Ökonometria

Beadandó feladat

Granát Marcell - AYCOPF

2020. december 27.

Tartalomjegyzék

Bevezetés

AZ 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.

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

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:

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

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.

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

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 termékcsoportot inferior (luxus) jószágnak tekinthetjük közgazdasági értelemben.

Modellbecslés és hipotézis vizsgálat

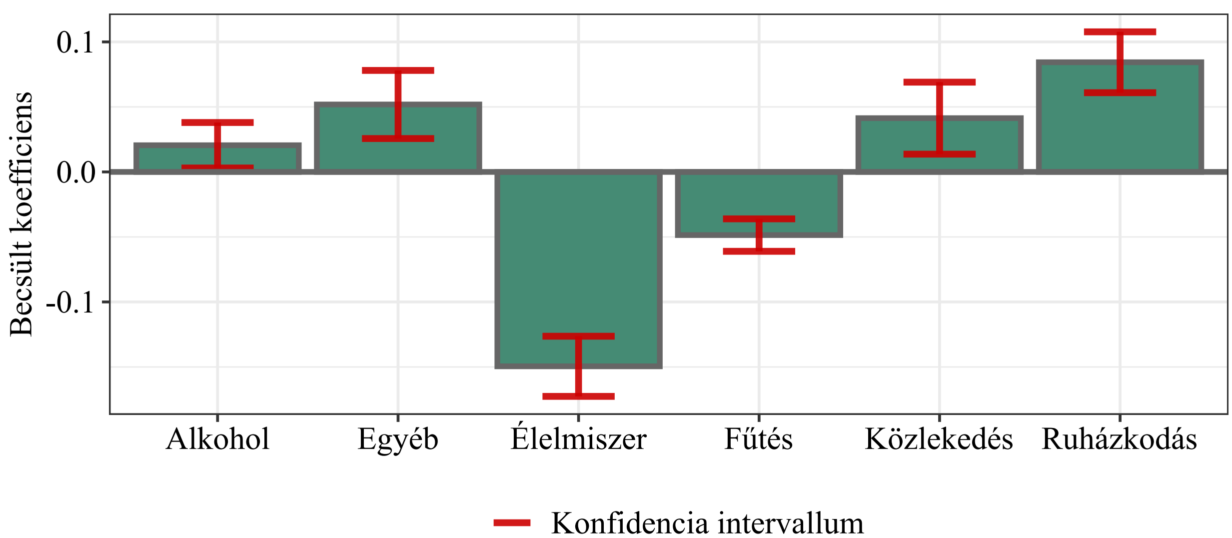
Becsült előjelek értelmezése

Az 1. egyenlet által leírt lineáris regresszió modellekben szereplő γ_j együtthatókat ismerteti az 1. ábra. Az ábráról leolvasható, hogy inferior jószágnak minősül az élelmiszer és a fűtés, míg az alkohol, a közlekedés, a ruházkodás és az egyéb termékcsoportba tartozó javak luxusjószágnak. Ez **az eredmény összhangban van azzal, amit empirikus vizsgálat nélkül mondanánk, ugyanis élelmiszerre és fűtésre mindenféle képpen szükséges egy bizonyos összeget kiadni.** Az alkohol fogyasztás és a ruházkodás szűkösség esetén visszaszorítható, a közlekedésnek is könnyen található olcsóbb alternatívája.

[1] Feladatkiírás által előírt megkötés.

Szignifikancia értelmezése

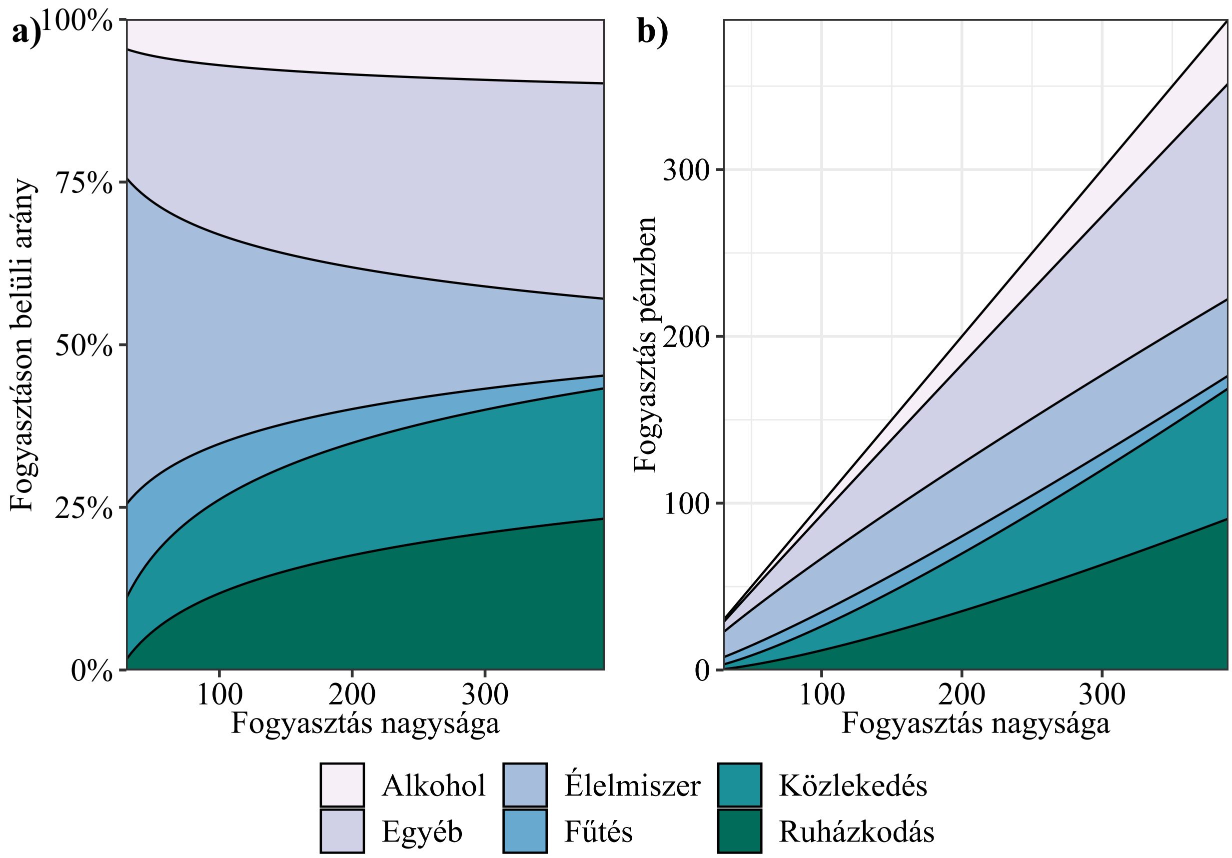
Statiszikai értemelben valamennyi meredekségi együttható szignifikánsnak bizonyult (belátható ez abból is, hogy a paraméterbecslésekhez tartozó 95%-os konfidencia intervallumok egyikse sem tartalmazza a zérust), a empirikus szignifikanciaszint mindössze némely tengelymetszet esetében nem különbözik szignifikánsan 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, míg 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.



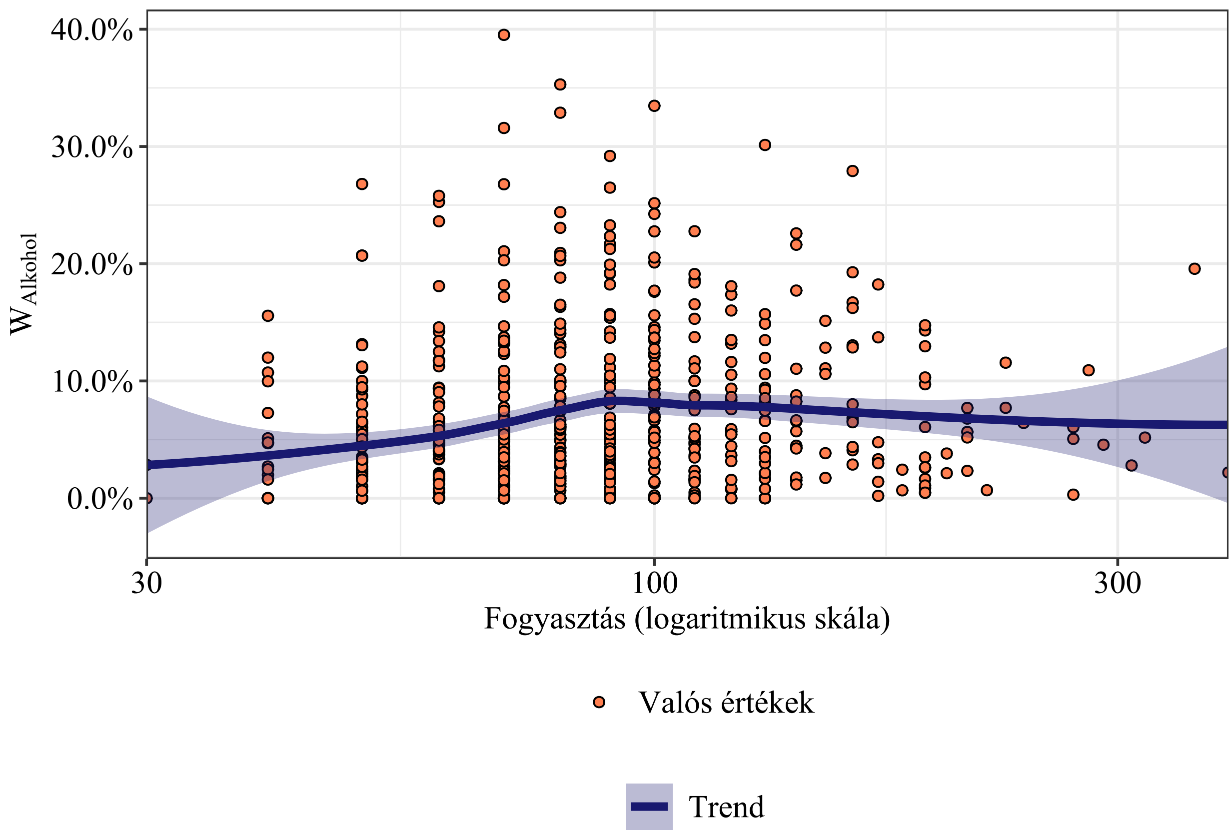
1. ábra. Termékcsoportonként becsült γ koefficiens

Együtthatók összegére tett előzetes feltevés ellenőrzése

A paraméterbecslések között fennálló algebrai kapcsolatot az alábbi módon elleőrzöm: (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 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 $99,99\%$).

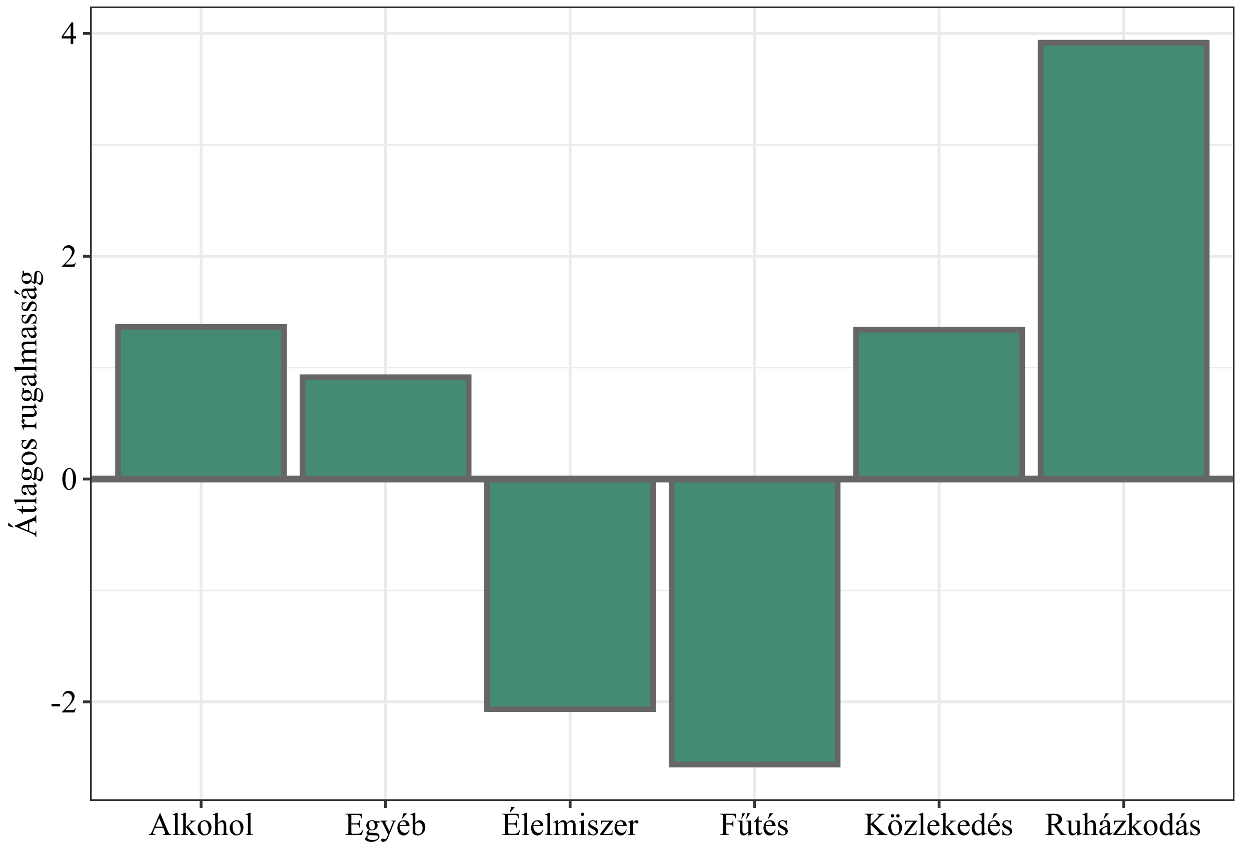


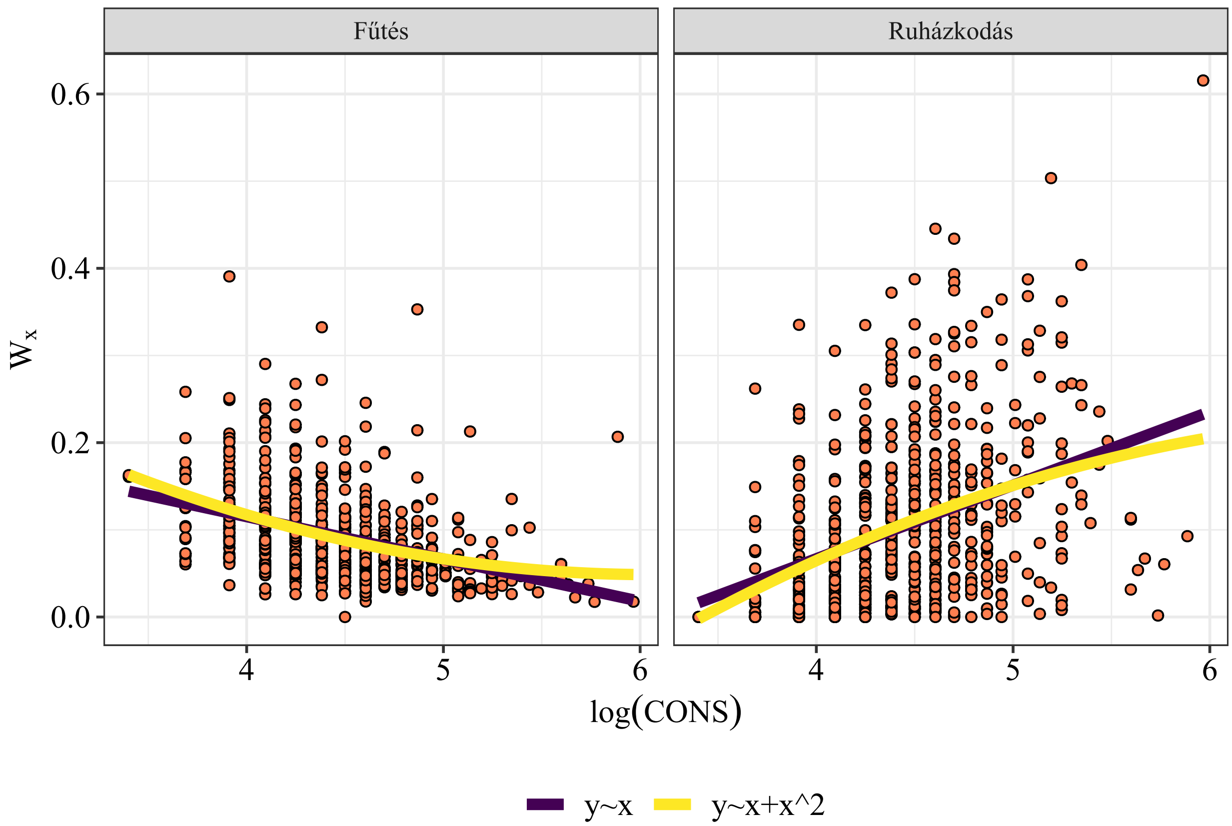
2. ábra. A becsült paraméterek alapján az egyes termékcsoportok fogyasztása a teljes fogyasztás függvényében



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

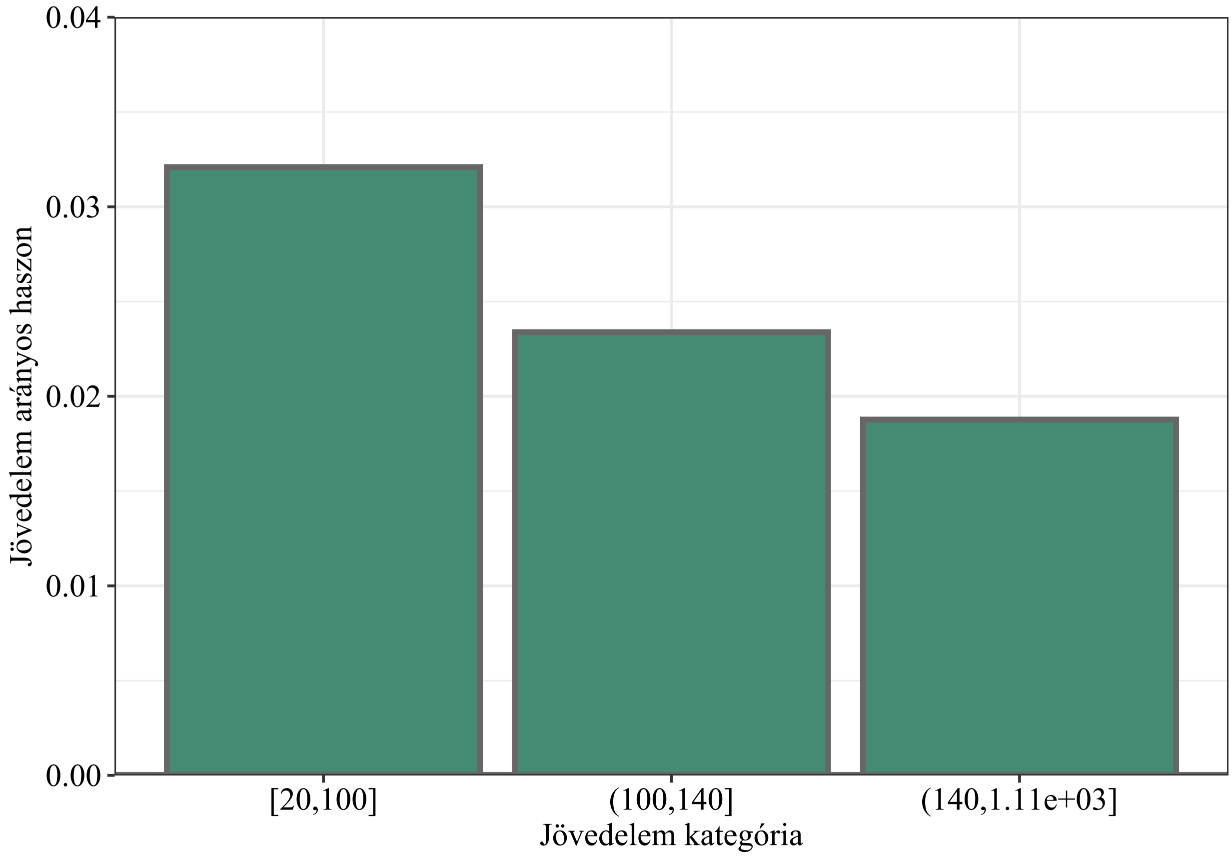
Átlagos kiadási rugalmasságok





Változó	Koefficiens	Standard hiba	T-statisztika	P-érték
konstans	-0,60	0,31	-1,93	5,42%
I(log(CONS))	0,23	0,14	1,68	9,31%
I(log(CONS)^2)	-0,02	0,01	-1,06	28,82%

Változó	Koefficiens	Standard hiba	T-statisztika	P-érték
konstans	0,65	0,16	3,93	0,01%
I(log(CONS))	-0,20	0,07	-2,73	0,65%
I(log(CONS)^2)	0,02	0,01	2,06	3,96%


0.04
0.03
0.02
0.01
0.00
Jövedelem arányos haszon
[20,100]
(100,140]
(140,1.11e+03]
Jövedelem kategória

R kódok

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

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'
  )
}
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'))

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)
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)
  }

f.ell <- function(y, x) {
```

```
  data.frame(y, x) %>%
    lm(formula = y ~ x) %>%
    {beta <- .$coefficients %>% .['x']
    broom::augment(.) %>% select(x, .fitted) %>%
      mutate(ell = x*beta/.fitted)
    } %>%
    pull(ell) %>%
    mean
}

dat %>%
  {select(., which(str_detect(names(.), 'W')))} %>%
  apply(2, function(y) f.ell(y = y, x = log(dat$CONS))) %>%
  {data.frame(x = f.products(names(.)), ell = .)} %>%
  ggplot(aes(x, ell)) + geom_col(fill = 'aquamarine4', color = 'grey40', size = 1) +
  geom_hline(yintercept = 0, color = 'grey40', size = 1.2) +
  labs(x = NULL, y = 'Átlagos rugalmasság')

dat %>%
  transmute(CONS = log(CONS), WCLOTH, WFUEL) %>%
  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) +
  labs(y = expression(W[x]), x = expression(log(CONS)), color = NULL) +
  scale_color_viridis_d()

dat %>%
  lm(formula = WCLOTH ~ I(log(CONS)) + I(log(CONS)^2)) %>%
  broom::tidy() %>%
  prtbl()

dat %>%
  lm(formula = WFUEL ~ I(log(CONS)) + I(log(CONS)^2)) %>%
  broom::tidy() %>%
  prtbl()

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))
```