

## mpg - wad

```
library(tidyverse)
library(rstatix)
library(ggpubr)

head(mpg)

## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv   cty   hwy fl   class
##   <chr>          <chr> <dbl> <int> <int> <chr>    <chr> <int> <int> <chr> <chr>
## 1 audi          a4      1.8  1999     4 auto(l5)  f      18    29 p   compa~
## 2 audi          a4      1.8  1999     4 manual(m5) f      21    29 p   compa~
## 3 audi          a4      2    2008     4 manual(m6) f      20    31 p   compa~
## 4 audi          a4      2    2008     4 auto(av)   f      21    30 p   compa~
## 5 audi          a4      2.8  1999     6 auto(l5)  f      16    26 p   compa~
## 6 audi          a4      2.8  1999     6 manual(m5) f      18    26 p   compa~

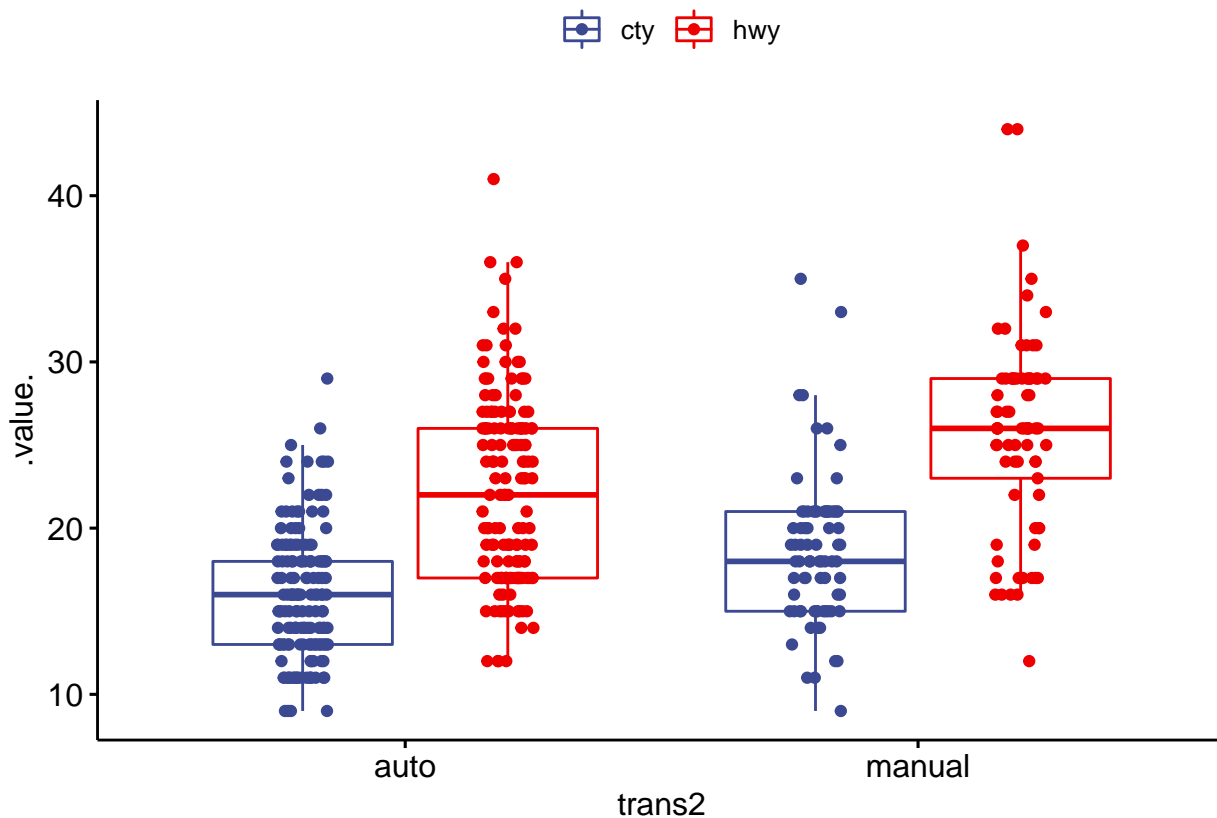
#mpg %>%
# mutate(trans2=str_sub(trans,1,4)) %>%
# view(mpg)

dt1 <- mpg %>%
  mutate(trans2=word(trans,1,sep = fixed("("))) %>%
  select(cty,hwy,trans2) %>%
  as.data.frame()

#statystyki opisowe; średnia i odchylenie standardowe
dt1 %>%
  group_by(trans2) %>%
  summarise_at(vars(cty,hwy), list(~mean(.),~sd(.)))

## # A tibble: 2 x 5
##   trans2 cty_mean hwy_mean cty_sd hwy_sd
##   <chr>    <dbl>    <dbl> <dbl> <dbl>
## 1 auto      16.0     22.3  3.85  5.62
## 2 manual    18.7     25.8  4.49  5.96

dt1 %>%
  ggboxplot(x="trans2",
            y=c("cty","hwy"),
            add="jitter",
            merge = T,
            palette = "aaas")
```



## Sprawdzenie założeń

Normalność wielowymiarowa w podgrupach war

```
dt1 %>%
  group_split(trans2, keep=F) %>%
  map(~mshapiro_test(.x))
```

```
## [[1]]
## # A tibble: 1 x 2
##   statistic p.value
##   <dbl>     <dbl>
## 1      0.971 0.00211
##
## [[2]]
## # A tibble: 1 x 2
##   statistic p.value
##   <dbl>     <dbl>
## 1      0.895 0.0000110
```

Brak wielowym normalności w obu grupach.

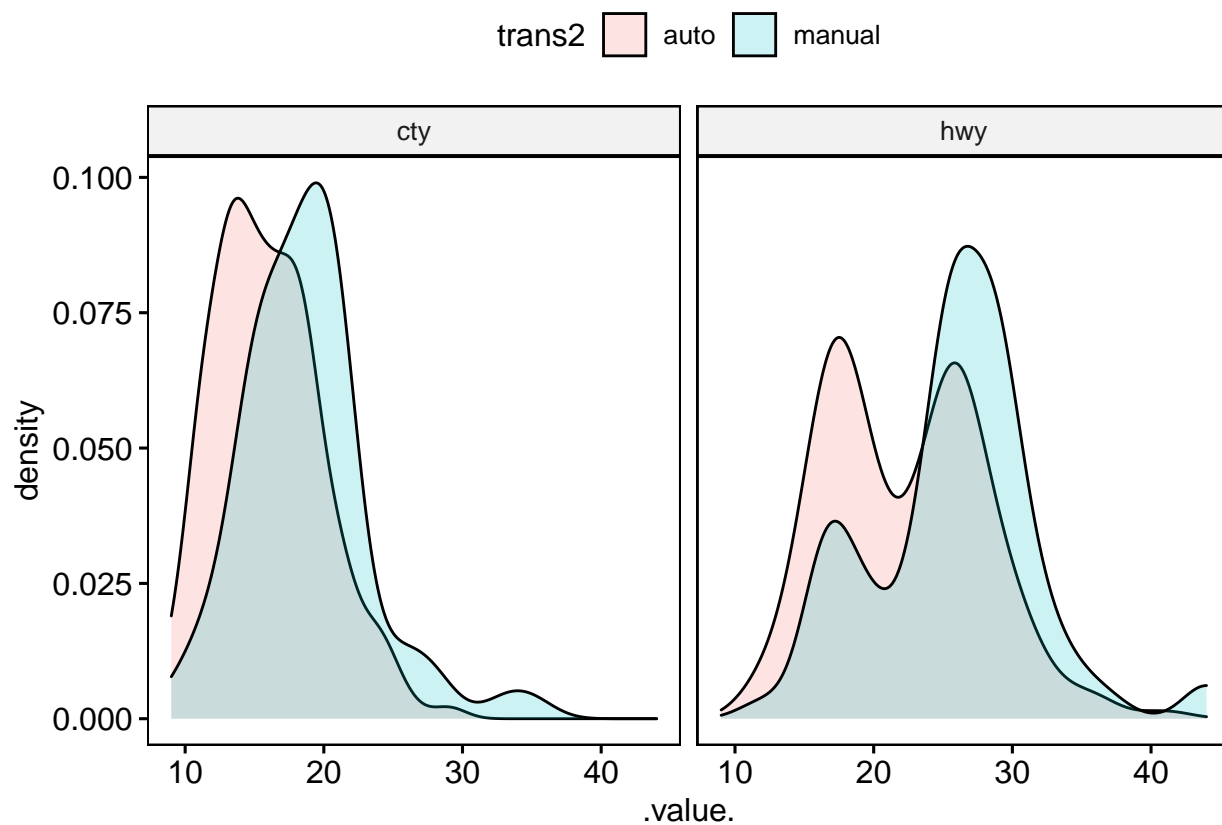
```
library(car)
summary(powerTransform(cbind(cty,hwy)~1,family = "bcPower", data=dt1 ))
```

```
## bcPower Transformations to Multinormality
##   Est Power Rounded Pwr Wald Lwr Bnd Wald Up Bnd
## cty   0.1727         0.0   -0.1356    0.4809
## hwy   0.5808         0.5    0.2265    0.9352
```

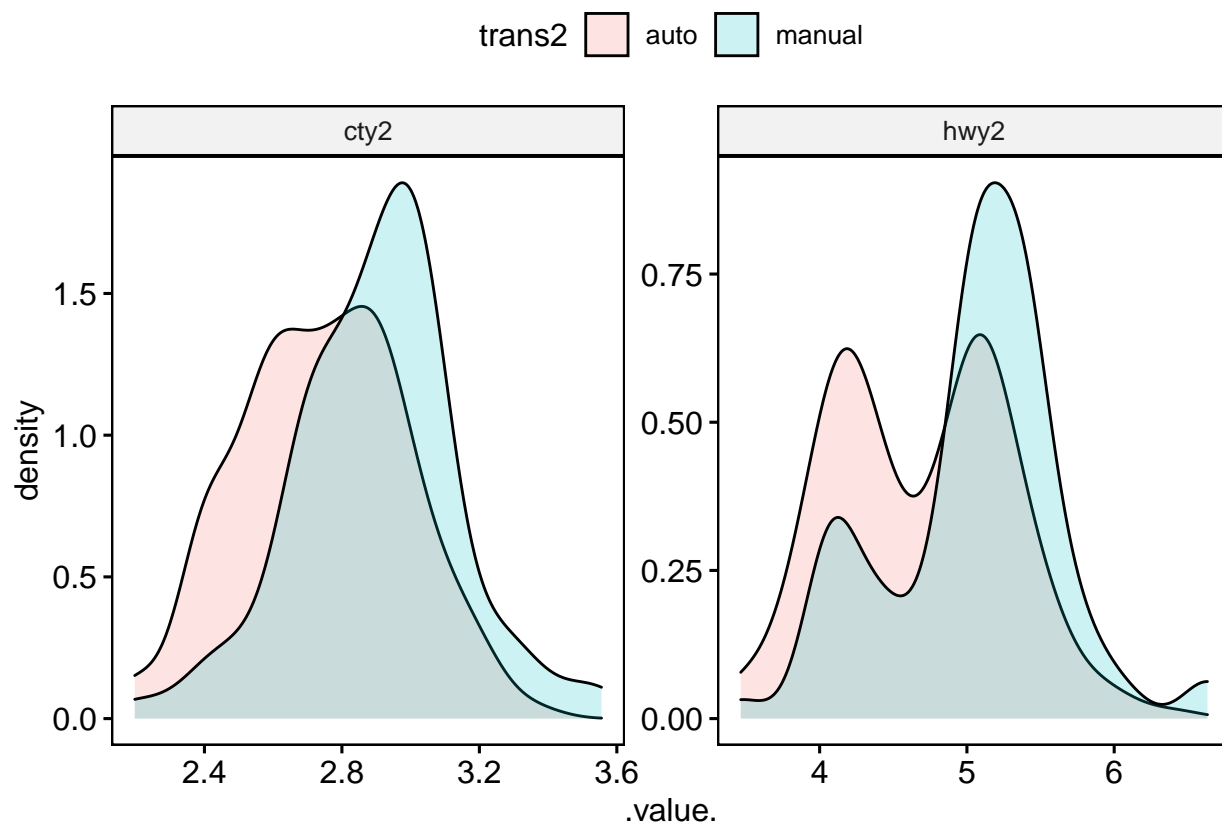
```
##
## Likelihood ratio test that transformation parameters are equal to 0
## (all log transformations)
##               LRT df      pval
## LR test, lambda = (0 0) 14.60934  2 0.00067239
##
## Likelihood ratio test that no transformations are needed
##               LRT df      pval
## LR test, lambda = (1 1) 37.80421  2 6.179e-09
```

```
dt2 <- dt1 %>%
  mutate(cty2=log(cty), hwy2=sqrt(hwy)) %>%
  as.data.frame()
```

```
dt2 %>%
  ggdensity(x=c("cty", "hwy"), combine = T, fill = "trans2", alpha = 0.2)
```



```
dt2 %>%
  ggdensity(x=c("cty2", "hwy2"), combine = T, fill = "trans2", alpha = 0.2, scales="free")
```



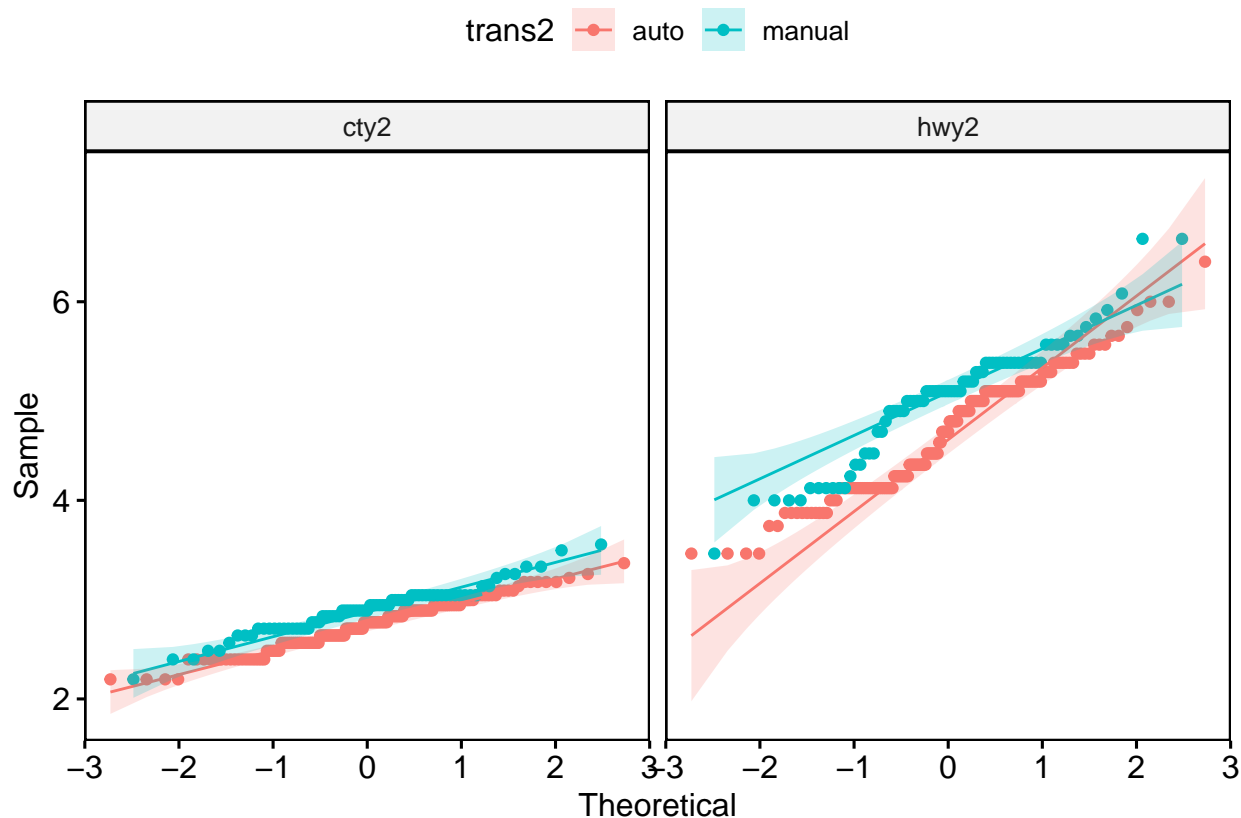
Spr czy po trans jest norm

```
dt2 %>%
  group_split(trans2, keep=F) %>%
  map(~mshapiro_test(.x))
```

```
## [[1]]
## # A tibble: 1 x 2
##   statistic p.value
##   <dbl>     <dbl>
## 1     0.689 8.19e-17
##
## [[2]]
## # A tibble: 1 x 2
##   statistic p.value
##   <dbl>     <dbl>
## 1     0.676 9.60e-12
```

wykr kwantylowy

```
ggqqplot(dt2, x=c("cty2", "hwy2"), combine = T, color = "trans2")
```



### Badanie jednorodności kowariancji

```
box_m(dt2[,4:5],dt2[,3])
```

```
## # A tibble: 1 x 4
##   statistic p.value parameter method
##   <dbl>     <dbl>     <dbl> <chr>
## 1      0.402   0.940         3 Box's M-test for Homogeneity of Covariance Matric~
```

Jest spełnione założenie o jednej wariancji

Jeśli próba jest dostatecznie duża (ta ma 232), to centralne tw graniczne nam zapewni to, że nawet jeśli oryginalne dane nie są z rozkładu normalnego, a po transformacji są to nadal jest ok dla dużej próby.

```
library(Hotelling)
wynikth <- hotelling.test(cbind(cty2,hwy2)~trans2,
                          data = dt2)
```

Odrzucamy hipotezę o równości wektorów średnich w obu grupach

### Testy posthoc

```
dt2 %>%
  pivot_longer(cols = cty2:hwy2) %>%
  group_by(name) %>%
  levene_test(value~trans2)
```

```
## # A tibble: 2 x 5
##   name    df1    df2 statistic      p
##   <chr> <int> <int>      <dbl> <dbl>
## 1 cty2      1    232      0.953 0.330
## 2 hwy2      1    232      3.07 0.0809
```

W przypadku obu zmiennych zależnych występuje jednorodność wariancji. Można testować testem tstudenta z varequal równym T, p.adjust.method bo robimy porównania wielokrotne.

```
dt2 %>%
  pivot_longer(cols = cty2:hwy2) %>%
  group_by(name) %>%
  t_test(value~trans2,
        data=.,
        var.equal = T,
        p.adjust.method = "bonferroni")
```

```
## # A tibble: 2 x 9
##   name .y. group1 group2    n1    n2 statistic    df      p
## * <chr> <chr> <chr> <chr> <int> <int>      <dbl> <dbl>    <dbl>
## 1 cty2 value auto  manual   157    77     -4.76   232 0.00000344
## 2 hwy2 value auto  manual   157    77     -4.34   232 0.0000214
```