Ökonometria

Beadandó feladat

Granát Marcell - AYCOPF

2021. január 2.

Tartalomjegyzék

Bevezetés **2**

Elméleti megfontolás **2**

Modellbecslés és hipotézis vizsgálat **2**

Becsült előjelek értelmezése 2

Szignifikancia értelmezése 3

Együtthatók összegére tett előzetes feltevés ellenőrzése 3

Átlagos kiadási rugalmasságok **5**

A modell bővítése **6**

R kódok **9**

Bevezetés

Ay alábbi rövid kézirat a brit **Family Expenditure Survey** adataiból kinyert fogyasztási jellemzőket tárgyalja. Az empirikus elemzés egyszerű lineáris regresszió alkalmazásával készül. Az adatok összesen 1519 család jövedelmét és kiadását tartalmazza, továbbá 6 termékcsoport (alkohol, ruházkodás, nem alkoholtartalmú élelmiszer, fűtés, közlekedés, egyéb) fogyasztáson belüli részarányát. Dolgozatom során én kizárólag az egy gyermekes családokra szűkítem az elemzést[1].

Elméleti megfontolás

Miután az egyes termékek fogyasztáson belüli aránya adott, így egy egyszerű lineáris regresszióban magyarázóváltozónak választva a fogyasztási kiadás logaritmusát jól interpretálható eredményeket kapunk. Ebben a felírásban (1. egyenlet) γ_j értelmezése az alábbi: a fogyasztás 1%-kal magasabb értéke mellett c.p. várhatóan hány $\%P$-al változik meg j-edik termékcsoportra fordított kiadás teljes fogyasztáson belüli megoszlása.

$$w_j = \beta_j + \gamma_j log(cons) + u_j, \quad j = 1, 2, ..., 6 \tag{1}$$

Miután az egyes termékcsoportokra fordított kiadások arányainak összegének 1-nek kell lennie és a hibatagok várható értéke 0, így az 1. egyenlet alapján adóddik a következő összefüggés:

$$\begin{aligned} \sum_{j=1}^{6} w_j &= 1 \\ \sum_{j=1}^{6} \beta_j + \gamma_j log(cons) &= 1 \end{aligned} \tag{2}$$

A 2. egyenletben β_j-k összege jelenít meg egy fogyasztás szintjétől független komponenst, míg a w_j becsült értékek egymáshoz viszonyított mértéke a fogyasztás függvényeként tud változni. Annak érdekében, hogy a fogyasztás nagyságának minden értéke mellett biztosítva legyen, hogy $\sum w_j = 1$, szükséges megkötés tehát, hogy a fogyasztás logaritmusához rendelt együtthatók összege 0 legyen, ezáltal pedig adódik, hogy a 6 termékre becsült egyenletekben szereplő konstansok összege 1 legyen.

$$\begin{aligned} \sum \gamma_j log(cons) &= 0 \\ \sum \beta_j &= 1 \end{aligned} \tag{3}$$

Mivel az egyes γ_j együttható megadják tehát, hogy a j-edik termék kereslete nő vagy csökken-e a fogyasztás növekedésével, így amennyiben a becsült együttható értéke 0 alatti (feletti), úgy a j-edik termékcsoportra költött összeg várhatóan nagyobb hányadát teszi ki a teljes fogyasztásnak, ha a fogyaszás szintje alacsonyabb (magasabb).

Modellbecslés és hipotézis vizsgálat

Becsült előjelek értelmezése

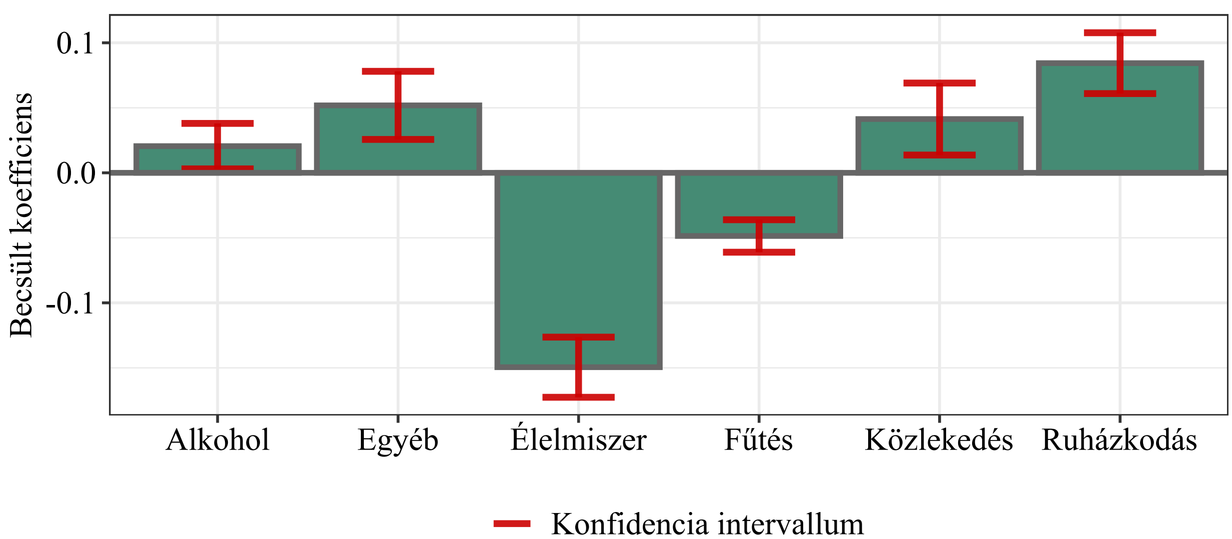
Az 1. egyenlet által leírt lineáris regresszió modellekben szereplő γ_j együtthatókat ismerteti az 1. ábra. Az ábráról leolvasható, hogy csökken a fogyasztáson belüli aránya az élelmiszernke és a fűtésnek, míg az alkohol,

[1] Feladatkiírás által előírt megkötés.

a közlekedés, a ruházkodás és az egyéb termékcsoportba tartozó javak aránya növekszik. Ez **az eredmény összhangban van azzal, amit empirikus vizsgálat nélkül mondanánk, ugyanis élelmiszerre és fűtésre mindenkor szükséges egy bizonyos összeget kiadni.** Az alkohol fogyasztás és a ruházkodás szűkösség esetén visszaszorítható, a közlekedésnek is könnyen található olcsóbb alternatívája.

Szignifikancia értelmezése

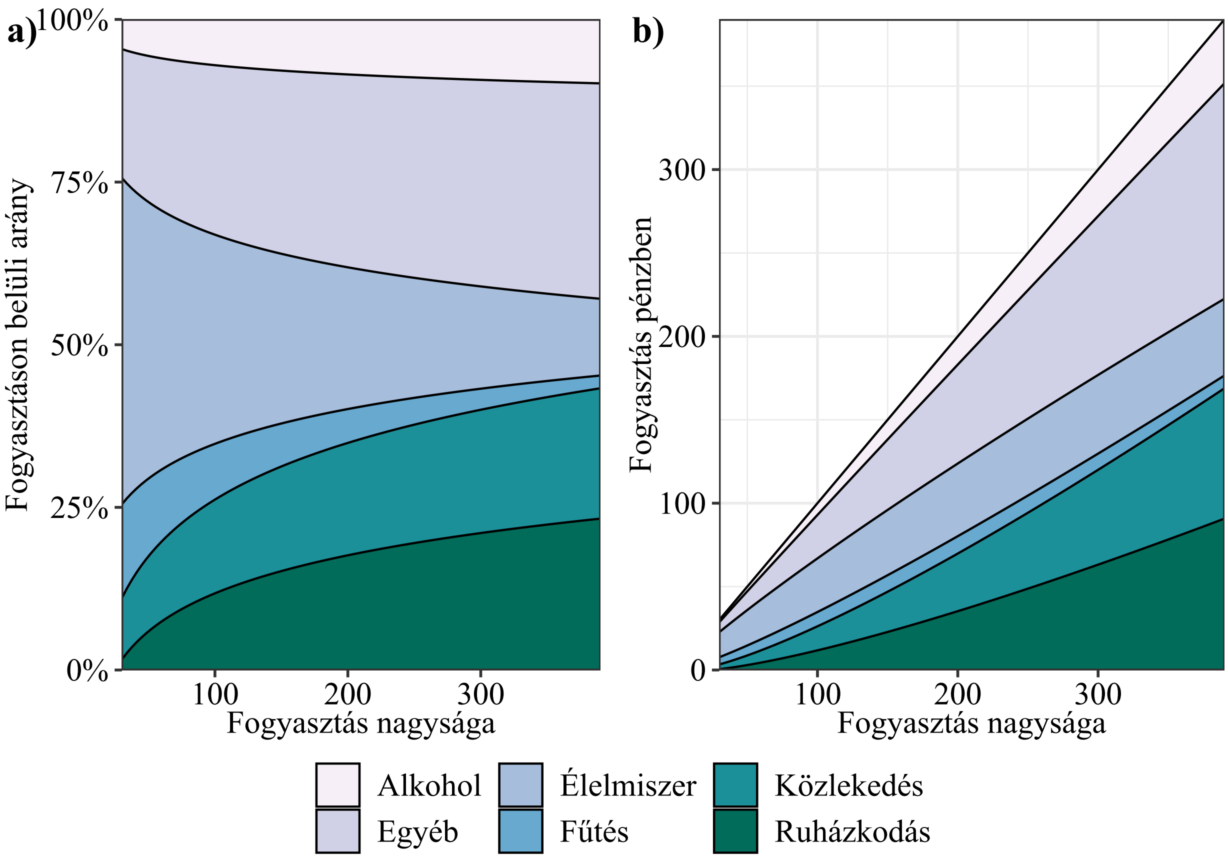
Statiszikai értemelben valamennyi meredekségi együttható szignifikánsnak bizonyult (belátható ez abból is, hogy a paraméterbecslésekhez tartozó 95%-os konfidencia intervallumok egyikse sem tartalmazza a zérust), az empirikus szignifikanciaszint mindössze némely tengelymetszet esetében nem különbözik szignifikánsan 0-tól. Közgazdasági értelemben vett szignifikancia (a becsült hatások kellően nagyok-e ahhoz, hogy figyelmet fordítsunk rá) értelmezéséhez a 2. ábra szolgáltat információt. A két diagram vízszintes tengelyén látható a mintában szereplő fogyasztási értékek intervalluma, az *a)* panel függőleges tengelyén a termékcsoportok fogyasztáson belüli megoszlásának modellbecslése, míg a *b)* panelen az egyes termékcsoportokra fordított kiadás pénzben (szintén modellbecslés). A diagramokból kivehető, hogy ezen becsült paraméterek mentén jelentős átrendeződés van a termékekre fordított kiadások arányai között, amíg a legkisebb kiadástól a legnagyobbig eljutunk. Az egyetlen viszonylag állandó fogyasztási aránnyal rendelkező terméktípus az alkohol. Ennek oka lehet, hogy az alkohol fogyasztás fogyasztási hányada nem lineáris függvénye a fogyasztás szintjének, ahogyan ezt a 3. ábra megvilágítja. Mind a kiemelten alacsony, mind a kiemelten magas fogyasztással rendelkező háztartások esetében alacsonyabb az alkoholra fordított kiadások aránya, mint a nem szélsőséges eseteknél. Előbbiek esetében valószínűleg nincs meg a forrás, amit szeszes italokra költenének, így vagy kevesebbet, vagy olcsóbb termékeket vásárolnak, míg utóbbiak esetében valószínűleg a kiemelkedő szocio-ökonómiai háttér folytán alacsonyabb a fogyasztás.



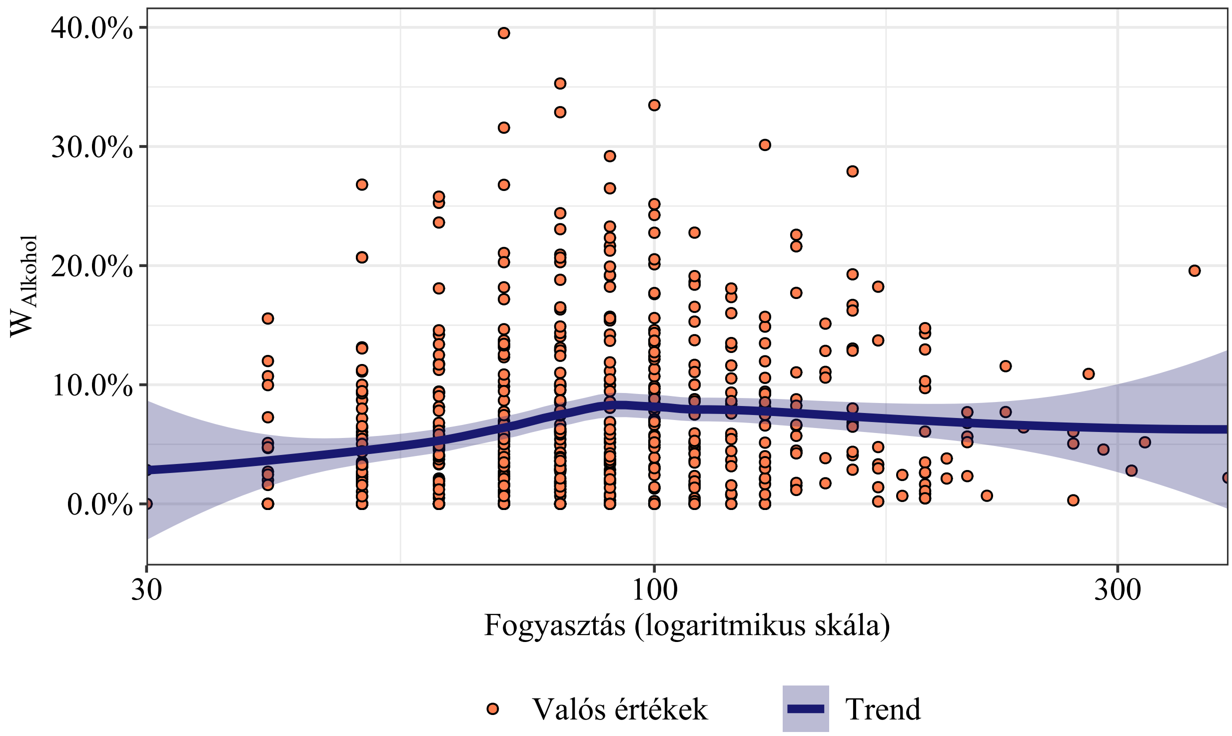
1. ábra. Termékcsoportonként becsült γ koefficiens

Együtthatók összegére tett előzetes feltevés ellenőrzése

A paraméterbecslések között fennálló algebrai kapcsolatot az alábbi módon ellenőrzöm: (1) Regresszió becslése, melyben eredményváltozó, hogy egy j-edik terméknek mekkora a megoszlása a fogyasztáson belül, magyarázó változó pedig, hogy mely termékcsoportról van szó (egyszerű dummy) és a fogyasztás logaritmusának termékcsoporttal vett interakciója. (2) Ezen a modellen pedig már ellenőrizhetőek az előzetesen együtthatók összegére feltett megszorítások Wald-féle F-próbával. A 2. egyenletben ismertetett



2. ábra. A becsült paraméterek alapján az egyes termékcsoportok fogyasztása a teljes fogyasztás függvényében



3. ábra. Az fogyasztás szintje és alkoholra fordított fogyasztási hányad megoszlása

algebrai kapcsolat fennállásának nullhipotézisét, miszerint mindegyik együttható összege 1 nem tudjuk elutasítani (F-próba p-értéke $49,47\%$). A 3. egyenletben látott H_0-t - miszerint a fogyasztás logaritmusának egyes termékcsoportokhoz becsült együtthatóinak összege 1 - szintén nem lehet elvetni (F-próba p-értéke $99,99\%$).

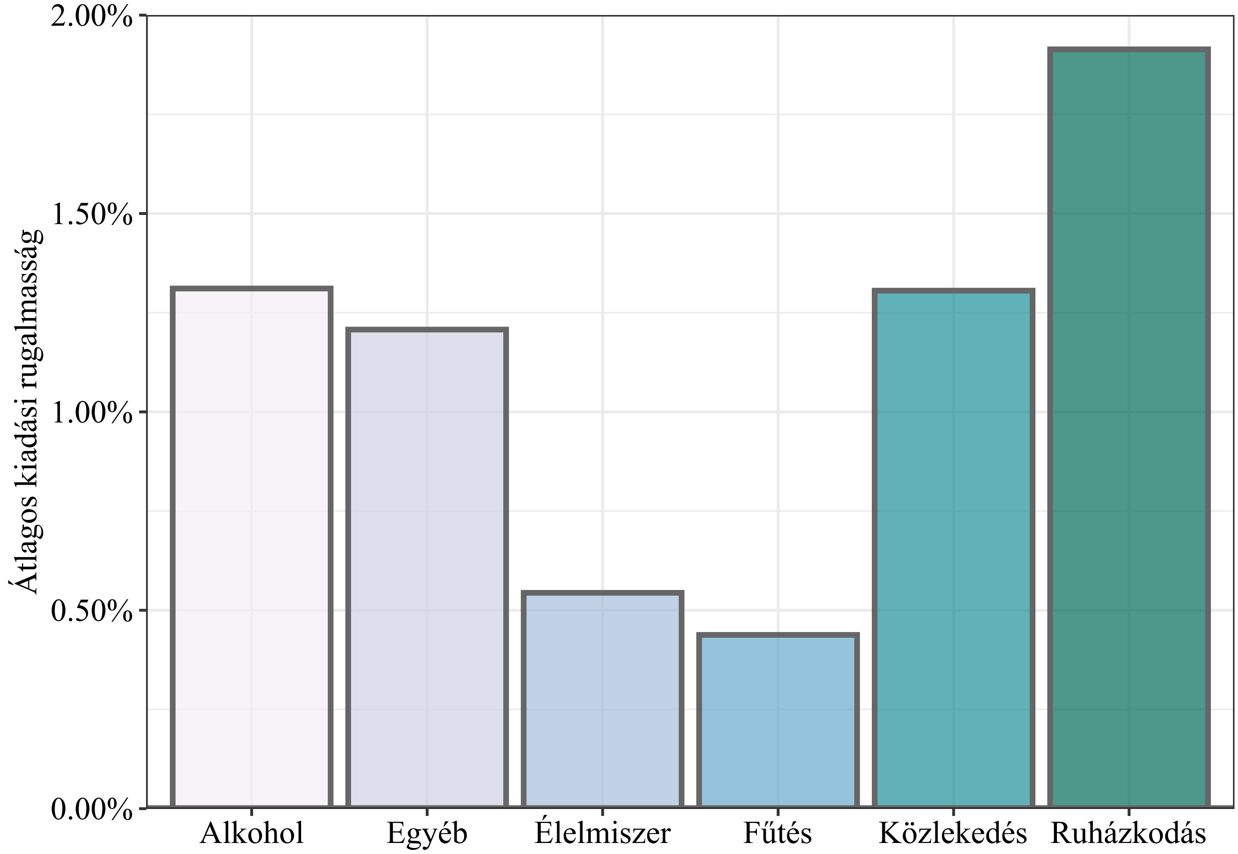
Átlagos kiadási rugalmasságok

Az előzőekben bemutatott lineráris modellben a fogyasztás logaritmusához rendelt együttható állandó volt, függetlenül attól, hogy mennyi az aktuális fogyasztás. Azonban itt a fogyasztáson belüli megoszlása a termékcsoportnak volt modellezve, nem a fogyasztásra fordított összeg. Érdemes megvizsgálni, hogy az egyes termékekre fordított összeg miként változik együtt a jöveledelemmel. Ehhez az átlagos kiadási rugalmasságot (ℓ) választom megfelelő eszköznek, ami megmutatja, hogy j-edik termékre fordított összeg hány %-kal változik meg várhatóan c. p., ha 1%-kal megnő a fogyasztás. Ez az érték viszont rögzített β_j és γ mellett sem lesz konstans a fogyasztás mértékének függvényében. A számszerűsítéséhez kiszámítom minden mintában lévő fogyasztóra ($i = 1, 2, ..., 594$) a modellek alapján becsült kiadási rugalmasságot, és ezeknek az értékeknek veszem számtani átlagát mindegyik fogyasztó esetében. A paraméterbecslésekből és adatokból való levezetést a 4. egyenlet írja le.

$$\begin{aligned} \hat{w}_{0ji} &= \hat{\beta}_j + \hat{\gamma}_j log(cons_i), \quad i = 1, 2, ..., 594 \\ \hat{w}_{1ji} &= \hat{\beta}_j + \hat{\gamma}_j log(cons_i \times 1,01) \end{aligned} \tag{4}$$

$$\ell_j = \frac{\sum_{i=1}^{594} \frac{\hat{w}_{1ji} \times 1,01 cons - \hat{w}_{0ji} \times cons}{\hat{w}_{0ji} \times cons}}{594}$$

A 6 termékcsoporthoz becsült átlagos kiadási rugalmasságot 4. ábra mutatja be. Mindegyik esetben 0-nál nagyobb értéket kaptunk, tehát mindegyik esetben **normál javakról** van szó, ahol a kiadás mértékének függvényében növekszik az egyes termékekre kiadott összeg.

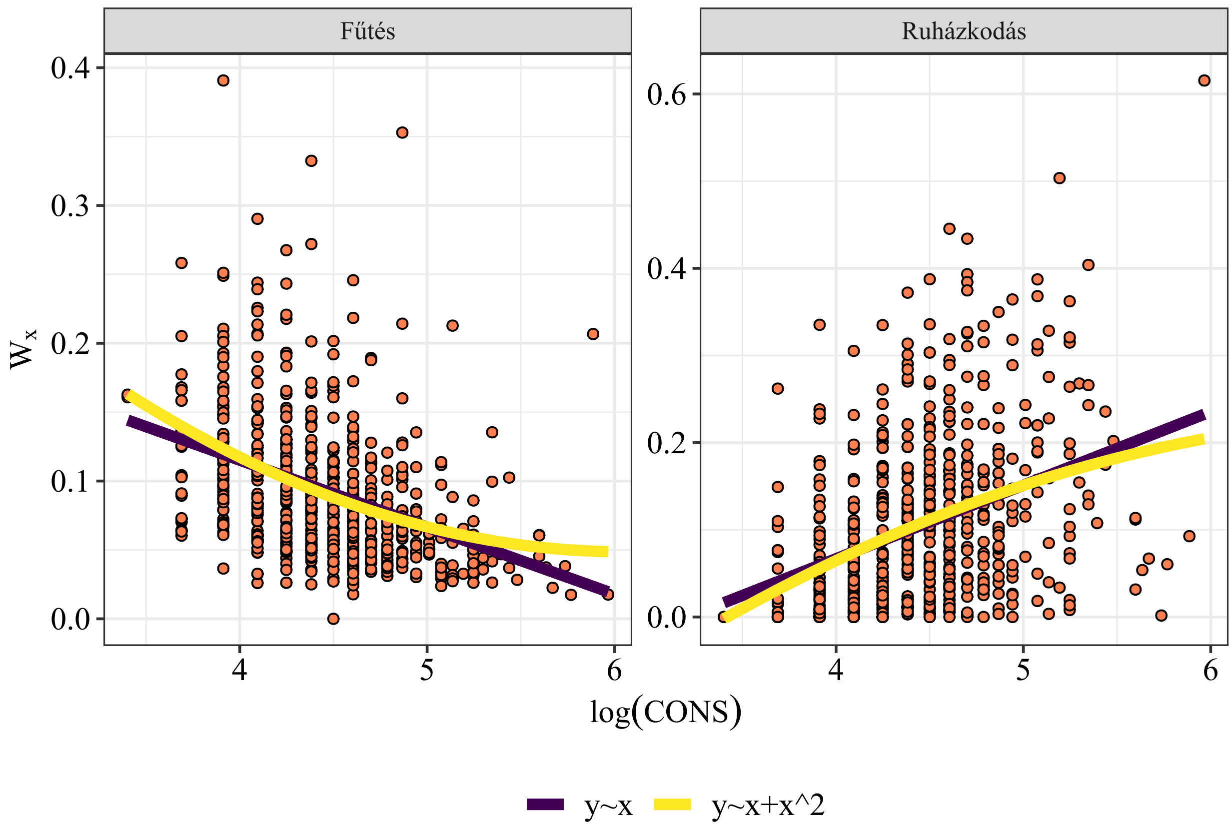


4. ábra. Regressziós modellek alapján becsült átlagos kiadási rugalmasság termékcsoportonként

A modell bővítése

A 3. ábra alapján belátható, hogy még a fogyasztás logaritmizálása után is felmerülhet, hogy a magyarázóváltozónak kvadratikus alakját is felhasználjuk a regresszióban. Dolgozatomban ezt a kiterjesztést a ruházkodás és az üzemanyag fogyasztásra fókuszálva vizsgálom el[2].

[2] Feladatleírás által megadott korlátozás


Fűtés
Ruházkodás
W_x
0.4
0.3
0.2
0.1
0.0
0.6
0.4
0.2
0.0
4
5
6
log(CONS)
y~x
y~x+x^2

2. táblázat. A fűtés fogyasztáson belüli arányára becsült eredeti modell (I.) és a bővített modell (II.)

Változó	I. modell	II. modell
Konstans	0,3095***	0,6459***
log(cons)	-0,0486***	-0,1974***
log(cons)^2	NA	0,0163**

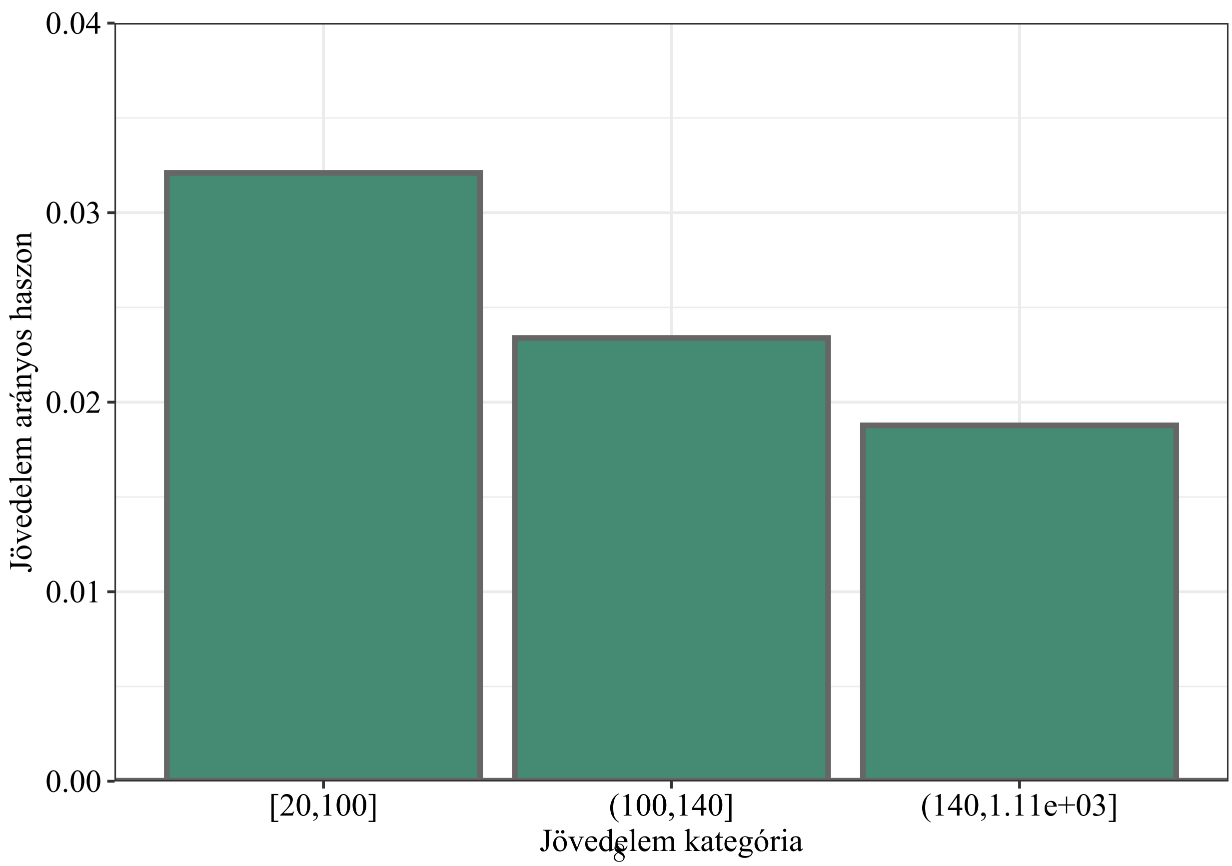
Jelölések:
* 10%-on szignifikáns
** 5%-on szignifikáns
*** 1%-on szignifikáns

1. táblázat: A ruházkodás fogyasztáson belüli arányára becsült eredeti modell (I.) és a bővített modell (II.)

Változó	I. modell	II. modell
Konstans	-0,2705***	-0,5963*
log(cons)	0,0843***	0,2284*
log(cons)^2	NA	-0,0158

3. táblázat: Átlagos parciális hatás az eredeti modellben (I. modell) és a kvadratikus komponenssel bővített modellben (II. modell)

Változó	I. modell	II. modell
Ruházkodás	0,0843	0,0843
Fűtés	-0,0486	-0,0486



R kódok

```
# SETUP ------------------------------------------------------------------------------

library(tidyverse)
library(granatlib) # my personal package: https://github.com/MarcellGranat/granatlib
theme_set(theme_granat())
dat <- rio::import('cons.xlsx') %>%
  filter(NK == 1)

# Frequently used functions ==========================================================

f.pvalues <- function(x) {
  case_when(
    x <= .1 & x > .05  ~ '10%',
    x <= .05 & x > .01  ~ '5%',
    x < .01 ~ '1%',
    T ~ '-'
  )
}

f.products <- function(x) {
  case_when(
    x == 'WALC' ~ 'Alkohol',
    x == 'WCLOTH' ~ 'Ruházkodás',
    x == 'WFOOD' ~ 'Élelmiszer',
    x == 'WFUEL' ~ 'Fűtés',
    x == 'WOTHER' ~ 'Egyéb',
    x == 'WTRANS' ~ 'Közlekedés'
  )
}
# PART I. ----------------------------------------------------------------------------

# a ==================================================================================

### Statistical significance #########################################################
dat %>%
  {select(., CONS, which(str_detect(names(.), 'W')))} %>%
  pivot_longer(-1, names_to = 'x', values_to = 'w') %>%
  mutate(CONS = log(CONS)) %>%
  group_by(x) %>%
  do(broom::tidy(lm(w ~ CONS, .), conf.int = T, conf.level = 0.99)) %>%
  filter(term == 'CONS') %>%
  mutate(x = f.products(x)) %>%
  ggplot +
  geom_col(aes(x, estimate), fill = 'aquamarine4', color = 'grey40', size = 1) +
  geom_errorbar(aes(x, ymin = conf.low, ymax = conf.high,
                    color = "Konfidencia intervallum"),
                width = 0.4, alpha = 0.9, size = 1.2) +
  geom_hline(yintercept = 0, color = 'grey40', size = 1) +
  labs(x = NULL, y = 'Becsült koefficiens', color = NULL) +
  scale_color_manual(values = c('red3'))

```

```r
### Economic significance ###############################################################
dat %>%
  {select(., which(str_detect(names(.), 'W')))} %>%
  apply(2, function(y) lm(data = data.frame(y, x = dat$CONS),
                          formula = y ~ I(log(x)))) %>%
  lapply(function(mod) predict.lm(object = mod,
                                  newdata = data.frame(x = seq(from = min(dat$CONS),
                                                               to = max(dat$CONS),
                                                               length.out = 100)))) %>%
  lapply(function(y) data.frame(y)) %>%
  reduce(cbind) %>%
  set_names(names(dat)[which(str_detect(names(dat), 'W'))]) %>%
  mutate(cons = seq(from = min(dat$CONS), to = max(dat$CONS),
                    length.out = 100)) %>%
  pivot_longer(-cons) %>%
  mutate(name = f.products(name)) %>%
  {
    ggpubr::ggarrange(
      ggplot(.) +
        geom_area(aes(x = cons, y = value, fill = name), position = 'stack',
                  color = 'black', size = .4) +
        scale_x_continuous(expand = c(0, 0)) +
        scale_y_continuous(expand = c(0, 0), labels = scales::percent) +
        scale_fill_brewer(palette = 10) +
        labs(x = 'Fogyasztás nagysága', y = 'Fogyasztáson belüli arány', fill = NULL),
      mutate(., value = value*cons) %>%
        ggplot +
        geom_area(aes(x = cons, y = value, fill = name), position = 'stack',
                  color = 'black', size = .4) +
        scale_x_continuous(expand = c(0, 0)) +
        scale_y_continuous(expand = c(0, 0)) +
        scale_fill_brewer(palette = 10) +
        labs(x = 'Fogyasztás nagysága', y = 'Fogyasztás pénzben', fill = NULL),
      nrow = 1, common.legend = T, legend = 'bottom', labels = c('a)', 'b)')
    )
  }

dat %>%
  select(CONS, WALC) %>%
  ggplot(aes(CONS, WALC)) +
  geom_point(shape = 21, mapping = aes(CONS, WALC, fill = 'Valós értékek'),
             color = 'black') +
  geom_smooth(mapping = aes(CONS, WALC, color = 'Trend'), size = 1.5,
              fill = 'midnightblue', alpha = .3) +
  scale_y_continuous(labels = scales::percent) +
  scale_x_log10(expand = c(0, 0)) +
  scale_color_manual(values = c('midnightblue')) +
  scale_fill_manual(values = c('coral')) +
  labs(x = 'Fogyasztás (logaritmikus skála)', y = expression(W[Alkohol]),
       color = NULL, fill = NULL) +
  theme(legend.box = 'horizontal')

# c =====================================================================================
```

```r

dat %>%
  select(CONS, which(str_detect(names(.), 'W'))) %>%
  pivot_longer(-1) %>%
  lm(formula = value ~ name + I(log(CONS)):name) %T>%
  {
    restrict <- paste(paste(coef(.) %>% names %>% tail(6), collapse = ' + '), '= 0')
    car::linearHypothesis(model = ., restrict) %>% print
  } %>%
  {
    restrict <- paste(paste(coef(.) %>% names, collapse = ' + '), '= 1')
    car::linearHypothesis(model = ., restrict)
  }

# b ==================================================================================
#
### Mean ellasticity to CONS ##########################################################
dat %>%
  {select(., which(str_detect(names(.), 'W')))} %>%
  apply(2, function(y) lm(data.frame(y, CONS = dat$CONS), formula = y ~ log(CONS))) %>%
  lapply(function(mod) {
    y0 <- predict.lm(object = mod, newdata = dat)*dat$CONS
    y1 <- predict.lm(object = mod, newdata = mutate(dat, CONS = CONS*1.01))*dat$CONS*1.01
    (y1-y0)/y0
  }
  ) %>%
  sapply(mean) %>%
  {data.frame(x = f.products(names(.)), ell = .)} %>%
  ggplot() + geom_col(aes(x, ell, fill = x), color = 'grey40',
                              alpha = .7, size = 1) +
  geom_hline(yintercept = 0, color = 'grey40', size = 1.2) +
  scale_y_continuous(labels = scales::percent, limits = c(0, .02), expand = c(0, 0)) +
  labs(x = NULL, y = 'Átlagos kiadási rugalmasság', fill = NULL) +
  scale_fill_brewer(palette = 10, guide = F)
# PART II. ---------------------------------------------------------------------------
dat %>%
  transmute(CONS = log(CONS), WCLOTH, WFUEL) %>% # neptun: AYCOPF
  pivot_longer(-1) %>%
  mutate(
    name = f.products(name)
  ) %>%
  ggplot(aes(CONS, value)) +
  geom_point(shape = 21, fill = 'coral') +
  geom_smooth(method = 'lm', formula = "y~x",
              aes(color = 'y~x'), se = F, size = 2) +
  geom_smooth(method = 'lm', formula = "y~poly(x, 2)", aes(color = 'y~x+x^2'),
              se = F, size = 2) +
  facet_wrap(~name, scales = 'free_y') +
  labs(y = expression(W[x]), x = expression(log(CONS)), color = NULL) +
  scale_color_viridis_d()

dat %>%
    {mod1 <- lm(., formula = WCLOTH ~ I(log(CONS)))
    mod2 <- lm(., formula = WCLOTH ~ I(log(CONS)) + I(log(CONS)^2))
```

```
    merge(broom::tidy(mod1), broom::tidy(mod2), by = ’term’, all = T)} %>%
  select(1, 2, 5, 6, 9) %>%
  mutate(
    p.value.x = case_when(
      p.value.x <= .1 & p.value.x > .05 ~ ’*’,
      p.value.x <= .05 & p.value.x > .01 ~ ’**’,
      p.value.x <= .01 ~ ’***’,
      T ~ ’’
    ),
    p.value.y = case_when(
      p.value.y <= .1 & p.value.y > .05 ~ ’*’,
      p.value.y <= .05 & p.value.y > .01 ~ ’**’,
      p.value.y <= .01 ~ ’***’,
      T ~ ’’
    )
  ) %>%
  transmute(term = (c(’Konstans’, ’log(cons)’, ’log(cons)^2’)),
            mod1 = str_c(format(estimate.x, decimal.mark = ’,’, digits = 3), p.value.x),
            mod2 = str_c(format(estimate.y, decimal.mark = ’,’, digits = 3), p.value.y)
            ) %>%
  knitr::kable(col.names = c(’Változó’, ’I. modell’, ’II. modell’),
               format.args = list(decimal.mark = ’,’), digits = 2,
               caption = ’A ruházkodás fogyasztáson belüli
               arányára becsült eredeti modell (I.) és a bővített modell (II.)’,
                align = c(’l’, ’c’, ’c’)) %>%
  kableExtra::footnote(symbol = c(’10%-on szignifikáns’, ’5%-on szignifikáns’,
                                  ’1%-on szignifikáns’),
                       symbol_manual = c(’*’, ’**’, ’***’),
                       symbol_title = ’Jelölések:’) # TODO footnote
dat %>%
    {mod1 <- lm(., formula = WFUEL ~ I(log(CONS)))
    mod2 <- lm(., formula = WFUEL ~ I(log(CONS)) + I(log(CONS)^2))
    merge(broom::tidy(mod1), broom::tidy(mod2), by = ’term’, all = T)} %>%
  select(1, 2, 5, 6, 9) %>%
  mutate(
    p.value.x = case_when(
      p.value.x <= .1 & p.value.x > .05 ~ ’*’,
      p.value.x <= .05 & p.value.x > .01 ~ ’**’,
      p.value.x <= .01 ~ ’***’,
      T ~ ’’
    ),
    p.value.y = case_when(
      p.value.y <= .1 & p.value.y > .05 ~ ’*’,
      p.value.y <= .05 & p.value.y > .01 ~ ’**’,
      p.value.y <= .01 ~ ’***’,
      T ~ ’’
    )
  ) %>%
  transmute(term = (c(’Konstans’, ’log(cons)’, ’log(cons)^2’)),
            mod1 = str_c(format(estimate.x, decimal.mark = ’,’, digits = 3), p.value.x),
            mod2 = str_c(format(estimate.y, decimal.mark = ’,’, digits = 3), p.value.y)
            ) %>%
  knitr::kable(col.names = c(’Változó’, ’I. modell’, ’II. modell’),
```

```
              format.args = list(decimal.mark = ','), digits = 2,
              format = 'latex',
              caption = 'A fűtés fogyasztáson belüli
              arányára becsült eredeti modell (I.) és a bővített modell (II.)',
               align = c('l', 'c', 'c')) %>%
  kableExtra::footnote(symbol = c('10%-on szignifikáns', '5%-on szignifikáns',
                                  '1%-on szignifikáns'),
                       symbol_manual = c('*', '**', '***'),
                       symbol_title = 'Jelölések:')
# PART III. ---------------------------------------------------------------------------

dat %>%
  select(WCLOTH, WFUEL) %>%
  apply(2, function(y) {
    mod1 <- lm(data = data.frame(y = y, x = log(dat$CONS)), formula = y ~ x)
    coef1 <- mod1$coeff['x']
    mod2 <- lm(data = data.frame(y = y, x = log(dat$CONS)), formula = y ~ x + x^2)
    coef2 <- mean(predict.lm(newdata = data.frame(x = log(dat$CONS) + 1), object = mod2)
                  - mod2$fitted)
    data.frame(coef1, coef2)
  }
  ) %>%
  reduce(rbind) %>%
  transmute(var = c('Ruházkodás', 'Fűtés'), coef1, coef2) %>%
  set_names('Változó', 'I. modell', 'II. modell') %>%
  knitr::kable(
    caption = 'Átlagos parciális hatás az eredeti modellben (I. modell) és a
    kvadratikus komponenssel bővített modellben (II. modell)',
    align = c('l', 'c', 'c'),
    format.args = list(digits = 3, decimal.mark = ',')
  )
# PART V. -----------------------------------------------------------------------------

dat %>%
  mutate(
    class = cut(INCOME, breaks = quantile(dat$INCOME, probs = (0:3)/3),
                include.lowest = T),
    cons_food = CONS*WFOOD,
    help = cons_food*.1/INCOME
  ) %>%
  group_by(class) %>%
  summarise(help = mean(help)) %>%
  ggplot(aes(factor(class), help)) +
  geom_col(fill = 'aquamarine4', color = 'grey40', size = 1) +
  geom_hline(yintercept = 0, color = 'grey40', size = 1.2) +
  labs(x = 'Jövedelem kategória', y = 'Jövedelem arányos haszon') +
  scale_y_continuous(limits = c(0, .04), expand = c(0, 0))
```