

Analysis of Variance (ANOVA) in R

Jens Schumacher

June 21, 2007

Die Varianzanalyse ist ein sehr allgemeines Verfahren zur statistischen Bewertung von Mittelwertunterschieden zwischen mehr als zwei Gruppen. Die Gruppeneinteilung kann dabei durch Unterschiede in experimentellen Bedingungen (Treatment = Behandlung) erzeugt worden sein, aber auch durch Untersuchung der gleichen Zielgröße an verschiedenen Orten usw. Allgemein kann man das von der Varianzanalyse umfasste Spektrum an Problemstellungen als Untersuchung des Einflusses einer oder mehrerer nominaler Einflussgrößen auf eine metrische Zielgröße charakterisieren. Zunächst beschäftigen wir uns mit der Situation einer Einflussgröße, der sogenannten einfachen Varianzanalyse (One-way ANOVA). Die Einflussgröße wird als Faktor bezeichnet. Die verschiedenen Werte der Einflussgröße nennt man Stufen des Faktors.

1 Einfache Varianzanalyse (One-way ANOVA)

Beispiel 1.1 *Die Daten stammen aus einem pflanzenphysiologischen Experiment, in dem der Effekt unterschiedlicher Zucker auf das Wachstum von Erbsen untersucht wurde. Die Werte stellen gemessene Längen (in "ocular units") dar. Insgesamt wurden fünf Gruppen untersucht, eine Kontrollgruppe sowie vier Gruppen, denen unterschiedliche Zucker bzw. eine Mischung von zwei Zuckern zugesetzt wurden. In jeder Gruppe wurden 10 Wiederholungen durchgeführt.*

Wiederholung	Gruppen (a = 5)				
	Kontrolle	2% Glukose	2% Fruktose	1% Glukose + 2% Saccharose	1% Fruktose
1	71	57	58	58	62
2	68	58	61	59	66
3	70	60	56	58	65
4	74	59	58	61	63
5	68	62	57	57	64
6	71	60	56	56	62
7	70	60	61	58	65
8	67	57	60	57	65
9	73	59	57	57	62
10	69	61	58	59	67

Zur Analyse dieser Daten laden wir nach dem Start von R die Datei "peas.csv".

```
> peas.data <- read.csv("peas.csv", sep = ";")  
> names(peas.data)
```

```
[1] "group" "length"
```

```
> attach(peas.data)
```

The following object(s) are masked from peas.data (position 3) :

group length

The following object(s) are masked from peas.data (position 4) :

group length

The following object(s) are masked from peas.data (position 5) :

group length

The following object(s) are masked from peas.data (position 6) :

group length

The following object(s) are masked from peas.data (position 7) :

group length

> *peas.data*

	group	length
1	Kontrolle	71
2	Kontrolle	68
3	Kontrolle	70
4	Kontrolle	74
5	Kontrolle	68
6	Kontrolle	71
7	Kontrolle	70
8	Kontrolle	67
9	Kontrolle	73
10	Kontrolle	69
11	2%Glukose	57
12	2%Glukose	58
13	2%Glukose	60
14	2%Glukose	59
15	2%Glukose	62
16	2%Glukose	60
17	2%Glukose	60
18	2%Glukose	57
19	2%Glukose	59
20	2%Glukose	61
21	2%Fruktose	58
22	2%Fruktose	61
23	2%Fruktose	56
24	2%Fruktose	58
25	2%Fruktose	57
26	2%Fruktose	56

27	2%Fruktose	61
28	2%Fruktose	60
29	2%Fruktose	57
30	2%Fruktose	58
31	1%Glukose&1%Fruktose	58
32	1%Glukose&1%Fruktose	59
33	1%Glukose&1%Fruktose	58
34	1%Glukose&1%Fruktose	61
35	1%Glukose&1%Fruktose	57
36	1%Glukose&1%Fruktose	56
37	1%Glukose&1%Fruktose	58
38	1%Glukose&1%Fruktose	57
39	1%Glukose&1%Fruktose	57
40	1%Glukose&1%Fruktose	59
41	2%Saccharose	62
42	2%Saccharose	66
43	2%Saccharose	65
44	2%Saccharose	63
45	2%Saccharose	64
46	2%Saccharose	62
47	2%Saccharose	65
48	2%Saccharose	65
49	2%Saccharose	62
50	2%Saccharose	67

Zunächst fällt auf, dass **R** die Daten nicht in der obigen übersichtlichen Tabellenform, sondern in nur zwei Spalten haben will. Das entspricht der Philosophie von **R** (und vieler anderer Statistik-Programme) - "jede Zeile in der Datenmatrix ist ein Fall". Jeder "Fall" ist in unserem Beispiel eine Erbse, die durch zwei Merkmale gekennzeichnet ist - die Zugehörigkeit zu einer Untersuchungsgruppe und die gemessene Länge.

Gegenstand der Untersuchung bei der Varianzanalyse ist der Zusammenhang zwischen dem nominalskalierten Merkmal "Gruppe" und dem metrischen Merkmal "Länge".

Die zunächst interessierende Fragestellung lautet:

Unterscheiden sich die Mittelwerte des Merkmals "Länge" in den einzelnen Gruppen, d.h., wirkt sich das Zuckerangebot auf das Wachstum aus?

1.1 Beschreibende Statistik

Bevor natürlicherweise vorhandene Unterschiede der Mittelwerte in den einzelnen Gruppen statistisch bewertet werden können, wollen wir zunächst einmal die in den Daten vorliegende Information durch die Berechnung statistischer Kenngrößen zusammenfassen. Dazu verwenden wir die **R** - Funktionen *table()* und *tapply()*. Durch den *table()*- Funktion können wir die Anzahl der in den verschiedenen Gruppen vorliegenden Messwerte überprüfen.

```
> table(group)
```

```
group
1%Glukose&1%Fruktose      2%Fruktose      2%Glukose      2%Saccharose
                10                10                10                10
```

Mit der Funktion *tapply()*-Funktion wird die selbe Funktion gleichzeitig auf durch die Gruppierungsvariable definierte Teildatensätze angewendet. Argumente dieser Funktion sind:

```
tapply(Zielgröße, Gruppierungsvariable, anzuwendende Funktion)
```

Wir verwenden sie zur Berechnung der Mittelwerte und der Standardabweichungen der Gruppen.

```
> tapply(length, group, mean)
```

1%Glukose&1%Fruktose	2%Fruktose	2%Glukose	2%Saccharose
58.0	58.2	59.3	64.1

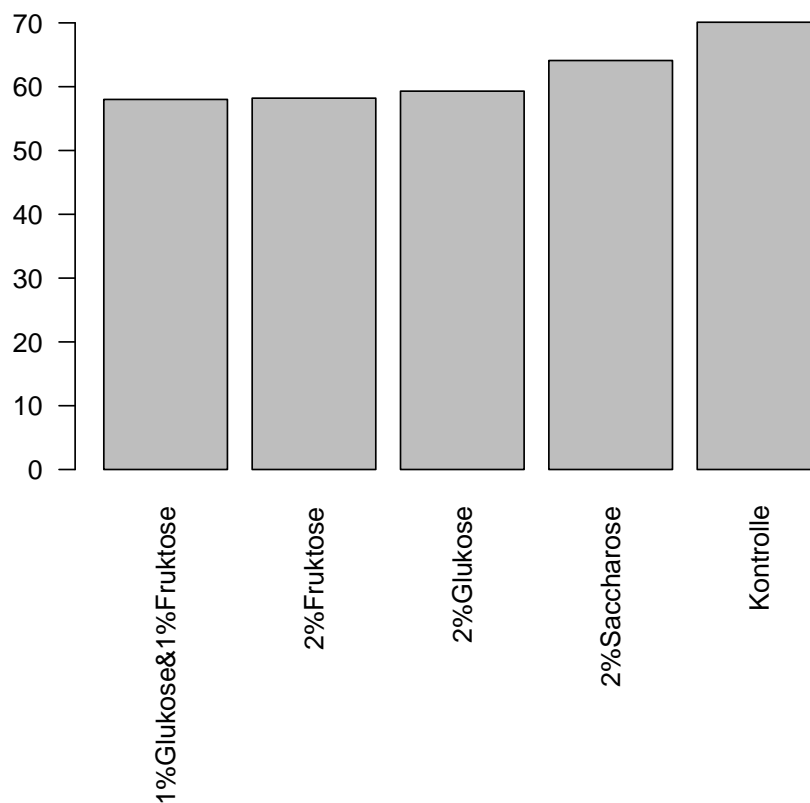
```
> tapply(length, group, sd)
```

1%Glukose&1%Fruktose	2%Fruktose	2%Glukose	2%Saccharose
1.414214	1.873796	1.636392	1.791957

1.2 Graphische Darstellung

Für die graphische Darstellung der Daten gibt es unterschiedliche Möglichkeiten. Zum Beispiel können die Mittelwerte der einzelnen Gruppen in einem Balkendiagramm dargestellt werden.

```
> peas.means <- tapply(length, group, mean)
> par(mar = c(12, 4, 4, 2))
> barplot(peas.means, las = 2)
```



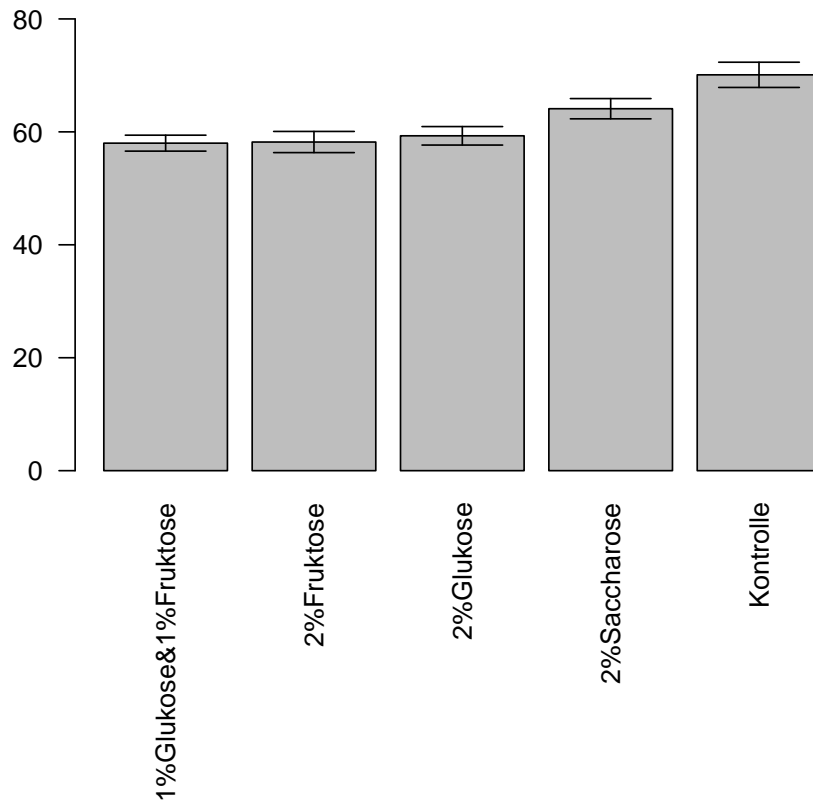
```
# der Grafikparameter mar (margin) wird so gesetzt, dass genug Platz für
die Beschriftung ist
# las=2 (label axis style) bewirkt, dass die Beschriftung senkrecht zur
x-Achse erfolgt
```

Etwas komplizierter wird es, wenn zusätzlich Information zur Variabilität innerhalb der Gruppen dargestellt werden soll.

```

> par(mar = c(12, 4, 4, 2))
> peas.sd <- tapply(length, group, sd)
> x.values <- barplot(peas.means, las = 2, ylim = c(0, 80))
> arrows(x.values, peas.means - peas.sd, x.values, peas.means + peas.sd, angle = 90, code =

```



```

# um die Fehlerbalken an die richtige Stelle zu bekommen, müssen die x-
# Koordinaten gespeichert werden
# # Argumente der arrows()-Funktion
# x1,y1,x2,y2
# angle=Winkel der Pfeilspitze
# code=3: Spitzen an beiden Enden

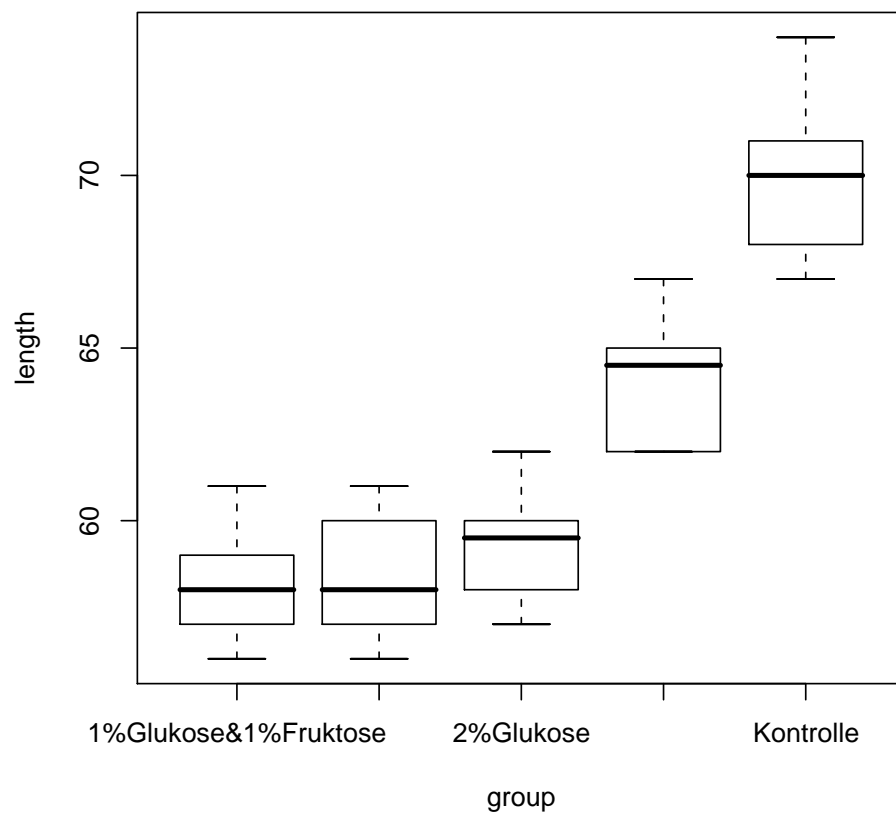
```

Als weitere Steigerung plädiere ich für die Darstellung sogenannter Boxplots oder Box-Whisker-Plots. Diese Darstellung ist die Standarddarstellung in **R** und daher auch am einfachsten zu realisieren.

```

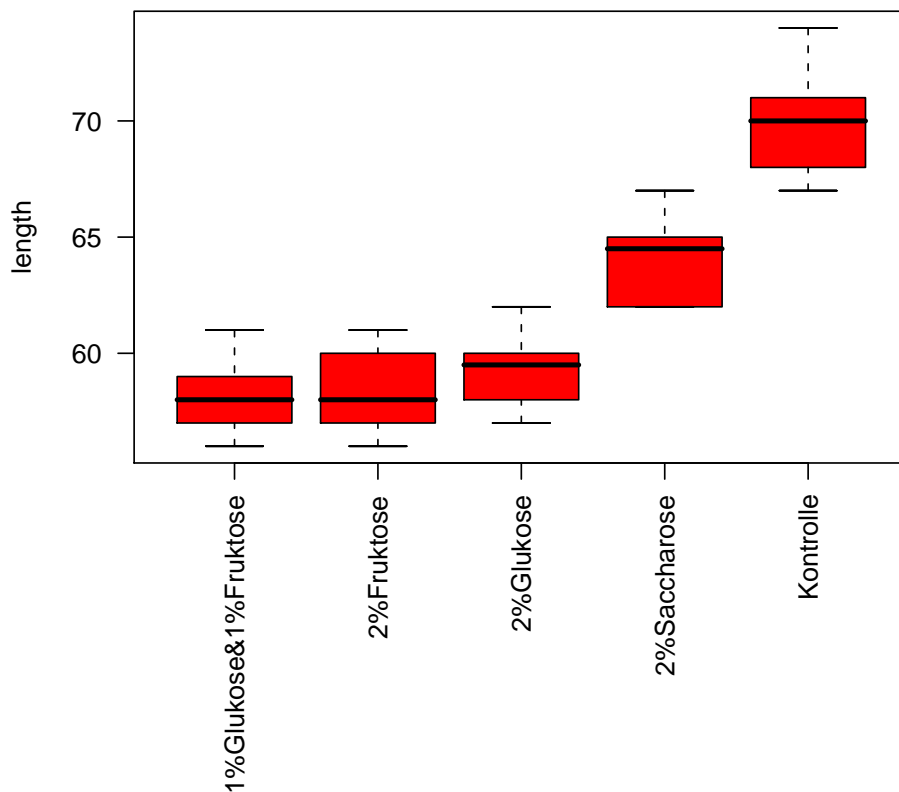
> plot(peas.data)

```



oder etwas schöner

```
> par(mar = c(12, 4, 4, 2))  
> plot(peas.data, las = 2, xlab = "", col = 2)
```



```
# no extra x-axis label, red color
```

Dieser einfache Aufruf der `plot()`-Funktion funktioniert nur, weil der Datensatz genau die zwei Variablen `group` und `length` enthält. Andere Möglichkeiten sind:

```
> plot(group, length)
> plot(length ~ group)
```

Der starke schwarze Strich kennzeichnet den Median in den Gruppen, die Box kennzeichnet den sogenannten Interquartilbereich, in dem 50% in unserem Beispiel die kleinsten bzw. größten Werte in den Gruppen. Im Allgemeinen reichen sie bis zum kleinsten bzw. größten Wert, der nicht mehr als die anderthalbfache Breite des Interquartilbereiches von der Box entfernt ist. Werte, die nicht innerhalb dieses Bereiches liegen, sind sehr extreme Werte und werden als "Ausreißer" extra gekennzeichnet.

Durchführung der Varianzanalyse

Die Varianzanalyse wird in **R** mit der `aov()`-Funktion realisiert.

```
> peas.aov <- aov(length ~ group, data = peas.data)
```

Die Ergebnisse werden in einer sogenannten ANOVA-Tabelle dargestellt.

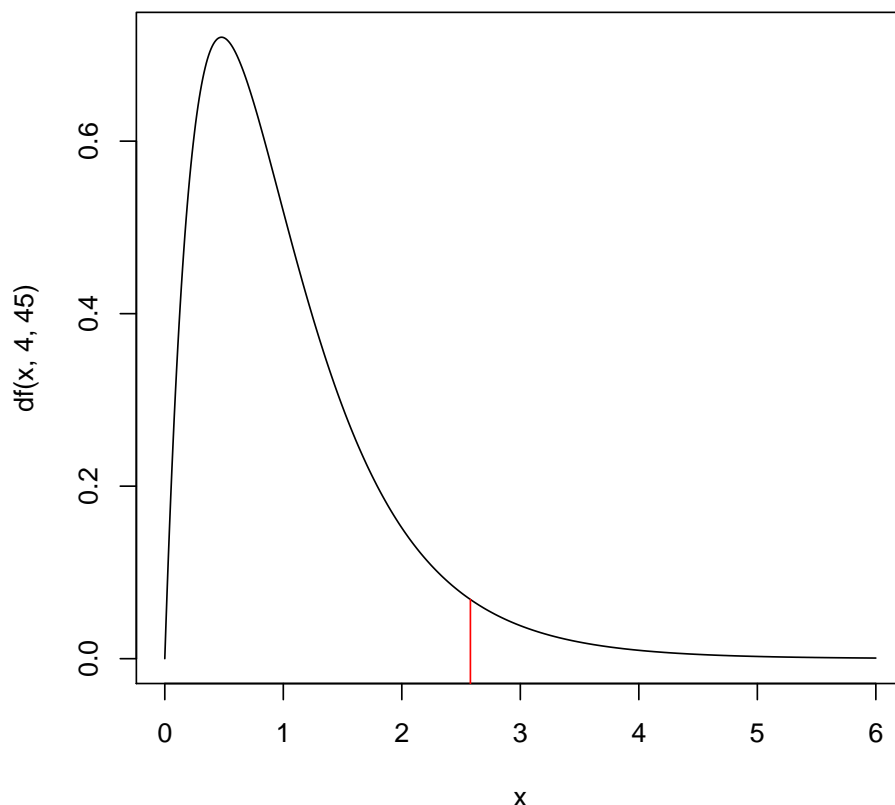
```
> summary(peas.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	4	1077.32	269.33	82.168	< 2.2e-16 ***

Residuals 45 147.50 3.28

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Die Gesamtvariabilität wird zerlegt in die Variabilität zwischen den Gruppen (Zeile group) und die Variabilität innerhalb der Gruppen (Residuals). Die Division der "Sums of Squares" durch die Anzahl der Freiheitsgrade ("Df") liefert die "Mean Squares". Diese sind unter der Annahme der Gleichheit der Gruppenmittelwerte unabhängige Schätzwerte für die Varianz in den Gruppen. Daher überprüfen wir mit Hilfe der F-Statistik, ob der Quotient dieser beiden Werte wesentlich größer als 1 ist. Diese Bewertung erfolgt mit Hilfe der F-Verteilung. Um zu verdeutlichen, wie extrem der berechnete Wert des F-Ratio ist, zeigt die folgende Grafik die für die vorliegende Situation relevante F-Verteilung mit 4 und 45 Freiheitsgraden.



Die Varianzanalyse führte uns also zu dem Ergebnis, dass zwischen den Mittelwerten der Gruppen statistisch signifikante Unterschiede bestehen. Dabei haben wir uns aber noch nicht darum gekümmert, ob die Voraussetzungen der Varianzanalyse erfüllt sind.

Überprüfung der Voraussetzungen

Die Voraussetzungen zur Durchführung der Varianzanalyse sind:

- **Normalverteilung:** Die Variabilität innerhalb der Gruppen lässt sich durch eine Normalverteilung beschreiben.
- **Varianzhomogenität:** Die Varianzen innerhalb der Gruppen sind gleich groß.
- **Unabhängigkeit:** Die Messwerte sind unabhängig voneinander.

Von diesen Voraussetzungen ist die Varianzhomogenität die wichtigste. Falls sie nicht erfüllt ist, können durch geeignete Transformationen der Zielgröße die Varianzen der verschiedenen Gruppen "gleicher" gemacht werden. Die Voraussetzung der Unabhängigkeit ist diejenige, die im Nachhinein kaum korrigiert werden kann. Deshalb ist vor der Datenerhebung (bei der Planung des Experiments beziehungsweise der Probennahme) zu sichern, dass die Messwerte innerhalb einer Gruppe tatsächlich die "richtige" Variabilität widerspiegeln.

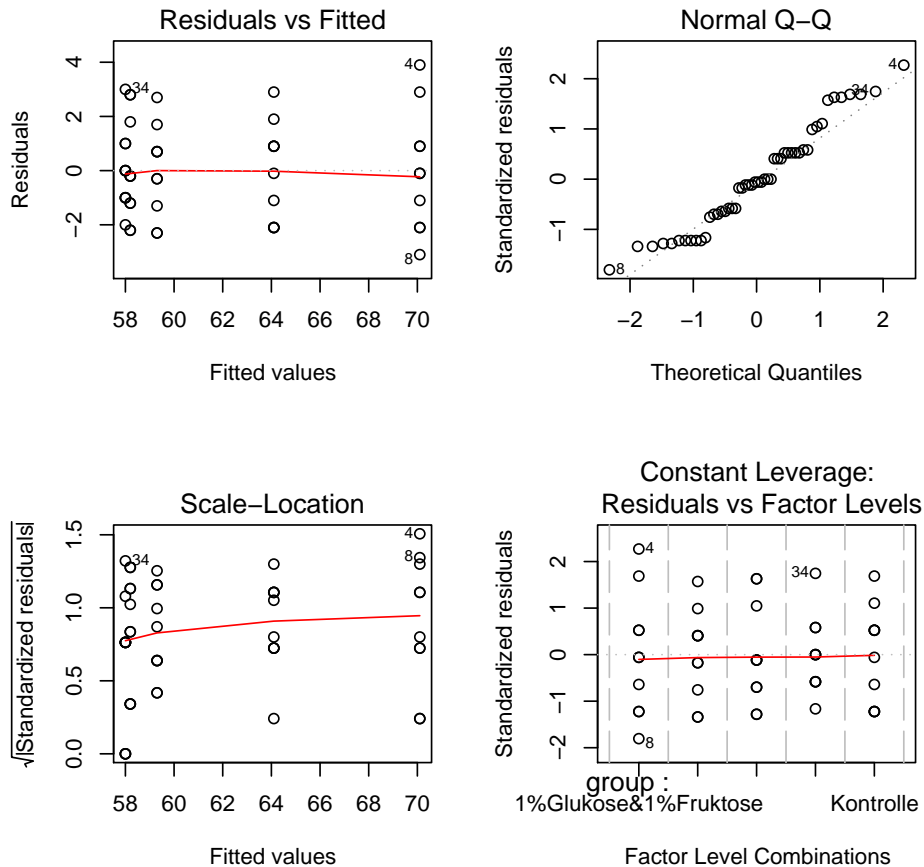
Beispiel 1.2 *Die Auswirkung von Konkurrenz auf die Größe der Blätter von Hieracium pilosella soll untersucht werden. Dazu wird Hieracium pilosella allein und mit unterschiedlich konkurrenzstarken Nachbarpflanzen in Töpfen kultiviert. Nimmt man als Messwerte jetzt die Blattfläche von 10 Blättern einer Hieracium pilosella-Pflanze, so stellen diese **keine** unabhängigen Messwerte dar. Dafür gibt es verschiedene Gründe:*

- *Die genetische Konstitution der Pflanze kann bewirken, dass alle Blätter relativ groß bzw. klein sind.*
- *Diese Einzelpflanze kann besonderen Konkurrenzsituationen begegnet sein, weil z.B. eine der konkurrenzstarken Nachbarpflanzen auf Grund eines Pathogenbefalls ausgefallen ist. Der verringerte Konkurrenzdruck wirkt sich auf **alle** Blätter der Pflanze gleichzeitig aus.*

Unabhängige Messwerte für diese Fragestellung bekommt man, wenn man Blätter verschiedener Individuen untersucht, die den gleichen Konkurrenzbedingungen unterlagen. Dazu kann man entweder aus allen Blättern die mittlere Blattfläche einer Pflanze bestimmen oder ein Blatt jeder Pflanze zufällig auswählen.

Diese Voraussetzungen gleichen den Forderungen, die bei der Regressionsanalyse gestellt werden. Eine graphische Überprüfung dieser Voraussetzungen ist wieder durch einen einfachen Aufruf der `plot()`-Funktion realisierbar.

```
> par(mfrow = c(2, 2))
> plot(peas.aov)
```



Ein möglicher formaler statistischer Test der Voraussetzung der Varianzhomogenität ist der LEVENE-Test, der allerdings in der Grundversion von R nicht verfügbar ist. Wir können aber die Gelegenheit nutzen und uns selbst eine eigene Funktion schreiben, die diesen Test realisiert. Die Grundidee des Tests besteht darin, dass bei größerer Variabilität innerhalb einer Gruppe die mittlere Abweichung vom Gruppenmittelwert größer ist. Daher überprüfen wir mit dem LEVENE-Test eigentlich die Hypothese, dass die mittlere Abweichung von Mittelwert in allen Gruppen gleich ist. Das ist ein Mittelwertvergleich zwischen verschiedenen Gruppen und kann daher durch eine einfache Varianzanalyse beantwortet werden.

```
levene <- function(y,group)
# unsere Funktion soll den Namen levene bekommen
# sie hat zwei Argumente: die Zielgröße, die wir formal mit y bezeichnen und die
# Gruppierungsvariable, die den formalen Namen group bekommt
{
# was in der Funktion passiert, wird in geschweifte Klammern eingeschlossen

group.means <- tapply(y,group,mean)
# berechne zunächst die Gruppenmittelwerte

ynew <- abs(y-group.means[group])
# eine neue Zielgröße wird berechnet:
# die absolute Abweichung vom richtigen Gruppenmittel
```

```

summary(aov(ynew ~ group))
# Das Ergebnis der ANOVA wird ausgegeben
}
# Ende der Funktion

```

Wenn wir diese Funktion definiert haben, können wir sie in **R** mit den passenden Argumenten aufrufen.

```
> levene(y = length, group = group)
```

```

          Df Sum Sq Mean Sq F value Pr(>F)
group      4  2.888   0.722   0.697  0.598
Residuals 45 46.612   1.036

```

Das Ergebnis ist eine ANOVA-Tabelle, der p-Wert größer als 0.05 bedeutet, dass die Nullhypothese nicht abgelehnt wird, die Variabilität innerhalb der fünf verschiedenen Gruppen kann also als etwa gleich angesehen werden.

(Natürlich ist das ein leichter Zirkelschluss - wir überprüfen die Voraussetzungen der ANOVA mit einer ANOVA. In der Praxis hat sich diese Vorgehensweise aber bewährt.)

Multiple Mittelwertvergleiche

Mit den bisherigen Betrachtungen wurde nur ein Einfluss der unterschiedlichen Nährlösungen auf das Wachstum der Erbsen festgestellt. Um die Art des Einflusses genauer zu analysieren, können im Anschluss an die Varianzanalyse verschiedene Mittelwerte konkreter miteinander verglichen werden.

Eine einfache Möglichkeit dazu besteht darin, sich eine Tabelle mit den geschätzten Parametern anzeigen zu lassen.

```
> summary.lm(peas.aov)
```

Call:

```
aov(formula = length ~ group, data = peas.data)
```

Residuals:

```

  Min      1Q  Median      3Q      Max
-3.1   -1.2   -0.1    0.9    3.9

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    58.0000    0.5725 101.307 < 2e-16 ***
group2%Fruktose  0.2000    0.8097   0.247   0.806
group2%Glukose  1.3000    0.8097   1.606   0.115
group2%Saccharose 6.1000    0.8097   7.534 1.66e-09 ***
groupKontrolle 12.1000    0.8097  14.944 < 2e-16 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 1.81 on 45 degrees of freedom

Multiple R-Squared: 0.8796, Adjusted R-squared: 0.8689

F-statistic: 82.17 on 4 and 45 DF, p-value: < 2.2e-16

Was bedeuten diese Parameter?

Zunächst fällt auf, dass die Gruppe "1% Fruktose + 1% Glukose" nicht in der Tabelle auftaucht. Diese wird von **R** als Referenzgruppe ausgewählt und der in der Zeile "Intercept" angegebene Wert gibt den Mittelwert in dieser Gruppe an (Vergleiche mit `peas.means`). Die anderen Parameter geben jeweils die Mittelwertdifferenz zwischen einer Gruppe und der Referenzgruppe an. Da die letzten beiden Zeilen p-Werte kleiner als 0.05 angeben, können wir die Unterschiede zwischen der Gruppe "1% Fruktose + 1% Glukose" und diesen beiden Gruppen als statistisch signifikant ansehen. Leider hatten wir keinen Einfluss auf die Auswahl der Referenzgruppe. Vielleicht wäre die "Kontrolle" eine bessere Referenz?

```
> groupnew <- relevel(group, ref = "Kontrolle")
> peas.aov1 <- aov(length ~ groupnew)
> summary.lm(peas.aov1)
```

Call:

```
aov(formula = length ~ groupnew)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.1	-1.2	-0.1	0.9	3.9

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	70.1000	0.5725	122.44	< 2e-16 ***
groupnew1%Glukose&1%Fruktose	-12.1000	0.8097	-14.94	< 2e-16 ***
groupnew2%Fruktose	-11.9000	0.8097	-14.70	< 2e-16 ***
groupnew2%Glukose	-10.8000	0.8097	-13.34	< 2e-16 ***
groupnew2%Saccharose	-6.0000	0.8097	-7.41	2.52e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.81 on 45 degrees of freedom

Multiple R-Squared: 0.8796, Adjusted R-squared: 0.8689

F-statistic: 82.17 on 4 and 45 DF, p-value: < 2.2e-16

Alle Nährlösungen mit Zuckerzusatz führen zu einem signifikant verringertem Wachstum der Erbsen.

Diese Vergleiche mit einer Referenzgruppe ergeben sich häufig auf natürliche Weise aus der Fragestellung des Experiments und werden *geplante Vergleiche* genannt.

Hat man allerdings keine konkreten Vorstellungen, zwischen welchen Gruppen man nach Unterschieden sucht und vergleicht im Anschluss an die Varianzanalyse alles mit allem, dann besteht die große Gefahr, dass man "scheinbar" signifikante Ergebnisse erhält. Das liegt daran, dass bei der Vorgehensweise des statistischen Testens selbst bei Gültigkeit der Nullhypothese (Mittelwerte gleich) zufällig in 5% der Fälle ein statistisch signifikantes Resultat herauskommt. Wenn man viele Vergleiche durchführt, erhöht sich die Wahrscheinlichkeit, mindestens ein signifikantes Resultat zu bekommen, drastisch. Bei fünf Gruppen gibt es bereits $5 * 4/2 = 10$ mögliche Vergleiche. Um die Wahrscheinlichkeit von 5% für mindestens ein zufällig signifikantes Resultat zu fixieren, wird daher für die einzelnen Vergleiche die kritische Grenze verringert, zum Beispiel indem man das klassische Signifikanzniveau durch die Anzahl der Vergleiche teilt:

$$\alpha^* = \frac{\alpha}{k} .$$

Diese Korrektur nennt man "Bonferroni adjustment". Alternativ werden die p-Werte adjustiert:

$$p^* = pk .$$

Es gibt mehrere Möglichkeiten, wie diese Adjustierung genau erfolgt, ich empfehle die Verwendung der Variante von HOLM (auch sequentielle Bonferroni-Korrektur genannt).

```
> pairwise.t.test(length, group, p.adj = "holm")
```

```
Pairwise comparisons using t tests with pooled SD
```

```
data: length and group
```

	1%Glukose&1%Fruktose	2%Fruktose	2%Glukose	2%Saccharose
2%Fruktose	0.81	-	-	-
2%Glukose	0.35	0.36	-	-
2%Saccharose	1.2e-08	1.9e-08	1.6e-06	-
Kontrolle	< 2e-16	< 2e-16	2.4e-16	1.5e-08

```
P value adjustment method: holm
```

In dieser Ergebnismatrix kann abgelesen werden, welche Gruppen sich statistisch signifikant unterscheiden.

Zweifache Varianzanalyse

Die Grundidee der Varianzanalyse, die Zerlegung der Gesamtvariabilität, kann auf Situationen erweitert werden, in denen mehrere Einflussgrößen vorliegen. Dazu wollen wir uns Biomassendaten aus dem Biodiversitätsexperiment anschauen. In diesem Experiment wurden Pflanzengemeinschaften etabliert, die sich in der Anzahl der Arten, aber auch in ihrer "funktionellen" Diversität, gemessen als Anzahl unterschiedlicher funktioneller Gruppen, unterscheiden.

```
> biomass.may2003 <- read.csv("biomass.may2003.csv", sep = ",")
```

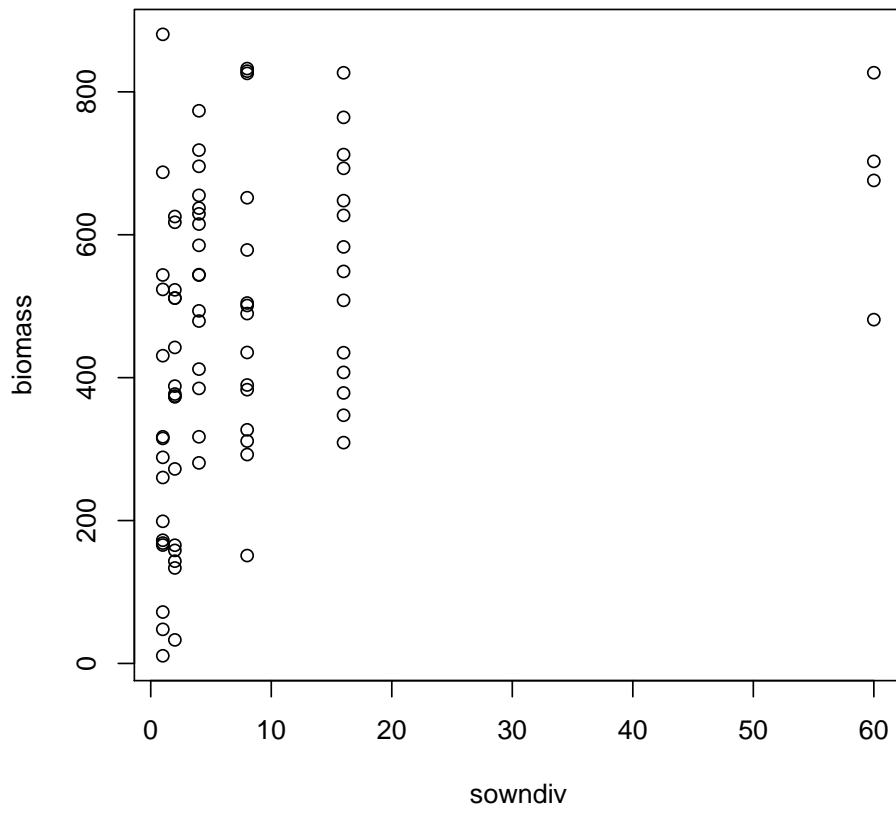
```
> names(biomass.may2003)
```

```
[1] "plotcode" "block" "sowndiv" "funcgr" "biomass"
```

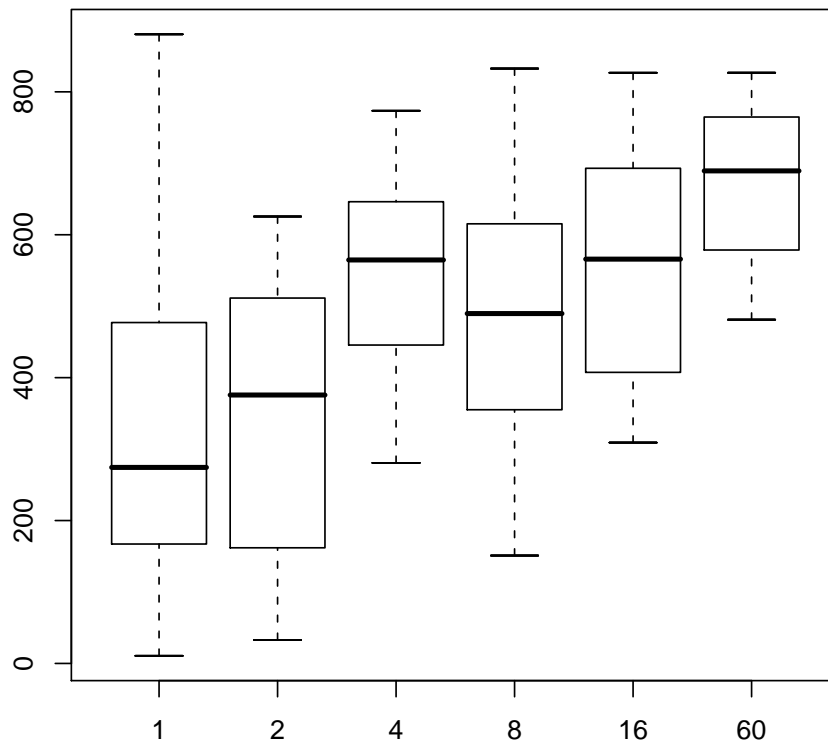
```
> attach(biomass.may2003)
```

Die beiden Einflussgrößen sind keine eigentlich nominalen Einflussgrößen, werden aber hier als unterschiedliche Stufen einer experimentellen Behandlung als solche betrachtet. Um das deutlich zu machen, verwendet man die Funktion *as.factor()*. Den Unterschied sieht man zum Beispiel in einer grafischen Darstellung:

```
> plot(sowndiv, biomass)
```



```
> plot(as.factor(sowndiv), biomass)
```



Beschreibende Statistiken sind jetzt für jede Kombination *Artenzahl* \times *Anzahl funktioneller Gruppen* interessant. Wir verwenden dazu wieder den *tapply()*-Befehl, allerdings in Verbindung mit *interaction()*.

```
> table(sowndiv, funcgr)
```

```

      funcgr
sowndiv 1  2  3  4
      1 16  0  0  0
      2  8  8  0  0
      4  4  4  4  4
      8  4  3  4  4
     16  2  4  4  4
     60  0  0  0  4

```

```
> tapply(biomass, interaction(sowndiv, funcgr), mean, na.rm = T)
```

```

      1.1      2.1      4.1      8.1      16.1      60.1      1.2      2.2      4.2      8.2
317.7031 346.0188 573.4063 434.7819 545.4375          NA          NA 360.1687 458.5375 334.1917 59
      60.3      1.4      2.4      4.4      8.4      16.4      60.4
      NA      NA      NA 616.6906 555.0969 612.5254 671.6838

```

```
> tapply(biomass, interaction(sowndiv, funcgr), sd, na.rm = T)
```

```

      1.1      2.1      4.1      8.1      16.1      60.1      1.2      2.2      4.2
241.09072 188.14436 217.23015 209.77780 235.90850          NA          NA 192.53640 98.22261 4
      8.3      16.3      60.3      1.4      2.4      4.4      8.4      16.4      60.4
229.44944 133.22856          NA          NA          NA 22.90278 212.79037 202.91770 142.97558

```

Die Durchführung der Varianzanalyse erfolgt wieder mit Hilfe der *aov()*-Funktion. Dabei sind jedoch verschiedene Aspekte zu beachten. Zunächst verwenden wir ein rein additives Modell. Damit gehen wir a priori davon aus, dass die Zunahme der Biomasse mit der Artenzahl gleich ist, egal ob 1,2,3 oder 4 funktionelle Gruppen in der Pflanzengemeinschaft vorkommen.

```
> biomass.may2003.aovfit2 <- aov(biomass ~ as.factor(sowndiv) + as.factor(funcgr), data = biomass.may2003)
> summary(biomass.may2003.aovfit2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(sowndiv)	5	963049	192610	5.2756	0.0003486 ***
as.factor(funcgr)	3	87671	29224	0.8004	0.4976674
Residuals	72	2628686	36510		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Das Ergebnis ist von der Reihenfolge der Einflussgrößen abhängig, da diese nicht unabhängig voneinander sind und daher die Anteile an der Gesamtvariabilität nicht eindeutig aufgeteilt werden können

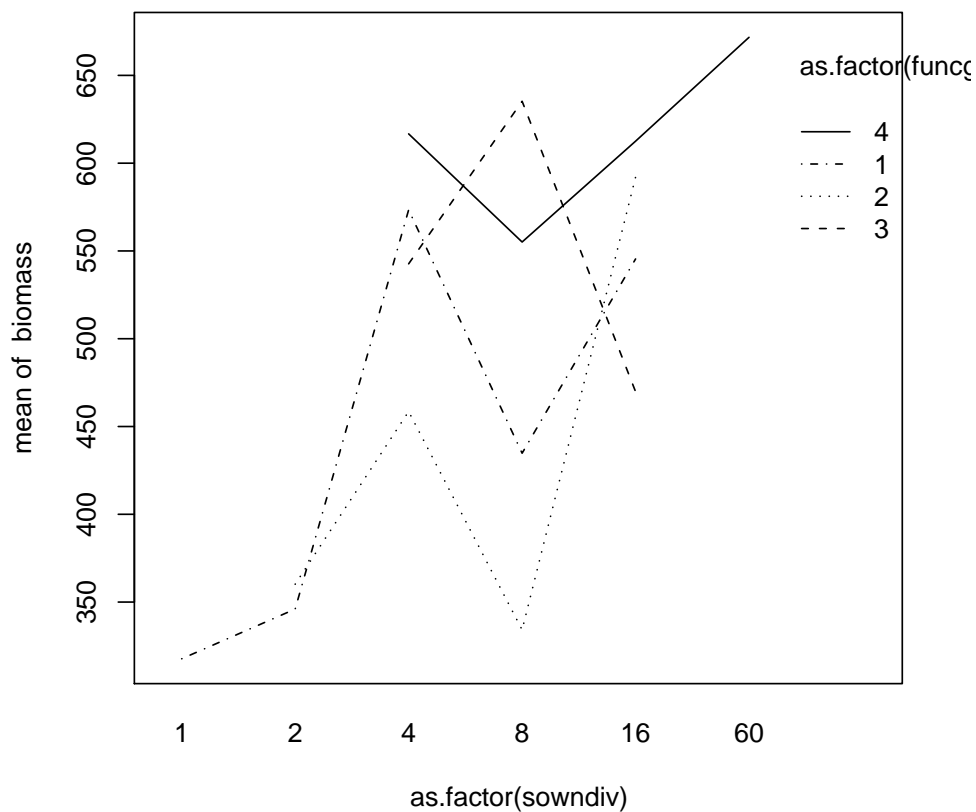
```
> biomass.may2003.aovfit3 <- aov(biomass ~ as.factor(funcgr) + as.factor(sowndiv), data = biomass.may2003)
> summary(biomass.may2003.aovfit3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(funcgr)	3	705848	235283	6.4444	0.0006305 ***
as.factor(sowndiv)	5	344872	68974	1.8892	0.1067985
Residuals	72	2628686	36510		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Wenn wir in einem Experiment zwei Einflussgrößen manipulieren, sind wir jedoch meist an der Frage interessiert, ob der Einfluss der zweiten Einflussgröße davon abhängt, welchen Wert die erste Einflussgröße gerade hat. Das ist die Frage nach einer Wechselwirkung (Interaktion). Diese können wir uns zunächst grafisch veranschaulichen.

```
> interaction.plot(as.factor(sowndiv), as.factor(funcgr), biomass)
```

Für die Spezifizierung in der *aov()*-Funktion gibt es zwei gleichwertige Möglichkeiten:

```
> biomass.may2003.aovfit4 <- aov(biomass ~ as.factor(sowndiv) + as.factor(funcgr) + as.factor(sowndiv):as.factor(funcgr), data = biomass.may2003)
> summary(biomass.may2003.aovfit4)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(sowndiv)	5	963049	192610	5.1537	0.0004823 ***
as.factor(funcgr)	3	87671	29224	0.7819	0.5083113
as.factor(sowndiv):as.factor(funcgr)	7	199425	28489	0.7623	0.6207231
Residuals	65	2429261	37373		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> biomass.may2003.aovfit5 <- aov(biomass ~ as.factor(sowndiv) * as.factor(funcgr), data = biomass.may2003)
> summary(biomass.may2003.aovfit5)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(sowndiv)	5	963049	192610	5.1537	0.0004823 ***
as.factor(funcgr)	3	87671	29224	0.7819	0.5083113
as.factor(sowndiv):as.factor(funcgr)	7	199425	28489	0.7623	0.6207231
Residuals	65	2429261	37373		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1