Additional Issues: Random effects diagnostics, multiple comparisons

Austin F. Frank, T. Florian Jaeger

April 30, 2009

The `dative` dataset

- Original analysis in Bresnan *et al* (2007)
- Data obtained from `languageR` (Baayen 2008)
- "Data describing the realization of the dative as NP or PP in the Switchboard corpus and the Treebank Wall Street Journal collection."
- 3,263 observations, 74% NP realization
- 75 verbs

The `dative` dataset

- Original analysis in Bresnan *et al* (2007)
- Data obtained from `languageR` (Baayen 2008)
- "Data describing the realization of the dative as NP or PP in the Switchboard corpus and the Treebank Wall Street Journal collection."
- 3,263 observations, 74% NP realization
- 75 verbs

The `dative` dataset

- Original analysis in Bresnan *et al* (2007)
- Data obtained from `languageR` (Baayen 2008)
- "Data describing the realization of the dative as NP or PP in the Switchboard corpus and the Treebank Wall Street Journal collection."
- 3,263 observations, 74% NP realization
- 75 verbs

The `dative` dataset

- Original analysis in Bresnan *et al* (2007)
- Data obtained from `languageR` (Baayen 2008)
- "Data describing the realization of the dative as NP or PP in the Switchboard corpus and the Treebank Wall Street Journal collection."
- 3,263 observations, 74% NP realization
- 75 verbs

The `dative` dataset

- Original analysis in Bresnan *et al* (2007)
- Data obtained from `languageR` (Baayen 2008)
- "Data describing the realization of the dative as NP or PP in the Switchboard corpus and the Treebank Wall Street Journal collection."
- 3,263 observations, 74% NP realization
- 75 verbs

Random intercept model

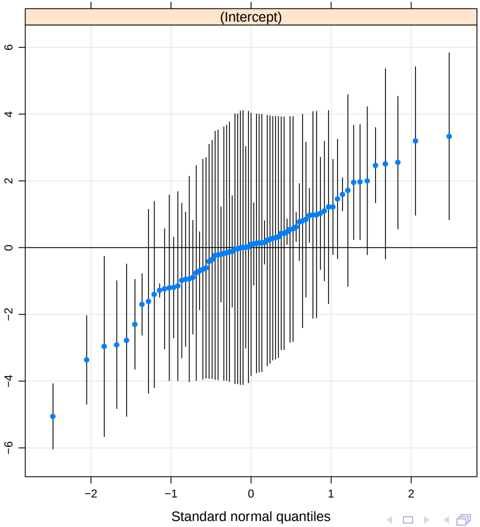

```
> dative.lmer <- lmer(RealizationOfRecipient ~
+        AccessOfRec + AnimacyOfRec + LengthOfRecipient +
+        AccessOfTheme + AnimacyOfTheme + LengthOfTheme +
+        PronomOfTheme + DefinOfTheme + SemanticClass + Modality +
+        (1 | Verb),
+        data = dative, family = ``binomial'')

Random effects:
 Groups Name        Variance Std.Dev.
 Verb   (Intercept) 4.3982   2.0972
Number of obs: 3263, groups: Verb, 75
```

Random intercept model: fixed effects


```
      Fixed effects:
                             Estimate Std. Error z value Pr(>|z|)
(Intercept)                   1.29306    0.65005   1.989   0.0467 *
AccessOfRecgiven             -2.46128    0.17761 -13.858  < 2e-16 ***
AccessOfRecnew                0.12462    0.24422   0.510   0.6099
AnimacyOfRecinanimate         2.24226    0.25864   8.669  < 2e-16 ***
LengthOfRecipient             0.41486    0.04754   8.727  < 2e-16 ***
AccessOfThemegiven            1.50539    0.25504   5.903 3.58e-09 ***
AccessOfThemenew             -0.41979    0.19067  -2.202   0.0277 *
AnimacyOfThemeinanimate      -0.86355    0.48283  -1.789   0.0737 .
LengthOfTheme                -0.23354    0.02766  -8.442  < 2e-16 ***
PronomOfThemepronominal       2.20502    0.24624   8.955  < 2e-16 ***
DefinOfThemeindefinite       -0.93294    0.19024  -4.904 9.39e-07 ***
SemanticClassc                0.38583    0.34928   1.105   0.2693
SemanticClassf                0.02208    0.57970   0.038   0.9696
SemanticClassp               -3.77591    1.47575  -2.559   0.0105 *
SemanticClasst                0.31043    0.20895   1.486   0.1374
Modalitywritten               0.85012    0.18536   4.586 4.51e-06 ***
```

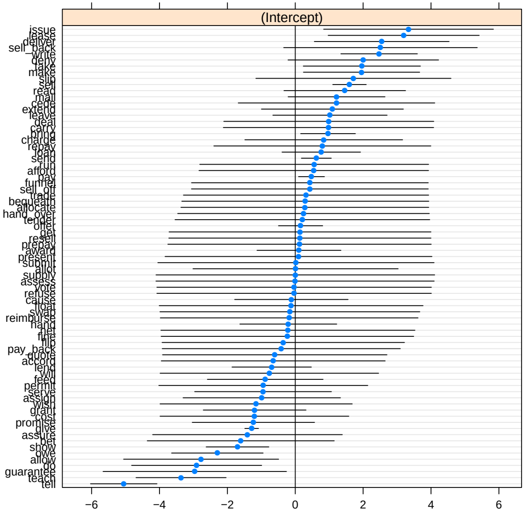
Random intercepts: QQ plot

```
> library(lattice)
> dative.ranef <- ranef(dative.lmer, postVar = TRUE)
> qqmath(dative.ranef)
```



Random intercepts: catepillar plot

```
> library(lattice)
> dative.ranef <- ranef(dative.lmer, postVar = TRUE)
> dotplot(dative.ranef)
```

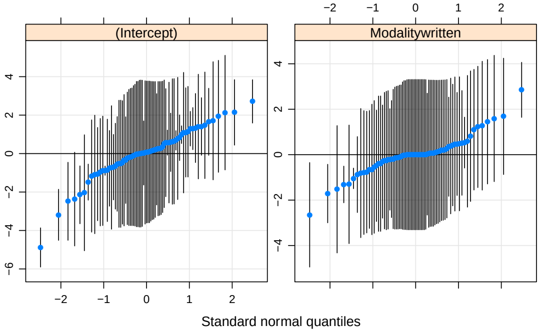


Random slope model

```
> dative.lmer.slope <- lmer(RealizationOfRecipient ~
+        AccessOfRec + AnimacyOfRec + LengthOfRecipient +
+        AccessOfTheme + AnimacyOfTheme + LengthOfTheme +
+        PronomOfTheme + DefinOfTheme + SemanticClass + Modality +
+        (1 + Modality | Verb),
      data = dative, family = ``binomial'')

Random effects:
 Groups Name             Variance Std.Dev. Corr
 Verb   (Intercept)      3.8098   1.9519
        Modalitywritten  2.8482   1.6877   0.000
Number of obs: 3263, groups: Verb, 75
```

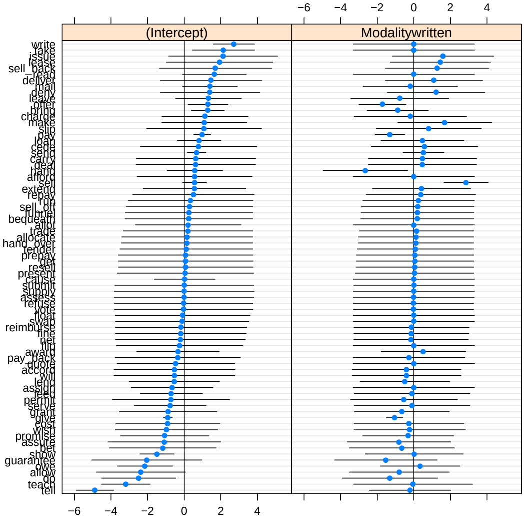
Random slope: QQ plot

```
> library(lattice)
> dative.ranef <- ranef(dative.lmer.slope, postVar = TRUE)
> qqmath(dative.ranef)
```



Random slope: catepillar plot

```
> library(lattice)
> dative.ranef <- ranef(dative.lmer.slope, postVar = TRUE)
> dotplot(dative.ranef)
```



Model comparison


```
> anova(dative.lmer, dative.lmer.slope)

                   Df     AIC     BIC  logLik  Chisq Chi Df Pr(>Chisq)
dative.lmer        17 1561.05 1664.59 -763.52
dative.lmer.slopes 19 1524.71 1640.42 -743.35 40.344      2  1.735e-09 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

We conclude that a random slope term for modality significantly improves the fit of the model to data based on a likelihood ratio test $\left(\chi^2_{\Delta_\Lambda}(2) = 40.3,\ p < .0001\right)$

Visualizing and testing random effects

Post hoc comparisons

Initial considerations

- Preferred method for hypothesis testing on regression parameters is to include relevant orthogonal contrasts in the original model
- Whether doing planned or *post hoc* comparisons on linear mixed models, distributional assumptions for hypothesis testing are murky at best (*c.f.* `http://wiki.r-project.org/rwiki/doku.php?id=guides:lmer-tests`, `https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html`)
- Hypothesis testing on regression parameters for generalized linear mixed models depends on the data matching distributional assumptions. Over- or underdispersion in the data makes standard error calculations on regression coefficients unreliable.
- Multiplicity corrections may be unnecessary, depending on your understanding of hypothesis testing (Gelman, Hill, & Yajima, in prep)
- "The difference between 'significant' and 'not significant' is not itself statistically significant" (Gelman & Stern 2006)

Initial considerations

- Preferred method for hypothesis testing on regression parameters is to include relevant orthogonal contrasts in the original model
- Whether doing planned or *post hoc* comparisons on linear mixed models, distributional assumptions for hypothesis testing are murky at best (*c.f.* `http://wiki.r-project.org/rwiki/doku.php?id=guides:lmer-tests`, `https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html`)
- Hypothesis testing on regression parameters for generalized linear mixed models depends on the data matching distributional assumptions. Over- or underdispersion in the data makes standard error calculations on regression coefficients unreliable.
- Multiplicity corrections may be unnecessary, depending on your understanding of hypothesis testing (Gelman, Hill, & Yajima, in prep)
- "The difference between 'significant' and 'not significant' is not itself statistically significant" (Gelman & Stern 2006)

Initial considerations

- Preferred method for hypothesis testing on regression parameters is to include relevant orthogonal contrasts in the original model
- Whether doing planned or *post hoc* comparisons on linear mixed models, distributional assumptions for hypothesis testing are murky at best (*c.f.* `http://wiki.r-project.org/rwiki/doku.php?id=guides:lmer-tests`, `https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html`)
- Hypothesis testing on regression parameters for generalized linear mixed models depends on the data matching distributional assumptions. Over- or underdispersion in the data makes standard error calculations on regression coefficients unreliable.
- Multiplicity corrections may be unnecessary, depending on your understanding of hypothesis testing (Gelman, Hill, & Yajima, in prep)
- "The difference between 'significant' and 'not significant' is not itself statistically significant" (Gelman & Stern 2006)

Initial considerations

- Preferred method for hypothesis testing on regression parameters is to include relevant orthogonal contrasts in the original model
- Whether doing planned or *post hoc* comparisons on linear mixed models, distributional assumptions for hypothesis testing are murky at best (*c.f.* http://wiki.r-project.org/rwiki/doku.php?id=guides:lmer-tests, https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html)
- Hypothesis testing on regression parameters for generalized linear mixed models depends on the data matching distributional assumptions. Over- or underdispersion in the data makes standard error calculations on regression coefficients unreliable.
- Multiplicity corrections may be unnecessary, depending on your understanding of hypothesis testing (Gelman, Hill, & Yajima, in prep)
- "The difference between 'significant' and 'not significant' is not itself statistically significant" (Gelman & Stern 2006)

Initial considerations

- Preferred method for hypothesis testing on regression parameters is to include relevant orthogonal contrasts in the original model
- Whether doing planned or *post hoc* comparisons on linear mixed models, distributional assumptions for hypothesis testing are murky at best (*c.f.* http://wiki.r-project.org/rwiki/doku.php?id=guides:lmer-tests, https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html)
- Hypothesis testing on regression parameters for generalized linear mixed models depends on the data matching distributional assumptions. Over- or underdispersion in the data makes standard error calculations on regression coefficients unreliable.
- Multiplicity corrections may be unnecessary, depending on your understanding of hypothesis testing (Gelman, Hill, & Yajima, in prep)
- "The difference between 'significant' and 'not significant' is not itself statistically significant" (Gelman & Stern 2006)

Post hoc comparisons: motivations

You may want to use *post hoc* comparisons if you . . .

- are interested in testing hypotheses about specific combinations of levels in a categorical predictor
- can't include all comparisons of interest in planned contrasts (*e.g.* because of collinearity)
- want to evaluate impact of different contrast coding schemes, balance assumptions
- seek to control **family-wise error rate** (probability of rejecting at least one true null hypothesis from a set of null hypotheses)

Post hoc comparisons: motivations

You may want to use *post hoc* comparisons if you ...

- are interested in testing hypotheses about specific combinations of levels in a categorical predictor
- can't include all comparisons of interest in planned contrasts (*e.g.* because of collinearity)
- want to evaluate impact of different contrast coding schemes, balance assumptions
- seek to control **family-wise error rate** (probability of rejecting at least one true null hypothesis from a set of null hypotheses)

Post hoc comparisons: motivations

You may want to use *post hoc* comparisons if you . . .

- are interested in testing hypotheses about specific combinations of levels in a categorical predictor
- can't include all comparisons of interest in planned contrasts (*e.g.* because of collinearity)
- want to evaluate impact of different contrast coding schemes, balance assumptions
- seek to control **family-wise error rate** (probability of rejecting at least one true null hypothesis from a set of null hypotheses)

Post hoc comparisons: motivations

You may want to use *post hoc* comparisons if you . . .

- are interested in testing hypotheses about specific combinations of levels in a categorical predictor
- can't include all comparisons of interest in planned contrasts (*e.g.* because of collinearity)
- want to evaluate impact of different contrast coding schemes, balance assumptions
- seek to control **family-wise error rate** (probability of rejecting at least one true null hypothesis from a set of null hypotheses)

Post hoc comparisons: motivations

You may want to use *post hoc* comparisons if you . . .

- are interested in testing hypotheses about specific combinations of levels in a categorical predictor
- can't include all comparisons of interest in planned contrasts (*e.g.* because of collinearity)
- want to evaluate impact of different contrast coding schemes, balance assumptions
- seek to control **family-wise error rate** (probability of rejecting at least one true null hypothesis from a set of null hypotheses)

Implementing post hoc comparisons

- Use the `multcomp` package in R (Hothorn 2006, 2008)
- Two orthogonal issues:
  1. Specifying contrasts
  2. Multiplicity corrections

Implementing post hoc comparisons

- Use the `multcomp` package in R (Hothorn 2006, 2008)
- Two orthogonal issues:
  1. Specifying contrasts
  2. Multiplicity corrections

Implementing post hoc comparisons

- Use the `multcomp` package in R (Hothorn 2006, 2008)
- Two orthogonal issues:
  1. Specifying contrasts
  2. Multiplicity corrections

Implementing post hoc comparisons

- Use the `multcomp` package in R (Hothorn 2006, 2008)
- Two orthogonal issues:
  1. Specifying contrasts
  2. Multiplicity corrections

Specifying contrasts in `multcomp`

Three methods

1. Named contrast coding scheme (*e.g.* Dunnett, Tukey, Helmert (see p. 648, Bretz *et al*, 2001)
2. String specification ("`A - B = 0`")
3. Build arbitrary contrast matrix from scratch

Specifying contrasts in `multcomp`

Three methods

1. Named contrast coding scheme (*e.g.* Dunnett, Tukey, Helmert (see p. 648, Bretz *et al*, 2001)
2. String specification ("`A - B = 0`")
3. Build arbitrary contrast matrix from scratch

Specifying contrasts in `multcomp`

Three methods

1. Named contrast coding scheme (*e.g.* Dunnett, Tukey, Helmert (see p. 648, Bretz *et al*, 2001)
2. String specification ("`A - B = 0`")
3. Build arbitrary contrast matrix from scratch

Specifying contrasts in `multcomp`

Three methods

1. Named contrast coding scheme (*e.g.* Dunnett, Tukey, Helmert (see p. 648, Bretz *et al*, 2001)
2. String specification (“`A - B = 0`”)
3. Build arbitrary contrast matrix from scratch

multcomp warning

multcomp contrast matrices are transposed from normal R contrast matrices

```
> contr.sum(3)

  [,1] [,2]
1    1    0
2    0    1
3   -1   -1

> contrMat(c(1,1,1), ``Dunnett'')

 Multiple Comparisons of Means: Dunnett Contrasts

       1 2 3
2 - 1 -1 1 0
3 - 1 -1 0 1
```

Use t() to change between orientations

multcomp warning

multcomp contrast matrices are transposed from normal R contrast matrices

```
> contr.sum(3)

  [,1] [,2]
1    1    0
2    0    1
3   -1   -1

> contrMat(c(1,1,1), ``Dunnett'')

 Multiple Comparisons of Means: Dunnett Contrasts

       1  2  3
2 - 1 -1  1  0
3 - 1 -1  0  1
```

Use t() to change between orientations

Worked example

```
> dative.lmer.access <-
+   lmer(RealizationOfRecipient ~
+         -1 + AccessOfTheme +
+         (0 + AccessOfTheme | Verb),
+         dative,
+         family = ``binomial'')

Random effects:
 Groups Name                    Variance Std.Dev. Corr
 Verb   AccessOfThemeaccessible 3.6343   1.9064
        AccessOfThemegiven      4.0210   2.0053   0.938
        AccessOfThemenew        5.3445   2.3118   0.929 0.919
Number of obs: 3263, groups: Verb, 75

Fixed effects:
                        Estimate Std. Error z value Pr(>|z|)
AccessOfThemeaccessible  -0.8172     0.2900  -2.818 0.004836 **
AccessOfThemegiven        1.3781     0.3590   3.839 0.000124 ***
AccessOfThemenew         -1.0370     0.3660  -2.833 0.004609 **
```

Aside: can you spot an error in this model specification?

Collinearity in accessibility model

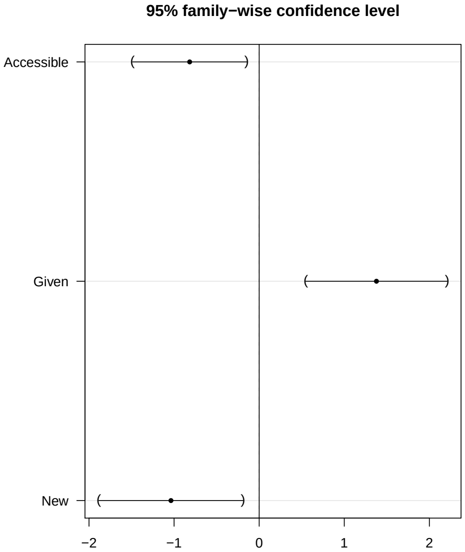

```
Correlation of Fixed Effects:
            AccssOfThmc AccssOfThmg
AccssOfThmg 0.659
AccssOfThmn 0.697       0.593

> collin.fnc(data.frame(dative.lmer.access@X))$cnumber

[1] 1.647076e+14
```

Treatment (dummy) coding


```
> K <- diag(length(fixef(dative.lmer.access)))
> rownames(K) <- c(``Accessible'', ``Given'', ``New'')
> colnames(K) <- c(``Accessible'', ``Given'', ``New'')

           Accessible Given New
Accessible          1     0   0
Given               0     1   0
New                 0     0   1
```

Treatment effects

```
> plot(confint(glht(dative.lmer.access, linfct = K)))
```

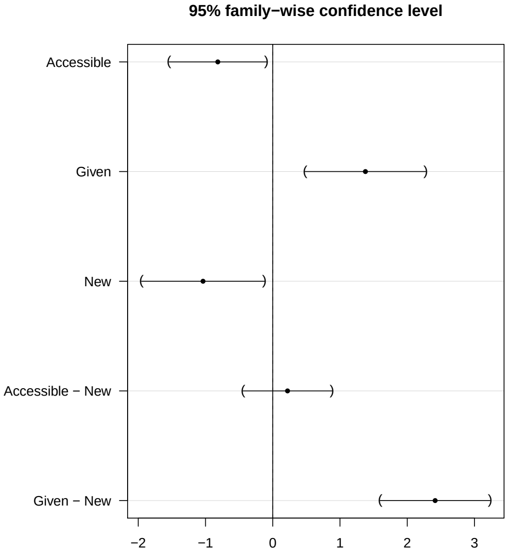


Contrast (Dunnett, sum) coding

```
> K <- rbind(K,
+            ``Accessible - New'' = c(1, 0, -1),
+            ``Given - New'' = c(0, 1, -1))

                  Accessible Given New
Accessible                 1     0   0
Given                      0     1   0
New                        0     0   1
Accessible - New           1     0  -1
Given - New                0     1  -1
```

Contrast effects

```
> plot(confint(glht(dative.lmer.access, linfct = K)))
```

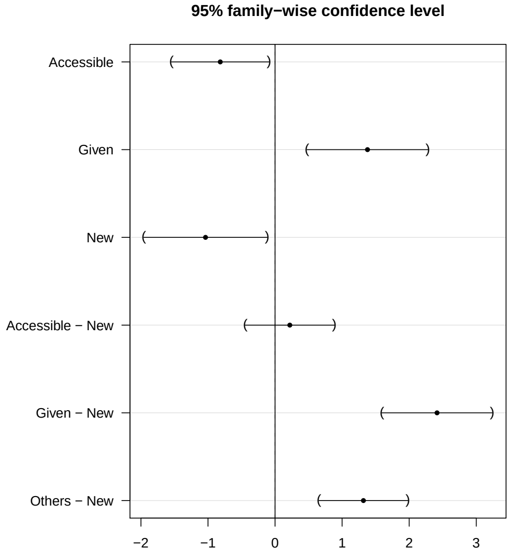


Arbitrary combinations


```
> K <- rbind(K, ``Others - New'' = c(.5, .5, -1))

                  Accessible Given New
Accessible               1.0   0.0   0
Given                    0.0   1.0   0
New                      0.0   0.0   1
Accessible - New         1.0   0.0  -1
Given - New              0.0   1.0  -1
Others - New             0.5   0.5  -1
```

Arbitrary comparison

```
> plot(confint(glht(dative.lmer.access, linfct = K)))
```



Hidden assumption: balanced data

The weights assigned in the contrasts coded tests and the arbitrary test assume that we have balanced data. If this assumption is violated, one condition will contribute more to the test than the others.
Is the accessibility data balanced?

```
> prop.table(table(dative$AccessOfRec))

accessible      given        new
 0.1884769  0.7054857  0.1060374
```

Discussion question / exercise: can we still use these coding strategies with unbalanced data? What should we change?

Multiplicity corrections

- Adjust criterion for rejecting H_0^i as a function of the number of hypotheses being tested.
- Suggested if analysis is being treated as "simultaneous inference".
- Correction procedures fall into two categories: **single step** and **stepwise**. Stepwise procedures require that all combinations of hypotheses including a particular H_0^i before it can be rejected.

Multiplicity corrections

- Adjust criterion for rejecting H_0^i as a function of the number of hypotheses being tested.
- Suggested if analysis is being treated as "simultaneous inference".
- Correction procedures fall into two categories: **single step** and **stepwise**. Stepwise procedures require that all combinations of hypotheses including a particular H_0^i before it can be rejected.

Multiplicity corrections

- Adjust criterion for rejecting H_0^i as a function of the number of hypotheses being tested.
- Suggested if analysis is being treated as "simultaneous inference".
- Correction procedures fall into two categories: **single step** and **stepwise**. Stepwise procedures require that all combinations of hypotheses including a particular H_0^i before it can be rejected.

No correction


```
> summary(glht(dative.lmer.access, linfct = K), test = adjusted(``none''))

 Simultaneous Tests for General Linear Hypotheses

Fit: glmer(formula = RealizationOfRecipient ~ -1 + AccessOfTheme +
    (0 + AccessOfTheme | Verb), data = dative, family = ``binomial'')

Linear Hypotheses:
                      Estimate Std. Error z value Pr(>|z|)
Accessible == 0        -0.8172     0.2900  -2.818 0.004836 **
Given == 0              1.3781     0.3590   3.839 0.000124 ***
New == 0               -1.0370     0.3660  -2.833 0.004609 **
Accessible - New == 0   0.2198     0.2646   0.831 0.406106
Given - New == 0        2.4151     0.3272   7.381 1.57e-13 ***
Others - New == 0       1.3174     0.2638   4.995 5.89e-07 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
(Adjusted p values reported -- none method)
```

Bonferroni correction


```
> summary(glht(dative.lmer.access, linfct = K), test = adjusted(``bon''))


 Simultaneous Tests for General Linear Hypotheses

Fit: glmer(formula = RealizationOfRecipient ~ -1 + AccessOfTheme +
    (0 + AccessOfTheme | Verb), data = dative, family = ``binomial'')

Linear Hypotheses:
                      Estimate Std. Error z value Pr(>|z|)
Accessible == 0        -0.8172     0.2900  -2.818 0.029017 *
Given == 0              1.3781     0.3590   3.839 0.000742 ***
New == 0               -1.0370     0.3660  -2.833 0.027655 *
Accessible - New == 0   0.2198     0.2646   0.831 1.000000
Given - New == 0        2.4151     0.3272   7.381 9.42e-13 ***
Others - New == 0       1.3174     0.2638   4.995 3.53e-06 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
(Adjusted p values reported -- bonferroni method)
```

Westfall correction

```
> summary(glht(dative.lmer.access, linfct = K), test = adjusted(``Westfall''))


 Simultaneous Tests for General Linear Hypotheses

Fit: glmer(formula = RealizationOfRecipient ~ -1 + AccessOfTheme +
    (0 + AccessOfTheme | Verb), data = dative, family = ``binomial'')

Linear Hypotheses:
                       Estimate Std. Error z value Pr(>|z|)
Accessible == 0         -0.8172     0.2900  -2.818   0.0123 *
Given == 0               1.3781     0.3590   3.839   <0.001 ***
New == 0                -1.0370     0.3660  -2.833   0.0123 *
Accessible - New == 0    0.2198     0.2646   0.831   0.4061
Given - New == 0         2.4151     0.3272   7.381   <0.001 ***
Others - New == 0        1.3174     0.2638   4.995   <0.001 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
(Adjusted p values reported -- Westfall method)
```