Robust Estimation Of Treatment Effect

Wim Krijnen
Lector Analyse Technieken voor Praktijk Onderzoek
Lectoraat Transparante Zorgverlening
Academie voor Gezondheisstudies
Eyssoniusplein 18
9714 CE Groningen

January 21, 2015

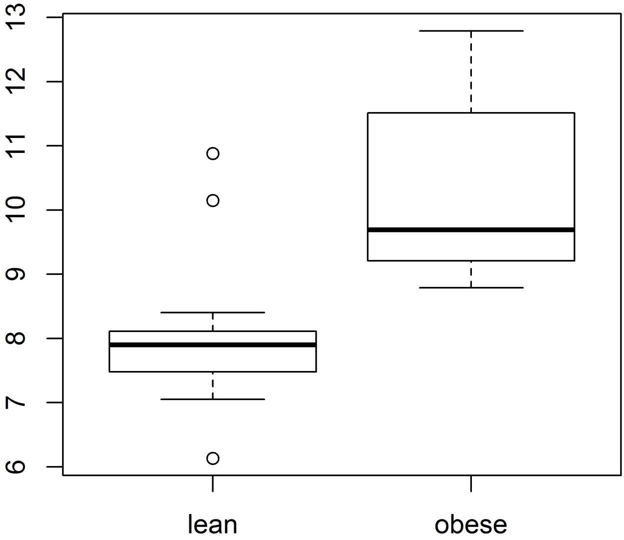
Outline

1. Robust Comparison of Two Independent Groups
   - Energy Expenditure Lean and Obese Women
   - Normality Tests
   - Descriptive Statistics
   - Two-sample t-test Without Assuming Equal Variance
   - Exact Wilcoxon Mann-Whitney Rank Sum Test
   - Medians Test
   - Robust Estimation of Effect by Linear Model
   - Overview of Results on two groups
2. Robust Comparison of More Than Two Independent Groups
   - Age of Walking of Children
   - Normality Tests
   - Descriptive Statistics
   - Kruskal-Wallis Rank Sum Test
   - Robust Estimation of Linear model

Energy Expenditure Lean and Obese Women

Experiment:

- Energy expenditure (MJ) measured during 24 hour
- 13 lean and 9 obese woman
- Reference: D.G. Altman (1991), Practical Statistics for Medical Research, Table 9.4, Chapman & Hall.
- See box-and-whiskers-plot



13
12
11
10
9
8
7
6
lean
obese

P-values from various Normality tests

Initial conclusion from boxplot:

1. Lean has outliers, Obese not
2. Lean seems symmetrically distributed, Obese not

P-values from various Normality tests

Initial conclusion from boxplot:

1. Lean has outliers, Obese not
2. Lean seems symmetrically distributed, Obese not

What do normality tests say?

P-values from various Normality tests

Initial conclusion from boxplot:

1. Lean has outliers, Obese not
2. Lean seems symmetrically distributed, Obese not

What do normality tests say?

	Lean	Obese
Shapiro-Wilk	0.0482	0.1426
Shapiro-Francia	0.0292	0.1549
Anderson-Darling	0.0186	0.1089
Cramer-von Mises	0.0139	0.0820
Lilliefors	0.0205	0.0820

Remark: Lilliefors is based on Kolmogorov-Smirnov

1. Conclusion: Normality violated by lean, not by Obese
2. What to do?

P-values from various Normality tests

Initial conclusion from boxplot:

1. Lean has outliers, Obese not
2. Lean seems symmetrically distributed, Obese not

What do normality tests say?

	Lean	Obese
Shapiro-Wilk	0.0482	0.1426
Shapiro-Francia	0.0292	0.1549
Anderson-Darling	0.0186	0.1089
Cramer-von Mises	0.0139	0.0820
Lilliefors	0.0205	0.0820

Remark: Lilliefors is based on Kolmogorov-Smirnov

1. Conclusion: Normality violated by lean, not by Obese
2. What to do? Let's look at descriptive statistics first.

Descriptive Statistics Energy Expenditure Lean and Obese Women

	lean	obese
mean	8.07	10.30
median	7.90	9.69
Huber	7.86	9.99
SD	1.24	1.40
MAD	0.62	0.74
IQR	0.47	1.70

Mean Absolute Deviation (MAD); Inter Quartile Range (IQR) both adapted for normal distribution

- small within group differences in mean, median, Huber
- within groups difference on SD, MAD, IQR considerable

For completeness: Definition of Huber mean

$$\min \rho(X_1, \cdots, X_k, \mu); \rho(x) = \begin{cases} x^2 & \text{if } |x| \leq k \\ 2k|x| - k^2 & \text{if } |x| > k \end{cases}$$

where $k = 1.345$

Robust Statistics - Theory and Methods, Ricardo Maronna, R. Douglas Martin and Victor Yohai, Wiley, 2006

Two-sample t-test Without Assuming Equal Variance

```
> t.test(expend ~ stature, var.equal = FALSE, data=ene
        Welch Two Sample t-test
data:  expend by stature
t = -3.8555, df = 15.919, p-value = 0.001411
alternative hypothesis: true difference in means is no
95 percent confidence interval:
 -3.459167 -1.004081
sample estimates:
 mean in group lean mean in group obese
```

- t-test robust against mild violation from normality
- t-test optimal under assumption of normality
- null hypothesis of no effect rejected
- effect size -2.23 with 95% CI (-3.45; -1.00)
- Effect size confirmed by tests robust against violations of normality?

Exact Wilcoxon Mann-Whitney Rank Sum Test

```
> wilcox_test(expend ~ stature, data=energy,
+       distribution = "exact", conf.int = TRUE)
        Exact Wilcoxon Mann-Whitney Rank Sum Test
data:  expend by stature (lean, obese)
Z = -3.1061, p-value = 0.001039
alternative hypothesis: true mu is not equal to 0
95 percent confidence interval:
 -3.56 -1.26
sample estimates: difference in location
                 -1.91
```

- Assumption of independent continuous measurements!
- SPPS does not give effect size or CI
- null hypothesis of no effect rejected
- effect size -1.91 with 95% CI (-3.56; -1.26)

Medians Test

```
> median_test(expend ~ stature, data=energy,
+      distribution = "exact", conf.int = TRUE)
        Exact Median Test
data:  expend by stature (lean, obese)
Z = 3.8129, p-value = 0.0002211
alternative hypothesis: true mu is not equal to 0
95 percent confidence interval:
 -5.31 -1.08
sample estimates: difference in location
                -1.825
```

- Assumption independent continuous measurements!
- SPPS does not give effect size or CI
- null hypothesis of no effect rejected
- effect size -1.825 with 95% CI (-5.31; -1.08)

Robust estimation of linear model

```
> library(robustbase)
> mod1 <- lmrob(expend ~ stature, data=energy) #MM met
> summary(mod1)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     7.7746     0.2481  31.334  < 2e-16 ***
statureobese    2.1087     0.6600   3.195  0.00455 **
```

- Assumption independent continuous measurements
- Extreme outliers are down weighted for maximum likelihood type of estimation (widely accepted)
- null hypothesis of no effect rejected
- effect size -2.10 with 95% CI (-3.4023; -0.8151)

Overview of Results

Method	Effect	CI
Independent t-test	-2.23	(-3.45; -1.00)
Median	-1.83	(-5.31; -1.08)
Wilcoxon	-1.91	(-3.56; -1.26)
Robust Regression	-2.10	(-3.40; -0.82)

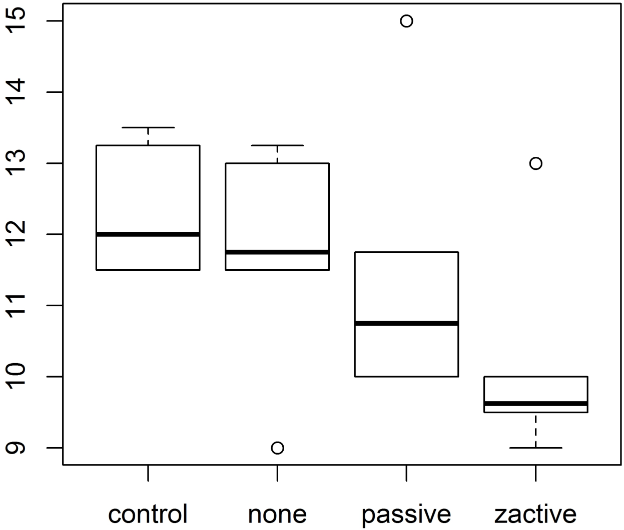
- identical conclusions existence of effect
- effect estimates relative close
- CI differ moderately

Outline

Age of Walking of children

- Age of walking in months
- Conditions: control (5), none (6), passive (6), zactive (6)
- Observe from box-and-wiskers-plot below that there are drastic outliers (deviation from normality)

P.R. Zelazo, N.A. Zelazo, and S. Kolb (1972), Walking in the newborn, Science, 176: 314315.


15
14
13
12
11
10
9
control
none
passive
zactive

P-values from various Normality tests

	control	none	passive	zactive
Shapiro-Wilk	0.1374	0.2939	0.0408	0.0087
Shapiro-Francia	0.1909	0.2116	0.0330	0.0084
Anderson-Darling				
Cramer-von Mises				
Lilliefors	0.4353	0.1520	0.2560	0.0110

- Some normality tests require at least 7 data points
- Partial non-normality for passive
- Clear non-normality for active

Descriptive Statistics Age of Walking of children

	control	none	passive	zactive
mean	12.35	11.71	11.38	10.12
median	12.00	11.75	10.75	9.62
Huber	12.34	11.92	10.98	9.69
SD	0.96	1.52	1.90	1.45
MAD	0.74	1.11	1.11	0.37
IQR	1.30	0.93	1.07	0.32

Remark: MAD and IQR adapted for normal distribution

- small within group differences in mean, median, Huber mean
- considerable within group differences in SD, MAD, IQR

Analysis of Variance

```
> anova(mod1)
Analysis of Variance Table

Response: x
          Df Sum Sq Mean Sq F value Pr(>F)
fac        3 14.778  4.9259  2.1422 0.1285
Residuals 19 43.690  2.2995
```

- No significant difference in Mean between groups

Kruskal-Wallis Rank Sum Test

```
> kruskal.test(x ~ fac, data=dfa)

        Kruskal-Wallis rank sum test

data:  x by fac
Kruskal-Wallis chi-squared = 6.8805, df = 3,
 p-value = 0.0758
```

- Null hypothesis of no effect not rejected!
- Note small sample size
- Neither CI nor effect size!

Robust Estimation of Linear model

```
> mod1 <- lmrob(x ~ fac, data=dfa)
> summary(mod1)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.3317     0.4403  28.010  < 2e-16 ***
facnone      -0.3814     0.7744  -0.492 0.628047
facpassive   -1.6311     0.5994  -2.721 0.013550 *
faczactive   -2.5537     0.5347  -4.776 0.000131 ***
```

- Unequal standard errors more realistic
- Passive group and active group have effect wrt control
- Confirms our intuition
- Better testing due to down weighting of outlying data points

Confidence Intervals of Effects Compared with Control

	Group Effect	2.5 %	97.5 %
(Intercept)	12.33	11.47	13.19
facnone	-0.38	-1.90	1.14
facpassive	-1.63	-2.81	-0.46
faczactive	-2.55	-3.60	-1.51

- CI of passive/ active group effect wrt control does not contain zero