Two-Way AOVA Test

Andrew Levine

December 2021

```
DP <- read.csv(file="/Users/andrewlevine/Downloads/Statistics II/DrugPrices.csv")
```

Summary Statistics and Box Plot:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
str(DP)
```

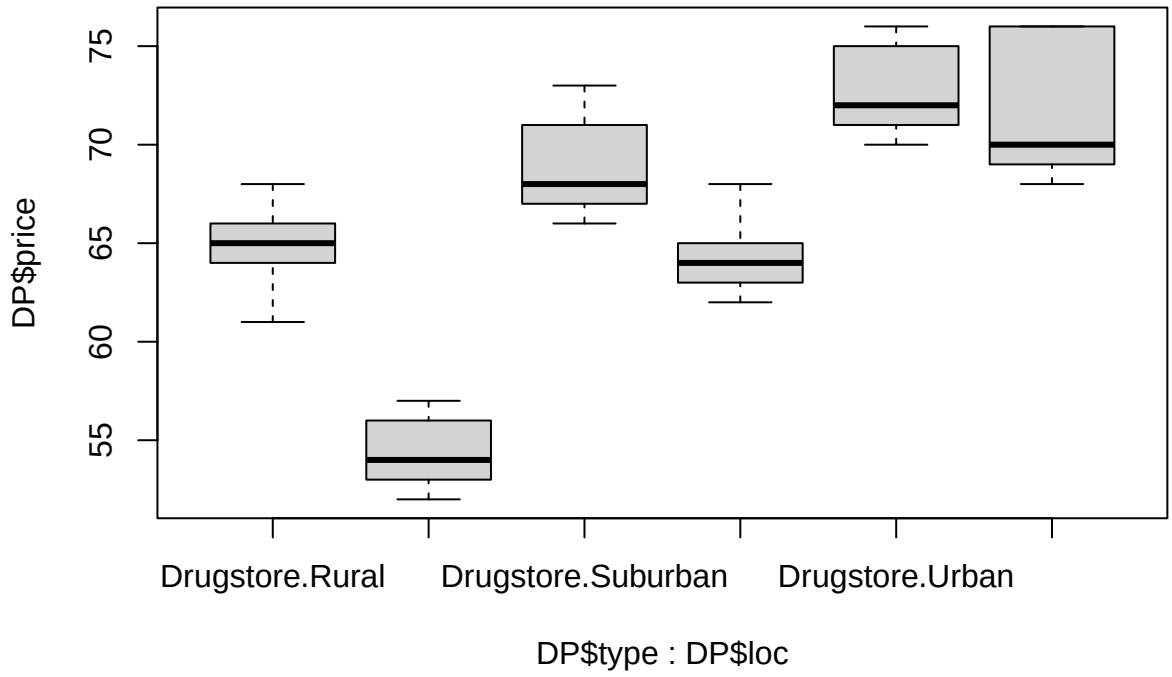
```
## 'data.frame':    30 obs. of  3 variables:
##  $ type : chr  "Drugstore" "Drugstore" "Drugstore" "Drugstore" ...
##  $ loc  : chr  "Urban" "Urban" "Urban" "Urban" ...
##  $ price: int  75 70 72 76 71 73 71 67 66 68 ...
```

```
DP$type <- factor(DP$type)
DP$loc <- factor(DP$loc)
grouped_DP <- group_by(DP, type, loc)
summarize(grouped_DP, mean(price), sd(price), n())
```

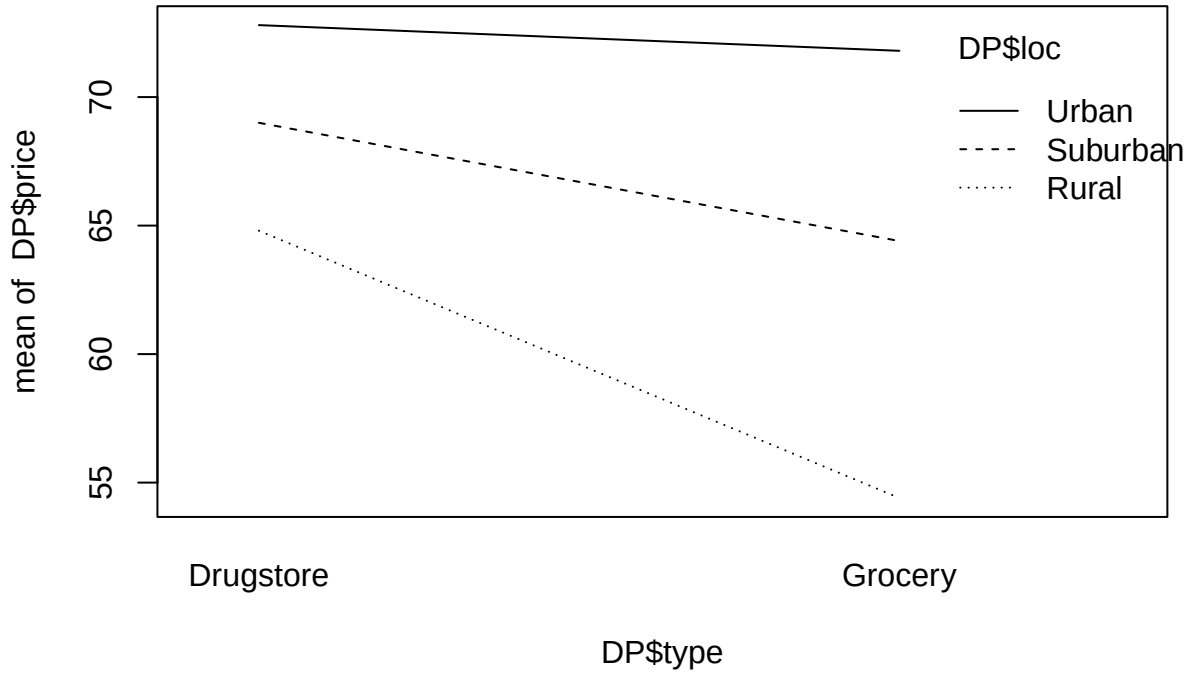
```
## `summarise()` has grouped output by 'type'. You can override using the `.groups` argument.

## # A tibble: 6 x 5
## # Groups:   type [2]
##   type      loc      `mean(price)` `sd(price)` `n()`
##   <fct>     <fct>            <dbl>       <dbl> <int>
## 1 Drugstore Rural             64.8        2.59     5
## 2 Drugstore Suburban          69          2.92     5
## 3 Drugstore Urban             72.8        2.59     5
## 4 Grocery   Rural             54.4        2.07     5
## 5 Grocery   Suburban          64.4        2.30     5
## 6 Grocery   Urban             71.8        3.90     5
```

```
boxplot(DP$price ~ DP$type*DP$loc)
```

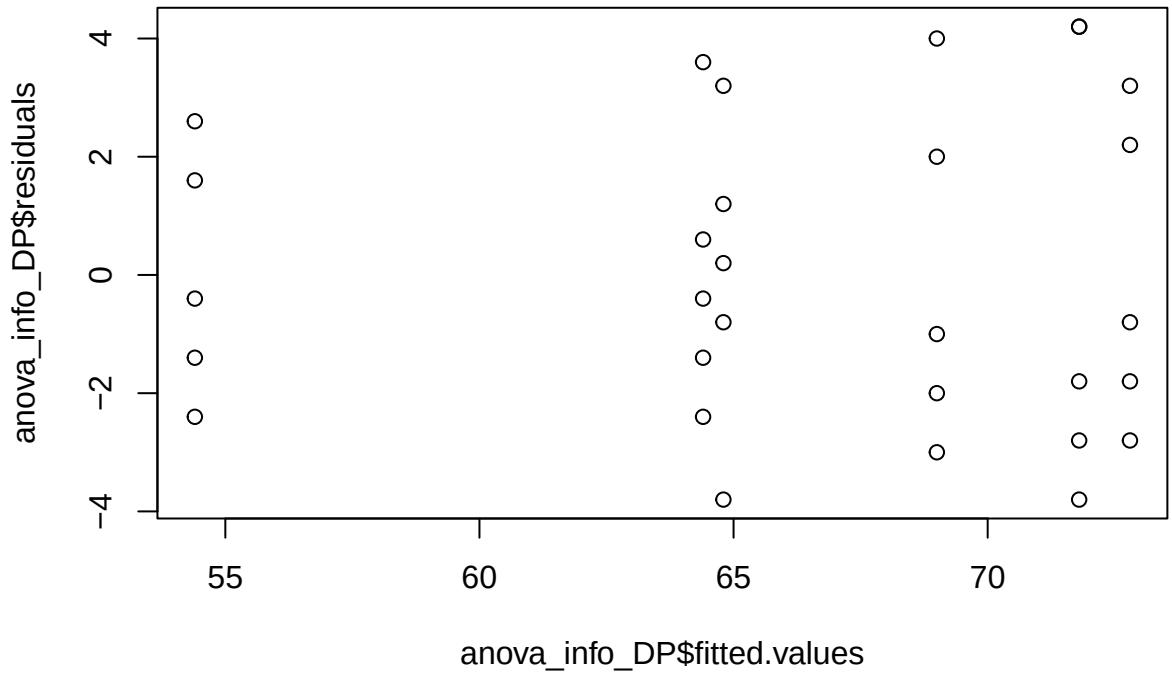



```
interaction.plot(DP$type, DP$loc, DP$price)
```

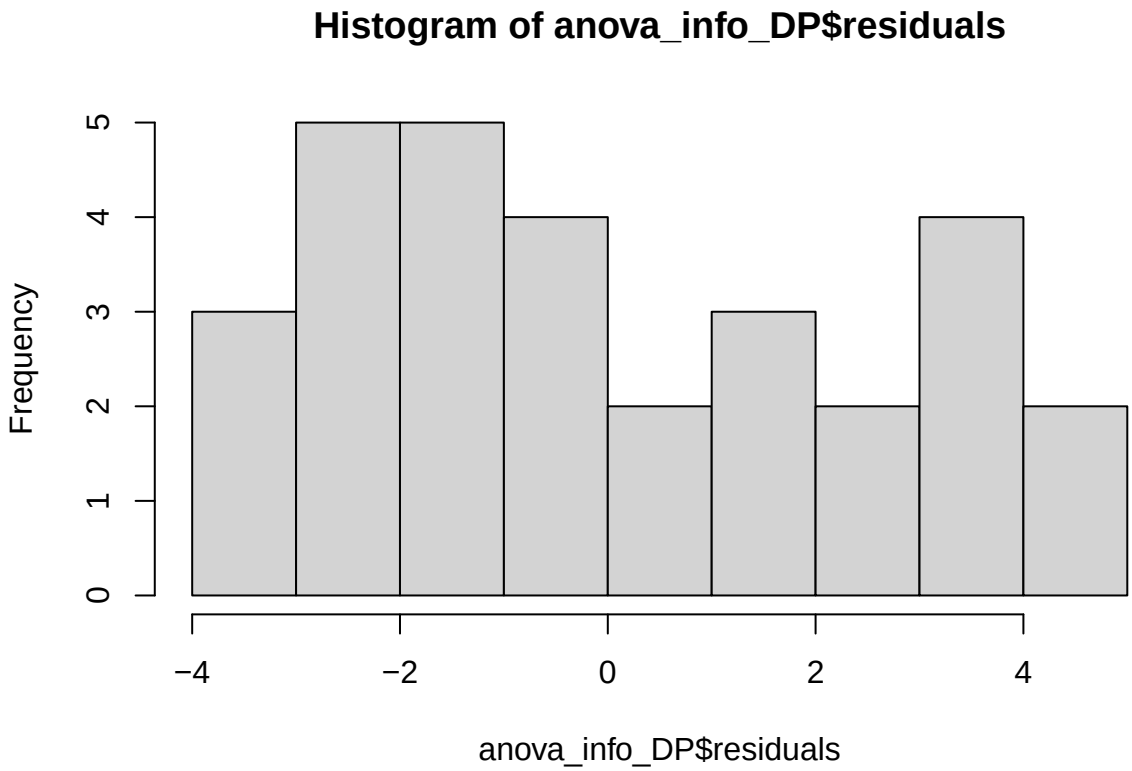


Plots for Linear Assumptions:

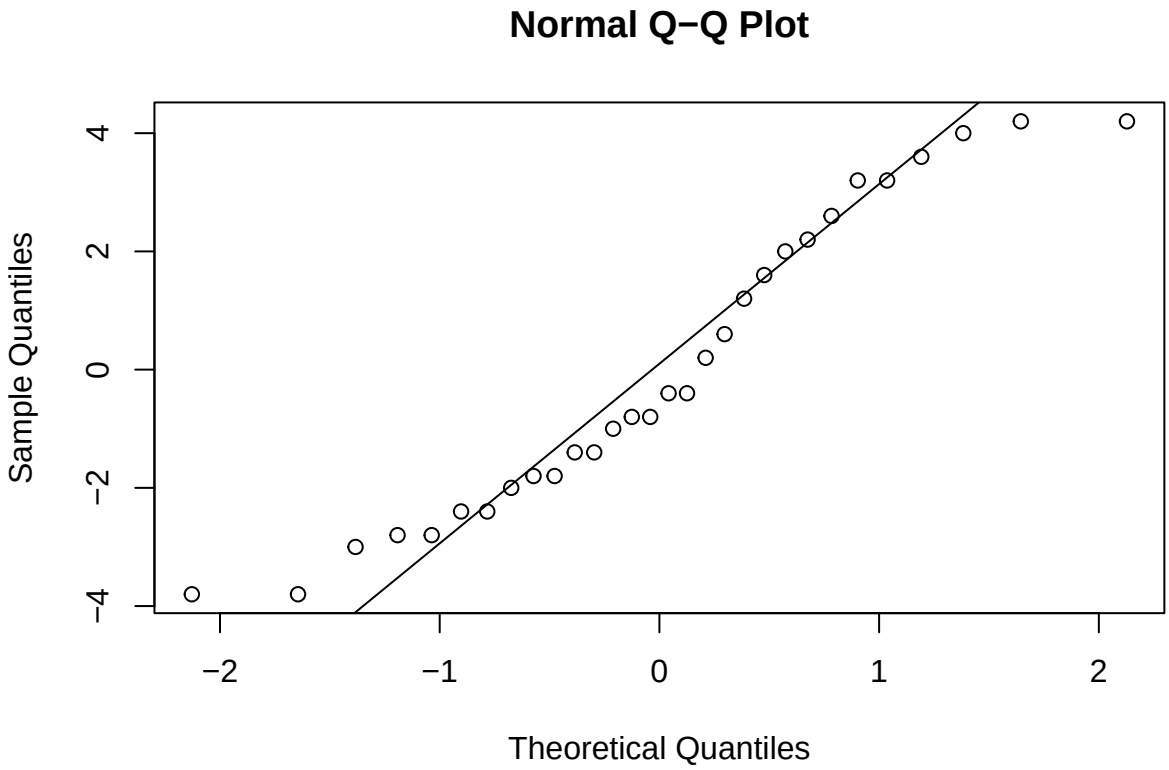
```
anova_info_DP <- aov(price ~ loc + type + loc*type, data = DP)
plot(x = anova_info_DP$fitted.values, y = anova_info_DP$residuals)
```




```
hist(anova_info_DP$residuals)
```




```
qqnorm(anova_info_DP$residuals)
qqline(anova_info_DP$residuals)
```



Levene Test for Equality of Variances:
Hypotheses and Significance Level:
H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2$
H_a: Not all population variances are the same.
α: 0.05

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

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

```
leveneTest(price ~ loc*type, data = DP)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  0.4078 0.8386
##       24
```

The Levene's test resulted in a p-value of 0.8386, which is larger than our α of 0.05. As a result, it is safe to assume equal variance of errors.

```
summary(anova_info_DP)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## loc          2  810.2   405.1  52.047 1.87e-09 ***
## type         1  213.3   213.3  27.409 2.29e-05 ***
## loc:type     2  112.5    56.2   7.225   0.0035 **
## Residuals   24  186.8     7.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hypotheses and Significance Level:
H_0: There is no difference in the mean drug price based on an interaction between store type and store location
H_a: There is a difference in the mean drug price based on an interaction between store type and store location
α: 0.05
The p-value of 2.29 * 10^{-5} is less than the α of 0.05, and so we reject H_0. At the 5% significance level, the mean price of the drug depends on the interaction between store type and store location.

It is appropriate to conduct a post-hoc procedure here, because we did indeed find that mean drug price does significantly vary based on the interaction between store type and store location. Since we have evidence suggesting that not all means are the same, we want to determine the ways in which the means differ and by how much. In order to do so, we will run Tukey's HSD test.

Tukey's HSD Test, ($\alpha = 0.05$):

```
library(DescTools)
```

```
##
## Attaching package: 'DescTools'
```

```
## The following object is masked from 'package:car':
##
##     Recode
```

```
PostHocTest(anova_info_DP, conf.level=0.95, method="hsd")
```

```
##
##   Posthoc multiple comparisons of means : Tukey HSD
##     95% family-wise confidence level
##
## $loc
##                 diff   lwr.ci    upr.ci    pval
## Suburban-Rural  7.1 3.984224 10.215776 2.1e-05 ***
## Urban-Rural    12.7 9.584224 15.815776 1.0e-09 ***
## Urban-Suburban  5.6 2.484224  8.715776 0.00043 ***
##
## $type
##                        diff    lwr.ci    upr.ci    pval
```

```
## Grocery-Drugstore -5.333333 -7.435855 -3.230812 2.3e-05 ***
##
## $`loc:type`
##                                       diff     lwr.ci      upr.ci    pval
## Suburban:Drugstore-Rural:Drugstore     4.2  -1.255601   9.6556012 0.20273
## Urban:Drugstore-Rural:Drugstore        8.0   2.544399  13.4556012 0.00169 **
## Rural:Grocery-Rural:Drugstore        -10.4 -15.855601  -4.9443988 5.9e-05 ***
## Suburban:Grocery-Rural:Drugstore      -0.4  -5.855601   5.0556012 0.99991
## Urban:Grocery-Rural:Drugstore          7.0   1.544399  12.4556012 0.00670 **
## Urban:Drugstore-Suburban:Drugstore     3.8  -1.655601   9.2556012 0.29507
## Rural:Grocery-Suburban:Drugstore     -14.6 -20.055601  -9.1443988 2.4e-07 ***
## Suburban:Grocery-Suburban:Drugstore   -4.6 -10.055601   0.8556012 0.13408
## Urban:Grocery-Suburban:Drugstore       2.8  -2.655601   8.2556012 0.61444
## Rural:Grocery-Urban:Drugstore        -18.4 -23.855601 -12.9443988 3.1e-09 ***
## Suburban:Grocery-Urban:Drugstore      -8.4 -13.855601  -2.9443988 0.00096 ***
## Urban:Grocery-Urban:Drugstore         -1.0  -6.455601   4.4556012 0.99231
## Suburban:Grocery-Rural:Grocery        10.0   4.544399  15.4556012 0.00010 ***
## Urban:Grocery-Rural:Grocery           17.4  11.944399  22.8556012 9.2e-09 ***
## Urban:Grocery-Suburban:Grocery         7.4   1.944399  12.8556012 0.00388 **
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The pairwise p-values for Tukey's HSD test are as follows:
$p\text{-}value_{Suburban:Drugstore-Rural:Drugstore} = 0.2027$
$p\text{-}value_{Urban:Drugstore-Rural:Drugstore} = 0.0017$
$p\text{-}value_{Rural:Grocery-Rural:Drugstore} = 5.9 * 10^{-5}$
$p\text{-}value_{Suburban:Grocery-Rural:Drugstore} = 0.9999$
$p\text{-}value_{Urban:Grocery-Rural:Drugstore} = 0.0067$
$p\text{-}value_{Urban:Drugstore-Suburban:Drugstore} = 0.2951$
$p\text{-}value_{Rural:Grocery-Suburban:Drugstore} = 2.4 * 10^{-7}$
$p\text{-}value_{Suburban:Grocery-Suburban:Drugstore} = 0.1341$
$p\text{-}value_{Urban:Grocery-Suburban:Drugstore} = 0.6144$
$p\text{-}value_{Rural:Grocery-Urban:Drugstore} = 3.1 * 10^{-9}$
$p\text{-}value_{Suburban:Grocery-Urban:Drugstore} = 0.001$
$p\text{-}value_{Urban:Grocery-Urban:Drugstore} = 0.9923$
$p\text{-}value_{Suburban:Grocery-Rural:Grocery} = 0.0001$
$p\text{-}value_{Urban:Grocery-Rural:Grocery} = 9.2 * 10^{-9}$
$p\text{-}value_{Urban:Grocery-Suburban:Grocery} = 0.0039$

Since the p-values for the pairs Urban:Drugstore-Rural:Drugstore (0.0017), Rural:Grocery-Rural:Drugstore ($5.9 * 10^{-5}$), Urban:Grocery-Rural:Drugstore (0.0067), Rural:Grocery-Suburban:Drugstore ($2.4 * 10^{-7}$), Rural:Grocery-Urban:Drugstore ($3.1 * 10^{-9}$), Suburban:Grocery-Urban:Drugstore (0.001), Suburban:Grocery-Rural:Grocery (0.0001), Urban:Grocery-Rural:Grocery ($9.2 * 10^{-9}$), and Urban:Grocery-Suburban:Grocery (0.0039) are all less than the α of 0.05, we can conclude that these pairs each have significant differences in mean drug prices. Judging by the signs of the differences, we can draw the following conclusions:
The mean drug price in urban drugstores is higher than the mean drug price in rural drugstores.
The mean drug price in rural grocery stores is less than the mean drug price in rural drugstores.
The mean drug price in urban grocery stores is higher than the mean drug price in rural drugstores.
The mean drug price in rural grocery stores is less than the mean drug price in suburban drugstores.
The mean drug price in rural grocery stores is less than the mean drug price in urban drugstores.

The mean drug price in suburban grocery stores is less than the mean drug price in urban drugstores.
The mean drug price in suburban grocery stores is more than the mean drug price in rural grocery stores.
The mean drug price in urban grocery stores is more than the mean drug price in rural grocery stores.
The mean drug price in urban grocery stores is more than the mean drug price in suburban grocery stores.

The pairwise confidence intervals for Tukey's HSD test are as follows:
$CI_{Suburban:Drugstore-Rural:Drugstore}$ = (-1.2556, 9.6556)
$CI_{Urban:Drugstore-Rural:Drugstore}$ = (2.5444, 13.4556)
$CI_{Rural:Grocery-Rural:Drugstore}$ = (-15.8556, -4.9444)
$CI_{Suburban:Grocery-Rural:Drugstore}$ = (-5.8556, 5.0556)
$CI_{Urban:Grocery-Rural:Drugstore}$ = (1.5444, 12.4556)
$CI_{Urban:Drugstore-Suburban:Drugstore}$ = (-1.6556, 9.2556)
$CI_{Rural:Grocery-Suburban:Drugstore}$ = (-20.0556, -9.1444)
$CI_{Suburban:Grocery-Suburban:Drugstore}$ = (-10.0556, 0.8556)
$CI_{Urban:Grocery-Suburban:Drugstore}$ = (-2.6556, 8.2556)
$CI_{Rural:Grocery-Urban:Drugstore}$ = (-23.8556, -12.9444)
$CI_{Suburban:Grocery-Urban:Drugstore}$ = (-13.8556, -2.9444)
$CI_{Urban:Grocery-Urban:Drugstore}$ = (-6.4556, 4.4556)
$CI_{Suburban:Grocery-Rural:Grocery}$ = (4.5444, 15.4556)
$CI_{Urban:Grocery-Rural:Grocery}$ = (11.9444, 22.8556)
$CI_{Urban:Grocery-Suburban:Grocery}$ = (1.9444, 12.8556)

Since the 95% confidence intervals for the pairs Urban:Drugstore-Rural:Drugstore (2.5444, 13.4556), Rural:Grocery-Rural:Drugstore (-15.8556, -4.9444), Urban:Grocery-Rural:Drugstore (1.5444, 12.4556), Rural:Grocery-Suburban:Drugstore (-20.0556, -9.1444), Rural:Grocery-Urban:Drugstore (-23.8556, -12.9444), Suburban:Grocery-Urban:Drugstore (-13.8556, -2.9444), Suburban:Grocery-Rural:Grocery (4.5444, 15.4556), Urban:Grocery-Rural:Grocery (11.9444, 22.8556), and Urban:Grocery-Suburban:Grocery (1.9444, 12.8556) all do not contain 0, we can conclude that these pairs each have significant differences in mean drug prices. Judging by the signs of the differences, we can draw the following conclusions:
The mean drug price in urban drugstores is higher than the mean drug price in rural drugstores.
The mean drug price in rural grocery stores is less than the mean drug price in rural drugstores.
The mean drug price in urban grocery stores is higher than the mean drug price in rural drugstores.
The mean drug price in rural grocery stores is less than the mean drug price in suburban drugstores.
The mean drug price in rural grocery stores is less than the mean drug price in urban drugstores.
The mean drug price in suburban grocery stores is less than the mean drug price in urban drugstores.
The mean drug price in suburban grocery stores is more than the mean drug price in rural grocery stores.
The mean drug price in urban grocery stores is more than the mean drug price in rural grocery stores.
The mean drug price in urban grocery stores is more than the mean drug price in suburban grocery stores.