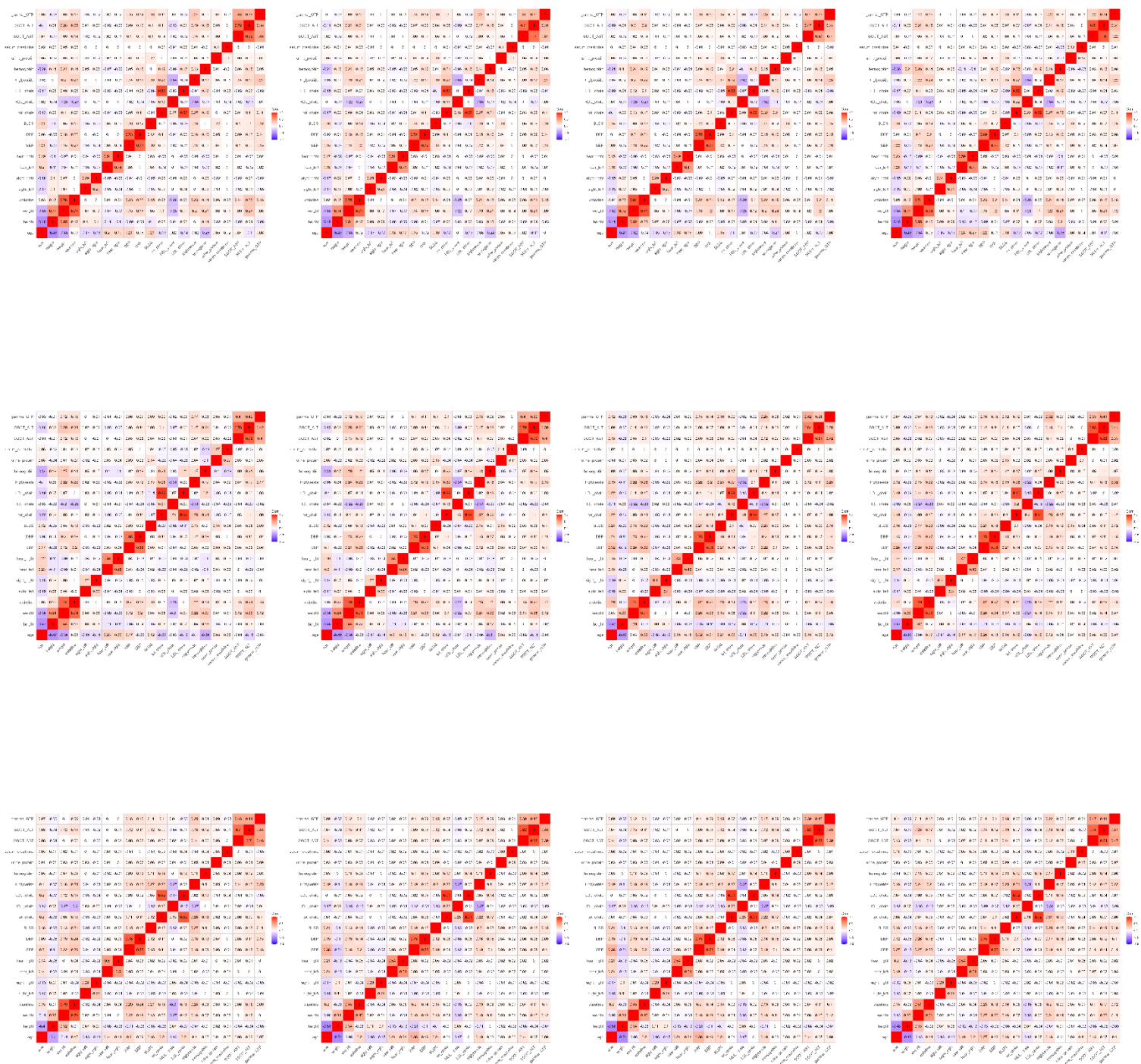| | hemoglobin |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| age | -0.020*** | -0.021*** | -0.019*** | -0.028*** | -0.031*** | -0.025*** | -0.001*** | -0.001 | 0.002* | 0.0004** | 0.0004** | -0.002 |
| | (0.001) | (0.0003) | (0.0003) | (0.0003) | (0.001) | (0.0005) | (0.0003) | (0.001) | (0.001) | (0.0002) | (0.0002) | (0.001) |
| weight | 0.009*** | 0.010*** | 0.012*** | 0.018*** | 0.018*** | 0.016*** | 0.004*** | 0.009*** | 0.007*** | 0.016*** | 0.016*** | 0.016*** |
| | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.002) | (0.0003) | (0.0003) | (0.002) |
| height | -0.007** | -0.007*** | -0.009*** | -0.011*** | -0.010*** | -0.009*** | -0.008*** | -0.006** | -0.004 | -0.006*** | -0.006*** | -0.008*** |
| | (0.003) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.002) | (0.0004) | (0.0004) | (0.003) |
| waistline | 0.011*** | 0.011*** | 0.012*** | 0.007*** | 0.010*** | 0.011*** | 0.007*** | 0.004 | 0.004* | -0.0002 | -0.0002 | 0.002* |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.003) | (0.002) | (0.0002) | (0.0002) | (0.001) |
| Constant | 15.765*** | 15.764*** | 15.879*** | 16.340*** | 16.331*** | 15.928*** | 13.775*** | 13.355*** | 13.242*** | 13.114*** | 13.114*** | 13.739*** |
| | (0.499) | (0.118) | (0.100) | (0.151) | (0.201) | (0.190) | (0.112) | (0.477) | (0.388) | (0.068) | (0.068) | (0.490) |
| Observations | 88,845 | 113,979 | 151,455 | 75,829 | 50,040 | 46,224 | 124,573 | 6,493 | 10,098 | 313,170 | 313,170 | 6,149 |
| Adjusted R^2 | 0.091 | 0.082 | 0.082 | 0.166 | 0.144 | 0.135 | 0.007 | 0.010 | 0.009 | 0.013 | 0.013 | 0.020 |

Note: *p<0.1; **p<0.05; ***p<0.01

1. Для деяких наборів даних (8, 9, 10, 11, 12) не всі змінні є попередньо значимими.
2. Наявні великі значення Constant.

Спробуємо зменшити значення Constant шляхом перебору наборів незалежних змінних.

Для кожного піддатасету побудуємо матрицю кореляцій:

На основі матриць кореляцій для кожного піддатасету було обрано незалежні змінні, таким чином, щоб кожна змінна була «попередньо значимою». «Попередня значущість» змінної оцінювалась згідно числової міри кореляції змінної із hemoglobin.

Отримані набори:

1. male_y_1: age + height + weight + waistline + DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT
2. male_y_2: age + height + weight + waistline + DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT
3. male_y_3: age + height + weight + waistline + DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT
4. male_n_1: age + height + weight + waistline + hear_left + hear_right + DBP + tot_chole + LDL_chole + triglyceride + SGOT_AST
5. male_n_2: age + height + weight + waistline + hear_left + hear_right + DBP + tot_chole + LDL_chole + triglyceride + urine_protein + serum_creatinine + SGOT_ALT
6. male_n_3: age + height + weight + waistline + DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT
7. female_y_1: DBP + SBP + BLDS + tot_chole + LDL_chole + triglyceride + SGOT_ALT
8. female_y_2: weight + waistline + SBP + DBP + BLDS + tot_chole + LDL_chole + triglyceride + SGOT_ALT
9. female_y_3: weight + waistline + SBP + DBP + BLDS + tot_chole + LDL_chole + triglyceride + SGOT_ALT
10. female_n_1: weight + SBP + DBP + BLDS + tot_chole + LDL_chole + triglyceride + SGOT_AST + SGOT_ALT + gamma_GTP
11. female_n_2: weight + SBP + DBP + BLDS + tot_chole + LDL_chole + triglyceride + SGOT_AST + SGOT_ALT + gamma_GTP
12. female_n_3: weight + waistline + SBP + DBP + BLDS + tot_chole + LDL_chole + triglyceride + SGOT_AST + SGOT_ALT

```r
models12 <- list()
models12[[1]] <- lm(hemoglobin ~ age + height + weight + waistline +
             DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT,
             data = data_12models[[1]])
models12[[2]] <- lm(hemoglobin ~ age + height + weight + waistline +
             DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT,
             data = data_12models[[2]])
models12[[3]] <- lm(hemoglobin ~ age + height + weight + waistline +
             DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT,
             data = data_12models[[3]])
models12[[4]] <- lm(hemoglobin ~ age + height + weight + waistline +
             hear_left + hear_right +
             DBP + tot_chole + LDL_chole + triglyceride + SGOT_AST,
             data = data_12models[[4]])
models12[[5]] <- lm(hemoglobin ~ age + height + weight + waistline +
             hear_left + hear_right + DBP + tot_chole + LDL_chole + triglyceride +
             urine_protein + serum_creatinine + SGOT_ALT,
             data = data_12models[[5]])
models12[[6]] <- lm(hemoglobin ~ age + height + weight + waistline +
             DBP + tot_chole + LDL_chole + triglyceride + SGOT_ALT,
             data = data_12models[[6]])
models12[[7]] <- lm(hemoglobin ~ DBP + SBP + BLDS + tot_chole + LDL_chole + triglyceride + SGOT_ALT,
             data = data_12models[[7]])
models12[[8]] <- lm(hemoglobin ~ weight + waistline + SBP + DBP + BLDS +
             tot_chole + LDL_chole + triglyceride + SGOT_ALT,
             data = data_12models[[8]])
models12[[9]] <- lm(hemoglobin ~ weight + waistline + SBP + DBP + BLDS +
             tot_chole + LDL_chole + triglyceride + SGOT_ALT,
             data = data_12models[[9]])
models12[[10]] <- lm(hemoglobin ~ weight + SBP + DBP + BLDS +
             tot_chole + LDL_chole + triglyceride + SGOT_AST + SGOT_ALT + gamma_GTP,
             data = data_12models[[10]])
models12[[11]] <- lm(hemoglobin ~ weight + SBP + DBP + BLDS +
             tot_chole + LDL_chole + triglyceride + SGOT_AST + SGOT_ALT + gamma_GTP,
             data = data_12models[[11]])
models12[[12]] <- lm(hemoglobin ~ weight + waistline + SBP + DBP + BLDS +
             tot_chole + LDL_chole + triglyceride + SGOT_AST + SGOT_ALT,
             data = data_12models[[12]])
stargazer(models12, out = "./output/Lab3_Regression/model12_manual.html", type = "html",
      label = "table:evals-reg",
      dep.var.labels = c("hemoglobin"),
      dep.var.caption = "",
      se = list(model_hc0[, 2]),
      no.space = TRUE,
      omit.stat = c("rsq", "f", "ser"))
```

	hemoglobin											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
age	-0.020***	-0.018***	-0.019***	-0.028***	-0.025***	-0.026***						
	(0.002)	(0.0003)	(0.0003)	(0.0004)	(0.001)	(0.001)						
height	-0.001	-0.002***	-0.004***	-0.007***	-0.004***	-0.003***						
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)						
weight	0.003***	0.006***	0.008***	0.014***	0.011***	0.009***		0.005*	0.005**	0.008***	0.008***	0.006***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		(0.003)	(0.002)	(0.0003)	(0.0003)	(0.002)
waistline	0.007***	0.007***	0.008***	0.005***	0.006***	0.007***		-0.001	-0.003			0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		(0.003)	(0.002)			(0.001)
hear_left				-0.045*	-0.028							
				(0.025)	(0.027)							
hear_right				-0.075***	-0.106***							
				(0.025)	(0.028)							
DBP	0.014***	0.013***	0.012***	0.016***	0.016***	0.013***	0.016***	0.017***	0.015***	0.018***	0.018***	0.020***
	(0.003)	(0.0003)	(0.0003)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.0003)	(0.0003)	(0.002)
SBP							-0.003***	-0.004**	-0.004***	-0.004***	-0.004***	-0.006***
							(0.0004)	(0.002)	(0.001)	(0.0002)	(0.0002)	(0.002)
BLDS							0.003***	0.004***	0.003***	0.002***	0.002***	0.003***
							(0.0002)	(0.001)	(0.001)	(0.0001)	(0.0001)	(0.001)
tot_chole	0.002***	0.002***	0.002***	0.004***	0.003***	0.002***	0.005***	0.002**	0.004***	0.006***	0.006***	0.005***
	(0.0004)	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.0004)	(0.0002)	(0.001)	(0.001)	(0.0001)	(0.0001)	(0.001)
LDL_chole	0.002	0.003***	0.002***	0.001	0.002***	0.003***	-0.001***	0.003**	0.0003	-0.002***	-0.002***	-0.001
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0005)	(0.0004)	(0.0002)	(0.001)	(0.001)	(0.0002)	(0.0002)	(0.001)
triglyceride	0.001	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.0003*	0.0002*	0.0004***	0.0004***	0.001***
	(0.0002)	(0.00004)	(0.00003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.00004)	(0.00004)	(0.0002)
urine_protein					-0.102***							
					(0.010)							
serum_creatinine					-0.404***							
					(0.011)							
SGOT_ALT	0.005	0.003***	0.002***		0.006***	0.003***	0.002***	0.003***	0.005***	0.009***	0.009***	0.008***
	(0.001)	(0.0001)	(0.0001)		(0.0003)	(0.0002)	(0.0001)	(0.001)	(0.001)	(0.0002)	(0.0002)	(0.001)
gamma_GTP										0.001***	0.001***	
										(0.0001)	(0.0001)	
SGOT_AST				0.001***						-0.004***	-0.004***	-0.003*
				(0.0001)						(0.0002)	(0.0002)	(0.002)
Constant	13.520***	13.479***	14.004***	14.192***	14.010***	13.848***	10.962***	11.016***	11.426***	10.419***	10.419***	10.816***
	(0.197)	(0.124)	(0.110)	(0.169)	(0.206)	(0.207)	(0.040)	(0.186)	(0.153)	(0.026)	(0.026)	(0.181)
Observations	76,205	110,054	136,216	68,320	49,124	42,391	102,375	5,366	7,906	297,385	297,385	5,534
Adjusted R²	0.138	0.129	0.125	0.204	0.223	0.183	0.052	0.059	0.058	0.072	0.072	0.085

Note: *p<0.1; **p<0.05; ***p<0.01

1. Спостерігаємо зменшення значень Constant
2. Для моделей (1, 4, 5, 8, 9, 12) деякі незалежні змінні не є значимими

Загалом зменшення Constant не є задовільним, маємо провести подальші дослідження.

Застосуємо AIC для автоматичного підбору незалежних параметрів моделі:

```
models12 <- list()
for (i in 1:length(data_12models)) {
  min_model <- lm(hemoglobin ~ 1, data = data_12models[[i]])
  model <- step(min_model, direction = "forward",
              scope = (~ age + height + weight + waistline +
                       sight_left + sight_right + hear_left + hear_right +
                       SBP + DBP + BLDS +
                       tot_chole + HDL_chole + LDL_chole + triglyceride +
                       urine_protein + serum_creatinine +
                       SGOT_AST + SGOT_ALT + gamma_GTP))
  model_hc0 <- coeftest(model, vcov. = vcovHC(model, type = "HC0"))
  models12[[i]] <- model
  print(sprintf("coeftest for model %d", i))
  print(model_hc0)
}
stargazer(models12, out = "./output/Lab3_Regression/model12_autobest.html", type = "html",
        label = "table:evals-reg",
        dep.var.labels = c("hemoglobin"),
        dep.var.caption = "",
        se = list(model_hc0[, 2]),
        no.space = TRUE,
        omit.stat = c("rsq", "f", "ser"))
```

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
							hemoglobin					
age	-0.019***	-0.018***	-0.020***	-0.025***	-0.023***	-0.023***	-0.007***	-0.007***	-0.005***	-0.002***	-0.002***	-0.007***
	(0.001)	(0.0003)	(0.0003)	(0.0003)	(0.001)	(0.0005)	(0.0003)	(0.001)	(0.001)	(0.0002)	(0.0002)	(0.001)
DBP	0.017***	0.018***	0.016***	0.021***	0.022***	0.020***	0.014***	0.012***	0.012***	0.017***	0.017***	0.017***
	(0.002)	(0.0005)	(0.0004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)
tot_chole	0.004***	0.005***	0.005***	0.006***	0.006***	0.004***	0.005***	0.005***	0.004***	0.005***	0.005***	0.004***
	(0.0004)	(0.0002)	(0.0002)	(0.0005)	(0.001)	(0.0001)	(0.0002)	(0.0004)	(0.0003)	(0.0003)	(0.0003)	(0.0004)
SGOT_ALT	0.006***	0.004***	0.002***	0.004***	0.009***	0.006***	0.002***	0.006***	0.007***	0.008***	0.008***	0.006***
	(0.001)	(0.0002)	(0.0001)	(0.0002)	(0.0004)	(0.0003)	(0.0002)	(0.001)	(0.001)	(0.0002)	(0.0002)	(0.001)
waistline	0.006***	0.006***	0.007***	0.004***	0.006***	0.007***	0.004***			-0.001***	-0.001***	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)			(0.0002)	(0.0002)	
HDL_chole	-0.002***	-0.005***	-0.005***	-0.004***	-0.003***	-0.004***		-0.002**		0.001**	0.001**	
	(0.0003)	(0.0003)	(0.001)	(0.001)	(0.0005)			(0.001)		(0.0003)	(0.0003)	
SGOT_AST	-0.003	-0.002***	-0.0003***	-0.002***	-0.005***	-0.005***	0.0004*	-0.003***	-0.002***	-0.004***	-0.004***	
		(0.0002)	(0.0001)	(0.0002)	(0.001)	(0.001)	(0.0003)	(0.001)	(0.001)	(0.0002)	(0.0002)	
serum_creatinine		-0.092***	0.017**	-0.110***	-0.400***	-0.074***	0.014**	0.282***	0.183***	-0.076***	-0.076***	-0.125
		(0.011)	(0.008)	(0.006)	(0.011)	(0.007)	(0.005)	(0.085)	(0.058)	(0.005)	(0.005)	(0.082)
SBP	-0.003**	-0.005***	-0.004***	-0.006***	-0.006***	-0.007***	-0.003***		-0.002*	-0.003***	-0.003***	-0.004**
	(0.002)	(0.003)	(0.003)	(0.0004)	(0.001)	(0.001)	(0.0004)		(0.001)	(0.0002)	(0.0002)	(0.002)
triglyceride	0.0003		0.0001**	0.0004***	0.0005***	0.001***	0.001***		0.0003**	0.001***	0.001***	0.001***
	(0.0002)		(0.00003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)	(0.0001)	(0.0002)
weight	0.002***	0.005***	0.006***	0.011***	0.010***	0.009***	-0.001**	0.002	0.002*	0.008***	0.008***	0.005***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.0003)	(0.0003)	(0.002)
LDL_chole		-0.001***	-0.001***	-0.002***	-0.001*		-0.001***			-0.001***	-0.001***	
		(0.0002)	(0.0002)	(0.0005)	(0.001)		(0.0002)			(0.0003)	(0.0003)	
BLDS	0.0005	0.001***	0.001***	0.001***	0.0004**	0.001***	0.003***	0.003***	0.003***	0.002***	0.002***	0.004***
	(0.001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.001)	(0.001)	(0.0001)	(0.0001)	(0.001)
hear_left	-0.069			-0.042*		-0.060*			-0.189*	-0.049***	-0.049***	
				(0.024)		(0.036)			(0.099)	(0.013)	(0.013)	
sight_right		0.014**										0.048**
		(0.005)										(0.021)
gamma_GTP		-0.0001**			-0.001***	-0.001***	0.001***		-0.0005***	0.001***	0.001***	-0.001
		(0.0001)			(0.0002)	(0.0001)	(0.0001)		(0.0002)	(0.0001)	(0.0001)	(0.0005)
urine_protein	0.022	0.015**	0.041***	-0.103***	-0.092***	-0.074***	0.019**		0.047*	-0.053***	-0.053***	
	(0.007)	(0.006)	(0.009)	(0.010)	(0.011)	(0.008)			(0.026)	(0.005)	(0.005)	
sight_left	0.013		0.009*							-0.006*	-0.006*	
			(0.005)							(0.003)	(0.003)	
height		-0.001**	-0.004***	-0.005***	-0.004***	-0.003***	-0.003***			-0.001*	-0.001*	
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)			(0.0004)	(0.0004)	
hear_right		-0.037**	-0.062***	-0.067***	-0.116***	-0.064*	0.126***			-0.026**	-0.026**	
		(0.018)	(0.020)	(0.025)	(0.023)	(0.036)	(0.029)			(0.013)	(0.013)	
Constant	13.837***	13.862***	14.276***	14.450***	14.273***	14.406***	11.582***	11.116***	11.587***	10.915***	10.915***	11.259***
	(0.181)	(0.123)	(0.105)	(0.156)	(0.203)	(0.198)	(0.120)	(0.158)	(0.156)	(0.071)	(0.071)	(0.161)
Observations	88,845	113,979	151,455	75,829	50,040	46,224	124,573	6,493	10,098	313,170	313,170	6,149
Adjusted R^2	0.139	0.135	0.128	0.220	0.228	0.192	0.045	0.056	0.057	0.069	0.069	0.084

Note: *p<0.1; **p<0.05; ***p<0.01

Constant, R^2 майже не відрізняються від значень моделей, параметри яких було обрано вручну.

Очікувалось, що завдяки розбиттю датасету на піддатасети значення Constant буде достатньо малим – адже розбиття гарантує урахування категоріальних змінних (sex, DRK_YN, SMK_stat_type_cd) у моделі.

Побудуємо 3 моделі без розбиття:

1. AIC, scope = age + height + weight + waistline + sight_left + sight_right + hear_left + hear_right + SBP + DBP + BLDS + tot_chole + HDL_chole + LDL_chole + triglyceride + urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT + gamma_GTP

2. AIC, scope = **sex** + age + height + weight + waistline + sight_left + sight_right + hear_left + hear_right + SBP + DBP + BLDS + tot_chole + HDL_chole + LDL_chole + triglyceride + urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT + gamma_GTP

3. AIC, scope = **SMK_Y + SMK_N + DRK_YN + sex** + age + height + weight + waistline + sight_left + sight_right + hear_left + hear_right + SBP + DBP + BLDS + tot_chole + HDL_chole + LDL_chole + triglyceride + urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT + gamma_GTP (категоріальна змінна SMK_stat_type_cd (1, 2, 3) замінена двома бінарними змінними SMK_Y, SMK_N).

```
data <- data %>%
  mutate(
    sex = as.numeric(as.factor(sex)), # male = 2, female = 1
    DRK_YN = as.numeric(as.factor(DRK_YN)), # 1 = no, 2 = yrs
    SMK_Y = (SMK_stat_type_cd == 3),
    SMK_N = (SMK_stat_type_cd == 1)
  ) %>%
  select(
    -SMK_stat_type_cd
  )
min_model <- lm(hemoglobin ~ 1, data = data)
model_all_nocat <- step(min_model, direction = "forward",
                        scope = (~ age + height + weight + waistline +
                                   sight_left + sight_right + hear_left + hear_right +
                                   SBP + DBP + BLDS +
                                   tot_chole + HDL_chole + LDL_chole + triglyceride +
                                   urine_protein + serum_creatinine +
                                   SGOT_AST + SGOT_ALT + gamma_GTP))
model_all_sexcat <- step(min_model, direction = "forward",
                         scope = (~ age + height + weight + waistline +
                                    sight_left + sight_right + hear_left + hear_right +
                                    SBP + DBP + BLDS +
                                    tot_chole + HDL_chole + LDL_chole + triglyceride +
                                    urine_protein + serum_creatinine +
                                    SGOT_AST + SGOT_ALT + gamma_GTP +
                                    sex))
model_all_allcat <- step(min_model, direction = "forward",
                         scope = (~ age + height + weight + waistline +
                                    sight_left + sight_right + hear_left + hear_right +
                                    SBP + DBP + BLDS +
                                    tot_chole + HDL_chole + LDL_chole + triglyceride +
                                    urine_protein + serum_creatinine +
                                    SGOT_AST + SGOT_ALT + gamma_GTP +
                                    sex + DRK_YN + SMK_Y + SMK_N))
stargazer(model_all_nocat, model_all_sexcat, model_all_allcat,
          out = "./output/Lab3_Regression/alldata.html", type = "html",
          label = "table:evals-reg",
          dep.var.labels = c("hemoglobin"),
          dep.var.caption = "",
          se = list(model_hc0[, 2]),
          no.space = TRUE,
          omit.stat = c("rsq", "f", "ser"))
```

	hemoglobin		
	(1)	(2)	(3)
sex		1.882***	1.807***
		(0.004)	(0.004)
height	0.067***	-0.004***	-0.005***
		(0.0002)	(0.0002)
weight	0.012***	0.010***	0.011***
	(0.002)	(0.0002)	(0.0002)
DBP	0.020***	0.017***	0.017***
	(0.002)	(0.0002)	(0.0002)
triglyceride	0.0005	0.0003***	0.0002***
	(0.0002)	(0.00002)	(0.00002)
SGOT_ALT	0.005***	0.004***	0.004***
	(0.001)	(0.0001)	(0.0001)
SMK_Y			0.241***
			(0.004)
LDL_chole	-0.001***	-0.002***	-0.001***
		(0.0001)	(0.0001)
gamma_GTP	0.002	0.0001***	-0.0002***
	(0.001)	(0.00003)	(0.00003)
HDL_chole	-0.009***	-0.003***	-0.003***
		(0.0001)	(0.0001)
tot_chole	0.005	0.006***	0.006***
	(0.0004)	(0.0001)	(0.0001)
serum_creatinine	0.109***	-0.074***	-0.070***
	(0.131)	(0.002)	(0.002)
BLDS	0.002	0.001***	0.001***
	(0.001)	(0.00005)	(0.00005)
age	-0.003	-0.014***	-0.013***
	(0.001)	(0.0001)	(0.0001)
waistline	0.004	0.001***	0.001***
		(0.0001)	(0.0001)
SMK_N			0.015***
			(0.004)
urine_protein	-0.068	-0.042***	-0.042***
		(0.003)	(0.003)
DRK_YN			0.040***
			(0.003)
SGOT_AST	-0.002	-0.001***	-0.001***
		(0.0001)	(0.0001)
sight_right	0.032		
	(0.018)		
sight_left	0.029		
SBP	-0.001	-0.003***	-0.003***
	(0.002)	(0.0001)	(0.0001)
hear_left	0.026	-0.049***	-0.047***
		(0.008)	(0.008)
hear_right	0.024	-0.032***	-0.032***
		(0.008)	(0.008)
Constant	0.076	10.164***	10.176***
	(0.181)	(0.037)	(0.038)
Observations	991,284	991,284	991,284
Adjusted R^2	0.375	0.508	0.511
Note:			*p<0.1; **p<0.05; ***p<0.01

Наявна закономірність: включення категоріальних змінних значно підвищує Constant. Неочікувано, модель (3) включає всі змінні датасету крім власне hemoglobin. При цьому значення Constant моделі (3) все ще є меншим за значення всіх моделей, побудованих «на розбитті».

Також враховуючи найбільше значення R^2 серед всіх побудованих моделей, в якості остаточної моделі оберемо модель (3), побудовану на всьому датасеті.

Перевірка на мультиколінеарність

```
> vif(model_all_allcat)
          sex           tot_chole          weight               age             DBP          SGOT_ALT
     3.268163           18.277348        3.508122          1.718001        2.324996          1.941503
        SMK_Y         triglyceride serum_creatinine               SBP            BLDS            height
     1.806842            3.793047        1.075557          2.483305        1.134132          3.564094
     SGOT_AST            HDL_chole     urine_protein            DRK_YN       LDL_chole         hear_left
     1.770981            3.590555        1.023614          1.354583       15.493765          1.425806
    gamma_GTP            waistline             SMK_N         hear_right
     1.369308            1.980847        2.536186          1.422976
```

Наявні суттєві ризики мультиколінеарних змінних. Приберемо змінну tot_chole:

```
model_all_allcat_no_tot_chole <- lm(hemoglobin ~ age + height + weight + waistline +
                            hear_left + hear_right +
                            SBP + DBP + BLDS +
                            HDL_chole + LDL_chole + triglyceride +
                            urine_protein + serum_creatinine +
                            SGOT_AST + SGOT_ALT + gamma_GTP +
                            sex + DRK_YN + SMK_Y + SMK_N,
                            data = data)
stargazer(model_all_allcat, model_all_allcat_no_tot_chole,
        out = "./output/Lab3_Regression/alldata_no_tot_chole.html", type = "html",
        label = "table:evals-reg",
        dep.var.labels = c("hemoglobin"),
        dep.var.caption = "",
        se = list(model_hc0[, 2]),
        no.space = TRUE,
        omit.stat = c("rsq", "f", "ser"))
vif(model_all_allcat_no_tot_chole)
```

```
> vif(model_all_allcat_no_tot_chole)
          age              height            weight          waistline       hear_left        hear_right
     1.716793            3.562698        3.506462          1.980416        1.425775          1.422926
          SBP                 DBP              BLDS          HDL_chole       LDL_chole      triglyceride
     2.483276            2.322921        1.134073          1.342993        1.024560          1.334250
urine_protein    serum_creatinine          SGOT_AST           SGOT_ALT       gamma_GTP               sex
     1.023583            1.075507        1.770458          1.940449        1.367957          3.267008
       DRK_YN               SMK_Y             SMK_N
     1.354324            1.806644        2.536008
>
```

	hemoglobin	
	(1)	(2)
sex	1.807***	1.803***
		(0.004)
tot_chole	0.006**	
	(0.0004)	
weight	0.011***	0.011***
	(0.002)	(0.0002)
age	-0.013***	-0.013***
	(0.001)	(0.0001)
DBP	0.017***	0.017***
	(0.002)	(0.0002)
SGOT_ALT	0.004***	0.005***
	(0.001)	(0.0001)
SMK_Y	0.241***	0.243***
		(0.004)
triglyceride	0.0002	0.001***
	(0.0002)	(0.00001)
serum_creatinine	-0.070***	-0.070***
	(0.131)	(0.002)
SBP	-0.003***	-0.003***
	(0.002)	(0.0001)
BLDS	0.001	0.001***
	(0.001)	(0.00005)
height	-0.005	-0.005***
		(0.0002)
SGOT_AST	-0.001	-0.001***
		(0.0001)
HDL_chole	-0.003	0.002***
		(0.0001)
urine_protein	-0.042	-0.042***
		(0.003)
DRK_YN	0.040	0.042***
		(0.003)
LDL_chole	-0.001	0.004***
		(0.00003)
hear_left	-0.047	-0.049***
		(0.008)
gamma_GTP	-0.0002	-0.0001***
	(0.001)	(0.00003)
waistline	0.001	0.001***
		(0.0001)
SMK_N	0.015	0.014***
		(0.004)
hear_right	-0.032	-0.034***
		(0.008)
Constant	10.176***	10.279***
	(0.181)	(0.038)
Observations	991,284	991,284
Adjusted R^2	0.511	0.510

Note: *p<0.1; **p<0.05; ***p<0.01

Видалення tot_chole майже не вплинуло на загальний характер моделі. Ризик мультиколінеарності зменшено.

Тестування гіпотез

Для кожної незалежної змінної було протестовано гіпотези:

H0: вплив змінної не є статично значущим

Ha: вплив змінної є статично значущим

Результати тестування:

```
> hypos <- c('age=0', 'height=0', 'weight=0', 'waistline=0',
+            'hear_left=0', 'hear_right',
+            'SBP=0', 'DBP=0', 'BLDS=0',
+            'HDL_chole=0', 'LDL_chole=0', 'triglyceride=0',
+            'urine_protein=0', 'serum_creatinine=0',
+            'SGOT_AST=0', 'SGOT_ALT=0', 'gamma_GTP=0',
+            'sex=0', 'DRK_YN=0', 'SMK_YTRUE=0', 'SMK_NTRUE=0')
> for (hypo in hypos) {
+   current_hypos <- hypos[!hypos %in% c(hypo)]
+   print(linearHypothesis(model_all_allcat_no_tot_chole,
+                 current_hypos,
+                 vcov. = hccm(model_all_allcat_no_tot_chole, type='hc1')))
+ }
Linear hypothesis test

Hypothesis:
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df    F    Pr(>F)
1 991282
2 991262 20 47365 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
```

```
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df    F    Pr(>F)
1 991282
2 991262 20 21188 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df    F    Pr(>F)
1 991282
2 991262 20 24676 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df    F    Pr(>F)
1 991282
2 991262 20 39270 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F     Pr(>F)
1 991282
2 991262 20 53115 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F     Pr(>F)
1 991282
2 991262 20 53098 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
```

```
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1 991282
2 991262 20 50575 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1 991282
2 991262 20 47480 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.
```

```
  Res.Df Df      F    Pr(>F)
1 991282
2 991262 20 52426 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F    Pr(>F)
1 991282
2 991262 20 46834 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F    Pr(>F)
1 991282
2 991262 20 52828 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
```

```
BLDS = 0
HDL_chole = 0
LDL_chole = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F      Pr(>F)
1 991282
2 991262 20 47363 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F      Pr(>F)
1 991282
2 991262 20 53340 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
```

```
        urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
        gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

   Res.Df Df      F      Pr(>F)
1 991282
2 991262 20 28129 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

   Res.Df Df      F      Pr(>F)
1 991282
2 991262 20 48290 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

   Res.Df Df      F      Pr(>F)
1 991282
2 991262 20 33187 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
```

```
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1 991282
2 991262 20 46674 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
DRK_YN = 0
SMK_YTRUE = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1 991282
2 991262 20 4624 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
SMK_YTRUE = 0
```

```
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1 991282
2 991262 20 42779 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_NTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1 991282
2 991262 20 40471 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Linear hypothesis test

Hypothesis:
age = 0
height = 0
weight = 0
waistline = 0
hear_left = 0
hear_right = 0
SBP = 0
DBP = 0
BLDS = 0
HDL_chole = 0
LDL_chole = 0
triglyceride = 0
urine_protein = 0
serum_creatinine = 0
SGOT_AST = 0
SGOT_ALT = 0
gamma_GTP = 0
sex = 0
DRK_YN = 0
SMK_YTRUE = 0

Model 1: restricted model
Model 2: hemoglobin ~ age + height + weight + waistline + hear_left +
    hear_right + SBP + DBP + BLDS + HDL_chole + LDL_chole + triglyceride +
    urine_protein + serum_creatinine + SGOT_AST + SGOT_ALT +
    gamma_GTP + sex + DRK_YN + SMK_Y + SMK_N

Note: Coefficient covariance matrix supplied.

  Res.Df Df     F    Pr(>F)
1 991282
2 991262 20 30434 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Для кожної змінної було відкинуто нульову гіпотезу. Тобто вплив кожного параметру є статично значущим, що є очікуваним.

Результати тестування моделі:

```
coeftest(model_all_allcat_no_tot_chole, vcoc. = hccm(model_all_allcat_no_tot_chole, type = 'hc0'))

t test of coefficients:

                      Estimate  Std. Error   t value  Pr(>|t|)
(Intercept)         1.0279e+01  3.7522e-02  273.9331 < 2.2e-16 ***
age                -1.2680e-02  1.0296e-04 -123.1612 < 2.2e-16 ***
height             -4.9152e-03  2.2658e-04  -21.6930 < 2.2e-16 ***
weight              1.0726e-02  1.6674e-04   64.3289 < 2.2e-16 ***
waistline           8.7672e-04  1.3233e-04    6.6252 3.469e-11 ***
hear_left          -4.8816e-02  7.6183e-03   -6.4077 1.478e-10 ***
hear_right         -3.3866e-02  7.7331e-03   -4.3793 1.190e-05 ***
SBP                -3.2424e-03  1.2075e-04  -26.8525 < 2.2e-16 ***
DBP                 1.7446e-02  1.7174e-04  101.5835 < 2.2e-16 ***
BLDS                1.3162e-03  4.9097e-05   26.8086 < 2.2e-16 ***
HDL_chole           2.4024e-03  8.4843e-05   28.3157 < 2.2e-16 ***
LDL_chole           4.3241e-03  3.2273e-05  133.9853 < 2.2e-16 ***
triglyceride        1.0678e-03  1.2818e-05   83.3061 < 2.2e-16 ***
urine_protein      -4.2377e-02  2.5763e-03  -16.4484 < 2.2e-16 ***
serum_creatinine   -6.9573e-02  2.4049e-03  -28.9299 < 2.2e-16 ***
SGOT_AST           -1.2005e-03  6.3139e-05  -19.0137 < 2.2e-16 ***
SGOT_ALT            4.5152e-03  5.9008e-05   76.5180 < 2.2e-16 ***
gamma_GTP          -1.1522e-04  2.5877e-05   -4.4526 8.486e-06 ***
sex                 1.8034e+00  4.0361e-03  446.8203 < 2.2e-16 ***
DRK_YN              4.2028e-02  2.5936e-03   16.2043 < 2.2e-16 ***
SMK_YTRUE           2.4268e-01  3.6409e-03   66.6531 < 2.2e-16 ***
SMK_NTRUE           1.3525e-02  3.6345e-03    3.7212 0.0001983 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```