Supplement 3. Analysis code for Pätsi et al.

Arho Toikka

2023-03-14

```
# Supplement 3: Analysis code for Pätsi et al.
library(tidyverse)
library(patchwork)
library(broom)
library(broom.mixed)
library(knitr)
library(kableExtra)
library(janitor)
library(glmmTMB)

library(readr)
tupakkadata <- read_delim("tupakan_myyntipaikat_postinumeroittain_valmis_anayysiaineisto.csv",
                          delim = ";", escape_double = FALSE,
                          locale = locale(decimal_mark = ",",
                                  grouping_mark = ".",
                                  encoding = "WINDOWS-1252"),
                          trim_ws = TRUE)

tupakkadata <- tupakkadata %>% filter(asukkaat_yhteensä_2020_he > 500)

tupakkadata <- tupakkadata |>
  mutate(myyntipaikat_log_prob = log(tupakan_myyntipaikkoja_per_1000_asukasta+1),
         myyntipaikat_log = ifelse(tupakan_myyntipaikkoja_per_1000_asukasta > 0,
                    log(tupakan_myyntipaikkoja_per_1000_asukasta), NA),
         myyntipaikat_no_0 = tupakan_myyntipaikkoja_per_1000_asukasta %>% na_if(0),
         myyntipaikka_ylipaansa = ifelse(tupakan_myyntipaikkoja_per_1000_asukasta > 0, 1, 0) %>%
                  as_factor() %>% fct_recode("No sales present"="0", "At least 1 sales location"="1"),
         mediaanitulot_1000e = asukkaiden_mediaanitulot_2020_hr/1000,
         alin_tuloluokka_pros = alin_tuloluokka_osuus*100,
         tyottymat_pros = tyottomienosuus*100,
         perusaste_pros = perusasteen_osuus*100,
         korkeakoulutettujen_osuus_pros = korkeakoulutettujen_osuus*100,
         yksinhuoltajia_lapsiperheista_pros = yhden_vanhemman_talous_per_lapsitaloudet*100,
         lapsitalous_per_kaikki_taloudet_pros = lapsitalous_per_kaikki_taloudet*100,
         asukastiheys = asukkaat_yhteensä_2020_he/1000 / (pinta_ala/1000000))

tee_jatkuvan_muuttujan_esittely <- function(aineisto, ...){
  taulukko <- aineisto %>%
    summarise(across(c(!!!quos(...)), list(
      mean = ~mean(.x, na.rm=T) %>% round_half_up(2),
      sd = ~sd(.x, na.rm=T) %>% round_half_up(2),
      min = ~min(.x, na.rm=T) %>% round_half_up(2),
```


Pätsi S-M, *et al. Tob Control* 2023;0:1–7. doi: 10.1136/tc-2022-057798

```
      max = ~max(.x, na.rm=T) %>% round_half_up(2)), .names = "{.col}.{.fn}")) %>%
    tidyr::pivot_longer(cols = contains("."), names_sep = "\\.", names_to  = c("Variable", ".value"))
  return(taulukko)
}

descriptives_table <- tupakkadata %>%
  tee_jatkuvan_muuttujan_esittely(tupakan_myyntipaikkoja_per_1000_asukasta,
                                  mediaanitulot_1000e,
                                  alin_tuloluokka_pros,
                                  tyottymat_pros,
                                  korkeakoulutettujen_osuus_pros,
                                  asukastiheys) %>% mutate(myyntipaikka_ylipaansa = "Total") %>%
  select(myyntipaikka_ylipaansa, everything())

names_desc <- c("Tobacco retailer density (per 1,000 people)",
                "Median income, 1000 euro", "% in lowest income category",
                "% unemployed", "% with higher education",
                "Population per square km, 1000s")

descriptives_table$Variable <- names_desc

grouped_descriptives <- tupakkadata %>% group_by(myyntipaikka_ylipaansa) %>%
  tee_jatkuvan_muuttujan_esittely(tupakan_myyntipaikkoja_per_1000_asukasta,
                                  mediaanitulot_1000e,
                                  alin_tuloluokka_pros,
                                  tyottymat_pros, korkeakoulutettujen_osuus_pros,
                                  asukastiheys)

grouped_descriptives$Variable <- c(names_desc, names_desc)

all_descriptives <- grouped_descriptives %>% bind_rows(descriptives_table)

all_descriptives %>% select(-myyntipaikka_ylipaansa) %>%
  kable(digits=1, caption="Descriptive statistics") %>%
  column_spec (c(1),border_left = T) %>%
  column_spec (c(5),border_right = T) %>%
  pack_rows(index=c("Postcodes with 0 sales locations"= 6,
                    "Postcodes with at least 1 sales location" = 6, "All postcodes"=6))
```

```
p1 <- tupakkadata |>
  ggplot(aes(y=tupakan_myyntipaikkoja_per_1000_asukasta,
             x=alin_tuloluokka_pros, shape=as_factor(myyntipaikka_ylipaansa))) +
  geom_smooth(data=subset(tupakkadata, tupakan_myyntipaikkoja_per_1000_asukasta>0),
              method="lm", color="black") +
  geom_point(alpha=0.15, size=1) + theme_bw() +
  labs(y="", x="% in lowest \n income category", shape="Presence of sales location") +ylim(-0.1, 10)
p2 <- tupakkadata |>
  ggplot(aes(y=tupakan_myyntipaikkoja_per_1000_asukasta,
             x=mediaanitulot_1000e, shape=as_factor(myyntipaikka_ylipaansa))) +
  geom_smooth(data=subset(tupakkadata, tupakan_myyntipaikkoja_per_1000_asukasta>0),
              method="lm", color="black") +
  geom_point(alpha=0.15, size=1) + theme_bw() +
  labs(y="", x="Median income, 1000€", shape="Presence of sales location") +ylim(-0.1, 10)
```


Pätsi S-M, *et al. Tob Control* 2023;0:1–7. doi: 10.1136/tc-2022-057798

Table 1: Descriptive statistics

Variable	mean	sd	min	max
Postcodes with 0 sales locations				
Tobacco retailer density (per 1,000 people)	0.0	0.0	0.0	0.0
Median income, 1000 euro	24.2	2.9	16.9	40.2
% in lowest income category	18.5	4.3	10.7	36.1
% unemployed	7.8	3.5	0.5	24.3
% with higher education	19.9	8.6	6.6	59.4
Population per square km, 1000s	0.1	0.4	0.0	3.2
Postcodes with at least 1 sales location				
Tobacco retailer density (per 1,000 people)	1.3	1.3	0.1	26.3
Median income, 1000 euro	22.6	3.3	11.4	36.4
% in lowest income category	19.9	4.9	8.2	61.0
% unemployed	10.1	4.2	2.1	35.9
% with higher education	21.9	10.9	4.6	60.2
Population per square km, 1000s	0.8	1.7	0.0	20.5
All postcodes				
Tobacco retailer density (per 1,000 people)	1.0	1.3	0.0	26.3
Median income, 1000 euro	23.0	3.3	11.4	40.2
% in lowest income category	19.6	4.8	8.2	61.0
% unemployed	9.6	4.1	0.5	35.9
% with higher education	21.4	10.4	4.6	60.2
Population per square km, 1000s	0.7	1.5	0.0	20.5

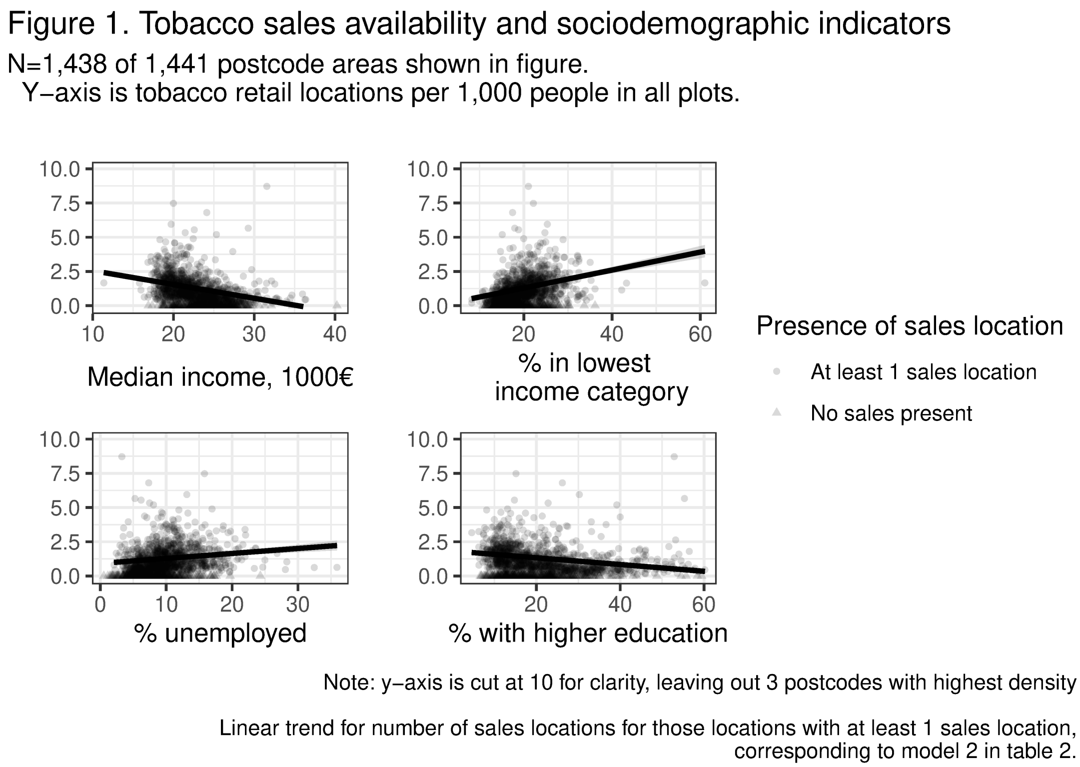
```
p3 <- tupakkadata |>
  ggplot(aes(y=tupakan_myyntipaikkoja_per_1000_asukasta,
             x=tyottymat_pros, shape=as_factor(myyntipaikka_ylipaansa))) +
  geom_smooth(data=subset(tupakkadata, tupakan_myyntipaikkoja_per_1000_asukasta>0),
              method="lm", color="black") +
  geom_point(alpha=0.15, size=1) + theme_bw() +
  labs(y="", x="% unemployed", shape="Presence of sales location") +ylim(-0.1, 10)
p4 <- tupakkadata |>
  ggplot(aes(y=tupakan_myyntipaikkoja_per_1000_asukasta,
             x=korkeakoulutettujen_osuus_pros, shape=as_factor(myyntipaikka_ylipaansa))) +
  geom_smooth(data=subset(tupakkadata, tupakan_myyntipaikkoja_per_1000_asukasta>0),
              method="lm", color="black") +
  geom_point(alpha=0.15, size=1) +
  theme_bw() +
  labs(y="", x="% with higher education", shape="Presence of sales location") +ylim(-0.1, 10)

fig1 <- (p2 + p1 + p3 + p4) & theme()

fig1 <- fig1 + plot_layout(guides = "collect")  +
  plot_annotation(
  title = 'Figure 1. Tobacco sales availability and sociodemographic indicators',
  subtitle = 'N=1,438 of 1,441 postcode areas shown in figure.
  Y-axis is tobacco retail locations per 1,000 people in all plots. \n',
  caption = 'Note: y-axis is cut at 10 for clarity, leaving out 3 postcodes with highest density\n
  Linear trend for number of sales locations for those locations with at least 1 sales location,\n correspondin
)
```


Pätsi S-M, *et al. Tob Control* 2023;0:1–7. doi: 10.1136/tc-2022-057798

```
fig1
```




```
# ggsave("fig1.png", plot=fig1)

# ggsave("fig1.tiff", device='tiff', dpi=700)
```

```
# Univariate models for the main paper


m1_lognormal <- tupakkadata %>% lm(data=., myyntipaikat_log~tyottymat_pros)
m2_lognormal <- tupakkadata %>% lm(data=., myyntipaikat_log~alin_tuloluokka_pros)
m3_lognormal <- tupakkadata %>% lm(data=., myyntipaikat_log~korkeakoulutettujen_osuus_pros)
m4_lognormal <- tupakkadata %>% lm(data=., myyntipaikat_log~mediaanitulot_1000e)

table_univariate_density <- tidy(m4_lognormal) %>%
  bind_rows(m2_lognormal %>% tidy) %>%
  bind_rows(m1_lognormal %>% tidy)%>%
  bind_rows(m3_lognormal %>% tidy)

m1_logistic <- tupakkadata %>%
  glm(data=., myyntipaikka_ylipaansa~tyottymat_pros,
      family = binomial(link="logit"))
m2_logistic <- tupakkadata %>%
  glm(data=., myyntipaikka_ylipaansa~alin_tuloluokka_pros,
```


Pätsi S-M, *et al. Tob Control* 2023;0:1–7. doi: 10.1136/tc-2022-057798

```
      family = binomial(link="logit"))
m3_logistic <- tupakkadata %>%
  glm(data=., myyntipaikka_ylipaansa~korkeakoulutettujen_osuus_pros,
      family = binomial(link="logit"))
m4_logistic <- tupakkadata %>%
  glm(data=., myyntipaikka_ylipaansa~mediaanitulot_1000e,
      family =  binomial(link="logit"))

table_univariate_logs <- tidy(m4_logistic) %>%
  bind_rows(m2_logistic %>% tidy) %>%
  bind_rows(m1_logistic %>% tidy)%>%
  bind_rows(m3_logistic %>% tidy)



univariate_tables_together <- table_univariate_logs %>%
  filter(term != "(Intercept)") %>% left_join(table_univariate_density, by=c("term")) %>%
  mutate(estimate.x = exp(estimate.x),
         estimate.y = exp(estimate.y),
         p.value.x = scales::pvalue(p.value.x),
         p.value.y = scales::pvalue(p.value.y)) %>%
  select(-statistic.x, -statistic.y)

univariate_tables_together$term <- c("Median income, 1000 euro",
                                     "% in lowest income category",
                                     "% unemployed",
                                     "% with higher education")
## Multivariate model for the main paper

main_log_normal_model <- tupakkadata %>%
  lm(data=., myyntipaikat_log~mediaanitulot_1000e+
       alin_tuloluokka_pros+tyottymat_pros+
       korkeakoulutettujen_osuus_pros+asukastiheys)

main_logistic_model <- tupakkadata %>%
  glm(data=., myyntipaikka_ylipaansa~mediaanitulot_1000e+
        alin_tuloluokka_pros+tyottymat_pros+
        korkeakoulutettujen_osuus_pros+asukastiheys, family = binomial)


multivariate_tables_together <- tidy(main_logistic_model) %>%
  filter(term != "(Intercept)") %>%
  left_join(tidy(main_log_normal_model), by=c("term")) %>%
  mutate(estimate.x = exp(estimate.x),
         estimate.y = exp(estimate.y), p.value.x = scales::pvalue(p.value.x),
         p.value.y = scales::pvalue(p.value.y)) %>% select(-statistic.x, -statistic.y)

multivariate_tables_together$term <- c("Median income, 1000 euro",
                                       "% in lowest income category",
                                       "% unemployed",
                                       "% with higher education",
                                       "Population density, 1000s per square km")
```

```
options(knitr.kable.NA = '')
all_tables_together <- univariate_tables_together %>%
  full_join(multivariate_tables_together, by=c("term"))

all_tables_together$term <- c("Median income, 1000 euro",
                              "% in lowest income category",
                              "% unemployed",
                              "% with higher education",
                              "Population density, 1000s per square km")
```

Split table to make space for Confidence Intervals & Effect sizes for the log-linear model

```
table_univariate_density <- tidy(m4_lognormal, conf.int = T) %>%
  bind_rows(m2_lognormal %>% tidy(conf.int = T)) %>%
  bind_rows(m1_lognormal %>% tidy(conf.int = T))%>%
  bind_rows(m3_lognormal %>% tidy(conf.int = T))

library(sjstats)
m1_effect_size <- anova_stats(m1_lognormal) %>% select(term, cohens.f)
m2_effect_size <- anova_stats(m2_lognormal) %>% select(term, cohens.f)
m3_effect_size <- anova_stats(m3_lognormal) %>% select(term, cohens.f)
m4_effect_size <- anova_stats(m4_lognormal) %>% select(term, cohens.f)

table_univariate_density <- table_univariate_density %>%
  left_join(m1_effect_size) %>% left_join(m2_effect_size, by="term") %>%
  left_join(m3_effect_size, by="term") %>% left_join(m4_effect_size, by="term") %>%
  mutate(cohens.f = coalesce(cohens.f.x, cohens.f.y, cohens.f.x.x, cohens.f.y.y), .keep="unused")


table_univariate_logs <- tidy(m4_logistic, conf.int = T) %>%
  bind_rows(m2_logistic %>% tidy(conf.int = T)) %>%
  bind_rows(m1_logistic %>% tidy(conf.int = T)) %>%
  bind_rows(m3_logistic %>% tidy(conf.int = T))



univariate_tables_together <- table_univariate_logs %>%
  filter(term != "(Intercept)") %>% left_join(table_univariate_density, by=c("term")) %>%
  mutate(estimate.x = exp(estimate.x),
         estimate.y = exp(estimate.y),
         p.value.x = scales::pvalue(p.value.x),
         p.value.y = scales::pvalue(p.value.y),
         conf.x = paste(round_half_up(exp(conf.low.x), 3), "to",
                        round_half_up(exp(conf.high.x), 3)),
          conf.y = paste(round_half_up(exp(conf.low.y), 3), "to",
                         round_half_up(exp(conf.high.y), 3))) %>%
  select(-statistic.x, -statistic.y, -std.error.x, -std.error.y) %>%
  select(term, estimate.x, conf.x, p.value.x, estimate.y, conf.y, p.value.y, cohens.f)

univariate_tables_together$term <- c("Median income, 1000 euro",
```


Pätsi S-M, *et al. Tob Control* 2023;0:1–7. doi: 10.1136/tc-2022-057798

```
                                     "% in lowest income category",
                                     "% unemployed",
                                     "% with higher education")
```

```
main_log_normal_model <- tupakkadata %>%
  lm(data=., myyntipaikat_log~mediaanitulot_1000e+
       alin_tuloluokka_pros+tyottymat_pros+
       korkeakoulutettujen_osuus_pros+asukastiheys)

main_model_effect_size <- anova_stats(main_log_normal_model) %>%
  select(term, cohens.f)

log_normal_table <- tidy(main_log_normal_model, conf.int=T, exponentiate=T) %>%
  left_join(main_model_effect_size)

main_logistic_model <- tupakkadata %>%
  glm(data=., myyntipaikka_ylipaansa~mediaanitulot_1000e+
        alin_tuloluokka_pros+tyottymat_pros+
        korkeakoulutettujen_osuus_pros+asukastiheys, family = binomial)


multivariate_tables_together <- tidy(main_logistic_model, conf.int = T, exponentiate = T) %>%
  filter(term != "(Intercept)") %>%
  left_join(log_normal_table, by=c("term")) %>%
  mutate(p.value.x = scales::pvalue(p.value.x),
         p.value.y = scales::pvalue(p.value.y),
         conf.x = paste(round_half_up(conf.low.x, 3), "to", round_half_up(conf.high.x, 3)),
          conf.y = paste(round_half_up(conf.low.y, 3), "to", round_half_up(conf.high.y, 3))) %>%
  select(-statistic.x, -statistic.y) %>%
  select(term, estimate.x, conf.x, p.value.x, estimate.y, conf.y, p.value.y, cohens.f)

multivariate_tables_together$term <- c("Median income, 1000 euro",
                                       "% in lowest income category",
                                       "% unemployed",
                                       "% with higher education",
                                       "Population density, 1000s per square km")
```

Table 2 for the main paper

```
all_tables_with_CIs <- univariate_tables_together %>%
  full_join(multivariate_tables_together, by=c("term"))

all_tables_with_CIs  %>%
  kable(digits=3,
        caption="Univariable and multivariable associations for
        presence of retailer and retailer density by SES",
        col.names = c("Variable", "Estimate", "95% CI", "P value",
                      "Estimate", "95% CI", "P value", "Cohens *f*",
                      "Estimate","95% CI", "P value",
                      "Estimate","95% CI", "P value", "Cohens *f*")) %>%
  column_spec (c(1),border_left = T)  %>%
```


Pätsi S-M, *et al. Tob Control* 2023;0:1–7. doi: 10.1136/tc-2022-057798

```
column_spec (c(15),border_right = T)  %>%
add_header_above(c(
  " " = 1, "Univariate logistic regression for presence of retailer" = 3,
  "Univariate linear regression for log of retailer density" = 4,
  "Multivariate logistic regression for presence of retailer" = 3,
  "Multivariate linear regression for log of retailer density"=4)) %>%
add_footnote("Estimates are exponentiated and can be
             interpreted as multiplicative increases in probability and density,
             given 1 unit change in predictor", notation="none") %>%
kableExtra::landscape() %>%
        kable_styling(latex_options = "scale_down")
```

Table 2: Univariable and multivariable associations for presence of retailer and retailer density by SES

	Univariate logistic regression for presence of retailer			Univariate linear regression for log of retailer density				Multivariate logistic regression for presence of retailer			Multivariate linear regression for log of retailer density			
Variable	Estimate	95% CI	P value	Estimate	95% CI	P value	Cohens *f*	Estimate	95% CI	P value	Estimate	95% CI	P value	Cohens *f*
Median income, 1000 euro	0.866	0.834 to 0.9	<0.001	0.907	0.896 to 0.918	<0.001	0.496	0.542	0.481 to 0.608	<0.001	0.956	0.929 to 0.984	0.002	0.506
% in lowest income category	1.071	1.041 to 1.103	<0.001	1.059	1.051 to 1.068	<0.001	0.422	0.884	0.836 to 0.933	<0.001	1.028	1.014 to 1.042	<0.001	0.068
% unemployed	1.186	1.142 to 1.234	<0.001	1.037	1.027 to 1.047	<0.001	0.214	1.063	1.01 to 1.12	0.020	0.976	0.964 to 0.987	<0.001	0.117
% with higher education	1.020	1.007 to 1.033	0.002	0.974	0.971 to 0.978	<0.001	0.433	1.139	1.102 to 1.178	<0.001	0.982	0.976 to 0.989	<0.001	0.144
Population density, 1000s per square km								2.903	1.843 to 4.909	<0.001	1.033	1.004 to 1.063	0.023	0.068

Estimates are exponentiated and can be interpreted as multiplicative increases in probability and density, given 1 unit change in predictor