

Practical_324

Robert Atkinson

07/12/2020

load libraries

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2     v purrr    0.3.4
## v tibble   3.0.3     v dplyr    1.0.2
## v tidyr    1.1.1     v stringr  1.4.0
## v readr    1.3.1     vforcats  0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

library(magrittr)

##
## Attaching package: 'magrittr'
## The following object is masked from 'package:purrr':
##   set_names
## The following object is masked from 'package:tidyr':
##   extract
library(psych)

##
## Attaching package: 'psych'
## The following objects are masked from 'package:ggplot2':
##   %+%, alpha
library(stargazer)

##
## Please cite as:
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
library(zoo)
```

```

## 
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
## 
##     as.Date, as.Date.numeric

library(car)

## Loading required package: carData

## 
## Attaching package: 'car'

## The following object is masked from 'package:psych':
## 
##     logit

## The following object is masked from 'package:dplyr':
## 
##     recode

## The following object is masked from 'package:purrr':
## 
##     some

library(lmtest)
library(lm.beta)

```

load data

```

leicester_2011OAC <- readr::read_csv("2011_OAC_Raw_uVariables_Leicester.csv")

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   OA11CD = col_character(),
##   LSOA11CD = col_character(),
##   LS011ANM = col_character(),
##   MSOA11CD = col_character(),
##   MSOA11NM = col_character(),
##   LAD11CD = col_character(),
##   LAD11NM = col_character(),
##   supgrpname = col_character(),
##   grpcode = col_character(),
##   grpname = col_character(),
##   subgrpcode = col_character(),
##   subgrpname = col_character()
## )
## See spec(...) for full column specifications.

```

Exercise 324.1

Question 324.1.1:

Create a model having as outcome variable the presence of people using private transport for commuting to work, and using a stepwise “both” approach, having all the variables created for the example above and related to the presence of people working in different industry sectors (perc_u141 to perc_u158) as scope.

```
# Select and
# normalise variables
leicester_20110AC_transp_q1 <-
leicester_20110AC %>%
dplyr::select(
OA11CD,
Total_Pop_No_NI_Students_16_to_74, Total_Employment_16_to_74,
u121, u141:u158
) %>%
# percentage method of travel
dplyr::mutate(
u121 = (u121 / Total_Pop_No_NI_Students_16_to_74) * 100
) %>%
# percentage across industry sector columns
dplyr::mutate(
dplyr::across(
u141:u158,
function(x){ (x / Total_Employment_16_to_74) * 100 }
)
) %>%
# rename columns
dplyr::rename_with(
function(x){ paste0("perc_", x) },
c(u121, u141:u158)
)

library(MASS)

##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##     select
# Selected variables
# perc_u120: Method of Travel to Work, Private Transport
# perc_u142: Industry Sector, Mining and quarrying
# perc_u144: Industry Sector, Electricity, gas, steam and air conditioning ...
# perc_u146: Industry Sector, Construction
# perc_u149: Industry Sector, Accommodation and food service activities
# Create model
commuting_model_q1 <-
leicester_20110AC_transp_q1 %$%
stepAIC(
  object =
lm(perc_u121 ~ 1),
scope =
```

```

perc_u121 ~
  perc_u141 + perc_u142 + perc_u143 + perc_u144 + perc_u145 + perc_u146 + perc_u147 + perc_u148 + perc_u149 + perc_u150 + perc_u151 + perc_u152 + perc_u153 + perc_u154 + perc_u155 + perc_u156 + perc_u157 + perc_u158
)

## Start: AIC=4826.63
## perc_u121 ~ 1
##
##          Df Sum of Sq    RSS    AIC
## + perc_u149  1    40534 100288 4499.7
## + perc_u146  1    17908 122913 4696.8
## + perc_u155  1    11982 128839 4742.5
## + perc_u151  1     9796 131025 4758.8
## + perc_u154  1     7574 133247 4775.1
## + perc_u158  1     5304 135517 4791.4
## + perc_u150  1     3835 136987 4801.9
## + perc_u153  1     2981 137841 4807.9
## + perc_u157  1     2648 138173 4810.2
## + perc_u142  1     2247 138574 4813.0
## + perc_u144  1     2120 138701 4813.9
## + perc_u147  1     1912 138909 4815.4
## + perc_u152  1     792  140030 4823.2
## + perc_u148  1     527  140295 4825.0
## + perc_u143  1     296  140525 4826.6
## <none>           140821 4826.6
## + perc_u145  1     161  140660 4827.5
## + perc_u156  1     144  140677 4827.6
## + perc_u141  1     125  140696 4827.8
##
## Step: AIC=4499.71
## perc_u121 ~ perc_u149
##
##          Df Sum of Sq    RSS    AIC
## + perc_u146  1    10880  89408 4390.4
## + perc_u154  1     5429  94859 4447.8
## + perc_u143  1     4648  95639 4455.7
## + perc_u155  1     3918  96370 4463.1
## + perc_u151  1     3869  96418 4463.6
## + perc_u150  1     1893  98395 4483.2
## + perc_u153  1     1584  98704 4486.3
## + perc_u158  1     1431  98856 4487.8
## + perc_u142  1     1107  99181 4491.0
## + perc_u144  1     1029  99259 4491.7
## + perc_u147  1      754  99533 4494.4
## + perc_u148  1      690  99598 4495.0
## + perc_u157  1      396  99891 4497.9
## + perc_u152  1      312  99975 4498.7
## <none>           100288 4499.7
## + perc_u156  1      27  100260 4501.4
## + perc_u145  1      10  100278 4501.6
## + perc_u141  1       8  100280 4501.6
## - perc_u149  1    40534 140821 4826.6
##
## Step: AIC=4390.43

```

```

## perc_u121 ~ perc_u149 + perc_u146
##
##          Df Sum of Sq    RSS    AIC
## + perc_u154  1     7666 81742 4305.6
## + perc_u151  1      6822 82586 4315.5
## + perc_u155  1      5062 84345 4336.0
## + perc_u153  1      4520 84888 4342.2
## + perc_u150  1      4178 85230 4346.1
## + perc_u144  1      1889 87519 4371.7
## + perc_u143  1      1375 88033 4377.4
## + perc_u158  1      1167 88241 4379.7
## + perc_u148  1      1054 88354 4380.9
## + perc_u142  1       834 88574 4383.4
## + perc_u145  1       770 88638 4384.1
## + perc_u152  1       467 88941 4387.4
## + perc_u156  1       364 89044 4388.5
## + perc_u147  1       248 89160 4389.7
## <none>
##          89408 4390.4
## + perc_u141  1        8 89400 4392.4
## + perc_u157  1        3 89405 4392.4
## - perc_u146  1     10880 100288 4499.7
## - perc_u149  1     33506 122913 4696.8
##
## Step: AIC=4305.57
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154
##
##          Df Sum of Sq    RSS    AIC
## + perc_u151  1     4035.4 77707 4258.5
## + perc_u155  1     3181.7 78561 4269.1
## + perc_u150  1     2203.0 79539 4281.1
## + perc_u153  1     1720.7 80021 4287.0
## + perc_u144  1     1456.2 80286 4290.2
## + perc_u158  1     1335.0 80407 4291.6
## + perc_u142  1     775.0 80967 4298.3
## + perc_u143  1     695.8 81046 4299.3
## + perc_u152  1     280.8 81461 4304.2
## + perc_u147  1     251.0 81491 4304.6
## + perc_u145  1     215.5 81527 4305.0
## + perc_u148  1     194.2 81548 4305.3
## <none>
##          81742 4305.6
## + perc_u156  1      98.2 81644 4306.4
## + perc_u157  1      92.1 81650 4306.5
## + perc_u141  1      19.5 81723 4307.3
## - perc_u154  1     7665.7 89408 4390.4
## - perc_u146  1     13116.3 94859 4447.8
## - perc_u149  1     30563.9 112306 4611.4
##
## Step: AIC=4258.52
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151
##
##          Df Sum of Sq    RSS    AIC
## + perc_u155  1     2596.3 75110 4227.6
## + perc_u150  1     1260.7 76446 4244.7
## + perc_u153  1     1139.5 76567 4246.2

```

```

## + perc_u158 1 1099.7 76607 4246.7
## + perc_u144 1 1049.3 76657 4247.3
## + perc_u142 1 647.5 77059 4252.4
## + perc_u143 1 415.3 77291 4255.3
## + perc_u152 1 308.7 77398 4256.7
## <none> 77707 4258.5
## + perc_u147 1 139.4 77567 4258.8
## + perc_u145 1 68.3 77638 4259.7
## + perc_u156 1 42.8 77664 4260.0
## + perc_u141 1 14.7 77692 4260.3
## + perc_u148 1 6.4 77700 4260.4
## + perc_u157 1 0.5 77706 4260.5
## - perc_u151 1 4035.4 81742 4305.6
## - perc_u154 1 4879.5 82586 4315.5
## - perc_u146 1 15140.9 92848 4429.0
## - perc_u149 1 25112.2 102819 4527.9
##
## Step: AIC=4227.59
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155
##
##          Df Sum of Sq   RSS   AIC
## + perc_u158 1 1102.2 74008 4215.3
## + perc_u150 1 1086.4 74024 4215.5
## + perc_u144 1 927.2 74183 4217.5
## + perc_u142 1 889.0 74221 4218.0
## + perc_u153 1 629.4 74481 4221.4
## + perc_u156 1 262.7 74848 4226.2
## + perc_u152 1 244.8 74866 4226.4
## <none> 75110 4227.6
## + perc_u143 1 83.3 75027 4228.5
## + perc_u148 1 48.4 75062 4229.0
## + perc_u141 1 37.0 75073 4229.1
## + perc_u157 1 32.4 75078 4229.2
## + perc_u145 1 29.5 75081 4229.2
## + perc_u147 1 0.3 75110 4229.6
## - perc_u155 1 2596.3 77707 4258.5
## - perc_u151 1 3450.1 78561 4269.1
## - perc_u154 1 3742.4 78853 4272.7
## - perc_u146 1 15660.3 90771 4409.1
## - perc_u149 1 20566.2 95677 4460.1
##
## Step: AIC=4215.26
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##     perc_u158
##
##          Df Sum of Sq   RSS   AIC
## + perc_u150 1 1115.7 72893 4202.5
## + perc_u144 1 879.8 73128 4205.7
## + perc_u142 1 856.4 73152 4206.0
## + perc_u153 1 781.9 73226 4207.0
## + perc_u143 1 267.6 73741 4213.8
## + perc_u152 1 185.6 73823 4214.8
## + perc_u156 1 179.4 73829 4214.9
## <none> 74008 4215.3

```

```

## + perc_u141 1      38.5 73970 4216.8
## + perc_u145 1      24.3 73984 4216.9
## + perc_u147 1      21.2 73987 4217.0
## + perc_u157 1      14.3 73994 4217.1
## + perc_u148 1      7.1 74001 4217.2
## - perc_u158 1     1102.2 75110 4227.6
## - perc_u155 1     2598.8 76607 4246.7
## - perc_u151 1     3234.0 77242 4254.7
## - perc_u154 1     3895.0 77903 4263.0
## - perc_u146 1     15307.3 89316 4395.4
## - perc_u149 1     18653.1 92661 4431.1
##
## Step: AIC=4202.54
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##           perc_u158 + perc_u150
##
##          Df Sum of Sq   RSS   AIC
## + perc_u142 1    772.6 72120 4194.2
## + perc_u144 1    701.2 72191 4195.2
## + perc_u153 1    479.1 72414 4198.2
## + perc_u156 1    367.9 72525 4199.6
## + perc_u152 1    182.5 72710 4202.1
## <none>            72893 4202.5
## + perc_u143 1    89.4 72803 4203.4
## + perc_u148 1    81.3 72811 4203.5
## + perc_u141 1    46.2 72846 4203.9
## + perc_u157 1    29.6 72863 4204.1
## + perc_u145 1    7.3 72885 4204.4
## + perc_u147 1    0.1 72893 4204.5
## - perc_u150 1   1115.7 74008 4215.3
## - perc_u158 1   1131.5 74024 4215.5
## - perc_u155 1   2422.1 75315 4232.2
## - perc_u151 1   2467.5 75360 4232.8
## - perc_u154 1   3170.9 76064 4241.8
## - perc_u146 1   16161.4 89054 4394.6
## - perc_u149 1   18087.3 90980 4415.3
##
## Step: AIC=4194.22
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##           perc_u158 + perc_u150 + perc_u142
##
##          Df Sum of Sq   RSS   AIC
## + perc_u144 1    724.9 71395 4186.4
## + perc_u153 1    531.2 71589 4189.1
## + perc_u156 1    371.4 71749 4191.2
## + perc_u152 1    232.3 71888 4193.1
## <none>            72120 4194.2
## + perc_u143 1    81.1 72039 4195.1
## + perc_u148 1    61.8 72058 4195.4
## + perc_u141 1    46.3 72074 4195.6
## + perc_u157 1    36.1 72084 4195.7
## + perc_u145 1    6.2 72114 4196.1
## + perc_u147 1    0.1 72120 4196.2
## - perc_u142 1    772.6 72893 4202.5

```

```

## - perc_u150 1 1031.8 73152 4206.0
## - perc_u158 1 1099.0 73219 4206.9
## - perc_u151 1 2367.6 74488 4223.5
## - perc_u155 1 2644.2 74764 4227.1
## - perc_u154 1 3143.4 75263 4233.6
## - perc_u146 1 15766.4 87886 4383.8
## - perc_u149 1 17532.6 89653 4403.1
##
## Step: AIC=4186.43
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##      perc_u158 + perc_u150 + perc_u142 + perc_u144
##
##          Df Sum of Sq   RSS   AIC
## + perc_u153 1    558.3 70837 4180.8
## + perc_u156 1    290.0 71105 4184.5
## + perc_u152 1    248.9 71146 4185.0
## <none>           71395 4186.4
## + perc_u143 1    79.2 71316 4187.4
## + perc_u157 1    79.1 71316 4187.4
## + perc_u148 1    56.2 71339 4187.7
## + perc_u141 1    48.8 71346 4187.8
## + perc_u145 1    4.5 71391 4188.4
## + perc_u147 1    1.9 71393 4188.4
## - perc_u144 1    724.9 72120 4194.2
## - perc_u142 1    796.2 72191 4195.2
## - perc_u150 1    857.1 72252 4196.0
## - perc_u158 1    1052.8 72448 4198.6
## - perc_u151 1    2179.4 73575 4213.6
## - perc_u155 1    2553.1 73948 4218.5
## - perc_u154 1    3093.4 74489 4225.5
## - perc_u146 1    16154.6 87550 4382.1
## - perc_u149 1    17206.9 88602 4393.7
##
## Step: AIC=4180.82
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##      perc_u158 + perc_u150 + perc_u142 + perc_u144 + perc_u153
##
##          Df Sum of Sq   RSS   AIC
## + perc_u156 1    709.1 70128 4173.1
## + perc_u152 1    194.8 70642 4180.2
## + perc_u148 1    172.5 70664 4180.5
## <none>           70837 4180.8
## + perc_u157 1    83.9 70753 4181.7
## + perc_u141 1    56.9 70780 4182.0
## + perc_u147 1    52.1 70785 4182.1
## + perc_u145 1    1.9 70835 4182.8
## + perc_u143 1    1.0 70836 4182.8
## - perc_u153 1    558.3 71395 4186.4
## - perc_u150 1    578.0 71415 4186.7
## - perc_u144 1    752.0 71589 4189.1
## - perc_u142 1    851.0 71688 4190.4
## - perc_u158 1    1174.1 72011 4194.8
## - perc_u151 1    1986.1 72823 4205.6
## - perc_u155 1    2098.8 72936 4207.1

```

```

## - perc_u154 1 2355.3 73192 4210.5
## - perc_u146 1 16706.5 87543 4384.0
## - perc_u149 1 17008.4 87845 4387.3
##
## Step: AIC=4173.07
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##     perc_u158 + perc_u150 + perc_u142 + perc_u144 + perc_u153 +
##     perc_u156
##
##          Df Sum of Sq   RSS   AIC
## + perc_u152 1    209.5 69918 4172.2
## + perc_u143 1    201.6 69926 4172.3
## <none>           70128 4173.1
## + perc_u157 1    136.7 69991 4173.2
## + perc_u141 1     69.3 70058 4174.1
## + perc_u148 1     49.7 70078 4174.4
## + perc_u147 1     15.2 70113 4174.9
## + perc_u145 1      7.4 70120 4175.0
## - perc_u144 1    630.3 70758 4179.7
## - perc_u156 1    709.1 70837 4180.8
## - perc_u150 1    721.2 70849 4181.0
## - perc_u142 1    877.3 71005 4183.1
## - perc_u153 1    977.4 71105 4184.5
## - perc_u158 1   1064.7 71192 4185.7
## - perc_u151 1   1669.9 71798 4193.9
## - perc_u155 1   2393.3 72521 4203.6
## - perc_u154 1   2820.0 72948 4209.3
## - perc_u146 1   16552.5 86680 4376.4
## - perc_u149 1   16957.2 87085 4380.9
##
## Step: AIC=4172.17
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##     perc_u158 + perc_u150 + perc_u142 + perc_u144 + perc_u153 +
##     perc_u156 + perc_u152
##
##          Df Sum of Sq   RSS   AIC
## + perc_u143 1    159.4 69759 4172.0
## + perc_u157 1    154.0 69764 4172.0
## <none>           69918 4172.2
## - perc_u152 1    209.5 70128 4173.1
## + perc_u141 1     76.7 69842 4173.1
## + perc_u148 1     54.8 69863 4173.4
## + perc_u147 1     14.4 69904 4174.0
## + perc_u145 1      5.8 69912 4174.1
## - perc_u144 1    642.2 70560 4179.0
## - perc_u156 1    723.8 70642 4180.2
## - perc_u150 1    728.7 70647 4180.2
## - perc_u153 1    912.8 70831 4182.7
## - perc_u142 1    925.3 70844 4182.9
## - perc_u158 1    993.7 70912 4183.8
## - perc_u151 1   1693.2 71611 4193.4
## - perc_u155 1   2368.9 72287 4202.5
## - perc_u154 1   2774.6 72693 4207.9
## - perc_u146 1   16609.8 86528 4376.7

```

```

## - perc_u149 1 16796.4 86715 4378.8
##
## Step: AIC=4171.96
## perc_u121 ~ perc_u149 + perc_u146 + perc_u154 + perc_u151 + perc_u155 +
##      perc_u158 + perc_u150 + perc_u142 + perc_u144 + perc_u153 +
##      perc_u156 + perc_u152 + perc_u143
##
##          Df Sum of Sq   RSS   AIC
## <none>           69759 4172.0
## - perc_u143 1     159.4 69918 4172.2
## - perc_u152 1     167.4 69926 4172.3
## + perc_u147 1     87.2 69672 4172.7
## + perc_u141 1     63.7 69695 4173.1
## + perc_u157 1     36.4 69722 4173.5
## + perc_u148 1     29.2 69730 4173.6
## + perc_u145 1     20.2 69739 4173.7
## - perc_u144 1     604.6 70363 4178.3
## - perc_u150 1     641.7 70401 4178.8
## - perc_u153 1     735.5 70494 4180.1
## - perc_u156 1     882.9 70642 4182.1
## - perc_u142 1     898.1 70657 4182.4
## - perc_u158 1    1114.5 70873 4185.3
## - perc_u151 1    1587.8 71347 4191.8
## - perc_u155 1    2167.4 71926 4199.6
## - perc_u154 1    2922.2 72681 4209.7
## - perc_u146 1   11004.2 80763 4311.9
## - perc_u149 1   15080.7 84840 4359.6

# Print summary
stargazer(commuting_model_q1, type = "text")

```

```

##
## -----
##                               Dependent variable:
## -----
##                               perc_u121
## -----
## perc_u149                  -1.281***  

##                               (0.089)
## 
## perc_u146                  1.248***  

##                               (0.102)
## 
## perc_u154                  -0.720***  

##                               (0.114)
## 
## perc_u151                  0.773***  

##                               (0.166)
## 
## perc_u155                  0.678***  

##                               (0.125)
## 
## perc_u158                  -0.555***  

##                               (0.142)
## 
```

```

## perc_u150          0.584***  

##                           (0.197)  

##  

## perc_u142          3.867***  

##                           (1.103)  

##  

## perc_u144          0.669***  

##                           (0.232)  

##  

## perc_u153          0.442***  

##                           (0.139)  

##  

## perc_u156          -0.234***  

##                           (0.067)  

##  

## perc_u152           0.446  

##                           (0.295)  

##  

## perc_u143           -0.080  

##                           (0.054)  

##  

## Constant            36.422***  

##                           (2.752)  

##  

## -----  

## Observations         969  

## R2                  0.505  

## Adjusted R2          0.498  

## Residual Std. Error    8.547 (df = 955)  

## F Statistic          74.834*** (df = 13; 955)  

## -----  

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

commuting_model_q1 %>%  

rstandard() %>%  

stats::shapiro.test()  

##  

## Shapiro-Wilk normality test  

##  

## data: .  

## W = 0.99663, p-value = 0.03668

Overall, we can say that the delay model computed above is fit ( $F(13, 955) = 74.83$ ,  $p < .001$ ), indicating that the model might account for 49.79% of the people who use private transport to commute. However the model is not robust. The residuals are not homoscedastic (Breusch-Pagan test, BP = 27.606,  $p < .01$ ), and are not independent (Durbin-Watson test, DW = 1.8471,  $p < 0.01$ ), but they are normally distributed (Shapiro-Wilk test, W = 0.99618,  $p < 0.01$ ).

commuting_model_q1 %>%  

lmtest::bptest()  

##  

## studentized Breusch-Pagan test  

##  

## data: .

```

```

## BP = 40.649, df = 13, p-value = 0.0001086
commuting_model_q1 %>%
lmtest::dwtest()

##
## Durbin-Watson test
##
## data: .
## DW = 1.8919, p-value = 0.04344
## alternative hypothesis: true autocorrelation is greater than 0

```

Question 324.1.2:

Is the presence of people using public transportation to commute to work statistically, linearly related to mean age (u020)?

```

# Select and normalise variables
leicester_20110AC_transp_q2 <-
leicester_20110AC %>%
dplyr::select(
OA11CD,
Total_Pop_No_NI_Students_16_to_74, Total_Employment_16_to_74,
u020, u120
) %>%
# percentage method of travel
dplyr::mutate(
perc_u020 = (u020/ Total_Pop_No_NI_Students_16_to_74) * 100,
perc_u120 = (u120/ Total_Pop_No_NI_Students_16_to_74) * 100
)

```

Check assumptions for Pearson's

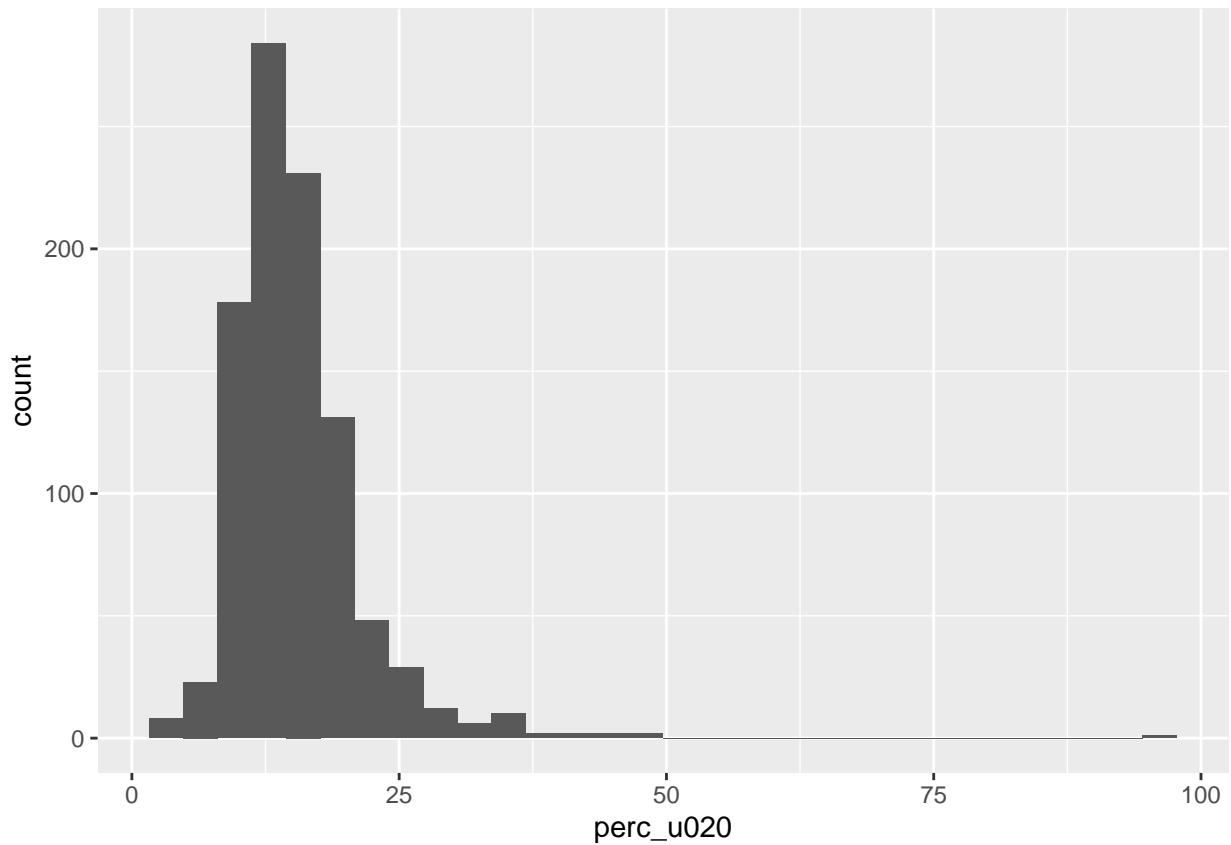
```

leicester_20110AC_transp_q2 %$% perc_u020 %>%
stats::shapiro.test()

##
## Shapiro-Wilk normality test
##
## data: .
## W = 0.80132, p-value < 2.2e-16

leicester_20110AC_transp_q2 %>%
ggplot2::ggplot(mapping = aes(x = perc_u020)) +
ggplot2::geom_histogram(bins = 30) +
ggplot2::theme_gray()

```

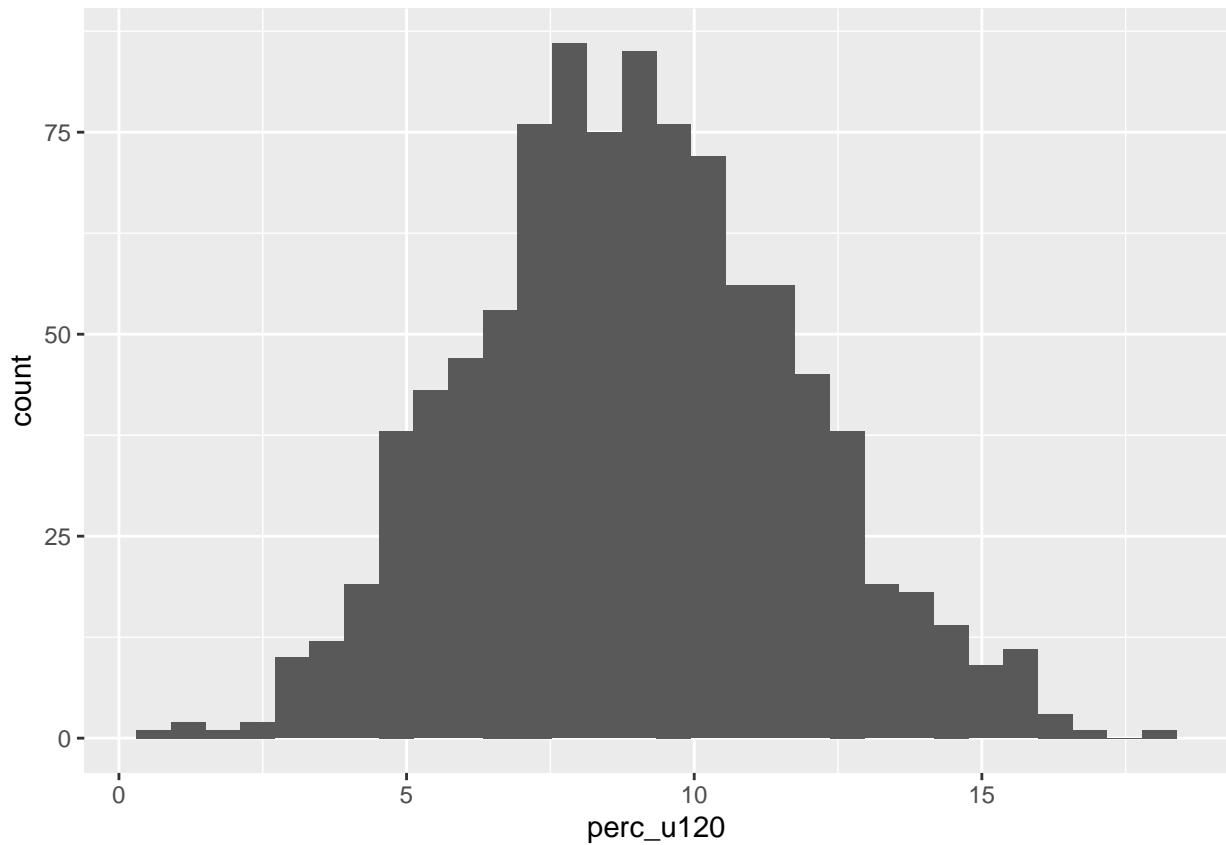


Not normally distributed.

```
leicester_20110AC_transp_q2 %$% perc_u120 %>%
  stats::shapiro.test()

## 
##  Shapiro-Wilk normality test
## 
## data: .
## W = 0.99743, p-value = 0.13

leicester_20110AC_transp_q2 %>%
  ggplot2::ggplot(mapping = aes(x = perc_u120)) +
  ggplot2::geom_histogram(bins = 30) +
  ggplot2::theme_gray()
```



Is normally distributed.

```
log_mean_age <- leicester_20110AC_transp_q2 %>%
  mutate(log10_u020 = log10(perc_u020))

log_public_commute <- leicester_20110AC_transp_q2 %>%
  mutate(log10_u120 = log10(perc_u120))

log_mean_age %$% log10_u020 %>%
  stats::shapiro.test()

## 
##  Shapiro-Wilk normality test
##
## data: .
## W = 0.97376, p-value = 3.208e-12
log_public_commute %$% perc_u120 %>%
  stats::shapiro.test()

## 
##  Shapiro-Wilk normality test
##
## data: .
## W = 0.99743, p-value = 0.13
```

Both variables are not normally distributed after logarithmic transformation.

Pearson's can't be used as both variables need to be normally distributed.

Check Spearman's assumptions

```
ties_check_mean_age <-
leicester_20110AC_transp_q2 %>%
dplyr::count(perc_u020) %>%
dplyr::filter(n > 1) %>%
dplyr::count(n()) %>%
dplyr::pull(n)
```

```
ties_check_mean_age
```

```
## integer(0)
ties_check_public_commute <-
leicester_20110AC_transp_q2 %>%
dplyr::count(perc_u120) %>%
dplyr::filter(n > 1) %>%
dplyr::count(n()) %>%
dplyr::pull(n)
```

```
ties_check_public_commute
```

```
## [1] 117
```

There are too many ties in the perc_u120 data. Spearman's cannot be used.

Kendall's Tau

```
kendall_mean_age_public_commute <-
leicester_20110AC_transp_q2 %$%
stats::cor.test(
perc_u020, perc_u120,
method = "kendall"
)
```

```
kendall_mean_age_public_commute
```

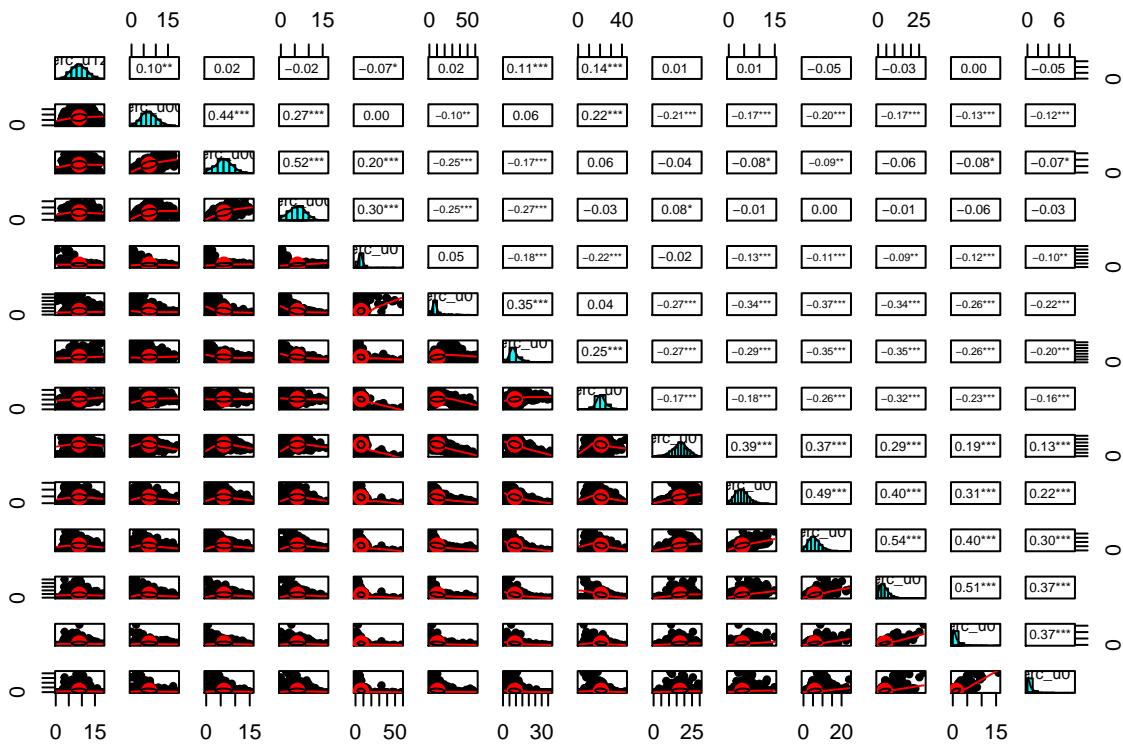
```
##
## Kendall's rank correlation tau
##
## data: perc_u020 and perc_u120
## z = 3.4962, p-value = 0.0004719
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.0750281
```

Mean age has a very small but significant linear relationship with people using public transportation to commute to work (Tau = 0.08, p < 0.001).

Question 324.1.3:

Is the presence of people using public transportation to commute to work statistically, linearly related to (a subset of) the age structure categories (u007 to u019)?

```
leicester_20110AC_transp_q3 <-  
leicester_20110AC %>%  
dplyr::select(  
  OA11CD,  
  Total_Population, Total_Pop_No_NI_Students_16_to_74, Total_Employment_16_to_74,  
  u007:u019, u120  
) %>%  
# percentage method of travel  
dplyr::mutate(  
  perc_u120 = (u120 / Total_Pop_No_NI_Students_16_to_74) * 100  
) %>%  
#percentage across industry sector columns  
dplyr::mutate(  
  dplyr::across(  
    u007:u019,  
    function(x){ (x / Total_Population) * 100 }  
)  
) %>%  
  
# rename columns  
dplyr::rename_with(  
  function(x){ paste0("perc_", x) },  
  c(u007:u019)  
)  
  
leicester_20110AC_transp_q3 %>%  
dplyr::select(perc_u120, perc_u007, perc_u008, perc_u009, perc_u010, perc_u011, perc_u012, perc_u013, p  
psych::pairs.panels(method = "kendal", stars = TRUE)
```



```

commuting_model_q3 <-
leicester_20110AC_transp_q3 %$%
stats::lm(
perc_u120 ~ perc_u013)

# Print summary
stargazer(commuting_model_q3, type = "text")

```

% Error: Unrecognized object type.

Age 25 - 29 (u012) has a small but significant linear relationship with people using public transportation to commute to work ($\text{Tau} = 0.11$, $p < 0.001$).

Age 30 - 44 (u013) also has a small but significant linear relationship with people using public transportation to commute to work ($\text{Tau} = 0.14$, $p < 0.001$).