afex – Analysis of Factorial Experiments in R

Henrik Singmann

Universität Zürich UZH
afex - overview

R package for convenient analysis of factorial experiments

Main functionality:

- works with data in the long format (i.e., one observation per row)
- ANOVA specification: aov_car(), ez_glm(), and aov4()
- Obtain $p$-values for generalized and linear mixed models (GLMMs and LMMs): mixed()
- Compare two vectors using different statistical tests: compare.2.vectors()

afex imitates commercial statistical packages by using effect/deviation coding (i.e., sum-to-zero coding) and type 3 sums of squares.
Standard analysis of variance (ANOVA) is somewhat neglected statistical procedure in (base) R:

"Although the methods encoded in procedures available in SAS and SPSS can seem somewhat oldfashioned, they do have some added value relative to analysis by mixed model methodology, and they have a strong tradition in several applied areas."

(Dalgaard, 2007, p. 2, R News)
ANOVA IN BASE R: `aov()`

Only for balanced designs (from `?aov`):
"`aov` is designed for balanced designs, and the results can be hard to interpret without balance: [...] If there are two or more error strata, the methods used are statistically inefficient without balance, and it may be better to use `lme` in package `nlme`.

Basically only supports "type 2" sums of squares

Cumbersome for within-subject factors (e.g.,
http://stats.stackexchange.com/q/6865/442)
DEFAULT CODING IN R

Categorical predictors (as for ANOVA) need to be transformed in \( k - 1 \) numerical predictors using coding scheme.

Default coding in R: treatment coding (= intercept corresponds to mean of the first group/factor level):

```r
> options("contrasts")
$contrasts
   unordered    ordered
  "contr.treatment" "contr.poly"
```

- Downside: main effects are simple effects when interactions included (i.e., effects of one variable when other is 0).

Usual coding for ANOVA is effects, deviation, or sum-to-zero coding (main effects interpretable in light of interactions):

```r
> options("contrasts")
$contrasts
[1] "contr.sum" "contr.poly"
```

Set contrasts globally to contrast coding (not necessary for afex functions): `set_sum_contrasts()`
ALTERNATIVES TO AOV()

car::Anova() from John Fox
  - can handle any number of between- and within-subjects factors
  - allows for so called "type 2" and "type 3" sums of squares.
  - but, relatively uncomfortable for within-subject factors, as data needs to be in wide format (i.e., one participant per row)

ez (by Mike Lawrence) provides a wrapper for car::Anova(), ezANOVA(), but does not replicate commercial packages without fine-tuning

afex is another car wrapper:
  - aov_car() provides an aov() like formula interface
  - aov_ez() specification of factors using character vectors
  - aov_4() specification using lme4::lmer type syntax.
  - afex automatically sets default contrasts to contr.sum (i.e., sum-to-zero or deviation coding)
EXAMPLE DATA

Reasoning experiment with 60 participants:
- Participants had to rate 24 syllogisms (i.e., 24 different contents)
  (Klauer & Singmann, 2013, JEP: LMC, Experiment 3)

Design:
- validity (2 levels, within-subjects) ×
- believability (3 levels, within-subjects) ×
- condition (2 levels, between-subjects)

Hypotheses: People like valid syllogisms more than invalid ones
(cf. Morsanyi & Handley, 2012, JEP: LMC)

Example item:

No hot things are vons.
Some vons are ice creams.
Therefore, some ice creams are not hot.

How much do you like the last statement?

Data comes with afex: data("ks2013.3")
Figure 3. Mean (filled symbols) and individual (nonfilled symbols) liking ratings in Experiment 3 for the group with fixed contents (left panel) and the group with randomized contents (right panel) as a function of validity/pseudo-validity and conclusion believability. A small amount of vertical jitter was added to individual liking ratings to avoid perfect overlap of two ratings.

Graph produced with `raw.means.plot2()` function (plotrix package).
> str(ks2013.3)
'data.frame': 1440 obs. of 6 variables:
$ id       : Factor w/ 60 levels "1","2","3","4",..: 1 1 1 1 1 1 ...
$ condition: Factor w/ 2 levels "fixed","random": 2 2 2 2 2 2 ...
$ validity : Factor w/ 2 levels "valid","invalid": 2 2 1 1 2 1 ...
$ believability: Factor w/ 3 levels "believable","abstract",..: 2 1 1 ...
$ content  : Factor w/ 24 levels "1","2","3","4",..: 21 4 1 ...
$ response : int 3 4 4 2 2 4 5 4 5 2 ...

> xtabs(~ believability + validity + id, data = d)

, , id = 1

<table>
<thead>
<tr>
<th>validity</th>
<th>believability</th>
<th>invalid</th>
<th>valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstract</td>
<td>believable</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>believable</td>
<td>believable</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>unbelievable</td>
<td>believable</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

[...]

ANOVA IN AFEX

```
aov_car(response ~ condition + Error(id/believability * validity), ks2013.3)
```

Differences to `aov()`:

- Error term mandatory (to specify id variable).
- within-subject factors only need to be present in Error term (but can be present outside of it, where they will be ignored).
- within-subject factors don't need to be enclosed in parentheses and are always fully crossed

```
aov_ez("id", "response", ks2013.3, between = "condition", within = c("believability", "validity"))
```

```
aov_4(response ~ condition + (believability * validity|id), ks2013.3)
```

Call `aov_car()` with the respective formula and produce identical output.
```r
ov_ez("id", "response", ks2013.3, between = "condition",
       within = c("believability", "validity"))
```

Rasts set to `contr.sum` for the following variables: `condition`

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>MSE</th>
<th>F</th>
<th>ges</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>1, 58</td>
<td>0.94</td>
<td>0.01</td>
<td>&lt;.0001</td>
<td>.90</td>
</tr>
<tr>
<td>believability</td>
<td>1.84,106.78</td>
<td>0.59</td>
<td>8.36</td>
<td>***</td>
<td>.05</td>
</tr>
<tr>
<td>condition:believability</td>
<td>1.84,106.78</td>
<td>0.59</td>
<td>0.29</td>
<td>.002</td>
<td>.73</td>
</tr>
<tr>
<td>validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>0.17</td>
<td>.0004</td>
<td>.68</td>
</tr>
<tr>
<td>condition:validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>2.07</td>
<td>.005</td>
<td>.16</td>
</tr>
<tr>
<td>believability:validity</td>
<td>1.85,107.52</td>
<td>0.28</td>
<td>8.29</td>
<td>***</td>
<td>.02</td>
</tr>
<tr>
<td>condition:believability:validity</td>
<td>1.85,107.52</td>
<td>0.28</td>
<td>3.58</td>
<td>*</td>
<td>.01</td>
</tr>
</tbody>
</table>

Error message:
```
ov.car(response ~ condition + Error(id/(believability * validity)), d): more than one observation per cell, aggregating the data using mean (i.e, aggregate = mean)!
```
```

ov_ez("id", "response", ks2013.3, between = "condition",
      within = c("believability", "validity"))

contrasts set to contr.sum for the following variables: condition

necessary: information about coding changes for between-subjects variables.

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>MSE</th>
<th>F</th>
<th>ges</th>
<th>p.value</th>
<th>&lt;</th>
<th>0.0001</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>1, 58</td>
<td>0.94</td>
<td>0.01</td>
<td>&lt;.0001</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>believability</td>
<td>1.84, 106.78</td>
<td>0.59</td>
<td>8.36</td>
<td>***</td>
<td>.05</td>
<td>.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>condition:believability</td>
<td>1.84, 106.78</td>
<td>0.59</td>
<td>0.29</td>
<td>.002</td>
<td>.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>0.17</td>
<td>.0004</td>
<td>.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>condition:validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>2.07</td>
<td>.005</td>
<td>.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>believability:validity</td>
<td>1.85, 107.52</td>
<td>0.28</td>
<td>8.29</td>
<td>***</td>
<td>.02</td>
<td>.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>condition:believability:validity</td>
<td>1.85, 107.52</td>
<td>0.28</td>
<td>3.58</td>
<td>*</td>
<td>.01</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

warning message:

ov.car(response ~ condition + Error(id/(believability * validity)), d): more than one observation per cell, aggregating the data using mean (i.e, aggregate = mean)!

```
```r
tov_ez("id", "response", ks2013.3, between = "condition",
within = c("believability", "validity"))

contrasts set to contr.sum for the following variables: condition

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>MSE</th>
<th>F</th>
<th>ges</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>1, 58</td>
<td>0.94</td>
<td>0.01</td>
<td>&lt;.0001</td>
<td>.90</td>
</tr>
<tr>
<td>believability</td>
<td>1.84, 106.78</td>
<td>0.59</td>
<td>8.36</td>
<td>***</td>
<td>.05</td>
</tr>
<tr>
<td>condition:believability</td>
<td>1.84, 106.78</td>
<td>0.59</td>
<td>0.29</td>
<td>.002</td>
<td>.73</td>
</tr>
<tr>
<td>validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>0.17</td>
<td>.0004</td>
<td>.68</td>
</tr>
<tr>
<td>condition:validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>2.07</td>
<td>.005</td>
<td>.16</td>
</tr>
<tr>
<td>believability:validity</td>
<td>1.85, 107.52</td>
<td>0.28</td>
<td>8.29</td>
<td>***</td>
<td>.02</td>
</tr>
<tr>
<td>condition:believability:validity</td>
<td>1.85, 107.52</td>
<td>0.28</td>
<td>3.58</td>
<td>*</td>
<td>.01</td>
</tr>
</tbody>
</table>

Warning message:
```
\begin{verbatim}
tov.car(response ~ condition + Error(id/(believability * validity)), d) : more than one observation per cell, aggregating the data using mean (i.e, aggregate = mean)!
\end{verbatim}
```

tov.car() automatically aggregates data for the within-subject factors (with warning). Warning can be suppressed by explicitly specifying the aggregation function.
The default output contains the "recommended effect size for repeated-measures design" (Bakeman, 2005, Behavior Research Methods), $\eta^2_G$.

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>MSE</th>
<th>F</th>
<th>ges</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>1, 58</td>
<td>0.94</td>
<td>0.01</td>
<td>&lt;.0001</td>
<td>.90</td>
</tr>
<tr>
<td>believability</td>
<td>1.84, 106.78</td>
<td>0.59</td>
<td>8.36 ***</td>
<td>.05</td>
<td>.0006</td>
</tr>
<tr>
<td>condition:believability</td>
<td>1.84, 106.78</td>
<td>0.59</td>
<td>0.29</td>
<td>.002</td>
<td>.73</td>
</tr>
<tr>
<td>validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>0.17</td>
<td>.0004</td>
<td>.68</td>
</tr>
<tr>
<td>condition:validity</td>
<td>1, 58</td>
<td>0.38</td>
<td>2.07</td>
<td>.005</td>
<td>.16</td>
</tr>
<tr>
<td>believability:validity</td>
<td>1.85, 107.52</td>
<td>0.28</td>
<td>8.29 ***</td>
<td>.02</td>
<td>.0006</td>
</tr>
<tr>
<td>condition:believability:validity</td>
<td>1.85, 107.52</td>
<td>0.28</td>
<td>3.58 *</td>
<td>.01</td>
<td>.03</td>
</tr>
</tbody>
</table>

*ov.car(response ~ condition + Error(id/(believability * validity)), d): more than one observation per cell, aggregating the data using mean (i.e., aggregate = mean)!*
ANOVA WITH AFEX

aov_car(), aov_ez(), aov_4() print nice ANOVA table as default

- Greenhouse-Geisser correction of df
- $\eta^2_G$ effect size

methods for returnend object (class "afex_aov"):

- nice() prints ANOVA table with rounded value (good for copy-paste).
- anova() prints standard R ANOVA table (without rounding).
- methods allow to specify:
  - df-correction: Greenhouse-Geisser (default), Huynh-Feldt, none
  - Specify effect size: $\eta^2_G$ (default) or $\eta^2_P$
- Can be passed to lsmeans for follow-up analysis (post-hoc contrasts)
require(lsmeans)
  <- aov_ez("id", "response", ks2013.3, between = "condition", within = c("believability","validity"))
smeans(a, ~believability)
E: Results may be misleading due to involvement in interactions

<table>
<thead>
<tr>
<th>believability</th>
<th>lsmean</th>
<th>SE</th>
<th>df</th>
<th>lower.CL</th>
<th>upper.CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>stract</td>
<td>3.106250</td>
<td>0.07485452</td>
<td>161.26</td>
<td>2.958428</td>
<td>3.254072</td>
</tr>
<tr>
<td>lievable</td>
<td>3.364583</td>
<td>0.07485452</td>
<td>161.26</td>
<td>3.216762</td>
<td>3.512405</td>
</tr>
<tr>
<td>believable</td>
<td>2.985417</td>
<td>0.07485452</td>
<td>161.26</td>
<td>2.837595</td>
<td>3.133238</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: cond, validity
Confidence level used: 0.95

pairs(lsmeans(a, ~believability))
E: Results may be misleading due to involvement in interactions

<table>
<thead>
<tr>
<th>contrast</th>
<th>estimate</th>
<th>SE</th>
<th>df</th>
<th>t.ratio</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>stract - believable</td>
<td>-0.2583333</td>
<td>0.09475594</td>
<td>116</td>
<td>-2.726</td>
<td>0.0201</td>
</tr>
<tr>
<td>stract - unbelievable</td>
<td>0.1208333</td>
<td>0.09475594</td>
<td>116</td>
<td>1.275</td>
<td>0.4120</td>
</tr>
<tr>
<td>lievable - unbelievable</td>
<td>0.3791667</td>
<td>0.09475594</td>
<td>116</td>
<td>4.002</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: cond, validity
Value adjustment: tukey method for a family of 3 means
(m <- lsmeans(a, ~validity:cond))

OTE: Results may be misleading due to involvement in interactions

<table>
<thead>
<tr>
<th>validity</th>
<th>cond</th>
<th>lsmean</th>
<th>SE</th>
<th>df</th>
<th>lower.CL</th>
<th>upper.CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>invalid</td>
<td>random</td>
<td>3.191667</td>
<td>0.085487</td>
<td>97.99</td>
<td>3.022019</td>
<td>3.361314</td>
</tr>
<tr>
<td>valid</td>
<td>random</td>
<td>3.125000</td>
<td>0.085487</td>
<td>97.99</td>
<td>2.955353</td>
<td>3.294647</td>
</tr>
<tr>
<td>invalid</td>
<td>fixed</td>
<td>3.086111</td>
<td>0.085487</td>
<td>97.99</td>
<td>2.916464</td>
<td>3.255758</td>
</tr>
<tr>
<td>valid</td>
<td>fixed</td>
<td>3.205556</td>
<td>0.085487</td>
<td>97.99</td>
<td>3.035908</td>
<td>3.375203</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: believability

value adjustment: holm method for 2 tests
```r
contrast(m, c, adjust = "holm")

<table>
<thead>
<tr>
<th>contrast</th>
<th>estimate</th>
<th>SE</th>
<th>df</th>
<th>t.ratio</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>val_random</td>
<td>0.11944</td>
<td>0.09138</td>
<td>58</td>
<td>1.307</td>
<td>0.3926</td>
</tr>
<tr>
<td>val_fixed</td>
<td>-0.06667</td>
<td>0.09138</td>
<td>58</td>
<td>-0.730</td>
<td>0.4686</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: believability value adjustment: holm method for 2 tests

require(multcomp)
summary(as.glht(contrast(m, c)), test=adjusted("free"))

Note: df set to 58

Simultaneous Tests for General Linear Hypotheses

Linear Hypotheses:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|---------|
| val_random == 0 | 0.11944 | 0.09138 | 1.307   | 0.352  |
| val_fixed == 0   | -0.06667 | 0.09138 | -0.730  | 0.469  |
(Adjusted p values reported -- free method)
POST-HOC CONTRASTS

1. estimate ANOVA with afex
2. pass returned object to lsmeans() using desired factors.
3. create contrasts on reference-grid (i.e., rows in lsmeans object)
4. obtain test on contrasts using contrast()
5. (pass contrast object to multcomp for advanced p-value corrections)

(see lsmeans vignette for more details)

Note: Do not use "aov" ANOVA!
BEYOND ANOVA: MIXED MODELS

Repeated-measures ANOVA has limitations (e.g., Keselman, et al., 2001, BJS&MP):
- Sphericity assumption: df correction known to be problematic
- Only one observation per cell of design and participant allowed
- No simultaneous analysis of multiple random effects (e.g., participant and item effects)

Linear Mixed Models (LMMs)
- overcome many of these limitations
- for multiple and crossed random effects
- for hierarchical or multilevel structures in the data.

afex contains convenience function mixed() for obtaining p-values for mixed models and fits them with lme4::lmer (package of choice for mixed models in R).
LINEAR MIXED MODELS (LMMS)

One interval scaled response variable $y$

$m$ predictors ($\beta$)

Linear Model (Observations are independent):
- $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_m x_m + \varepsilon$,
  where $\varepsilon \sim N(0, \sigma^2)$

Non-independent observations:
- Participants see all levels of $\beta_1$ (i.e., within-subjects factor), and the effect of $\beta_1$ may be different for each participant $P$
- $I = $ Each Item may also have specific effects

$y = \beta_0 + P_0 + I_0 + (\beta_1 + P_1) x_1 + \ldots + \beta_m x_m + \varepsilon$,
where $\varepsilon \sim N(0, \sigma^2)$,
$(P_0, P_1) \sim N(0, [\ldots])$,
$I_0 \sim N(0, \omega^2)$
LINEAR MIXED MODELS (LMMS)

Random intercepts

Non-independent observations:
- Participants see all levels of β₁ (i.e., within-subjects factor), and the effect of β₁ may be different for each participant P
- I = Each item may also have specific effects

\[ y = \beta_0 + P_0 + I_0 + (\beta_1 + P_1)x_1 + ... + \beta_mx_m + \varepsilon, \]
where \( \varepsilon \sim N(0, \sigma^2) \),
\( (P_0, P_1) \sim N(0, [\ldots]) \),
\( I_0, \sim N(0, \omega^2) \)
Obtaining $p$ values for lme4 models is not trivial:

- sampling distribution of NULL hypothesis problematic
- correct number of denominator degrees of freedoms unknown

mixed() implements "best" options (according to lme4 FAQ) to overcome this:

- for LMMs: Kenward-Rogers approximation for df (method = "KR", default) [also offered in car::Anova(..., test = "F")]
- for GLMMs and LMMs: Parametric bootstrap (method = "PB")
- for GLMMs and LMMs: Likelihood-ratio tests (method = "LRT")
- first two options achieved through package pbkrtest (Halekoh & Hojsgaard, 2012).
mixed()

mixed() wrapper of lme4::lmer() with additional arguments:

- **type**: type of "sums of squares" (i.e., how should effects be calculated), default is 3

- **method**:
  - Kenward-Rogers ("KR", default, may needs lots of RAM)
  - parametric bootstrap ("PB", can be parallelized using the parallel package)
  - LRTs ("LRT")

- **args.test**: further arguments passed to pbkrtest.

\[
m1 <- \text{mixed}(\text{response} \sim \text{condition} * \text{validity} * \text{believability} + (\text{believability} * \text{validity}|\text{id}) + (1|\text{content}), \text{ks2013.3}, \text{method} = \text{"LRT"})
\]
1 <- mixed(response ~ condition * validity * believability + (believability | id) + (1 | content), ks2013.3, method = "LRT")

contrasts set to contr.sum for the following variables: condition, validity, believability, id, content

L argument to lmer() set to FALSE for method = 'PB' or 'LRT'

Fitting 8 (g)lmer() models:

......

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>Chisq</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>condition</td>
<td>1</td>
<td>0.02</td>
<td>.90</td>
</tr>
<tr>
<td>validity</td>
<td>1</td>
<td>0.03</td>
<td>.87</td>
</tr>
<tr>
<td>believability</td>
<td>2</td>
<td>6.43</td>
<td>* .04</td>
</tr>
<tr>
<td>condition:validity</td>
<td>1</td>
<td>1.90</td>
<td>.17</td>
</tr>
<tr>
<td>condition:believability</td>
<td>2</td>
<td>0.47</td>
<td>.79</td>
</tr>
<tr>
<td>validity:believability</td>
<td>2</td>
<td>5.94</td>
<td>+ .05</td>
</tr>
<tr>
<td>condition:validity:believability</td>
<td>2</td>
<td>0.83</td>
<td>.66</td>
</tr>
</tbody>
</table>
mixed() – return value

returns S3 object of class "mixed" with methods:
- print()/nice() prints ANOVA table with rounded value (good for copy-paste).
- anova() prints standard R ANOVA table (without rounding).
- summary() prints summary() of lme4 object

> str(m1, 1)

List of 4
$ anova_table :Classes ‘anova’ and 'data.frame’: 7 obs. of 4 variables:...
- attr(*, "heading")= chr [1:5] "Mixed Model Anova Table (Type 3 tests)\n"Model: response ~ condition * validity * believability + (believability * "Model: validity | id) + (1 | content)" "Data: ks2013.3" ...
$ full.model :Formal class 'lmerMod' [package "lme4"] with 13 slots
$ restricted.models:List of 7
$ tests :List of 7
  - attr(*, "class")= chr "mixed"
  - attr(*, "type")= num 3
  - attr(*, "method")= chr "LRT"
NOTE: Results may be misleading due to involvement in interactions

<table>
<thead>
<tr>
<th>validity</th>
<th>cond</th>
<th>lsmean</th>
<th>SE</th>
<th>df</th>
<th>asymp.LCL</th>
<th>asymp.UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>invalid</td>
<td>random</td>
<td>3.201079</td>
<td>0.0958</td>
<td>NA</td>
<td>3.0132</td>
<td>3.3892</td>
</tr>
<tr>
<td>valid</td>
<td>random</td>
<td>3.115587</td>
<td>0.0969</td>
<td>NA</td>
<td>2.9256</td>
<td>3.3056</td>
</tr>
<tr>
<td>invalid</td>
<td>fixed</td>
<td>3.091634</td>
<td>0.1006</td>
<td>NA</td>
<td>2.8943</td>
<td>3.2889</td>
</tr>
<tr>
<td>valid</td>
<td>fixed</td>
<td>3.200033</td>
<td>0.1017</td>
<td>NA</td>
<td>3.0007</td>
<td>3.3993</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: believability
Confidence level used: 0.95

> contrast(means, c, adjust="holm")

<table>
<thead>
<tr>
<th>contrast</th>
<th>estimate</th>
<th>SE</th>
<th>df</th>
<th>z.ratio</th>
<th>p.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>val_random</td>
<td>-0.085492</td>
<td>0.0895</td>
<td>NA</td>
<td>-0.9551</td>
<td>0.6364</td>
</tr>
<tr>
<td>val_fixed</td>
<td>0.108399</td>
<td>0.1086</td>
<td>NA</td>
<td>0.9981</td>
<td>0.6364</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: believability
P value adjustment: holm method for 2 tests
P values are asymptotic
afex provides convenience functions for specifying statistical models for factorial experimental designs:

- ANOVA: `aov_ez()`, `aov_car()`, and `aov_4()`
- `mixed()` for LMMs and GLMMs (i.e., models with potentially crossed random effects), see Barr, Levy, Scheepers, & Tily (2013). *Keep it maximal.* Journal of Memory and Language.

Returned objects can be passed to `lsmeans` for contrasts and further inspection (and from there to `multcomp`)

Two vectors (unpaired or paired) can be compared with `compare.2.vectors` using t-, (Welch-), Wilcoxon-, and permutation-test
THANK YOU FOR YOUR ATTENTION
GLMMs

Suppose dependent variable was not interval scaled, but binary (i.e., if <= 3, 0, else 1).

Need to extend LMM to model with binomial residual distribution and link function (default binomial link function is logit).

```r
m2 <- mixed(resp2 ~ cond * validity * believability + (believability * validity|id) + (1|content), d, family = binomial, method = "LRT")
```
## GLMM — RESULTS

```r
> m2

Effect df.large df.small chisq df p.value
1   cond 34   33  0.17   1 0.68
2 validity 34   33  0.07   1 0.79
3 believability 34   32  8.22   2 0.02
4        cond:validity 34   33  1.48   1 0.22
5     cond:believability 34   32  2.62   2 0.27
6     validity:believability 34   32  7.44   2 0.02
7  cond:validity:believability 34   32  2.50   2 0.29

Warning messages:
1: In print.mixed(list(anova.table = list(Effect = c("cond", "validity"), : lme4 reported (at least) the following warnings for 'full':
  * failure to converge in 10000 evaluations
  * Model failed to converge with max|grad| = 0.00439336 (tol = 0.001, component 16)
2: In print.mixed(list(anova.table = list(Effect = c("cond", "validity"), : lme4 reported (at least) the following warnings for 'cond':
  * failure to converge in 10000 evaluations
  * Model failed to converge with max|grad| = 0.00578346 (tol = 0.001, component 16)
3: In print.mixed(list(anova.table = list(Effect = c("cond", "validity"), : [...]
```
COMPARE.2.VECTORS()

compares two vectors using various tests:

> compare.2.vectors(1:10, c(7:20, 200))

$parametric
   test test.statistic test.value test.df  p
1    t          -1.325921  23.0000  0.1978842
2  Welch  t           -1.632903 14.1646  0.1245135

$nonparametric
   test     test.statistic test.value test.df  p
1    W            8.000000     NA  0.0002228503
2 permutation  Z         -1.305464     NA  0.0979700000
3    coin::Wilcoxon  Z         -3.719353     NA  0.0000200000
4     median  Z          3.545621     NA  0.0005600000

default uses 100,000 Monte Carlo samples to estimate approximation of exact conditional distribution (for last three tests) using coin (Hothorn, Hornik, van de Wiel, & Zeileis, 2008, JSS)
Generalized Linear Mixed Models (GLMMs)

One interval scaled response variable $y$

$m$ predictors ($\beta$), repeated measures on $\beta_1$, and $P$ and $I$ effects

$$y = \beta_0 + P_0 + I_0 + (\beta_1 + P_1)x_1 + \ldots + \beta_mx_m + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$, $(P_0, P_1) \sim N(0, [...]')$, $I_0, \sim N(0, \omega^2)$.

The dependent variable $dv$ directly corresponds to the predicted variable $y$.

For e.g., binomial (i.e., 0,1) data this is not the case and we need a function that links $y$ to $dv$, which would be the logit function.

(In addition to the link function we also need to specify the distribution of $\epsilon$)
mixed() obtains p-values of effects in LMMs and GLMMs by fitting different versions of model (using \texttt{lmer}) and comparing those with larger model (via \texttt{pbkrtest} or \texttt{anova}).

Type 3 tests: full model is compared with a model in which only the effect is excluded.

Type 2 tests: For each effect a model in which all higher order effects are excluded is tested against one in which all higher and this effects are excluded.

Note, effects are excluded by directly altering the model matrix (and not by excluding it via R formula).
WHY ARE TYPE 3 TESTS STANDARD?

Type 2 tests assume no higher order effects for any effect, and tests of lower order effects are meaningless if higher-order effects are present.

Type 3 tests do not have this requirements, they calculate tests of lower-order effects in presence of higher-order effects.

Many statisticians prefer Type 2 tests as
- they are more powerful (Lansgrund, 2003),
- do not violate marginality (Venables, 2000),
- and most notably if interactions are present, main effects are per se not interpretable.