960:584 Project 1 (CrohnD)

Author: Vahe Nersisyan, Eric Lam

Load library

Hide

```
require(robustbase)
require(faraway)
require(glmnet)
require(Rfit)
library(ggplot2)
library(DescTools)
require(MASS)
library(table1)
require(psych)
```

Load data

Hide

```
data("CrohnD")
dt <- CrohnD
summary(dt)
```

```
      ID               nrAdvE            BMI            height          country  sex         age
 Min.   :19908    Min.   : 0.000    Min.   :16.00    Min.   :124.0    c1:78    M: 17    Min.   :19.00
 1st Qu.:23909    1st Qu.: 0.000    1st Qu.:23.05    1st Qu.:157.0    c2:39    F:100    1st Qu.:48.00
 Median :25919    Median : 1.000    Median :25.15    Median :162.0                     Median :56.00
 Mean   :34103    Mean   : 2.034    Mean   :26.06    Mean   :162.7                      Mean   :54.66
 3rd Qu.:51909    3rd Qu.: 3.000    3rd Qu.:28.40    3rd Qu.:166.0                      3rd Qu.:62.00
 Max.   :54937    Max.   :12.000    Max.   :44.06    Max.   :182.0                      Max.   :75.00
     weight            treat
 Min.   : 36.00    placebo:39
 1st Qu.: 59.00    d1     :39
 Median : 68.00    d2     :39
 Mean   : 69.03
 3rd Qu.: 76.00
 Max.   :117.00
```

Hide

```
head(dt)
```

	ID <int>	nrAdvE <int>	BMI <dbl>	height <int>	country <fctr>	sex <fctr>	age <int>	weight <int>	treat <fctr>
1	19908	4	25.22	163	c1	F	47	67	placebo

	ID <int>	nrAdvE <int>	BMI <dbl>	height <int>	country <fctr>	sex <fctr>	age <int>	weight <int>	treat <fctr>
2	19909	4	23.80	164	c1	F	53	64	d1
3	19910	1	23.05	164	c1	F	68	62	placebo
4	20908	1	25.71	165	c1	F	48	70	d2
5	20909	2	25.95	170	c1	F	67	75	placebo
6	20910	2	28.70	168	c1	F	54	81	d1

6 rows

Table of demographics

Hide

```
tbl1 <- table1(~nrAdvE + BMI + height + country +
        sex + age + weight|treat, dt,topclass="Rtable1-grid")
tbl1
```

	placebo (N=39)	d1 (N=39)	d2 (N=39)	Overall (N=117)
nrAdvE				
Mean (SD)	2.41 (3.15)	1.56 (2.33)	2.13 (2.65)	2.03 (2.73)
Median [Min, Max]	1.00 [0, 12.0]	0 [0, 9.00]	1.00 [0, 9.00]	1.00 [0, 12.0]
BMI				
Mean (SD)	25.5 (3.49)	25.8 (5.88)	26.9 (5.29)	26.1 (4.99)
Median [Min, Max]	24.8 [19.5, 33.3]	25.3 [16.0, 44.1]	25.2 [18.2, 43.0]	25.2 [16.0, 44.1]
height				
Mean (SD)	163 (7.71)	161 (10.6)	163 (7.86)	163 (8.78)
Median [Min, Max]	163 [149, 182]	162 [124, 182]	162 [150, 182]	162 [124, 182]
country				
c1	26 (66.7%)	26 (66.7%)	26 (66.7%)	78 (66.7%)
c2	13 (33.3%)	13 (33.3%)	13 (33.3%)	39 (33.3%)
sex				
M	7 (17.9%)	5 (12.8%)	5 (12.8%)	17 (14.5%)
F	32 (82.1%)	34 (87.2%)	34 (87.2%)	100 (85.5%)
age				
Mean (SD)	57.0 (10.0)	54.3 (11.4)	52.6 (10.4)	54.7 (10.7)
Median [Min, Max]	58.0 [35.0, 74.0]	56.0 [19.0, 73.0]	53.0 [29.0, 75.0]	56.0 [19.0, 75.0]
weight				
Mean (SD)	68.3 (11.6)	66.8 (14.0)	72.1 (16.6)	69.0 (14.2)
Median [Min, Max]	67.0 [48.0, 103]	65.0 [36.0, 110]	70.0 [43.0, 117]	68.0 [36.0, 117]

Table of summary by sex and BMI(abberant/normal)

<div align="right">Hide</div>

```
dt2 <- dt
dt2$BMI_new <- ifelse(dt2$BMI >= 18.5 & dt2$BMI <= 24.9, 1, 0)
label(dt2$treat) <- c("Placebo Group", "Treatment Group 1", "Treatment  Group 2")
label(dt2$sex) = c("Male", "Female")
dt2$BMI_new <- factor(dt2$BMI_new, levels = c(0, 1), labels = c("Abberant BMI", "Normal BMI"))
tbl2 <- table1(~nrAdvE + factor(treat) + country|sex * BMI_new, dt2, topclass="Rtable1-grid")
tbl2
```

	M		F		Overall	
	Abberant BMI (N=9)	**Normal BMI (N=8)**	**Abberant BMI (N=58)**	**Normal BMI (N=42)**	**Abberant BMI (N=67)**	**Normal BMI (N=50)**
nrAdvE						
Mean (SD)	4.33 (3.57)	0.875 (1.36)	2.00 (2.56)	1.81 (2.77)	2.31 (2.80)	1.66 (2.61)
Median [Min, Max]	6.00 [0, 9.00]	0 [0, 3.00]	1.00 [0, 12.0]	0.500 [0, 12.0]	1.00 [0, 12.0]	0 [0, 12.0]
factor(treat)						
placebo	4 (44.4%)	3 (37.5%)	15 (25.9%)	17 (40.5%)	19 (28.4%)	20 (40.0%)
d1	3 (33.3%)	2 (25.0%)	20 (34.5%)	14 (33.3%)	23 (34.3%)	16 (32.0%)
d2	2 (22.2%)	3 (37.5%)	23 (39.7%)	11 (26.2%)	25 (37.3%)	14 (28.0%)
country						
c1	7 (77.8%)	6 (75.0%)	39 (67.2%)	26 (61.9%)	46 (68.7%)	32 (64.0%)
c2	2 (22.2%)	2 (25.0%)	19 (32.8%)	16 (38.1%)	21 (31.3%)	18 (36.0%)

Plots

<div align="right">Hide</div>

```
boxplot(sqrt(dt$nrAdvE)~dt$treat, col = "gray",
        main = "Boxplot of Numbers of Adverse Events per Treatment Group")
```
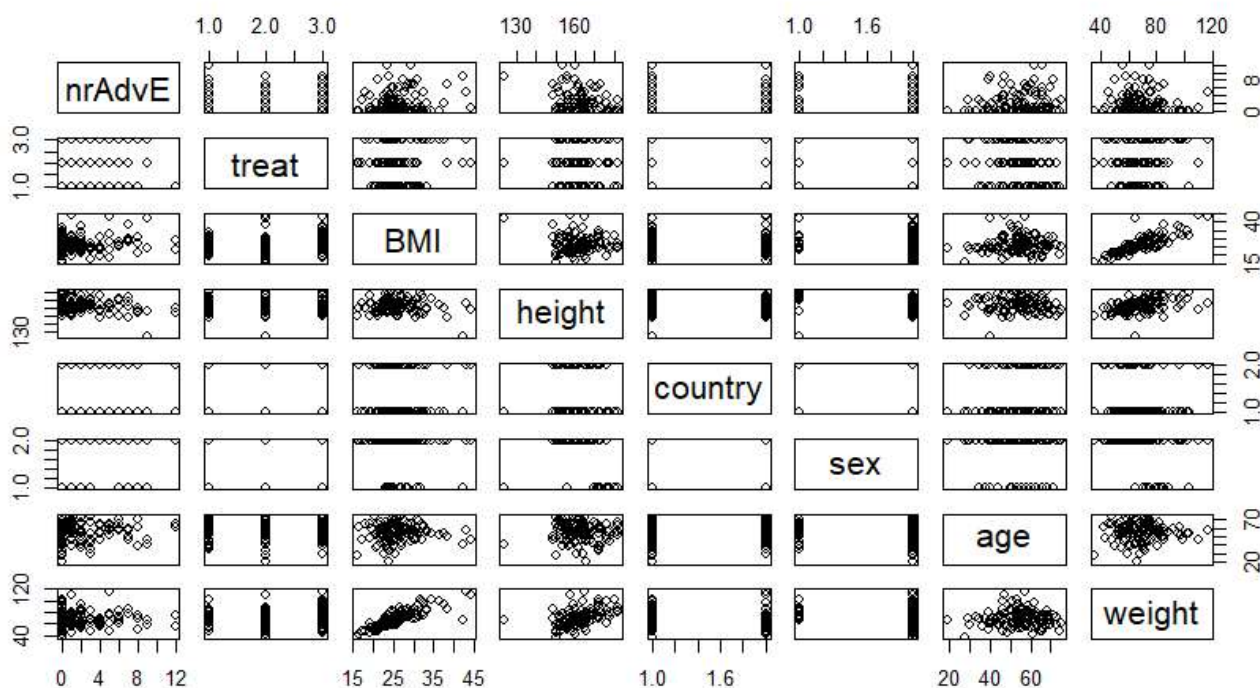
Boxplot of Numbers of Adverse Events per Treatment Group
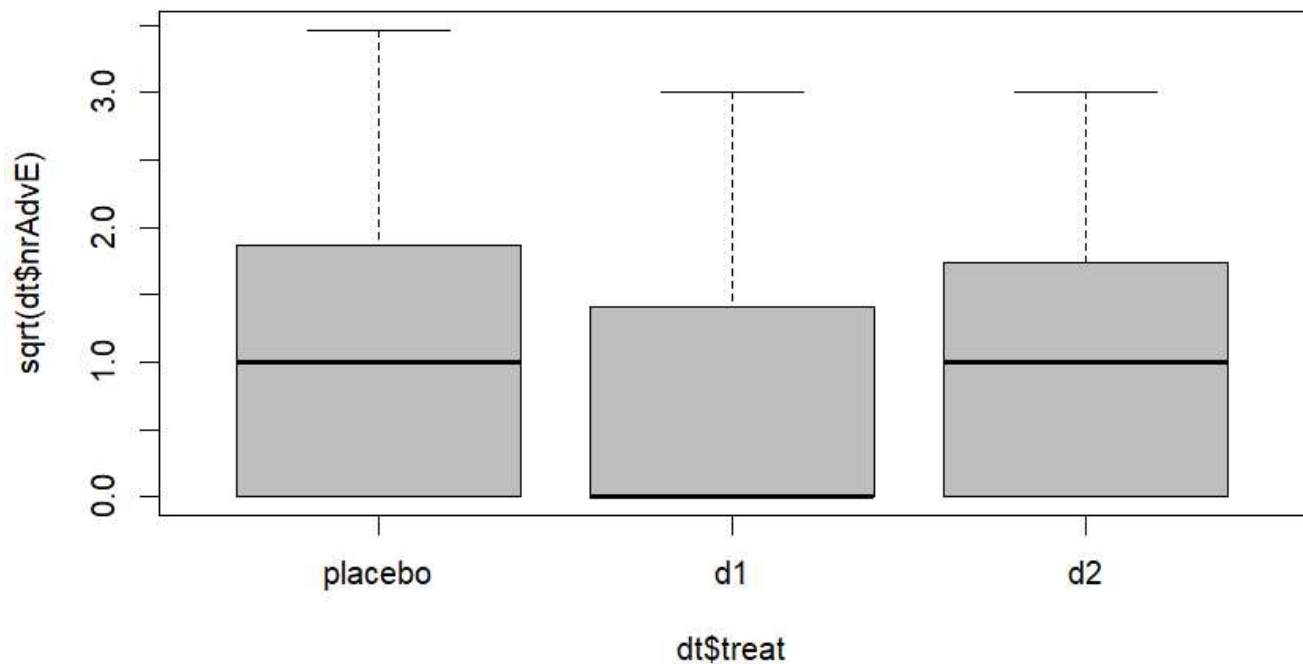
```
scater_plot <- pairs(nrAdvE~treat + BMI + height + country
          + sex + age + weight, data=dt,
          main = "Scatterplot of Number of Adverse Events vs. Predictor Variables")
```

Scatterplot of Number of Adverse Events vs. Predictor Variables

Fit regression model

```
#Robust Fitting of Linear Models
model <- rlm(sqrt(nrAdvE)~factor(treat) + BMI + height +
             country + sex + age + weight, data=dt)
summary(model)
```

```
Call: rlm(formula = sqrt(nrAdvE) ~ factor(treat) + BMI + height + country +
    sex + age + weight, data = dt)
Residuals:
      Min       1Q    Median       3Q       Max
-1.37690 -0.84226 -0.09792  0.72238   2.41353

Coefficients:
               Value    Std. Error t value
(Intercept)    -13.5377  9.9862    -1.3556
factor(treat)d1 -0.3909  0.2186    -1.7881
factor(treat)d2 -0.0518  0.2221    -0.2334
BMI              0.3378  0.1627     2.0760
height           0.0856  0.0613     1.3969
countryc2        0.3368  0.1913     1.7607
sexF            -0.3612  0.3076    -1.1744
age              0.0094  0.0086     1.0962
weight          -0.1214  0.0638    -1.9020

Residual standard error: 1.209 on 108 degrees of freedom
```

```
model2 <- lmrob(nrAdvE ~ factor(treat) + BMI + height +
              country + sex + age + weight, data = dt)
summary(model2)
```

```
Call:
lmrob(formula = nrAdvE ~ factor(treat) + BMI + height + country + sex + age +
    weight, data = dt)
 \--> method = "MM"
Residuals:
    Min      1Q  Median      3Q     Max
-2.1697 -1.0108 -0.1273  1.6053 11.0046


Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   -50.055364  24.377686  -2.053  0.04246 *
factor(treat)d1 -0.264844   0.483623  -0.548  0.58508
factor(treat)d2  0.040040   0.525500   0.076  0.93941
BMI             1.095860   0.333832   3.283  0.00139 **
height          0.312758   0.150639   2.076  0.04025 *
countryc2       0.462614   0.683422   0.677  0.49991
sexF           -0.795177   1.278389  -0.622  0.53524
age             0.006582   0.018585   0.354  0.72393
weight         -0.403574   0.141405  -2.854  0.00518 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Robust residual standard error: 1.406
Multiple R-squared:  0.2884,    Adjusted R-squared:  0.2357
Convergence in 29 IRWLS iterations


Robustness weights:
 4 observations c(23,49,51,96) are outliers with |weight| = 0 ( < 0.00085);
 10 weights are ~= 1. The remaining 103 ones are summarized as
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0139  0.8272  0.9402  0.8362  0.9705  0.9984
Algorithmic parameters:
       tuning.chi              bb         tuning.psi        refine.tol           rel.tol
        1.548e+00       5.000e-01          4.685e+00         1.000e-07         1.000e-07
         scale.tol       solve.tol           zero.tol        eps.outlier             eps.x
         1.000e-10       1.000e-07          1.000e-10         8.547e-04         3.311e-10
warn.limit.reject warn.limit.meanrw
         5.000e-01        5.000e-01
        nResample          max.it           best.r.s          k.fast.s             k.max        maxit.scale
              500              50                  2                 1               200                200
        trace.lev             mts         compute.rd   fast.s.large.n
                0            1000                  0               2000
               psi               subsampling                         cov compute.outlier.stats
        "bisquare"             "nonsingular"               ".vcov.avar1"                       "SM"
seed : int(0)
```

```
model3 <- glm(sqrt(nrAdvE)~factor(treat), data=dt, family = "poisson")
```

```
Warning: non-integer x = 1.414214Warning: non-integer x = 1.414214Warning: non-integer x = 1.732
051Warning: non-integer x = 2.236068Warning: non-integer x = 1.414214Warning: non-integer x = 2.
449490Warning: non-integer x = 2.236068Warning: non-integer x = 2.828427Warning: non-integer x =
1.732051Warning: non-integer x = 1.414214Warning: non-integer x = 1.732051Warning: non-integer x
= 1.732051Warning: non-integer x = 1.414214Warning: non-integer x = 1.414214Warning: non-integer
x = 1.414214Warning: non-integer x = 1.732051Warning: non-integer x = 1.414214Warning: non-integ
er x = 3.464102Warning: non-integer x = 2.828427Warning: non-integer x = 1.414214Warning: non-in
teger x = 2.236068Warning: non-integer x = 1.414214Warning: non-integer x = 1.414214Warning: non
-integer x = 1.414214Warning: non-integer x = 2.645751Warning: non-integer x = 2.449490Warning:
non-integer x = 2.645751Warning: non-integer x = 2.449490Warning: non-integer x = 1.414214Warnin
g: non-integer x = 1.414214Warning: non-integer x = 2.236068Warning: non-integer x = 2.645751War
ning: non-integer x = 3.464102Warning: non-integer x = 1.732051Warning: non-integer x = 2.645751
Warning: non-integer x = 1.732051Warning: non-integer x = 2.449490Warning: non-integer x = 2.828
427Warning: non-integer x = 2.236068Warning: non-integer x = 1.414214Warning: non-integer x = 1.
414214
```

Hide

```
summary(model3)
```

```
Call:
glm(formula = sqrt(nrAdvE) ~ factor(treat), family = "poisson",
    data = dt)

Deviance Residuals:
     Min       1Q    Median        3Q       Max
-1.52526  -1.26726  -0.05156   0.61517   1.87463

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)       0.1512     0.1485   1.018    0.309
factor(treat)d1  -0.3706     0.2323  -1.595    0.111
factor(treat)d2  -0.1001     0.2154  -0.465    0.642

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 148.37  on 116  degrees of freedom
Residual deviance: 145.66  on 114  degrees of freedom
AIC: Inf

Number of Fisher Scoring iterations: 5
```

Forest Plot for Treatment Effects on Adverse Events
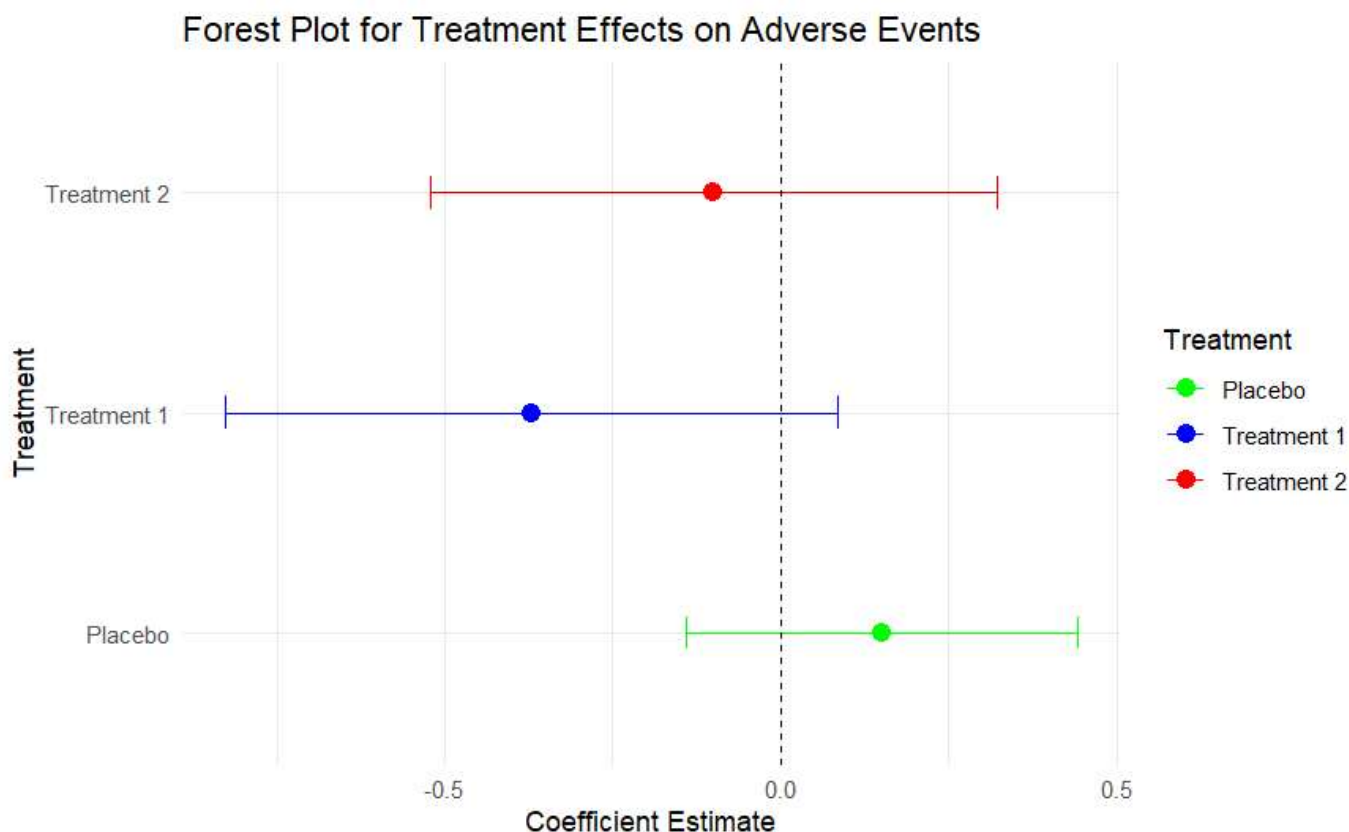
Hide

```
# Create a data frame with coefficients, standard errors, and confidence intervals
forest_data <- data.frame(
      Treatment = c("Placebo", "Treatment 1", "Treatment 2"),
      Coefficient = c(0.1512, -0.3706, -0.1001 ),   # Replace with your extracted coefficients
      Std_Error = c(0.1485,  0.2323, 0.2154),      # Replace with your extracted standard errors
      Lower_CI = c(-0.1398109, -0.8259636, -0.5222866),    # Replace with your lower confidence
interval bounds
      Upper_CI = c(0.44217980, 0.08472773, 0.32214986)        # Replace with your upper confiden
ce interval bounds
)

# Create a color mapping for treatments
color_mapping <- c( "Placebo" = "green","Treatment 1" = "blue", "Treatment 2" = "red")  # Adjust
colors as needed

# Create the forest plot
a <- ggplot(forest_data, aes(x = Coefficient, y = Treatment, color = Treatment)) +
  geom_point(aes(x = Coefficient), size = 3) +
  geom_errorbarh(aes(xmin = Lower_CI, xmax = Upper_CI), height = 0.15) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  labs(title = "Forest Plot for Treatment Effects on Adverse Events",
       x = "Coefficient Estimate",
       y = "Treatment") +
  scale_color_manual(values = color_mapping, name = "Treatment") +  # Add a legend
theme_minimal()

print(a)
```


Forest Plot for Treatment Effects on Adverse Events

Test weither there are significan difference between treatment groups

<div align="right">Hide</div>

```
# treatment 1 vs placebo
wilcox.test(x = dt$nrAdvE[which(dt$treat == "d1")],
            dt$nrAdvE[which(dt$treat == "placebo")],
            alternative = "two.sided", paired = T)
```

```
Warning: cannot compute exact p-value with tiesWarning: cannot compute exact p-value with zeroes
```

```
	Wilcoxon signed rank test with continuity correction

data:  dt$nrAdvE[which(dt$treat == "d1")] and dt$nrAdvE[which(dt$treat == "placebo")]
V = 140.5, p-value = 0.1547
alternative hypothesis: true location shift is not equal to 0
```

<div align="right">Hide</div>

```
# treatment 2 vs placebo
wilcox.test(x = dt$nrAdvE[which(dt$treat == "d2")],
            dt$nrAdvE[which(dt$treat == "placebo")],
            alternative = "two.sided", paired = T)
```

```
Warning: cannot compute exact p-value with tiesWarning: cannot compute exact p-value with zeroes
```

```
	Wilcoxon signed rank test with continuity correction

data:  dt$nrAdvE[which(dt$treat == "d2")] and dt$nrAdvE[which(dt$treat == "placebo")]
V = 258.5, p-value = 0.6991
alternative hypothesis: true location shift is not equal to 0
```

<div align="right">Hide</div>

```
# treatment 1 vs treatment 2
wilcox.test(x = dt$nrAdvE[which(dt$treat == "d1")],
            dt$nrAdvE[which(dt$treat == "d2")],
            alternative = "two.sided", paired = T)
```

```
Warning: cannot compute exact p-value with tiesWarning: cannot compute exact p-value with zeroes
```

```
	Wilcoxon signed rank test with continuity correction

data:  dt$nrAdvE[which(dt$treat == "d1")] and dt$nrAdvE[which(dt$treat == "d2")]
V = 143, p-value = 0.2713
alternative hypothesis: true location shift is not equal to 0
```

<div align="right">Hide</div>

```
# Kruskal-Wallis rank sum test
kruskal.test(nrAdvE~treat, dt)
```

```
	Kruskal-Wallis rank sum test

data:  nrAdvE by treat
Kruskal-Wallis chi-squared = 2.3991, df = 2, p-value = 0.3013
```