Stat 205: Introduction to Nonparametric Statistics

Lecture 06: Linear Model, ANOVA

Instructor David Donoho; TA: Yu Wang

Nonparametric Test of Specified Linear Fit Single Predictor x

- Pairs (Y_i, X_i) and model

$$Y_i = bX_i + e_i$$

- $H_0 : b = b_0$; (e_i) stochastically independent of (X_i).
- Deviations

$$D_i(b) = Y_i - b \cdot X_i$$

- Theil's test Statistic

$$C(b) = \sum_{i<j} \text{sign}(D_i(b) - D_j(b))$$

- Under H_0, $C(b_0)$ is distribution-free.

Properties of Theil's Test Statistic

- Under H_0: (X_i) formally independent of $(Y_i - b_0 X_i)$:
  - Distribution-free when (e_i) are iid with strictly monotone CDF.
  - Approximate normality

$$C(b_0) \approx_D N(0, \frac{n(n-1)(2n+5)}{18}), \qquad n \to \infty.$$

- Normalized Statistic:

$$\tilde{C} = \frac{C(b_0)}{\sqrt{Var_0(S(b_0))}} = 3 \cdot \frac{C(b_0)}{\sqrt{n(n-1)(n+5/2)}}$$

- Approximate Level-α two-sided test:

$$\text{Reject } H_0 \text{ if } |\tilde{C}| > \mathfrak{z}_{1-\alpha/2}$$

- Exact distributions, tests by using *permutation inference*.
- Asymptotic Relative efficiency under iid Normal (e_i)

$$\text{ARE}(C(b_0), b_{OLS}|\text{bivariate Normal}) = \rho^2 \cdot e(Wilcoxon, Mean|\text{univariate Normal}).$$

where

$$\rho = \rho_{Pearson}(\{X_{(1)}, \ldots X_{(n)}\}, \{1, \ldots, n\})$$

Nonparametric Linear Fit with Single Predictor x

- Pairs (Y_i, X_i) and model

$$Y_i = a + bX_i + e_i$$

- Pairwise Slopes

$$b_{ij} = \frac{Y_i - Y_j}{X_i - X_j}$$

- Theil's Slope Estimate

$$\hat{b} = \text{median}_{i<j} b_{ij}$$

- Ordered Pairwise Slopes, $N = \binom{n}{2}$

$$b^1 \le b^2 \le \cdots \le b^{N-1} \le b^N$$

- Theil Confidence Statement

$$b \in (b^{c_-}, b^{c_+}), \qquad c_\pm \approx \frac{N}{2} \pm \mathfrak{z}_{\alpha/2} \sqrt{\frac{n(n-1)(n+2/5)}{9}}$$

- Intercept estimate (recall $D_i(b) = Y_i - bX_i$):

$$\hat{a} = \text{median}_i D_i(\hat{b}).$$

Properties of Theil's Slope Estimate

- Affine-Equivariance
  - $X_i^{'} = a' + b' X_i$ *affine transformation*
  - $\hat{b}_{Theil}(\{(Y_i, X_i^{'})\}) \cdot b' = \hat{b}_{Theil}(\{(Y_i, X_i)\})$
- Relation to Kendall τ

$$\tau_K(\{(D_i(\hat{b}_{Theil}), X_i)\}) \approx 0.$$

- High-breakdown

$$\text{breakdown fraction}(\hat{b}_{Theil}) = 0.29\%.$$

- High Asymptotic Relative Efficiency at standard normal ≈ 0.91.
- Applications
  - Astronomy (generalization to censored data)
  - Environmental data
  - Climatology
  - Environmental Monitoring

Example: Granato (2006), 1

Kendall-Theil Robust Line (KTRLine—version 1.0)—A Visual Basic Program for Calculating and Graphing Robust Nonparametric Estimates of Linear-Regression Coefficients Between Two Continuous Variables

By Gregory E. Granato

Chapter 7
Section A, Statistical Analysis,
Book 4, Hydrologic Analysis and Interpretation

Example: Granato (2006), 2

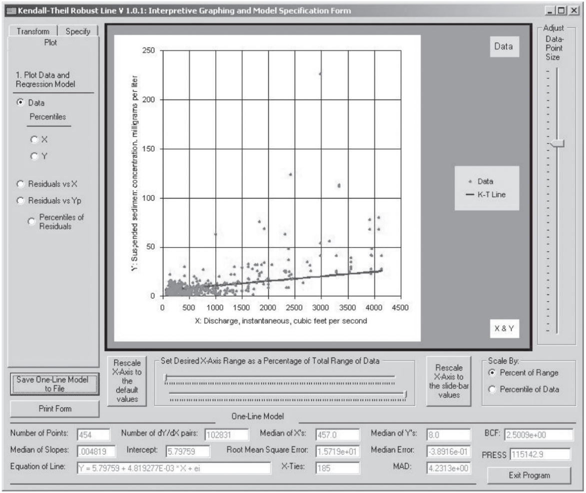
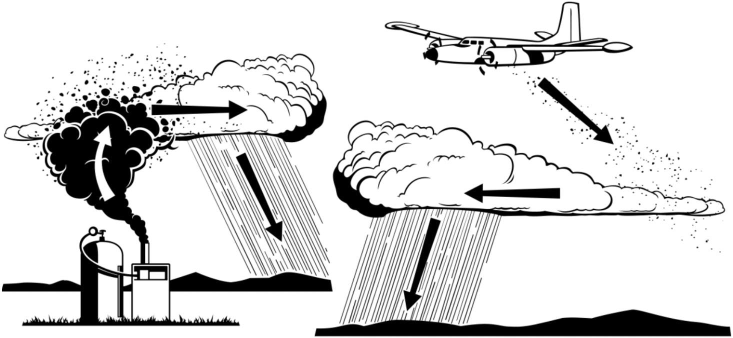


Figure 8. Example of the Kendall-Theil Robust Line Interpretive Graphing and Model Specification Form with the plot menu selected.

Example: Cloud Seeding



- Experiment in Australia's Snowy Mountains
- Measure Effect of Cloud Seeding on Rainfall

Smith (1967)

Cloud Seeding, 2

- T: Rainfall in Target Area
- Q: Rainfall in Control Area
- $[T/Q]$ Rainfall ratio
- Double Ratio

$$y_i = \frac{[T/Q][Seeded]}{[T/Q][Unseeded]}$$

- x_i: Years seeded so far $1 \leq x_i \leq 5$
- Slope $b_0 =$ effect of more years of seeding previously.
- $b_0 = 0$ is theory of no effect

Cloud Seeding, 3

- T: Rainfall in Target Area; Q: Rainfall in Control Area
- $[T/Q]$ Rainfall ratio
- Double Ratio

$$y_i = \frac{[T/Q][Seeded]}{[T/Q][Unseeded]}$$

- x_i: Years seeded so far $1 \le x_i \le 5$

x_i	y_i
1	1.26
2	1.27
3	1.12
4	1.16
5	1.03

Table: Double Ratio in Snowy Mountains Seeding Experiment

Calculation of Test Statistic

R command:

```
theil(x,y, beta.0=0, type="l")
```

When $b_0 = 0$, $D_i(0) = Y_i$; hence

$$C(0) = \sum_{i<j} \text{sign}(Y_i - Y_j) = -6; \qquad \tilde{C} = -0.6.$$

exact p-value for one sided alternative $b_0 < 0$: $p = 0.117$

Alternate Approach

```
ken = cor.test(year,doubleRatio,method="kendall",alternative = "two.sided")
ken$p.value
```

```
## [1] 0.2333333
```

No evidence for cloud seeding impacting rainfall

exact p-value for two-sided alternative $b_0 \neq 0$: $p = 0.234$

Example: Engel Data


```
> library(Rfit)
> data(engel)
> plot(engel)
> abline(rfit(foodexp~income,data=engel))
> abline(lm(foodexp~income,data=engel),lty=2)
> legend("topleft",c('R','LS'),lty=c(1,2))
```

The command **rfit** obtains robust R estimates for the linear regression models, for example (4.1). To examine the coefficients of the fit, use the **summary** command. Critical values and p-values based on a Student t distribution with $n-2$ degrees of freedom recommended for inference. For this example, **Rfit** used the t-distribution with 233 degrees of freedom to obtain the p-value.

```
> fit<-rfit(foodexp~income,data=engel)
> coef(summary(fit))
```

```
               Estimate  Std. Error   t.value       p.value
(Intercept) 103.7667620 12.78877598  8.113893  2.812710e-14
income        0.5375705  0.01150719 46.716038 2.621879e-120
```

Example: Engel Data, 2

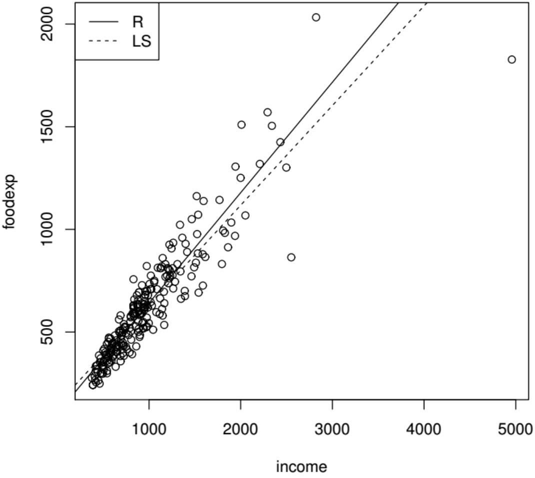
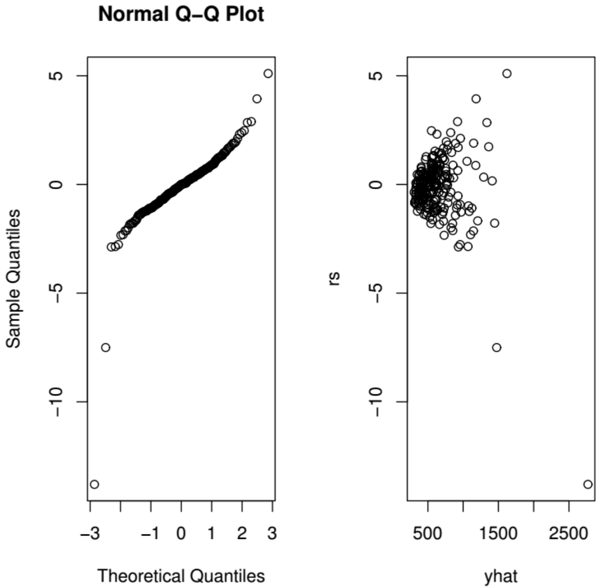


FIGURE 4.1
Scatterplot of Engel data with overlaid regression lines.

Example: Engel Data, 3



Aside: Ladder of Transformations, 1

LADDER OF POWERS
(modified from Velleman and Hoaglin, 1981; Helsel and Hirsch, 2002)

Use	Power	Transformation	Name	Comment
		• • •		higher powers can be used
for (-) skewness	3	x^3	cube	
	2	x^2	square	
———	1	x	original units	no transformation
	1/2	$\sqrt{x}$	square root	commonly used
for (+) skewness	1/3	$\sqrt[3]{x}$	cube root	commonly used
	0	$\log(x)$	logarithim	commonly used; holds the place of x^0
	-1/2	$-1/\sqrt{x}$	reciprocal root	minus sign preserves order of observations
	-1	$-1/x$	reciprocal	
	-2	$-1/x^2$	reciprocal square	
		• • •		lower powers can be used

Figure 5. The ladder of powers for use in transforming the independent (X) and(or) dependent (Y) variables to improve a regression model. (Modified from Helsel and Hirsch, 2002.) All powers except for the reciprocal root and reciprocal square are available in the Kendall-Theil Robust Line software. The line separates transformations for negative (-) and positive (+) skewness.

Aside: Ladder of Transformations, 2

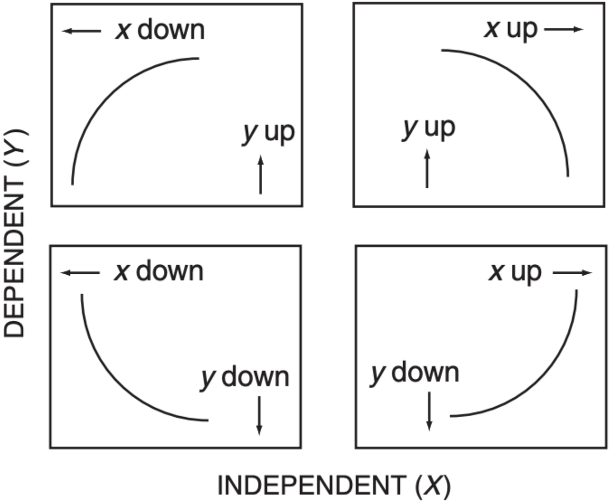
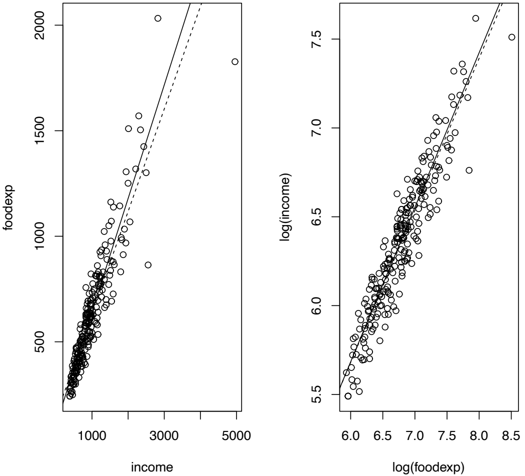


Figure 6. The bulging rule for transforming curvature to linearity. (Modified from Helsel and Hirsch, 2002.)

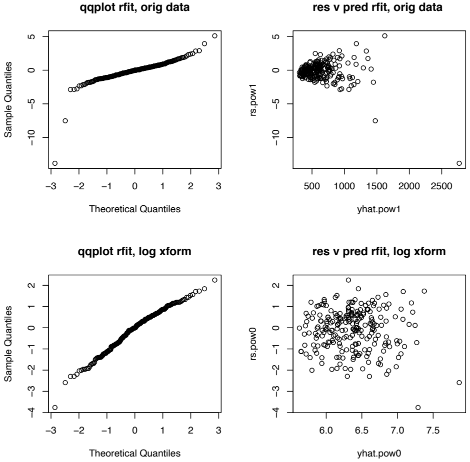
Example: Transformed Engel Data, 1



Example: Transformed Engel Data, 1

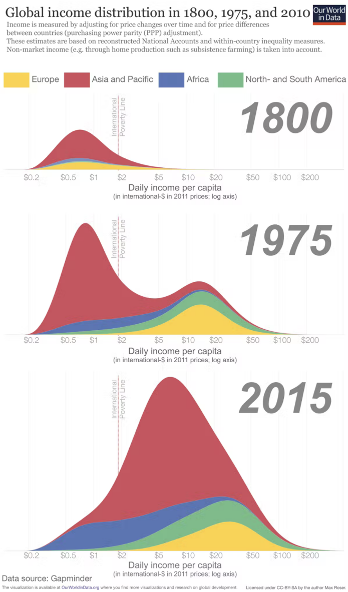

```
data(engel)
par(mfrow=c(1,2))
plot(engel)
abline(rfit(foodexp ~ income,data=engel))
abline(lm(foodexp ~ income,data=engel),lty=2)
plot(log(engel),ylab="log(income)",xlab="log(foodexp)")
abline(rfit(log(foodexp) ~ log(income),data=engel))
abline(lm(log(foodexp) ~ log(income),data=engel),lty=2)
```

Example: Transformed Engel Data, 3



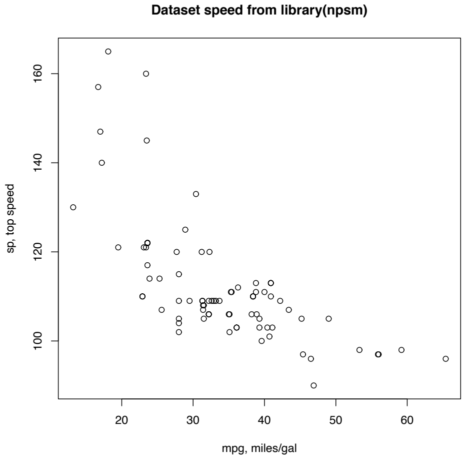
Example: Hans Rosling's argument

5. Global income inequality has gone down



Max Roser, CC BY-SA

Example: Speed Data



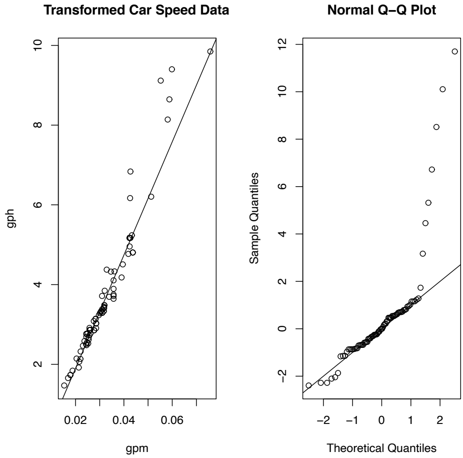
Transforming Speed Data, 1

```
library(devtools)
install_github('kloke/npsm')
library(npsm)
par(mfrow=c(1,1))
plot(sp ~ mpg, data=speed,
     main="Dataset speed from library(npsm)",
     ylab="sp, top speed",
     xlab="mpg, miles/gal")
```

Plotting Speed Data

```
gpm <- 1/speed$mpg
gph <- gpm * speed$sp
rfit(galh ~ gpm) -> fit
par(mfrow=c(1,2))
plot(gpm,gph,main="Transformed Car Speed Data")
abline(coef(fit))
rs <- rstudent(fit)
qqnorm(rs)
abline(0,1)
```

Transforming Speed Data



Multiple Predictor Setting

- $Y = (Y_1, \ldots, Y_n)^T$ $n \times 1$ column vector of responses
- $X = (X_{ij})$ $n \times p$ *centered* matrix of predictors.
- $\mathbf{1}$ $n \times 1$ column vector of ones
- Linear Model

$$Y = 1\alpha + X\beta + e$$

- Here $\alpha \in \mathbf{R}$ viewed as different kind of parameter than $\beta \in \mathbf{R}^p$

Rank-Based Estimator

- Would-be coefficients b ($p \times 1$); would-be residuals

$$v(b) = Y - Xb,$$

  (note these are not necessarily centered)
- Normalized ranks $R_i v = Rank(v_i(b)|(v_j(b)))/(n+1)$
- Rank Discrepancy (convex!)

$$\|Y - Xb\|_\phi = \sum_{i=1}^{n} \phi(R_i v) v_i$$

  - $\sum_{i=1}^{n} \phi(\frac{i}{n+1}) = 0$
  - $\phi(t)$ nondecreasing.
- Examples:
  - Sign Scores: $\phi(u) = \text{sign}(u - 1/2)$
  - Wilcoxon Scores: $\phi(u) = \sqrt{12} \cdot (u - 1/2)$
- **Rank estimator** (Jaeckel, Jureckova, Hettmansperger-McKean)

$$\hat{\beta} = \text{argmin}_b \|Y - Xb\|_\phi .$$

Properties of Rank-Based Estimator

- Properties: convex objective, can be solved!
- Gradient $\nabla_b \|Y - Xb\|_\phi = X^T\phi(Rv(b))$; minimized where:

$$0 = \nabla_{\hat{\beta}} \|Y - X\hat{\beta}\|_\phi \Longrightarrow 0 = X^T\phi(R(Y - X\hat{\beta}))$$

- Asymptotic Normality:

$$\hat{\beta}_\phi \sim_{approx} N(\beta, \tau_\phi^2 \cdot (X'X)^{-1})$$

  Note: τ_ϕ is *not* Kendall's τ_K. [1]
- Compare standard least squares

$$\hat{\beta}_{ls} \sim_{approx} N(\beta, Var(e_i) \cdot (X'X)^{-1})$$

- Asymptotic Relative Efficiency

$$\text{ARE}(\hat{\beta}_\phi, \hat{\beta}_{ls}|F) = \frac{Var(e_i)}{\tau_\phi^2}$$

- (Asymptotic) Standard Errors:

$$se([\hat{\beta}_\phi]_j) = \tau_\phi \cdot [(X'X)^{-1}]_{jj}$$

- Pro-Forma t-statistics

$$t([\hat{\beta}_\phi]_j) = \frac{[\hat{\beta}_\phi]_j}{se([\hat{\beta}_\phi]_j)}$$

[1] τ_ϕ defined in (3.19) in Kloke and McKean.

Inference for Nested Linear Models

- $H_0 : M\beta = 0$ vs $H_0 : M\beta \neq 0$.
- Wald Test Statistic

$$\frac{Q(M\hat{\beta}_\phi; M'(X'X)^{-1}M)/dim(span(M))}{\tau_\phi^2} > F_{1-\alpha,q,n-p-1}$$

$Q(m, S) \equiv m' S^{-1} m.$

- Recall Jaeckel Discrepancy

$$D(b) = \|Y - Xb\|_\phi$$

$$D(Full) = \min_b D(b);$$

- Reduced model:

$$X^{Full} = [X^{Red} X^{Extra}], \qquad b^{restricted} \equiv [b^{Red} 0]$$

$$D(Red) = \min_{b^{restricted}} D(b^{restricted});$$

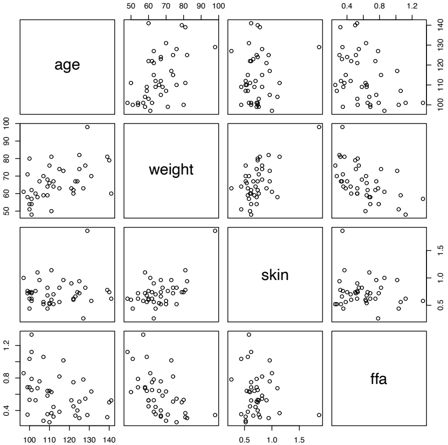
- Drop in dispersion test:

$$RD = D(Red) - D(Full) \qquad (\geq 0)$$

- Significance of Drop in Dispersion

$$F_\phi \equiv \frac{RD/q}{\tau_\phi/2} > F_{1-\alpha,q,n-p-1}$$

Example: Free Fatty Acid data



```
> print(cor(ffa,method="spearman"),digits=3)
           age weight    skin     ffa
age     1.0000  0.504  0.0428 -0.4168
weight  0.5041  1.000  0.3852 -0.6032
skin    0.0428  0.385  1.0000 -0.0102
ffa    -0.4168 -0.603 -0.0102  1.0000
> print(cor(ffa,method="pearson"),digits=3)
          age weight   skin    ffa
age     1.000  0.488  0.101 -0.378
weight  0.488  1.000  0.566 -0.542
skin    0.101  0.566  1.000 -0.149
ffa    -0.378 -0.542 -0.149  1.000
```

```
>  rfitF <-rfit(ffa ~ age+weight+skin,data=ffa)
> rfitR <-rfit(ffa ~weight, data=ffa)
> drop.test(rfitF,rfitR)

Drop in Dispersion Test
F-Statistic     p-value
     2.1735      0.1281
> print(summary(rfitF))
Call:
rfit.default(formula = ffa ~ age + weight + skin, data = ffa)

Coefficients:
              Estimate Std. Error t.value   p.value
(Intercept)  1.4905899  0.2676129  5.5699 2.401e-06 ***
age         -0.0011337  0.0026178 -0.4331 0.6674769
weight      -0.0153484  0.0038216 -4.0163 0.0002779 ***
skin         0.2747982  0.1333516  2.0607 0.0464133 *
---

Multiple R-squared (Robust): 0.3773118
Reduction in Dispersion Test: 7.47326 p-value: 0.00049
```

```
> lfitF <-lm(ffa ~ age+weight+skin,data=ffa)
> lfitR <-lm(ffa ~weight, data=ffa)
> print(summary(lfitF))

Call:
lm(formula = ffa ~ age + weight + skin, data = ffa)

Residuals:
     Min       1Q   Median       3Q      Max
-0.24277 -0.17080 -0.04435  0.10698  0.59315

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.702428   0.326988   5.206 7.44e-06 ***
age         -0.002101   0.003269  -0.643  0.52441
weight      -0.015246   0.004773  -3.194  0.00286 **
skin         0.204574   0.166541   1.228  0.22706
---

Residual standard error: 0.2153 on 37 degrees of freedom
Multiple R-squared:  0.3379,Adjusted R-squared:  0.2842
F-statistic: 6.295 on 3 and 37 DF,  p-value: 0.001467

> anova(lfitF,lfitR)
Analysis of Variance Table

Model 1: ffa ~ age + weight + skin
Model 2: ffa ~ weight
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     37 1.7158
2     39 1.8295 -2   -0.1137 1.2259 0.3051
```

```
> lfitM <-lm(ffa ~ weight+skin,data=ffa)
> anova(lfitM,lfitR)
Analysis of Variance Table

Model 1: ffa ~ weight + skin
Model 2: ffa ~ weight
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     38 1.7350
2     39 1.8295 -1  -0.09455 2.0709 0.1583
> rfitM <-rfit(ffa ~ weight+skin,data=ffa)
> drop.test(rfitM,rfitR)

Drop in Dispersion Test
F-Statistic     p-value
   4.086830    0.050302
```

Robust analysis essentially can reject weight-only model in favor of weight+skin.

Least-squares analysis clearly cannot.

One-Way ANOVA, 1

- Study would-be effect of single factor on response
- Factor varies through k levels

i	$j=1$	$j=2$	$\ldots$	$j=k$
1	Y_{11}	Y_{12}	$\ldots$	Y_{1k}
2	Y_{21}	Y_{22}	$\ldots$	Y_{2k}
$\ldots$				
n_1	$Y_{n_1 1}$	$Y_{n_1 2}$	$\ldots$	$Y_{n_1 k}$
$\ldots$				
n_k		$Y_{n_k 2}$	$\ldots$	$Y_{n_k k}$
$\ldots$				
n_2		$Y_{n_2 2}$		

Example of ragged array where $n_1 < n_k < n_2$.

Assumptions enabling inference

- Standard *randomized design*, n subjects randomly selected from reference population
- n_j randomly assigned to treatment j, $j = 1, \ldots, k$
- Y_{ij} response of i-th individual to j-th treatment; $i = 1, \ldots, n_j$.
- **Assumptions**
  - Independence of responses
  - Treatment induces shift in location

Rank test in 1-way analysis of variance

- Total sample size $n = \sum_{j=1}^{k} n_j$.
- Rank R_{ij} of response Y_{ij} among all n observations; ranked without respect to treatment status
- $R_{\cdot j}$ average rank of j-th treatment group
- Kruskal-Wallis statistic

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} n_j (R_{\cdot j} - \frac{n+1}{2})^2$$

- Null hypothesis: all observations iid w/o regard to treatment group
- Distribution-free under null hypothesis; exact distribution available by permutation inference.
- Approx χ^2 distributed with $k-1$ degrees of freedom.

Motivation for χ^2 approximation

- Kruskal-Wallis statistic

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} n_j (R_{\cdot j} - \frac{n+1}{2})^2$$

- Derivation: under null hyp, each rank is a random sample without replacement from $1, \ldots, n$

$$E_0(R_{ij}) = \frac{n+1}{2}.$$

$$Var_0(R_{i,j}) = (n^2 - 1)/12$$

- The mean rank in a group has a variance $\approx 1/n_j$ as large as any individual rank:

$$Var_0(\bar{R}_{\cdot j}) = n_j^{-1} Var_0(R_{i,j})$$

Define $Z_j \equiv \sqrt{n_j}(\bar{R}_{\cdot j} - \frac{n+1}{2})/\sqrt{Var_0(\bar{R}_{\cdot j})}$; it is approximately standardized; since $\sum_j Z_j \equiv 0$ the vector $(Z_j)_{j=1}^k$ has only $k-1$ degrees of freedom.

- Kruskal-Wallis statistic is approximately the sum of k standardized statistics, squared:

$$H = \sum_{j=1}^{k} (Z_j)^2$$

- Approx χ^2 distributed with $k-1$ degrees of freedom.

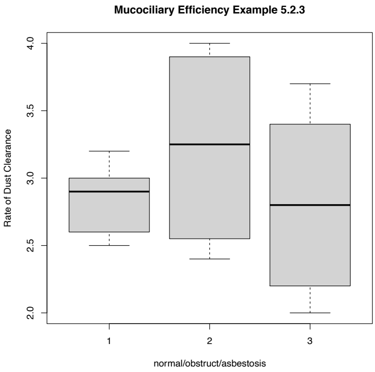
$$\text{Reject } H_0 \text{ for } H \gtrsim (k-1) + \sqrt{2(k-1)}\mathfrak{z}_{1-\alpha}$$

Example of Kruskal-Wallis test

Example 5.2.3. Mucociliary Efficiency
Efficiency of self-clearing mechanism of respiratory tract
Three groups:

- Normal subjects,
- Subjects with obstructive airway disease, and
- Subjects with asbestosis

Responses: measurements of clearance half-lives
Sample Sizes: $n_1 = n_3 = 5$ and $n_2 = 4$
Null hypothesis: no difference between class-conditional distributions.


Mucociliary Efficiency Example 5.2.3
Rate of Dust Clearance
2.0
2.5
3.0
3.5
4.0
1
2
3
normal/obstruct/asbestosis

```
> normal = c(2.9,3.0,2.5,2.6,3.2)
> obstruct = c(3.8,2.7,4.0,2.4)
> asbestosis = c(2.8,3.4,3.7,2.2,2.0)
> x = c(normal,obstruct,asbestosis)
> g = c(rep(1,5),rep(2,4),rep(3,5))
> boxplot( x ~ g,main="Mucociliary Efficiency Example 5.2.3",xlab="normal/obstruct/asbestosis",y
> test = kruskal.test(x,g)
> print(test,digits=5)

Kruskal-Wallis rank sum test

data:  x and g
Kruskal-Wallis chi-squared = 0.771, df = 2, p-value = 0.68
```

Formal Hypotheses

- Distribution of responses $Y_{ij} \sim_{iid} F_j$, $e_{ij} \sim_{iid} F$.

$$F_j(t) = F(t - \mu_j), \qquad -\infty < t < \infty.$$

- Null hypothesis of *no difference*

$$H_0 : \mu_1 = \cdots = \mu_k;$$

$$H_0 : \Delta_{21} = \Delta_{31} = \cdots = \Delta_{k1} = 0.$$

- Alternative of *some difference*

$$H_A : \mu_1, \ldots, \mu_k \text{ not all equal }.$$

$$H_0 : max_j|\Delta_{j1}| > 0.$$

Model Parametrizations, 1

- One-way layout Model

$$Y_{ij} = \mu_i + e_{ij}, \qquad i = 1, \ldots, n_j; \qquad j = 1, \ldots, k.$$

  - μ_i location
  - $e_{ij} \sim_{iid} F$

- Alternate 'reference-level' Parametrization (used by R)

$$Y_{ij} = \mu_1 + \Delta_{j1} + e_{ij}, \qquad i = 1, \ldots, n_j; \qquad j = 1, \ldots, k.$$

  - $\Delta_{j1} \equiv \mu_j - \mu_1$
  - Reference level μ_1

Model Parametrizations, 2

Linear Model Parameterization

$$vec(Y) = X\beta + vec(e)$$

- $vec(Y)$, $vec(e)$ are $n \times 1$ column vectors indexed by pairs (i, j) taken from the array Y in row-major order:
$$(1,1), (2,1), \ldots, (n_1, 1), (1,2), (2,2), \ldots, (n_2, 2), \ldots, (n_k, k)$$
- X is $n \times k$ matrix with row id's given by pairs (i, j), $i = 1, \ldots, n_j$.
$$X_{(i,1),1} = 1; \qquad X_{(i,j),\ell} = 1_{\{\ell=j\}}, \qquad \ell = 2, \ldots, k$$
- $\beta = (\beta_\ell)$ is $k \times 1$ vector.
$$\beta_1 = \mu_1, \qquad \beta_j = \Delta_{j,1}, \qquad j = 2, \ldots, k.$$

$$\begin{aligned} vec(Y)_{(i,j),1} &= \sum_{\ell=1}^{k} X_{(i,j),\ell}\beta_\ell + vec(e)_{(i,j),1} \\ &= \mu_1 + \Delta_{j,1} + e_{ij} \end{aligned}$$

- Reduced model: μ_1 arbitrary, $\Delta_{j,1} = 0$, $j = 2, \ldots, k$.
- Full Model: μ_1 arbitrary, $max_{j=2,\ldots,k}|\Delta_{j,1}|$.
- Reduction in dispersion $RD_\phi = D_\phi(Red) - D_\phi(Full)$.
- Drop in dispersion statistic

$$F_\phi = \frac{RD_\phi/(k-1)}{\hat{\tau}_\phi/2}$$

- $\hat{\tau}_\phi$ estimate of scale.
- Specifically, for Wilcoxon rank scores write W subscripts, not ϕ.

$$F_W = \frac{RD_W/(k-1)}{\hat{\tau}_W/2}$$

Multiple Comparisons

$(1-\alpha)\cdot 100\%$ CI for effect $\mu_j - \mu_{j'}$:

$$\hat{\Delta}_{jj'} \pm \mathfrak{z}_{\alpha/2} \cdot \hat{\tau} \cdot \sqrt{\frac{1}{n_j} + \frac{1}{n_j'}}$$

- There are $\binom{k}{2}$ such CI's.
- Expected number of failures to cover: $\binom{k}{2} \cdot \alpha$.
- This includes failures to cover 0, when 0 is true.
- Familywise error rate FWER= $P\{$ one of the CI's does not cover 0 $|H_0\}$
- If $k \gg 14$ and $\alpha = .05$ we expect several (or many) failures.
- Tukey-Kramer rule instead adjusts CI lengths so that FWER is α. (in some cases; in others, approximately α)

```
> robfit= with(quail,oneway.rfit(ldl,treat))
> robfit
Call:
oneway.rfit(y = ldl, g = treat)

Overall Test of All Locations Equal

Drop in Dispersion Test
F-Statistic     p-value
   3.916944    0.016394


    Pairwise comparisons using Rfit

data:  ldl and treat

  1      2      3
2 0.0046 -      -
3 0.6315 0.0157 -
4 0.5599 0.0243 0.9069

P value adjustment method: none
```

Wilcoxon reduction in dispersion $F_W = 3.92$ w/ p-value 0.016

```
> summary(robfit,method="tukey")

Multiple Comparisons
Method Used  tukey

  I J Estimate  St Err Lower Bound CI Upper Bo
1 1 2      -25 8.26704       -47.29541      -2
2 1 3       -4 8.26704       -26.29541      18
3 1 4       -5 8.49358       -27.90636      17
4 2 3      -21 8.26704       -43.29541       1
5 2 4      -20 8.49358       -42.90636       2
6 3 4        1 8.49358       -21.90636      23
```

Drug compounds I and II are declared different by Tukey/Kramer
after accounting for multiple comparisons

Two-Way ANOVA, 1

- Study would-be effect of two factors A, B (say) on response
- Factor A varies through a levels, B through b levels

$$Y_{ijk} = \mu_{ij} + e_{ijk}, \qquad i = 1, \ldots, a; \quad j = 1, \ldots, b; k = 1, \ldots, n_{ij}$$

Example:

- Serum Luteinizing Hormone Data:
  Y_{ijk} is nanograms/ml of luteinizing hormone in blood
- 2×5 factorial design;
  effect of *light* on release of *luteinizing hormone*.
  - $a = 2$ light regimes (24-hour light vs 14 on, 10 off)
  - $b = 5$ dosage levels of LRF
- $n_{ij} = 6$ replicates (mice) per treatment combination

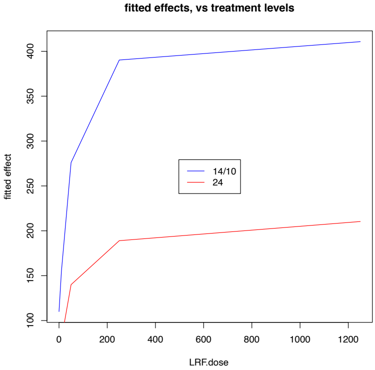
```
> head(serumLH)
  serum light.regime LRF.dose
1    72     Constant        0
2    64     Constant        0
3    78     Constant        0
4    20     Constant        0
5    56     Constant        0
6    70     Constant        0
> tail(serumLH)
   serum light.regime LRF.dose
55   296 Intermittent     1250
56   545 Intermittent     1250
57   630 Intermittent     1250
58   418 Intermittent     1250
59   396 Intermittent     1250
60   227 Intermittent     1250
>
```

```
> raov(serum~ light.regime+LRF.dose,data=serumLH)

Robust ANOVA Table
                     DF        RD   Mean RD        F p-value
light.regime          1 1642.3333 1642.3333 58.28334 0.00000
LRF.dose              4 3027.6735  756.9184 26.86162 0.00000
light.regime:LRF.dose  4  451.4553  112.8638  4.00533 0.00678
> summary(aov(serum~ light.regime+LRF.dose+light.regime*LRF.dose,data=serumLH))
                     Df Sum Sq Mean Sq F value   Pr(>F)
light.regime          1 242189  242189  40.223 6.41e-08 ***
LRF.dose              4 545549  136387  22.652 1.02e-10 ***
light.regime:LRF.dose  4  55099   13775   2.288   0.0729 .
Residuals            50 301055    6021
---
>
```

Conclusions *differ*

Interaction *significant* by Rank AOV; *not significant* by Usual AOV



Dose-Response less with 24H light vs 14H light

Summary

- Transformations can be important (transformative!)
- Rank-based analysis can be done for univariate and multivariate regression
- Similar UX to classical methods + outlier-resistant + distribution-free
- Can be more sensitive to detect subtle effects.
- One-Way Layout/Kruskal-Wallis/Linear Model
- Two-Way Layout/Linear Model

Generalizations: k-Way layout.