RegressionSummary

David

2024-03-21

Regression functions

```
Regression1 <-  function(df) {

  regr1 <- felm(filled ~ trans_cost + mediancost_insublist +
                  mediancost_outsidesublist | req_id | 0 | req_id, data = df)
  summary_regr1 <- summary(regr1, cluster = c("req_id"))
  return(summary_regr1)
}
#regression2: filled ~ trans_cost mincost_insublist mincost_outsidesublist
Regression2 <- function(df){

  regr2 <- felm(filled ~ trans_cost + mincost_insublist +
                  mincost_outsidesublist | req_id | 0 | req_id, data = df)
  summary_regr2 <- summary(regr2, cluster = c("req_id"))

  return(summary_regr2)
}
winsorize_trans_cost <- function(df){
    df$trans_cost <- Winsorize(df$trans_cost, minval = quantile(df$trans_cost, 0.005, na.rm = TRUE), ma
    return(df)
    }
```

Winsorizing trans_cost data

```
subset_allothersublist <- winsorize_trans_cost(subset_largestsublist)
subset_largestsublist <- winsorize_trans_cost(subset_allothersublist)
```

Detailed summary of trans_cost

```
sum_stats_sub <- subset%>%
  summarise(
    Mean = mean(    trans_cost,       na.rm = TRUE) %>% round(3),
    SD   = sd(      trans_cost,       na.rm = TRUE) %>% round(3),
    p1   = quantile(trans_cost, 0.01, na.rm = TRUE) %>% round(3),
    p5   = quantile(trans_cost, 0.05, na.rm = TRUE) %>% round(3),
```

```
    p10  = quantile(trans_cost, 0.10, na.rm = TRUE) %>% round(3),
    p50  = quantile(trans_cost, 0.50, na.rm = TRUE) %>% round(3),
    p90  = quantile(trans_cost, 0.90, na.rm = TRUE) %>% round(3),
    p95  = quantile(trans_cost, 0.95, na.rm = TRUE) %>% round(3),
    p99  = quantile(trans_cost, 0.99, na.rm = TRUE) %>% round(3))
sum_stats_sub
```

```
## # A tibble: 1 x 9
##    Mean    SD     p1     p5    p10   p50   p90   p95   p99
##   <dbl> <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 0.009 0.038 -0.007 -0.002 -0.001 0.005 0.022 0.033 0.065
```

Regression1 on subset_largestsublist

```
summary1_largestsublist <- Regression1(subset_largestsublist)
print(summary1_largestsublist)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost + mediancost_insublist + mediancost_outsidesublist |      req_
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11144 -0.20513  0.01673  0.19670  1.63338
##
## Coefficients:
##                            Estimate Cluster s.e.  t value Pr(>|t|)
## trans_cost                -13.3411       0.1289 -103.537  < 2e-16 ***
## mediancost_insublist        1.2479       0.3583    3.483 0.000499 ***
## mediancost_outsidesublist  -0.3227       0.6443   -0.501 0.616460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3519 on 188840 degrees of freedom
##   (101568 observations deleted due to missingness)
## Multiple R-squared(full model): 0.5173   Adjusted R-squared: 0.4995
## Multiple R-squared(proj model): 0.1435   Adjusted R-squared: 0.1119
## F-statistic(full model, *iid*):29.03 on 6972 and 188840 DF, p-value: < 2.2e-16
## F-statistic(proj model):  3591 on 3 and 6969 DF, p-value: < 2.2e-16
```

Regression2 on subset_largestsublist

```
summary2_largestsublist <- Regression2(subset_largestsublist)
print(summary2_largestsublist)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost + mincost_insublist + mincost_outsidesublist |      req_id | 0
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11173 -0.20500  0.01675  0.19685  1.63159
##
## Coefficients:
##                           Estimate Cluster s.e.  t value Pr(>|t|)
## trans_cost              -13.348562     0.128899 -103.559   <2e-16 ***
## mincost_insublist       -0.017229     0.008992   -1.916   0.0554 .
## mincost_outsidesublist   0.023845     0.020692    1.152   0.2492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3519 on 188840 degrees of freedom
##   (101568 observations deleted due to missingness)
## Multiple R-squared(full model): 0.5172   Adjusted R-squared: 0.4994
## Multiple R-squared(proj model): 0.1434   Adjusted R-squared: 0.1118
## F-statistic(full model, *iid*):29.02 on 6972 and 188840 DF, p-value: < 2.2e-16
## F-statistic(proj model):  3578 on 3 and 6969 DF, p-value: < 2.2e-16
```

Regression1 on subset_allothersublist

```
summary1_allothersublist <- Regression1(subset_allothersublist)
print(summary1_allothersublist)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost + mediancost_insublist + mediancost_outsidesublist |      req_
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11144 -0.20513  0.01673  0.19669  1.63334
##
## Coefficients:
##                            Estimate Cluster s.e.  t value Pr(>|t|)
## trans_cost                 -13.3405       0.1289 -103.532  < 2e-16 ***
## mediancost_insublist         1.2480       0.3583    3.483 0.000498 ***
## mediancost_outsidesublist  -0.3227       0.6443   -0.501 0.616501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3519 on 188840 degrees of freedom
##   (101568 observations deleted due to missingness)
## Multiple R-squared(full model): 0.5173   Adjusted R-squared: 0.4995
## Multiple R-squared(proj model): 0.1435   Adjusted R-squared: 0.1119
## F-statistic(full model, *iid*):29.03 on 6972 and 188840 DF, p-value: < 2.2e-16
## F-statistic(proj model):  3591 on 3 and 6969 DF, p-value: < 2.2e-16
```

Regression2 on subset_allothersublist

```
summary2_allothersublist <- Regression2(subset_allothersublist)
print(summary2_allothersublist)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost + mincost_insublist + mincost_outsidesublist |      req_id | 0
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11173 -0.20501  0.01676  0.19684  1.63156
##
## Coefficients:
##                          Estimate Cluster s.e.  t value Pr(>|t|)
## trans_cost             -13.347934     0.128899 -103.554   <2e-16 ***
## mincost_insublist       -0.017228     0.008992   -1.916   0.0554 .
## mincost_outsidesublist   0.023845     0.020693    1.152   0.2492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3519 on 188840 degrees of freedom
##   (101568 observations deleted due to missingness)
## Multiple R-squared(full model): 0.5172   Adjusted R-squared: 0.4994
## Multiple R-squared(proj model): 0.1434   Adjusted R-squared: 0.1118
## F-statistic(full model, *iid*):29.02 on 6972 and 188840 DF, p-value: < 2.2e-16
## F-statistic(proj model):  3578 on 3 and 6969 DF, p-value: < 2.2e-16
```

Simple regression model of filled on trans_cost for non-dealer HY requests

without list fixed effects

```
model1_fe <- lm(filled ~ trans_cost, data = subset_largestsublist)
summary(model1_fe)
```

```
##
## Call:
## lm(formula = filled ~ trans_cost, data = subset_largestsublist)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7885 -0.5017  0.2931  0.4006  1.6344
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.643529   0.001053   611.3   <2e-16 ***
## trans_cost  -15.435108   0.068701  -224.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4605 on 282674 degrees of freedom
##   (14705 observations deleted due to missingness)
## Multiple R-squared:  0.1515, Adjusted R-squared:  0.1515
## F-statistic: 5.048e+04 on 1 and 282674 DF,  p-value: < 2.2e-16
```

Simple regression model of filled on trans_cost for non-dealer HY requests

with list fixed effects

This is run with felm, because lm with req_id fixed effect is computationally very intensive (vector memory is exceeded)

```
model2_nofe <- felm(filled ~ trans_cost | req_id, data = subset_largestsublist)
summary(model2_nofe)
```

```
##
## Call:
##    felm(formula = filled ~ trans_cost | req_id, data = subset_largestsublist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.10960 -0.21631  0.00433  0.20107  1.62454
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## trans_cost -13.22138    0.06245  -211.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3535 on 271211 degrees of freedom
##   (14705 observations deleted due to missingness)
## Multiple R-squared(full model): 0.5204   Adjusted R-squared: 0.5001
## Multiple R-squared(proj model): 0.1418   Adjusted R-squared: 0.1055
## F-statistic(full model):25.67 on 11464 and 271211 DF, p-value: < 2.2e-16
## F-statistic(proj model): 4.482e+04 on 1 and 271211 DF, p-value: < 2.2e-16
```