Multi-Way Analysis of Variance

Statistics II (LIX002X05)

University of Groningen, Faculty of Arts, Information Science

Wilbert Heeringa

Introduction

- **One-factor ANOVA**:
  is used in order to test whether population means of **three** or more groups differ from each other, when the groups are classified according to **one** factor.
- **Example**: is there a difference in effectivity between **three** of more reading methods for children of the second grade?
- ANOVA = ANalysis Of VAriance. ANOVA tests have test statistics F.
- **Two-factor ANOVA / Multi-way ANOVA**:
  is used when groups are classified according to **two or more** factors.
- **Example**: is there a difference in effectivity between three reading methods for children of the second grade **and** is there also a difference between boys and girls.
- Three advantages!

First advantage

- Experiment Van Bezooijen & Heeringa (2006): measure intuitions of non-linguists about dialects in the Netherlands and Flanders.
- Task: subjects give a dialect distance compared to standard Dutch for each province in the Netherlands and Flanders. 0=no distance, 100=maximal distance.
- Subjects from four regions: North (Groningen, Friesland, Drenthe), East (Overijssel, Gelderland), West (Noord-Holland, Zuid-Holland, Utrecht), South (Noord-Brabant, Limburg).
- We study the ratings the subjects gave for the dialect of the province of Noord-Brabant:
  - Do the ratings differ per region?
  - Do the ratings differ per sex?

First advantage

- Two one-factor ANOVA-tests?
- Research region:

	North	East	West	South
men	16	16	16	16

- Research sex:

	men	women
North	16	16

- In total we need $(4 \times 16 + 2 \times 16)$ - $16 = 80$ subjects.

First advantage

- Two-factor design:

	North	East	West	South	totaal
men	8	8	8	8	32
women	8	8	8	8	32
total	16	16	16	16	64

- The number of subjects per region is the same (16) and per sex even larger (32). Yet, the total number of subjects is smaller (64 instead of 80).
- Therefore, it is more efficient to study several factors at once instead of studying each of them separately.

First advantage

- In our example there are 140 subjects:

	North	East	West	South	totaal
men	11	23	10	25	69
women	12	16	26	17	71
total	23	39	36	42	140

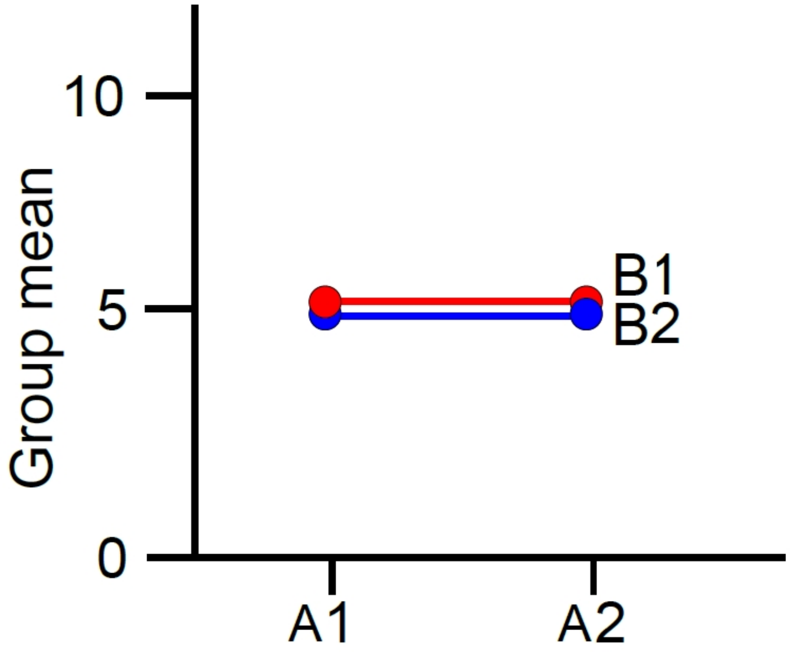
Second advantage

- DATA = FIT + RESIDUAL
- In a one-factor design with factor $region$ the variation due to variation in sex will be assigned to the RESIDUAL part (variation **within** groups).
- In a two-factor ANOVA sex is processed as a second factor, therefore variation in sex will be included in the FIT part (variation **between** groups) of the model.
- Each time the variation can be moved from RESIDUAL to FIT, the variance of the model will decrease and the power of discrimination will increase.
- The pooled variance(=MSE) in a
  - one-factor ANOVA (with factor $region$) is: 45438;
  - two-factor ANOVA (with factors $region$ and sex) is: 42825.

Third advantage

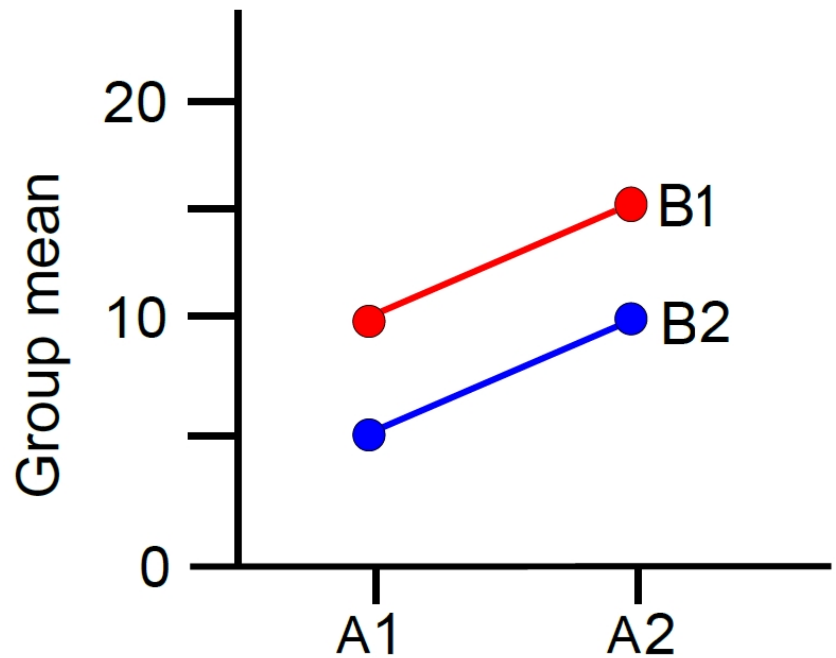
- The interaction between factors can be studied.
- Example: we investigate the reading scores of second graders.
- Two factors:
  - Factor A represents the distinction in reading method, having levels 1 (reading method 1) and 2 (reading method 2).
  - Factor B represents sex, having levels 1 (boys) and girls (2).
- The main effects A and B and the interaction $A \times B$ can be visualized in interaction plots.

Interaction type I



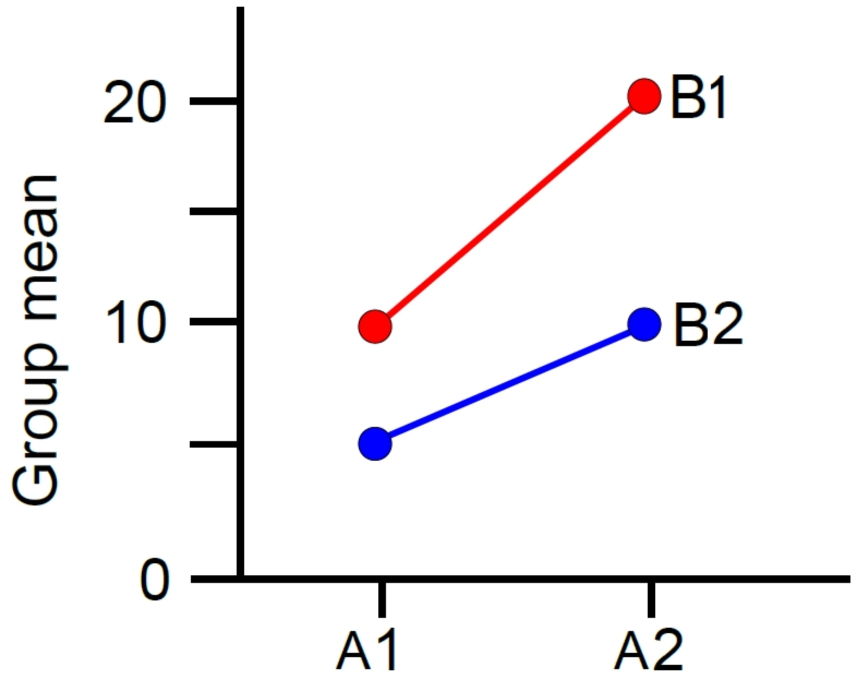
No main effect for factor A, no main effect for factor B, no interaction.

Interaction type II



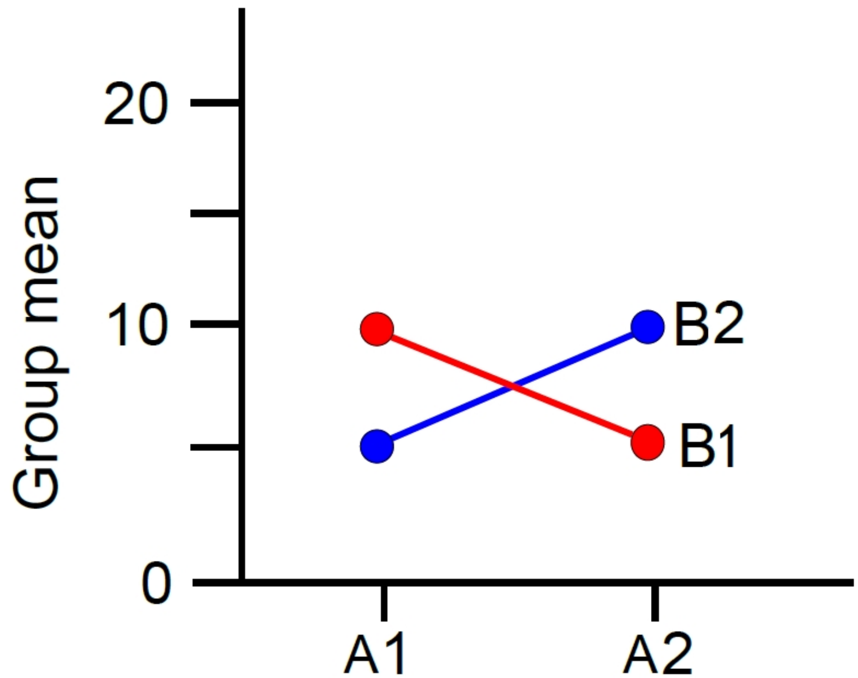
The A_1s have lower averages than the A_2s, therefore, there is a main effect for factor A. The red line lies higher than the blue line, therefore there is a main effect for factor B. There is no interaction, because the two lines run parallel.

Interaction type III



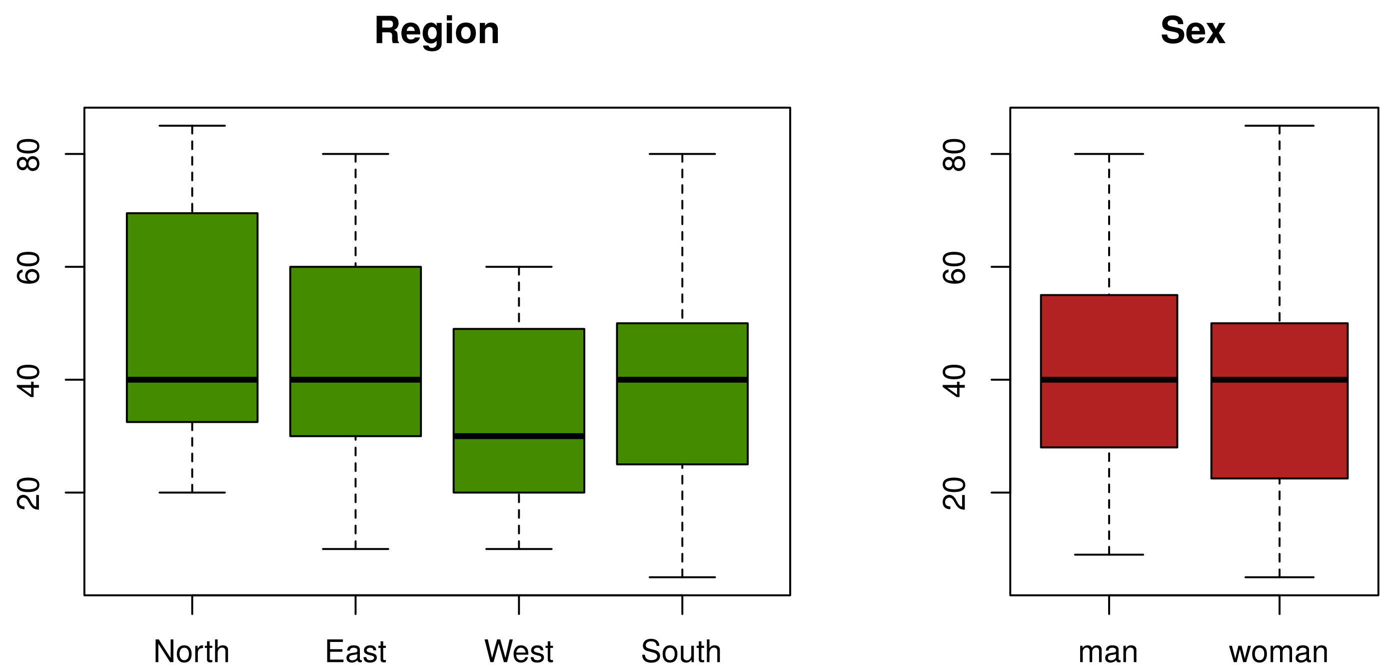
There are main effects for factors A and B. There is also an interaction between A and B, since B_1 and B_2 are closer to each other for A_1 than for A_2. The result is that the red line and the blue line do not run parallel.

Interaction type IV



There is an interaction between factors A and B, but there are no main effects. When the red line and the blue line are taken together, they 'neutralize' each other.

Comparing the groups



The hypotheses of two-factor ANOVA

- Assume we have two factors A and B.
  Factor A gives a division in I groups, and factor B gives a division in J groups.
- Main effect for factor A:
  H_0: $\mu_{A_1} = \mu_{A_2} = \ldots = \mu_{A_I}$
  H_a: not all μ_As are the same
- Main effect for factor B:
  H_0: $\mu_{B_1} = \mu_{B_2} = \ldots = \mu_{B_J}$
  H_a: not all μ_Bs are the same
- Interaction between factor A and factor B:
  H_0: $\mu_{AB_{11}} = \mu_{AB_{12}} = \ldots = \mu_{AB_{IJ}}$
  H_a: not all μ_{AB}'s are the same

The two-factor ANOVA-model

- DATA = FIT + RESIDUAL
- We have independent SRSs with size n_{ij}, one from each of $I \times J$ *normally* distributed populations.
- The population means may u_{ij} differ, but all populations have the **same** standard deviation σ. The μ_{ij}'s and σ are the unknown parameters.
- Assume x_{ijk} is the kth observation belonging to the population defined by level i for factor A and level j for factor B.
- Assume the x_{ijk}s vary around their population mean μ_{ij}:
  $x_{ijk} = \mu_{ij} + \epsilon_{ijk}$, where $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $k = 1, \ldots, n_{ij}$.
- The ϵ_{ijk}s are an SRS taken from the $N(0, \sigma)$ distribution.
- μ_{ij}s: FIT, ϵ_{ijk}s: RESIDUAL

Assumptions

- 1. Independence:
  observations are independent of each other.
- 2. Interval scale:
  the dependent variable is measured on at least an interval scale.
- 3. Normality:
  each sample is drawn from a normally distributed population. Use normal quantile plots and the Shapiro-Wilk test.
- 4. Homogeneity of variance:
  the groups (in our case three groups) have the same variance. Use Levene's test and Hartley's test.

1. Independence

- All the values of the dependent variable $intuitive\ distance$ are independent of each other.

2. Interval scale

- The values of the dependent variable *intuitive distance* are measured on at least the interval scale (namely the ratio scale).

3. Normality

- For each group defined by *region* and *sex* we perform the Shapiro-Wilk test:

region	sex	p value
north	men	0.155
east	men	0.210
west	men	0.211
south	men	0.184
north	women	0.088
east	women	0.370
west	women	0.047
south	women	0.827

- Most p values are larger than 0.05, one p value is nearly 0.05.

4. Homogeneity of variance

- **Homogeneity of variance**: assumption that the spread of scores is roughly the same in different groups of cases.
- Assumption can be tested with **Levene's test** and the **Hartley's Homogeneity of Variance Test**.
- If the result is not significant, homogeneity of variance can be assumed.
- In that case each sample standard deviation is an approximation of σ. We want to combine the standard deviations of the samples to one estimate.

4. Homogeneity of variance

- Results of Levene's test:

```
Levene's Test for Homogeneity of Variance (center = median)
         Df   F value   Pr(>F)
group     7    0.6236   0.7357
        132
```

- We find $p < 0.05$, therefore it looks as if the groups do not have the same variance.

4. Homogeneity of variance

- Variances for each of six groups defined by the two factors:

region	sex	p value
north	men	348.9
east	men	329.9
west	men	323.3
south	men	390.0
north	women	411.2
east	women	349.7
west	women	184.9
south	women	338.7

- Largest variance (411.2) / smallest variance (184.9) $=$ 2.22.

4. Homogeneity of variance

- We perform Hartley's Homogeneity of Variance Test.
- Df is $n - 1$ of smallest group: 26-1=25. Number of groups k=8. $\alpha = 0.05$.
- Find critical value at: `http://www.csulb.edu/~acarter3/course-biostats/tables/table-Fmax-values.pdf`.
- The critical value is 3.12 (actually for $df = 30$). Since $2.22 < 3.12$, we may assume that the groups have the same variances.

Pooled standard deviation

- The **pooled sample variance**

$$s_p^2 = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^{I} \sum_{j=1}^{J} (n_{ij} - 1)}$$

  is an unbiased estimate of σ^2.

- The **pooled standard deviation**

$$s_p = \sqrt{s_p^2}$$

  is the estimate of σ.

Main effects and interactions

- SSA: variation between means of groups distinguished by factor A.
  DFA $= I - 1$: I observations are compared with 1 mean.
  MSA = SSA / DFA.

- SSB: variation between means of groups distinguished by factor B.
  DFB $= J - 1$: J observations are compared to 1 mean.
  MSB = SSB / DFB.

- SSAB: variation between means of groups distinguished by factors A and B.
  DFAB $= (I - 1)(J - 1)$
  MSAB = SSAB / DFAB.

Model, Error and Total

- SSM: model variation = SSA+SSB+SSAB.
  DFM = DFA+DFB+DFAB = $(I \times J) - 1$.
  MSM = SSM / DFM.

- SSE: the deviations among individual observations compared to their group means.
  DFE = $N - (I \times J)$: N observations compared to $I \times J$ sample means.
  MSE = SSE / DFE = s_p^2 (pooled variance).

- SST: total variance = SSM+SSE.
  DFT = DFM+DFE = $N - 1$.
  MST = SST / DST.

Two-factor ANOVA-table

Source	Sum of squares	Degrees of freedom	Mean sum of squares	F
Model	SSM	DFM$=(I \times J)-1$	MSM=SSM/DFM	F=MSM/MSM
A	SSA	DFA$=I-1$	MSA=SSA/DFA	F_A=MSA/MSE
B	SSB	DFB$=J-1$	MSB=SSB/DFB	F_B=MSB/MSE
AB	SSAB	DFAB$=(I-1)\times(J-1)$	MSAB=SSAB/DFAB	F_{AB}=MSAB/MSE
Error	SSE	DFE$=N-(I \times J)$	MSE=SSE/DFE	
Totaal	SST	DFT$=N-1$	MST=SST/DFT	

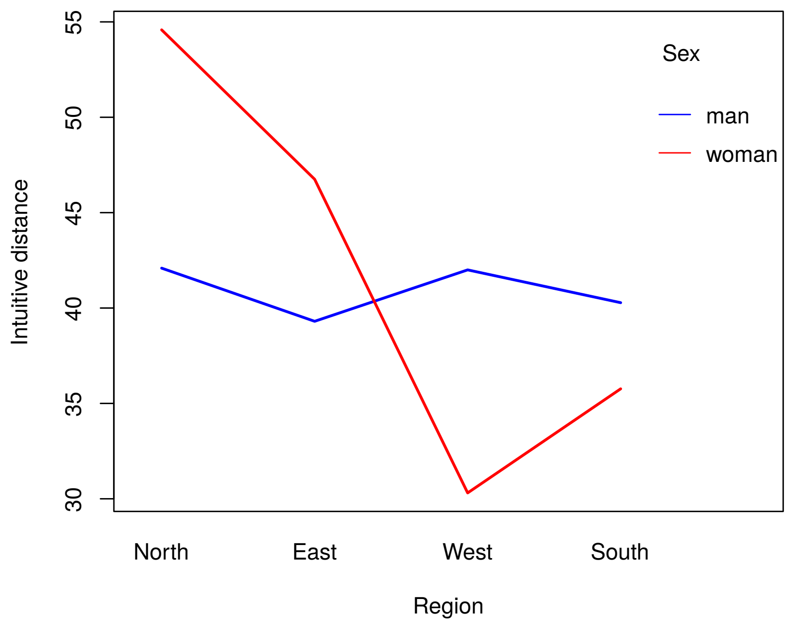
General form of a two-factor ANOVA-table with test statistics F for main effects A and B and interaction A $\times$ B.

ANOVA table

```
             Df Sum Sq Mean Sq F value Pr(>F)
REGIO         3   3517  1172.3   3.613 0.0151 *
SEKSE         1      4     4.3   0.013 0.9086
REGIO:SEKSE   3   2608   869.4   2.680 0.0496 *
Residuals   132  42825   324.4
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Interaction plot

- We found an interaction between the factor $region$ and the factor sex. The graph confirms this:



Effect size

- In a one-factor ANOVA the **determination coefficient** is defined as:

$$R^2 = \frac{SSG}{SST}$$

- In a two-factor ANOVA the **determination coefficient** is defined as:

$$R^2 = \frac{SSM}{SST}$$

- R^2 is also refered to as η^2 (eta squared)

Effect size

- In two factor ANOVA we also want to calculate the η^2 per factor and per interaction.
- However, we cannot divide by SST, since SST represents the variance of all factors together, while we want to keep the factors separated.
- Solution: we calculate the **partial** eta squared per factor and interaction. Partial squared eta for factor A:

$$\eta_p^2 = \frac{SS_A}{SS_A + SSE}$$

Effect size

- In our example:
  - Model:

$$R^2 = \eta^2 = \frac{SSM}{SST} = \frac{6129}{6129 + 42825} = 0.13$$

  - region:

$$\eta^2_{region} = \frac{SS_{region}}{SS_{region} + SSE} = \frac{3517}{2517 + 42825} = 0.08$$

  - sex:

$$\eta^2_{sex} = \frac{SS_{sex}}{SS_{sex} + SSE} = \frac{4}{4 + 42825} = 0.00$$

  - region × sex:

$$\eta^2_{region \times sex} = \frac{SS_{region \times sex}}{SS_{region \times sex} + SSE} = \frac{2608}{2608 + 42825} = 0.06$$

Effect size

- Rule of thumb:

$$\begin{array}{rcll} 0.01 \leq & \eta_p^2 & < 0.06 & \text{small effect} \\ 0.06 \leq & \eta_p^2 & < 0.14 & \text{medium effect} \\ & \eta_p^2 & \geq 0.14 & \text{large effect} \end{array}$$

- The distinction according to region explains 8% of the variation of the measurements, the distinction according to sex explains 0%, and the interaction region$\times$sex explains 6%.

Effect size

- R squared is 0.13, SPSS gives also an adjusted R squared which is 0.08. Use the adjusted R Squared instead of R Squared!
- R squared (or Eta squared) is a estimate of the degree of association for the *sample*, i.e. it estimates only the effect size in the sample. Therefore it is biased.
- An unbiased estimate is Omega squared. See Gray & Kinnear (2012) p. 284 for how to calculate complete omega squared and partial omega squared's for the factors and the interaction.
- The bias of R squared is small (e.g., 65.16% vs. 64.98%). Depending on your situation you can use this. As the sample size gets larger the amount of bias gets smaller.

Multiple comparisons

- Results with the Holm-Bonferroni method:

```
Pairwise comparisons using t tests with pooled SD

      North East  West
East  0.587 -     -
West  0.015 0.170 -
South 0.170 0.587 0.587

P value adjustment method: holm
```

- There is a difference between the North and the West.

Robust alternatives

- Brown & Forsythe's F or Welch's F are not available in SPSS as robust alternatives for multi-way ANOVA, neither can the KruskalWallis test be used.