Statistical Analysis & Reporting in R

Jacob O. Wobbrock, Ph.D.
The Information School | DUB Group
University of Washington
wobbrock@uw.edu
# Table of Analyses

## Tests of Proportion & Association

<table>
<thead>
<tr>
<th>Samples</th>
<th>Categories</th>
<th>Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>One-sample $\chi^2$ test, binomial test</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>One-sample $\chi^2$ test, multinomial test</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>Two-sample $\chi^2$ test, $G$-test, Fisher’s exact test</td>
</tr>
</tbody>
</table>

## Significance Tests

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within</th>
<th>Parametric Tests</th>
<th>Nonparametric Tests &amp; Generalized Linear Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>B</td>
<td>Independent-samples $t$-test</td>
<td>Mann-Whitney $U$ test</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>W</td>
<td>Paired-samples $t$-test</td>
<td>Wilcoxon signed-rank test</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>B</td>
<td>One-way ANOVA</td>
<td>Kruskal-Wallis test</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>W</td>
<td>One-way repeated measures ANOVA</td>
<td>Friedman test</td>
</tr>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>B</td>
<td>Factorial ANOVA Linear Model (LM)</td>
<td>Aligned Rank Transform (ART)</td>
</tr>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>W</td>
<td>Factorial repeated measures ANOVA Linear Mixed Model (LMM)</td>
<td>Generalized Linear Model (GLM)</td>
</tr>
</tbody>
</table>

## Tests of ANOVA Assumptions & Tests of Data Distributions

**Normality:**
- Shapiro-Wilk
- Anderson-Darling

**Homoscedasticity:**
- Levene
- Brown-Forsythe

**Sphericity:**
- Mauchly

**Distribution tests:**
- Kolmogorov-Smirnov
- $\chi^2$ Goodness-of-Fit

---

© Jacob O. Wobbrock, Ph.D.
Analysis Categories

<table>
<thead>
<tr>
<th>Independent Variable (X)</th>
<th>Dependent Variable (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>continuous</td>
<td>continuous</td>
</tr>
<tr>
<td>categorical</td>
<td>categorical</td>
</tr>
</tbody>
</table>

- analysis of variance
- linear regression†
- contingency analysis‡
- logistic regression

Credit: This table is adapted from the “Fit Y by X” dialog in SAS® JMP™. [http://jmp.com/](http://jmp.com/)

†All of the parametric tests in the previous table are forms of linear regression.
‡See the tests of proportion in the previous table.
Tests of Proportion & Association

./data/Proportions/
# Tests of Proportion & Association

## One sample

<table>
<thead>
<tr>
<th>Samples</th>
<th>Response Categories</th>
<th>Test Name</th>
<th>Exact Test?</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | 2                   | Binomial test | Yes, use with N≤200 | # df is a long-format data table w/columns for subject (S) and 2-category outcome (Y)
# df$S = factor(df$S) # Subject id is nominal (unused)
df$Y = factor(df$Y) # Y is an outcome of 2 categories
xt = xtabs(~ Y, data=df) # make counts
binom.test(xt, p=0.5, alternative="two.sided") |
| 1       | ≥2                  | Multinomial test | Yes, use with N≤200 | # df is a long-format data table w/columns for subject (S) and N-category outcome (Y)
library(XNomial) # for xmulti
df$S = factor(df$S) # Subject id is nominal (unused)
df$Y = factor(df$Y) # Y is an outcome of ≥2 categories
xt = xtabs(~ Y, data=df) # make counts
xmulti(xt, rep(1/length(xt), length(xt)), statName="Prob")
# the following gives the same result
library(RVAideMemoire) # for multinomial.test
multinomial.test(df$Y) |

One-sample Pearson Chi-Squared test | No, use with N>200 | # df is a long-format data table w/columns for subject (S) and N-category outcome (Y)
df$S = factor(df$S) # Subject id is nominal (unused)
df$Y = factor(df$Y) # Y is an outcome of ≥2 categories
xt = xtabs(~ Y, data=df) # make counts
chisq.test(xt) |
# Tests of Proportion & Association

## One sample

<table>
<thead>
<tr>
<th>Samples</th>
<th>Response Categories</th>
<th>Test Name</th>
<th>Exact Test?</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Binomial test</td>
<td>Yes, use with $N \leq 200$</td>
<td>“Out of 60 outcomes, 19 were ‘x’ and 41 were ‘y’. A two-sided exact binomial test indicated that these proportions were statistically significantly different from chance ($p &lt; .05$).”</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>Multinomial test</td>
<td>Yes, use with $N \leq 200$</td>
<td>“Out of 60 outcomes, 17 were ‘x’, 8 were ‘y’, and 35 were ‘z’. An exact multinomial test indicated that these proportions were statistically significantly different from chance ($p &lt; .0001$).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One-sample Pearson Chi-Squared test</td>
<td>No, use with $N&gt;200$</td>
<td>“Out of 60 outcomes, 17 were ‘x’, 8 were ‘y’, and 35 were ‘z’. A one-sample Pearson Chi-Squared test indicated that these proportions were statistically significantly different from chance ($\chi^2(2, N=60) = 18.90, p &lt; .0001$).”</td>
</tr>
</tbody>
</table>
## Tests of Proportion & Association

### Two samples

<table>
<thead>
<tr>
<th>Samples</th>
<th>Response Categories</th>
<th>Test Name</th>
<th>Exact Test?</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 2       | ≥2                  | Fisher’s exact test          | Yes, use with \( N \leq 200 \) | # df is a long-format data table w/subject (S), categorical factor (X) and outcome (Y)  
  df$S = factor(df$S) # Subject id is nominal (unused)  
  df$X = factor(df$X) # X is a factor of \( m \geq 2 \) levels  
  df$Y = factor(df$Y) # Y is an outcome of \( n \geq 2 \) categories  
  xt = xtabs(~ X + Y, data=df) # make \( m \times n \) crosstabs  
  fisher.test(xt) |
|         |                     | G-test                       | No, use with \( N > 200 \) | # df is a long-format data table w/subject (S), categorical factor (X) and outcome (Y)  
  library(RVAideMemoire) # for G.test  
  df$S = factor(df$S) # Subject id is nominal (unused)  
  df$X = factor(df$X) # X is a factor of \( m \geq 2 \) levels  
  df$Y = factor(df$Y) # Y is an outcome of \( n \geq 2 \) categories  
  xt = xtabs(~ X + Y, data=df) # make \( m \times n \) crosstabs  
  G.test(xt) |
|         |                     | Two-sample Pearson Chi-Squared test | No, use with \( N > 200 \) | # df is a long-format data table w/subject (S), categorical factor (X) and outcome (Y)  
  df$S = factor(df$S) # Subject id is nominal (unused)  
  df$X = factor(df$X) # X is a factor of \( m \geq 2 \) levels  
  df$Y = factor(df$Y) # Y is an outcome of \( n \geq 2 \) categories  
  xt = xtabs(~ X + Y, data=df) # make \( m \times n \) crosstabs  
  chisq.test(xt) |
## Tests of Proportion & Association

### Two samples

<table>
<thead>
<tr>
<th>Samples</th>
<th>Response Categories</th>
<th>Test Name</th>
<th>Exact Test?</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>≥2</td>
<td>Fisher’s exact test</td>
<td>Yes, use with $N\leq200$</td>
<td>“Table 1 shows the counts of the ‘x’, ‘y’, and ‘z’ outcomes for each of ‘a’ and ‘b’. Fisher’s exact test indicated a statistically significant association between X and Y ($p &lt; .0001$).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G-test</td>
<td>No, use with $N&gt;200$</td>
<td>“Table 1 shows the counts of the ‘x’, ‘y’, and ‘z’ outcomes for each of ‘a’ and ‘b’. A $G$-test indicated a statistically significant association between X and Y ($G(2) = 21.40, p &lt; .0001$).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two-sample Pearson Chi-Squared test</td>
<td>No, use with $N&gt;200$</td>
<td>“Table 1 shows the counts of the ‘x’, ‘y’, and ‘z’ outcomes for each of ‘a’ and ‘b’. A two-sample Pearson Chi-Squared test indicated a statistically significant association between X and Y ($\chi^2(2, N=60) = 19.88, p &lt; .0001$).”</td>
</tr>
<tr>
<td>X</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>3</td>
<td>26</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>14</td>
<td>9</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>
Tests of Proportion & Association

Post hoc comparisons
## Tests of Proportion & Association

*Post hoc* tests – One sample

<table>
<thead>
<tr>
<th>Samples</th>
<th>Response Categories</th>
<th>Omnibus Test</th>
<th>Contrast Test</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥ 2                 | Multinomial test             | Binomial test | `# xt is a table of counts for each category of Y
library(RVAideMemoire) # for multinomial.multcomp
multinomial.multcomp(xt, p.method="holm") # xt shows levels`                                                                                                                                           |
| 1       | ≥ 2                 | One-sample Pearson Chi-Squared test | Chi-Squared test | `# xt is a table of counts for each category of Y
library(RVAideMemoire) # for chisq.multcomp
chisq.multcomp(xt, p.method="holm") # xt shows levels
# for the Chi-Squared values, use qchisq(1-p, df=1), where p is the pairwise p-value.`                                                                                                      |
| 1       | ≥ 2                 | Multinomial test, One-sample Pearson Chi-Squared test | Binomial tests | `# A different kind of post hoc test for one sample. For Y's response categories (x,y,z),
# test each proportion against chance.
x = binom.test(sum(df$Y == "x"), nrow(df), p=1/3) # proportion of "x" rows
y = binom.test(sum(df$Y == "y"), nrow(df), p=1/3) # proportion of "y" rows
z = binom.test(sum(df$Y == "z"), nrow(df), p=1/3) # proportion of "z" rows
p.adjust(c(x$p.value, y$p.value, z$p.value), method="holm")`                                                                                       |
# Tests of Proportion & Association

*Post hoc tests – One sample*

<table>
<thead>
<tr>
<th>Samples</th>
<th>Response Categories</th>
<th>Omnibus Test</th>
<th>Contrast Test</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Multinomial test</td>
<td>Binomial test</td>
<td>“Pairwise comparisons using exact binomial tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the proportions of ‘x’ and ‘z’ and ‘y’ and ‘z’ were statistically significantly different ( (p &lt; .05) ), but that the proportions of ‘x’ and ‘y’ were not.”</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>One-sample Pearson Chi-Squared test</td>
<td>Chi-Squared</td>
<td>“Pairwise comparisons using Pearson Chi-Squared tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the proportions of ‘x’ and ‘z’ were statistically significantly different ( (\chi^2(1, N=52) = 5.02, p &lt; .05) ), as were the proportions of ‘y’ and ‘z’ ( (\chi^2(1, N=43) = 14.96, p &lt; .001) ), but that the proportions of ‘x’ and ‘y’ were only marginal ( (\chi^2(1, N=25) = 3.24, p = .072) ).”</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>Multinomial test, One-sample Pearson Chi-Squared test</td>
<td>Binomial tests</td>
<td>“Three <em>post hoc</em> binomial tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the proportions of ‘y’ and ‘z’ were statistically significantly different from chance ( (p &lt; .05) ), but not the proportion of ‘x’. Specifically, the proportion of ‘y’ at 8 was significantly lower than 20, and the proportion of ‘z’ at 35 was significantly higher than 20.”</td>
</tr>
<tr>
<td>Samples</td>
<td>Response Categories</td>
<td>Omnibus Test</td>
<td>Contrast Test</td>
<td>R Code</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>--------</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>Fisher’s exact test for 2×2 crosstabs</td>
<td>Fisher’s exact test for 2×2 crosstabs</td>
<td># xt is an m × n crosstabs with categories X and Y library(RVAideMemoire) # for fisher.multcomp fisher.multcomp(xt, p.method=&quot;holm&quot;)</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>G-test</td>
<td>G-test</td>
<td># xt is an m × n crosstabs with categories X and Y library(RVAideMemoire) # for G.multcomp G.multcomp(xt, p.method=&quot;holm&quot;) # xt shows levels</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>Two-sample Pearson Chi-Squared test</td>
<td>Chi-Squared test</td>
<td># xt is an m × n crosstabs with categories X and Y library(RVAideMemoire) # for chisq.multcomp chisq.multcomp(xt, p.method=&quot;holm&quot;) # xt shows levels</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>Fisher’s exact test, G-test, Two-sample Pearson Chi-Squared test</td>
<td>Binomial tests</td>
<td># A different kind of post hoc test for two samples. For X’s categories (a,b) and Y’s response categories (x,y,z), test each proportion of Y within each level of X against chance. # df is a long-format data table w/subject (S), categorical factor (X) and outcome (Y). ax = binom.test(sum(df[df$X == &quot;a&quot;,]$Y == &quot;x&quot;), nrow(df[df$X == &quot;a&quot;],), p=1/3) ay = binom.test(sum(df[df$X == &quot;a&quot;,]$Y == &quot;y&quot;), nrow(df[df$X == &quot;a&quot;],), p=1/3) az = binom.test(sum(df[df$X == &quot;a&quot;,]$Y == &quot;z&quot;), nrow(df[df$X == &quot;a&quot;],), p=1/3) p.adjust(c(ax$p.value, ay$p.value, az$p.value), method=&quot;holm&quot;) bx = binom.test(sum(df[df$X == &quot;b&quot;,]$Y == &quot;x&quot;), nrow(df[df$X == &quot;b&quot;],), p=1/3) by = binom.test(sum(df[df$X == &quot;b&quot;,]$Y == &quot;y&quot;), nrow(df[df$X == &quot;b&quot;],), p=1/3) bz = binom.test(sum(df[df$X == &quot;b&quot;,]$Y == &quot;z&quot;), nrow(df[df$X == &quot;b&quot;],), p=1/3) p.adjust(c(bx$p.value, by$p.value, bz$p.value), method=&quot;holm&quot;)</td>
</tr>
</tbody>
</table>

Tests of Proportion & Association

*Post hoc tests – Two samples*
## Tests of Proportion & Association

*Post hoc* tests – Two samples

<table>
<thead>
<tr>
<th>Samples</th>
<th>Response Categories</th>
<th>Omnibus Test</th>
<th>Contrast Test</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>≥2</td>
<td>Fisher’s exact test</td>
<td>Fisher’s exact test for 2×2 crosstabs</td>
<td>“Three <em>post hoc</em> Fisher’s exact tests, one for each 2×2 crosstabulation of X and Y, corrected with Holm’s sequential Bonferroni procedure, indicated that there were statistically significant associations for (a,b) × (x,y) (p &lt; .001) and (a,b) × (y,z) (p &lt; .01), but not for (a,b) × (x,z).”</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>G-test</td>
<td>G-test</td>
<td>“Pairwise comparisons using <em>G</em>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,x) vs. (a,y), (a,z) vs. (b,x), (a,z) vs. (a,y), (b,y) vs. (a,y), and (b,z) vs. (a,y) were all statistically significantly different (p &lt; .05). Furthermore, (a,x) vs. (b,x) (p = .055) was marginal, as was (a,z) vs. (b,y) (p = .060). The eight other possible pairwise comparisons were not statistically significant.”</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>Two-sample Pearson Chi-Squared test</td>
<td>Chi-Squared test</td>
<td>“Pairwise comparisons using Pearson Chi-Square tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,x) vs. (a,y), (a,z) vs. (b,x), (a,z) vs. (a,y), (b,y) vs. (a,y), (b,z) vs. (a,y) were all statistically significantly different (p &lt; .05). Furthermore, (a,x) vs. (b,x) (p = .076) was marginal. The nine other possible pairwise comparisons were not statistically significant.”</td>
</tr>
<tr>
<td>2</td>
<td>≥2</td>
<td>Fisher’s exact test, <em>G</em>-test, Two-sample Pearson Chi-Squared test</td>
<td>Binomial tests</td>
<td>“Three <em>post hoc</em> binomial tests, corrected with Holm’s sequential Bonferroni procedure, indicated that, within ‘a’, the proportions of ‘x’, ‘y’, and ‘z’ were all significantly different from chance (i.e., 10 each) (p &lt; .01). Three similar tests within ‘b’ indicated that none of the proportions of ‘x’, ‘y’, or ‘z’ significantly differed from chance (i.e., 10 each).”</td>
</tr>
</tbody>
</table>
Assumptions & Distributions

./data/Assumptions/
## Tests of ANOVA Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Test Name</th>
<th>Context of Use</th>
<th>R Code</th>
</tr>
</thead>
</table>
| Normality    | Shapiro-Wilk test (on response in each condition) | t-test, ANOVA, LM, LMM             | # assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y  
shapiro.test(df[df$X1 == "a" & df$X2 == "a",]$Y) # condition a,a  
shapiro.test(df[df$X1 == "a" & df$X2 == "b",]$Y) # condition a,b  
shapiro.test(df[df$X1 == "b" & df$X2 == "a",]$Y) # condition b,a  
shapiro.test(df[df$X1 == "b" & df$X2 == "b",]$Y) # condition b,b |
| Normality    | Shapiro-Wilk test (on residuals)         | t-test, ANOVA, LM, LMM             | # assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y  
m = aov(Y ~ X1*X2, data=df) # make anova model  
shapiro.test(residuals(m))  
qqnorm(residuals(m)); qqline(residuals(m)) # Q-Q plot |
| Normality    | Anderson-Darling test (on response in each condition) | t-test, ANOVA, LM, LMM             | # assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y  
library(nortest) # for ad.test  
ad.test(df[df$X1 == "a" & df$X2 == "a",]$Y) # condition a,a  
ad.test(df[df$X1 == "a" & df$X2 == "b",]$Y) # condition a,b  
ad.test(df[df$X1 == "b" & df$X2 == "a",]$Y) # condition b,a  
ad.test(df[df$X1 == "b" & df$X2 == "b",]$Y) # condition b,b |
| Normality    | Anderson-Darling test (on residuals)      | t-test, ANOVA, LM, LMM             | # assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y  
library(nortest) # for ad.test  
m = aov(Y ~ X1*X2, data=df) # make anova model  
ad.test(residuals(m))  
qqnorm(residuals(m)); qqline(residuals(m)) # Q-Q plot |
# Tests of ANOVA Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Test Name</th>
<th>Context of Use</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normality</td>
<td>Shapiro-Wilk test (on response in each condition)</td>
<td>t-test, ANOVA, LM, LMM</td>
<td>“To test the assumption of conditional normality, a Shapiro-Wilk test was run on the response Y for each combination of levels of factors X1 and X2. All combinations were found to be statistically non-significant except condition (b,b), which showed a statistically significant deviation from normality ($W = .794, p &lt; .01$).”</td>
</tr>
<tr>
<td>Normality</td>
<td>Shapiro-Wilk test (on residuals)</td>
<td>t-test, ANOVA, LM, LMM</td>
<td>“To test the normality assumption, a Shapiro-Wilk test was run on the residuals of a between-subjects full-factorial ANOVA model. The test was statistically non-significant ($W = .988, p = .798$), indicating compliance with the normality assumption. A Q-Q plot of residuals visually confirms the same (Figure 1).”</td>
</tr>
<tr>
<td>Normality</td>
<td>Anderson-Darling test (on response in each condition)</td>
<td>t-test, ANOVA, LM, LMM</td>
<td>“To test the assumption of conditional normality, an Anderson-Darling test was run on the response Y for each combination of levels of factors X1 and X2. All combinations were found to be statistically non-significant except condition (b,b), which showed a statistically significant deviation from normality ($A = 1.417, p &lt; .001$).”</td>
</tr>
<tr>
<td>Normality</td>
<td>Anderson-Darling test (on residuals)</td>
<td>t-test, ANOVA, LM, LMM</td>
<td>“To test the normality assumption, an Anderson-Darling test was run on the residuals of a between-subjects full-factorial ANOVA model. The test was statistically non-significant ($A = 0.329, p = .510$), indicating compliance with the normality assumption. A Q-Q plot of residuals visually confirms the same (Figure 1).”</td>
</tr>
</tbody>
</table>
Figure 1

Normal Q-Q Plot

2F2LBs_normal.csv
<table>
<thead>
<tr>
<th>Assumption</th>
<th>Test Name</th>
<th>Context of Use</th>
<th>R Code</th>
</tr>
</thead>
</table>
| Homoscedasticity, a.k.a. homogeneity of variance | Levene’s test                  | t-test, ANOVA, LM, LMM    | # assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y  
  library(car) # for leveneTest  
  leveneTest(Y ~ X1*X2, data=df, center=mean)  
  # if a violation occurs and only a t-test is needed, use a Welch t-test  
  t.test(Y ~ X1, data=df, var.equal=FALSE) # Welch t-test |
| Homoscedasticity, a.k.a. homogeneity of variance | Brown-Forsythe test            | t-test, ANOVA, LM, LMM    | # assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y  
  library(car) # for leveneTest  
  leveneTest(Y ~ X1*X2, data=df, center=median)  
  # if a violation occurs and only a t-test is needed, use a Welch t-test  
  t.test(Y ~ X1, data=df, var.equal=FALSE) # Welch t-test |
| Sphericity                                    | Mauchly’s test of sphericity   | repeated measures ANOVA  | # assume df has subjects (S), one between-Ss factor (X1), and one within-Ss factor (X2)  
  library(ez) # for ezANOVA  
  df$S = factor(df$S) # Subject id is nominal  
  m = ezANOVA(dv=Y, between=c(X1), within=c(X2), wid=S, data=df) # use c() for multiple factors  
  m$Mauchly # p<.05 indicates a sphericity violation for within-Ss effects |
## Tests of ANOVA Assumptions

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Test Name</th>
<th>Context of Use</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homoscedasticity, a.k.a. homogeneity of variance</td>
<td>Levene’s test</td>
<td>t-test, ANOVA, LM, LMM</td>
<td>“To test the homoscedasticity assumption, Levene’s test was run on a between-subjects full-factorial ANOVA model. The test was statistically significant (F(3, 56) = 3.97, ( p &lt; .05 )), indicating a departure from homoscedasticity.”</td>
</tr>
<tr>
<td>Homoscedasticity, a.k.a. homogeneity of variance</td>
<td>Brown-Forsythe test</td>
<td>t-test, ANOVA, LM, LMM</td>
<td>“To test the homoscedasticity assumption, the Brown-Forsythe test was run on a between-subjects full-factorial ANOVA model. The test was statistically significant (F(3, 56) = 3.75, ( p &lt; .05 )), indicating a departure from homoscedasticity.”</td>
</tr>
<tr>
<td>Sphericity</td>
<td>Mauchly’s test of sphericity</td>
<td>repeated measures ANOVA</td>
<td>“To test the sphericity assumption for repeated measures ANOVA, Mauchly’s test of sphericity was run on a mixed factorial ANOVA model with a between-subjects factor X1 and a within-subjects factor X2. The test was statistically significant for both X2 (( W = .637, p &lt; .01 )) and X1×X2 (( W = .637, p &lt; .01 )), indicating sphericity violations. Accordingly, the Greenhouse-Geisser correction was used when reporting these ANOVA results.”</td>
</tr>
</tbody>
</table>
### Distribution Tests

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameterization</th>
<th>R Distribution Funcs</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>mean (µ): mean</td>
<td>_norm:</td>
<td># assume df has one factor (X) w/two levels (a,b) and continuous response Y</td>
</tr>
<tr>
<td></td>
<td>standard deviation (σ): sd</td>
<td>dnorm pnorm qnorm rnorm</td>
<td>library(MASS) # for fitdistr</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fa = fitdistr(df[df$X == &quot;a&quot;,]$Y, &quot;normal&quot;)$estimate # create fit for X.a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ks.test(df[df$X == &quot;a&quot;,]$Y, &quot;pnorm&quot;, mean=fa[1], sd=fa[2])</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fb = fitdistr(df[df$X == &quot;b&quot;,]$Y, &quot;normal&quot;)$estimate # create fit for X.b</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ks.test(df[df$X == &quot;b&quot;,]$Y, &quot;pnorm&quot;, mean=fb[1], sd=fb[2])</td>
<td></td>
</tr>
<tr>
<td>Lognormal</td>
<td>mean (µ): meanlog standard deviation (σ): sdlog</td>
<td>_lnorm:</td>
<td># assume df has one factor (X) w/two levels (a,b) and continuous response Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dlnorm plnorm qlnorm rlnorm</td>
<td>library(MASS) # for fitdistr</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fa = fitdistr(df[df$X == &quot;a&quot;,]$Y, &quot;lognormal&quot;)$estimate # create fit for X.a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ks.test(df[df$X == &quot;a&quot;,]$Y, &quot;plnorm&quot;, meanlog=fa[1], sdlog=fa[2])</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fb = fitdistr(df[df$X == &quot;b&quot;,]$Y, &quot;lognormal&quot;)$estimate # create fit for X.b</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ks.test(df[df$X == &quot;b&quot;,]$Y, &quot;plnorm&quot;, meanlog=fb[1], sdlog=fb[2])</td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td>lambda (λ): lambda</td>
<td>_pois:</td>
<td># assume df has one factor (X) w/two levels (a,b) and nonnegative integer response Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dpois ppois qpois rpois</td>
<td>library(fitdistrplus) # for fitdist, gofstat</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fa = fitdist(df[df$X == &quot;a&quot;,]$Y, &quot;pois&quot;) # create fit for X.a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>gofstat(fa)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fb = fitdist(df[df$X == &quot;b&quot;,]$Y, &quot;pois&quot;) # create fit for X.b</td>
<td></td>
</tr>
<tr>
<td>gofstat(fa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>theta (θ): size mu (µ): mu</td>
<td>_nbinom:</td>
<td># assume df has one factor (X) w/two levels (a,b) and nonnegative integer response Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dnbibnom pnbibnom qnbibnom rnbibnom</td>
<td>library(fitdistrplus) # for fitdist, gofstat</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fa = fitdist(df[df$X == &quot;a&quot;,]$Y, &quot;nbinom&quot;) # create fit for X.a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>gofstat(fa)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fb = fitdist(df[df$X == &quot;b&quot;,]$Y, &quot;nbinom&quot;) # create fit for X.b</td>
<td></td>
</tr>
<tr>
<td>gofstat(fa)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Distribution Tests

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameterization</th>
<th>R Distribution Funcs</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>mean (µ): mean</td>
<td>_norm: dnorm</td>
<td>“Figure 2 shows the distributions of response Y for both levels of factor X. To test whether these distributions were normally distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .158, p = .404$), as was the test for level ‘b’ ($D = .104, p = .867$), indicating non-detectable deviations from a normal distribution for both levels.”</td>
</tr>
<tr>
<td></td>
<td>standard deviation (σ): sd</td>
<td>pnorm qnorm rnorm</td>
<td></td>
</tr>
<tr>
<td>Lognormal</td>
<td>mean (µ): meanlog</td>
<td>_lnorm: dlnorm</td>
<td>“Figure 3 shows the distributions of response Y for both levels of factor X. To test whether these distributions were lognormally distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .096, p = .918$), as was the test for level ‘b’ ($D = .161, p = .375$), indicating non-detectable deviations from a lognormal distribution for both levels.”</td>
</tr>
<tr>
<td></td>
<td>standard deviation (σ): sdlog</td>
<td>plnorm qlnorm rlnorm</td>
<td></td>
</tr>
<tr>
<td>Poisson</td>
<td>lambda (λ): lambda</td>
<td>_pois: dpois</td>
<td>“Figure 4 shows the distributions of response Y for both levels of factor X. To test whether these distributions were Poisson distributed, a Chi-Squared goodness-of-fit test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($\chi^2(3, N=30) = 2.62, p = .454$), as was the test for level ‘b’ ($\chi^2(4, N=30) = 2.79, p = .593$), indicating non-detectable deviations from a Poisson distribution for both levels.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ppois qpois rpois</td>
<td></td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>theta (θ): size</td>
<td>_nbinom: dnbilnom</td>
<td>“Figure 5 shows the distributions of response Y for both levels of factor X. To test whether these distributions were negative binomially distributed, a Chi-Squared goodness-of-fit test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($\chi^2(4, N=30) = 1.74, p = .783$), as was the test for level ‘b’ ($\chi^2(3, N=30) = 1.27, p = .737$), indicating non-detectable deviations from a negative binomial distribution for both levels.”</td>
</tr>
<tr>
<td></td>
<td>mu (µ): mu</td>
<td>qnbilnom rnbilnom</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2

Histograms of Y by X

Count

Y

Histogram for Y frequency distribution by X categories a and b.
Figure 3

Histories of Y by X

Counts

Y

0 10 20 30 40 50 60 70 80

1F2LBs_lognormal.csv
Figure 4

Histograms of Y by X

Count

Y

X

a

b

1F2LBs_poisson.csv
Figure 5

Histograms of Y by X

Count

Y

X

a

b

1F2LBs_negbin.csv
## Distribution Tests

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameterization</th>
<th>R Distribution Fns</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>rate (λ): rate</td>
<td>dexp, pexp, qexp, rexp</td>
<td># assume df has one factor (X) with two levels (a,b) and continuous response Y library(MASS) # for fitdistr fa = fitdistr(df[df$X == &quot;a&quot;,]$Y, &quot;exponential&quot;)$estimate # create fit for X.a ks.test(df[df$X == &quot;a&quot;,]$Y, &quot;pexp&quot;, rate=fa[1]) fb = fitdistr(df[df$X == &quot;b&quot;,]$Y, &quot;exponential&quot;)$estimate # create fit for X.b ks.test(df[df$X == &quot;b&quot;,]$Y, &quot;pexp&quot;, rate=fb[1])</td>
</tr>
<tr>
<td>Gamma</td>
<td>shape (α): shape rate (β): rate</td>
<td>dgamma, pgamma, qgamma, rgamma</td>
<td># assume df has one factor (X) with two levels (a,b) and continuous response Y library(MASS) # for fitdistr fa = fitdistr(df[df$X == &quot;a&quot;,]$Y, &quot;gamma&quot;)$estimate # create fit for X.a ks.test(df[df$X == &quot;a&quot;,]$Y, &quot;pgamma&quot;, shape=fa[1], rate=fa[2]) fb = fitdistr(df[df$X == &quot;b&quot;,]$Y, &quot;gamma&quot;)$estimate # create fit for X.b ks.test(df[df$X == &quot;b&quot;,]$Y, &quot;pgamma&quot;, shape=fb[1], rate=fb[2])</td>
</tr>
</tbody>
</table>
Distribution Tests

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameterization</th>
<th>R Distribution Fns</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>rate (λ): rate</td>
<td>_exp: dexp pexp qexp rexp</td>
<td>“Figure 6 shows the distributions of response Y for both levels of factor X. To test whether these distributions were exponentially distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .107, p = .849$), as was the test for level ‘b’ ($D = .119, p = .742$), indicating non-detectable deviations from an exponential distribution for both levels.”</td>
</tr>
<tr>
<td>Gamma</td>
<td>shape (α): shape rate (β): rate</td>
<td>_gamma: dgamma pgamma qgamma rgamma</td>
<td>“Figure 7 shows the distributions of response Y for both levels of factor X. To test whether these distributions were gamma distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .116, p = .773$), as was the test for level ‘b’ ($D = .143, p = .526$), indicating non-detectable deviations from a gamma distribution for both levels.”</td>
</tr>
</tbody>
</table>
Figure 6

Histograms of Y by X

Count

0 5 10 15 20 25 30 35 40 45 50

X

Y

0 2 4 6 8 10 12 14

a

b

1F2LBs_exponential.csv
Figure 7

Histograms of Y by X

Count

Y

1F2LBs_gamma.csv
Parametric Tests
./data/Parametric/
## Parametric Tests

### One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Between</td>
<td>Independent-samples t-test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Within</td>
<td>Paired-samples t-test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 ≥2</td>
<td></td>
<td>Between</td>
<td>One-way ANOVA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### One factor

- **Between**
  - **Independent-samples t-test**
    - `df` has subjects (S), one between-Ss factor (X) w/levels (a,b), and continuous response (Y)
    - `df$S = factor(df$S)` # Subject id is nominal (unused)
    - `df$X = factor(df$X)` # X is a 2-level factor
    - `t.test(Y ~ X, data=df, var.equal=TRUE)` # use var.equal=FALSE if heteroscedastic

- **Paired-samples t-test**
  - `df$S = factor(df$S)` # Subject id is nominal
  - `df$X = factor(df$X)` # X is a 2-level factor
  - `df2 <- dcast(df, S ~ X, value.var="Y")` # make wide-format table
  - `t.test(df2$a, df2$b, paired=TRUE, var.equal=TRUE)` # use var.equal=FALSE if heteroscedastic

- **One-way ANOVA**
  - `df` has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and continuous response (Y)
  - `df$S = factor(df$S)` # Subject id is nominal (unused)
  - `df$X = factor(df$X)` # X is a 3-level factor
  - `m = aov(Y ~ X, data=df)` # fit model
  - `anova(m)`
<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Between</td>
<td>Independent-samples t-test</td>
<td>“The mean of ‘a’ was 14.63 ($SD = 2.13$) and of ‘b’ was 11.01 ($SD = 1.75$). This difference was statistically significant according to an independent-samples t-test ($t(58) = 7.18$, $p &lt; .0001$).”</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Within</td>
<td>Paired-samples t-test</td>
<td>“The mean of ‘a’ was 13.15 ($SD = 2.53$) and of ‘b’ was 14.37 ($SD = 2.16$). This difference was statistically significant according to a paired-samples t-test ($t(29) = -2.14$, $p &lt; .05$).”</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>Between</td>
<td>One-way ANOVA</td>
<td>“The mean of ‘a’ was 13.74 ($SD = 2.84$), of ‘b’ was 14.15 ($SD = 2.65$), and of ‘c’ was 9.08 ($SD = 4.29$). These differences were statistically significant according to a one-way ANOVA ($F(2, 57) = 14.18$, $p &lt; .0001$).”</td>
</tr>
</tbody>
</table>

Note: “$SD$” stands for “standard deviation,” i.e., the spread of values around the mean.
## Parametric Tests

### One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Within                      | One-way repeated measures ANOVA | `# df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y)`
|         |        |                             |           | `library(ez) # for ezANOVA` |
|         |        |                             |           | `df$S = factor(df$S) # Subject id is nominal` |
|         |        |                             |           | `df$X = factor(df$X) # X is a 3-level factor` |
|         |        |                             |           | `m = ezANOVA(dv=Y, within=c(X), wid=S, data=df) # use c() for multiple factors` |
|         |        |                             |           | `m$Mauchly # p<.05 indicates a sphericity violation` |
|         |        |                             |           | `m$ANOVA # use if no violation` |
|         |        |                             |           | `# if there is a sphericity violation, report the Greenhouse-Geisser or Huynh-Feldt correction` |
|         |        |                             |           | `p = match(m$Sphericity$Effect, m$ANOVA$Effect) # positions of within-Ss effects in m$ANOVA` |
|         |        |                             |           | `m$Sphericity$GGe.DFn = m$Sphericity$GGe * m$ANOVA$DFn[p] # Greenhouse-Geisser DFs` |
|         |        |                             |           | `m$Sphericity$GGe.DFd = m$Sphericity$GGe * m$ANOVA$DFd[p]` |
|         |        |                             |           | `m$Sphericity$HFe.DFn = m$Sphericity$HFe * m$ANOVA$DFn[p] # Huynh-Feldt DFs` |
|         |        |                             |           | `m$Sphericity$HFe.DFd = m$Sphericity$HFe * m$ANOVA$DFd[p]` |
|         |        |                             |           | `m$Sphericity # show results` |
|         |        |                             |           | `# the following also performs the equivalent repeated measures ANOVA, but does not address sphericity` |
|         |        |                             |           | `m = aov(Y ~ X + Error(S/X), data=dfs)` |
|         |        |                             |           | `summary(m)` |
## Parametric Tests
### One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Within</td>
<td>One-way repeated measures ANOVA</td>
<td>“The mean of ‘a’ was 14.04 (SD = 2.98), of ‘b’ was 11.95 (SD = 1.98), and of ‘c’ was 11.40 (SD = 2.75). Mauchly’s test of sphericity indicated no sphericity violation ($W = .926, p = .499$), allowing for an uncorrected repeated measures ANOVA, which showed statistically significant differences ($F(2, 38) = 6.57, p &lt; .01$).”</td>
</tr>
</tbody>
</table>

---

### Within One-way repeated measures ANOVA

The mean of ‘a’ was 14.04 (SD = 2.98), of ‘b’ was 11.95 (SD = 1.98), and of ‘c’ was 11.40 (SD = 2.75). Mauchly’s test of sphericity indicated no sphericity violation ($W = .926, p = .499$), allowing for an uncorrected repeated measures ANOVA, which showed statistically significant differences ($F(2, 38) = 6.57, p < .01$).
## Parametric Tests
### Multiple between-Ss factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Between</td>
<td>Factorial ANOVA</td>
<td># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(ez) # for ezANOVA df$S = factor(df$S) # Subject id is nominal df$X1 = factor(df$X1) # X1 is a 2-level factor df$X2 = factor(df$X2) # X2 is a 2-level factor m = ezANOVA(dv=Y, between=c(X1,X2), wid=S, data=df) # use c() for multiple factors m$ANOVA # no sphericity concerns with only between-Ss factors # the following also performs the equivalent factorial ANOVA m = aov(Y ~ X1*X2, data=df) anova(m)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Linear Model (LM)</td>
<td># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) df$S = factor(df$S) # Subject id is nominal (unused) df$X1 = factor(df$X1) # X1 is a 2-level factor df$X2 = factor(df$X2) # X2 is a 2-level factor m = lm(Y ~ X1*X2, data=df) anova(m)</td>
</tr>
<tr>
<td>Factors</td>
<td>Levels</td>
<td>Between or Within Subjects</td>
<td>Test Name</td>
<td>Report</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>-----------------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Between</td>
<td>Factorial ANOVA</td>
<td>“Figure 8 shows an interaction plot with ±1 standard deviation error bars for X1 and X2. A factorial ANOVA indicated a statistically significant effect on Y of X1 (F(1, 56) = 9.35, p &lt; .01) and X2 (F(1, 56) = 4.30, p &lt; .05), but not of the X1×X2 interaction (F(1, 56) = 0.42, n.s.).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Linear Model (LM)</td>
<td>“Figure 8 shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance indicated a statistically significant effect on Y of X1 (F(1, 56) = 9.35, p &lt; .01) and X2 (F(1, 56) = 4.30, p &lt; .05), but not of the X1×X2 interaction (F(1, 56) = 0.42, n.s.).”</td>
</tr>
</tbody>
</table>
Figure 8

Y by X1, X2

X1

a

b

X2

Diamond a

Circle b

13.35 (±3.63)

11.57 (±1.51)

13.99 (±0.98)

14.93 (±3.04)

2F2LBs.csv
## Parametric Tests

### Multiple within-Ss factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Within</td>
<td>Factorial repeated measures ANOVA</td>
<td># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(ez) # for ezANOVA df$S = factor(df$S) # Subject id is nominal df$X1 = factor(df$X1) # X1 is a 2-level factor df$X2 = factor(df$X2) # X2 is a 2-level factor m = ezANOVA(dv=Y, within=c(X1,X2), wid=S, data=df) # use c() for multiple factors m$Mauchly # p&lt;.05 indicates a sphericity violation m$ANOVA # use if no violation # if there is a sphericity violation, report the Greenhouse-Geisser or Huynh-Feldt correction p = match(m$Sphericity$Effect, m$ANOVA$Effect) # positions of within-Ss effects in m$ANOVA m$Sphericity$GGe.DFn = m$Sphericity$GGe * m$ANOVA$DFn[p] # Greenhouse-Geisser DFs m$Sphericity$GGe.DFd = m$Sphericity$GGe * m$ANOVA$DFd[p] m$Sphericity$HFe.DFn = m$Sphericity$HFe * m$ANOVA$DFn[p] # Huynh-Feldt DFs m$Sphericity$HFe.DFd = m$Sphericity$HFe * m$ANOVA$DFd[p] m$Sphericity # show results # the following also performs the equivalent repeated measures ANOVA, but does not address sphericity m = aov(Y ~ X1<em>X2 + Error(S/(X1</em>X2)), data=df) summary(m)</td>
</tr>
</tbody>
</table>
### Parametric Tests

**Multiple within-Ss factors**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Within</td>
<td>Factorial repeated measures ANOVA</td>
<td>“Figure 9 shows an interaction plot with ±1 standard deviation error bars for X1 and X2. A factorial repeated measures ANOVA indicated a statistically significant effect on Y of X1 (F(1, 14) = 5.45, p &lt; .05), but not of X2 (F(1, 14) = 0.18, n.s.), or of the X1×X2 interaction (F(1, 14) = 0.12, n.s.).”</td>
</tr>
</tbody>
</table>

---

40
## Parametric Tests

### Multiple within-Ss factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Within</td>
<td>Linear Mixed Model (LMM)*</td>
<td><code># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y)</code>&lt;br&gt;<code>library(lme4)</code> # for lmer&lt;br&gt;<code>library(lmerTest)</code>&lt;br&gt;<code>library(car)</code> # for Anova&lt;br&gt;<code>df$S = factor(df$S)</code> # Subject id is nominal&lt;br&gt;<code>df$X1 = factor(df$X1)</code> # X1 is a 2-level factor&lt;br&gt;<code>df$X2 = factor(df$X2)</code> # X2 is a 2-level factor&lt;br&gt;<code>contrasts(df$X1) &lt;- &quot;contr.sum&quot;</code>&lt;br&gt;<code>contrasts(df$X2) &lt;- &quot;contr.sum&quot;</code>&lt;br&gt;`m = lmer(Y ~ X1*X2 + (1</td>
</tr>
</tbody>
</table>

*The LMM sample code uses a random intercept for subject (S). There are also random slope models, which are used when the response changes at different rates for each subject over a repeated factor. A 2-minute random slope example of county population growth over time can be seen here ([https://www.youtube.com/watch?v=YDe6F7CXjWw](https://www.youtube.com/watch?v=YDe6F7CXjWw)). A free webinar on the topic of random intercept and random slope models is available here ([https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/](https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/)).
## Parametric Tests

### Multiple within-Ss factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Within</td>
<td>Linear Mixed Model (LMM)</td>
<td>“Figure 9 shows an interaction plot with ±1 standard deviation error bars for X1 and X2. A linear mixed model analysis of variance indicated a statistically significant effect on Y of X1 (F(1, 42) = 4.42, ( p &lt; .05 )), but not of X2 (F(1, 42) = 0.22, n.s.), or of the X1×X2 interaction (F(1, 42) = 0.12, n.s.).”</td>
</tr>
</tbody>
</table>
Figure 9

Y by X1, X2

2F2LWs.csv
Parametric Tests

Post hoc comparisons
# Parametric Tests

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | One-way ANOVA| Independent samples *t*-test | Btwn  | # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and continuous response (Y)
library(multcomp) # for glht, mcp
summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm")) # m is from aov
# the following also performs the same contrast tests
library(emmeans) # for emm, emmeans
summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm"))
# or, using the Tukey HSD correction instead of Holm's
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |
| 1       | ≥2     | One-way repeated measures ANOVA | Paired samples *t*-test | Within | # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y)
library(reshape2) # for dcast
df2 <- dcast(df, S ~ X, value.var="Y") # make wide-format table
ab = t.test(df2$a, df2$b, paired=TRUE, var.equal=TRUE) # a vs. b
ac = t.test(df2$a, df2$c, paired=TRUE, var.equal=TRUE) # a vs. c
bc = t.test(df2$b, df2$c, paired=TRUE, var.equal=TRUE) # b vs. c
p.adjust(c(ab$p.value, ac$p.value, bc$p.value), method="holm") |
## Parametric Tests

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
</table>
| 1       | ≥2     | One-way ANOVA | Independent samples *t*-test | Btwn | “Pairwise comparisons using independent-samples *t*-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ and ‘c’ (*t*(57) = 4.41, *p* < .0001) and ‘b’ and ‘c’ (*t*(57) = 4.80, *p* < .0001) were statistically significantly different, but that ‘a’ and ‘b’ were not (*t*(57) = -0.39, n.s.).”

  “Simultaneous pairwise comparisons using Tukey’s HSD test indicated that ‘a’ and ‘c’ (*t*(57) = 4.41, *p* < .0001) and ‘b’ and ‘c’ (*t*(57) = 4.80, *p* < .0001) were statistically significantly different, but that ‘a’ and ‘b’ were not (*t*(57) = -0.39, n.s.).”

| 1       | ≥2     | One-way repeated measures ANOVA | Paired samples *t*-test | Within | “Three *post hoc* paired-samples *t*-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ and ‘b’ (*t*(19) = 3.14, *p* < .05) and ‘a’ and ‘c’ (*t*(19) = 3.39, *p* < .01) were statistically significantly different, but that ‘b’ and ‘c’ were not (*t*(19) = 0.65, n.s.).”

Note: “Tukey’s HSD” test stands for “Tukey’s Honestly Significant Difference” test.
<table>
<thead>
<tr>
<th>Factors Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2 ≥2</td>
<td>Factorial ANOVA, Linear Model (LM)</td>
<td>Independent samples t-test</td>
<td>Btw</td>
<td># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1<em>X2)), test=adjusted(type=&quot;holm&quot;)) # m is from aov or lm # or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1</em>X2, adjust=&quot;tukey&quot;, mode=&quot;linear.predictor&quot;, type=&quot;Score&quot;))</td>
</tr>
<tr>
<td>≥2 ≥2</td>
<td>Factorial repeated measures ANOVA</td>
<td>Paired samples t-test</td>
<td>Within</td>
<td># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(reshape2) # for dcast df2 &lt;- dcast(df, S ~ X1 + X2, value.var=&quot;Y&quot;) # make wide-format table aa_ab = t.test(df2$a_a, df2$a_b, paired=TRUE, var.equal=TRUE) # aa vs. ab aa_ba = t.test(df2$a_a, df2$b_a, paired=TRUE, var.equal=TRUE) # aa vs. ba aa_bb = t.test(df2$a_a, df2$b_b, paired=TRUE, var.equal=TRUE) # aa vs. bb ab_ba = t.test(df2$a_b, df2$b_a, paired=TRUE, var.equal=TRUE) # ab vs. ba ab_bb = t.test(df2$a_b, df2$b_b, paired=TRUE, var.equal=TRUE) # ab vs. bb ba_bb = t.test(df2$b_a, df2$b_b, paired=TRUE, var.equal=TRUE) # ba vs. bb p.adjust(c(aa_ab$p.value, aa_ba$p.value, aa_bb$p.value, ab_ba$p.value, ab_bb$p.value, ba_bb$p.value), method=&quot;holm&quot;)</td>
</tr>
<tr>
<td>≥2 ≥2</td>
<td>Linear Mixed Model (LMM)</td>
<td>Paired samples t-test</td>
<td>Within</td>
<td># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1<em>X2)), test=adjusted(type=&quot;holm&quot;)) # m is from lmer # or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1</em>X2, adjust=&quot;tukey&quot;, mode=&quot;kenward-roger&quot;, type=&quot;Score&quot;))</td>
</tr>
</tbody>
</table>
## Parametric Tests

**Post hoc pairwise comparisons – Multiple factors**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
</table>
| ≥2      | ≥2     | Factorial ANOVA, Linear Model (LM) | Independent samples t-test | Btwn | “Pairwise comparisons using independent-samples t-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (b,a) vs. (a,b) showed a statistically significant difference (t(56) = 3.63, p < .01), and that (a,b) vs. (b,b) was marginal (t(56) = -2.62, p = .057). The other four comparisons showed no statistically significant differences.”

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that that (b,a) vs. (a,b) showed a statistically significant difference (t(56) = 3.63, p < .01), and that (a,b) vs. (b,b) was marginal (t(56) = -2.62, p = .054). The other four comparisons showed no statistically significant differences.” |

| ≥2 | ≥2 | Factorial repeated measures ANOVA | Paired samples t-test | Within | “Six post hoc paired-samples t-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,a) vs. (b,b) showed a statistically significant difference (t(14) = 3.10, p < .05), and that (a,b) vs. (b,b) was marginal (t(14) = 2.65, p = .096). The other four comparisons showed no statistically significant differences.” |

| ≥2 | ≥2 | Linear Mixed Model (LMM) | Paired samples t-test | Within | “Pairwise comparisons using paired-samples t-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no pairs were statistically significantly different.”

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no pairs were statistically significantly different.” |

Note: “Tukey’s HSD” test stands for “Tukey’s Honestly Significant Difference” test.
Variance-Covariance Structures
# Variance-Covariance Structures

## Optional when fitting linear mixed models (LMMs)

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name</th>
<th>Description</th>
<th>R Code</th>
</tr>
</thead>
</table>
| ID           | Scaled identity       | All variances are equal, and all covariances are zero. Default for nlme::lme. | # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y)  
library(nlme) # for lme  
library(car) # for Anova  
library(emmeans) # for emmeans  
m = lme(Y ~ X, random=~1|S, data=df, weights=varIdent(form=~X)) # ID  
getVarCov(m, type="conditional") # get VCV matrix  
anova(m) # for F-test  
Anova(m, type=3, test.statistic="Chisq") # for chisq test  
summary(emmeans(m, pairwise ~ X, adjust="holm", mode="containment")) # post hoc tests |
| DIAG         | Diagonal              | All variances can differ, and all covariances are zero. Default in IBM® SPSS™. | # See R Code for ID. Only the blue model building line changes to...  
m = lme(Y ~ X, random=~1|S, data=df, weights=varIdent(form=~1|X)) # DIAG |
| CS           | Compound symmetry     | All variances are equal, and all covariances are equal. Default for SAS® JMP™. | # See R Code for ID. Only the blue model building line changes to...  
m = lme(Y ~ X, random=~1|S, data=df, correlation=corCompSymm(form=~1|S)) # CS |
| CSH          | Heterogeneous compound symmetry | All variances can differ; otherwise, like CS. | # See R Code for ID. Only the blue model building line changes to...  
m = lme(Y ~ X, random=~1|S, data=df, correlation=corCompSymm(form=~1|S), weights=varIdent(form=~1|X)) # CSH |


**Note:** The correlation parameter sets covariances (matrix off-diagonal), and the weights parameter sets variances (matrix on-diagonal). Above, when weights=varIdent(form=~X), the variances are required to be equal, which is the default when weights is unspecified. When weights=varIdent(form=~1|X), the variances are allowed to differ.
# Variance-Covariance Structures

Optional when fitting linear mixed models (LMMs)

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name</th>
<th>Description</th>
<th>R Code</th>
</tr>
</thead>
</table>
| AR1          | First-order autoregressive | All variances are equal, and all covariances decrease the further apart they are. | # See R Code for ID. Only the blue model building line changes to...  
  m = lme(Y ~ X, random=~1|S, data=df, correlation=corAR1(form=~1|S)) # AR1 |
| ARH1         | Heterogeneous first-order autoregressive | All variances can differ; otherwise, like AR1. | # See R Code for ID. Only the blue model building line changes to...  
  m = lme(Y ~ X, random=~1|S, data=df, correlation=corAR1(form=~1|S), weights=varIdent(form=~1|X)) # ARH1 |
| ARMA11       | Autoregressive moving average | All variances are equal, and all covariances decrease the further apart they are, influenced by a moving average. | # See R Code for ID. Only the blue model building line changes to...  
  m = lme(Y ~ X, random=~1|S, data=df, correlation=corARMA(form=~1|S, p=1, q=1)) # ARMA11  
  # Note that (p,q)=(1,0) would be AR1. The 'q' parameter determines the moving average. |
| TP           | Toeplitz | All variances are equal, and covariances are equal across adjacent pairs, equal again across skip-adjacent pairs, and so on. | # See R Code for ID. Only the blue model building line changes to...  
  m = lme(Y ~ X, random=~1|S, data=df, correlation=corARMA(form=~1|S, p=2, q=0)) # TP |
| TPH          | Heterogeneous Toeplitz | All variances can differ; otherwise, like TP. | # See R Code for ID. Only the blue model building line changes to...  
  m = lme(Y ~ X, random=~1|S, data=df, correlation=corARMA(form=~1|S, p=2, q=0), weights=varIdent(form=~1|X)) # TPH |
| UN           | Unstructured | All variances and covariances can differ. A completely unpatterned matrix. Default for lme4::lmer. | # See R Code for ID. Only the blue model building line changes to...  
  m = lme(Y ~ X, random=~1|S, data=df, correlation=corSymm(form=~1|S), weights=varIdent(form=~1|X)) |

Note: See ?nlme::corClasses for additional variance-covariance structures usable by nlme::lme.
Nonparametric Tests
## Nonparametric Tests

### One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | 2      | Between                     | Mann-Whitney U test         | # df has subjects (S), one between-Ss factor (X) w/levels (a,b), and continuous response (Y)  
library(coin) 
df$S = factor(df$S) # Subject id is nominal (unused)  
df$X = factor(df$X) # X is a 2-level factor  
wilcox_test(Y ~ X, data=df, distribution="exact")                                                                                                           |
| 1       | 2      | Within                      | Wilcoxon signed-rank test   | # df has subjects (S), one within-Ss factor (X) w/levels (a,b), and continuous response (Y)  
library(coin) 
df$S = factor(df$S) # Subject id is nominal  
df$X = factor(df$X) # X is a 2-level factor  
wilcoxsign_test(Y ~ X | S, data=df, distribution="exact")                                                                                             |
| 1       | ≥2     | Between                     | Kruskal-Wallis test         | # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and continuous response (Y)  
library(coin) 
df$S = factor(df$S) # Subject id is nominal (unused)  
df$X = factor(df$X) # X is a 3-level factor  
kruskal_test(Y ~ X, data=df, distribution="asymptotic")                                                                                                                                 |
| 1       | ≥2     | Within                      | Friedman test               | # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y)  
library(coin) 
df$S = factor(df$S) # Subject id is nominal  
df$X = factor(df$X) # X is a 3-level factor  
friedman_test(Y ~ X | S, data=df, distribution="asymptotic")                                                                                           |

Note: The Mann-Whitney U test is also known as the Wilcoxon rank-sum test, not to be confused with the Wilcoxon signed-rank test.
## Nonparametric Tests
### One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Between</td>
<td>Mann-Whitney <em>U</em> test</td>
<td>“The median of ‘a’ was 13.26 (<em>IQR</em> = 5.01) and of ‘b’ was 14.98 (<em>IQR</em> = 6.19). This difference was statistically significant according to a Mann-Whitney <em>U</em> test (<em>Z</em> = -2.44, <em>p</em> &lt; .05).”</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Within</td>
<td>Wilcoxon signed-rank test</td>
<td>“The median of ‘a’ was 11.92 (<em>IQR</em> = 2.35) and of ‘b’ was 10.74 (<em>IQR</em> = 2.29). This difference was statistically significant according to a Wilcoxon signed-rank test (<em>Z</em> = 2.11, <em>p</em> &lt; .05).”</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>Between</td>
<td>Kruskal-Wallis test</td>
<td>“The median of ‘a’ was 13.96 (<em>IQR</em> = 5.98), of ‘b’ was 11.50 (<em>IQR</em> = 3.49), and of ‘c’ was 9.62 (<em>IQR</em> = 1.61). These differences were statistically significant according to a Kruskal-Wallis test (<em>χ²</em>(2, <em>N</em>=60) = 17.77, <em>p</em> &lt; .001).”</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>Within</td>
<td>Friedman test</td>
<td>“The median of ‘a’ was 8.37 (<em>IQR</em> = 1.71), of ‘b’ was 8.97 (<em>IQR</em> = 2.21), and of ‘c’ was 9.97 (<em>IQR</em> = 3.56). These differences were statistically significant according to a Friedman test (<em>χ²</em>(2, <em>N</em>=60) = 7.90, <em>p</em> &lt; .05).”</td>
</tr>
</tbody>
</table>

Note: “*IQR*” stands for “interquartile range,” i.e., the distance between the top and bottom of the box in a boxplot (25% - 75% quartile range).
## Nonparametric Tests

### Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Between</td>
<td>Aligned Rank Transform (ART)</td>
<td># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(ARTool) df$S = factor(df$S) # Subject id is nominal (unused) df$X1 = factor(df$X1) # X1 is a 2-level factor df$X2 = factor(df$X2) # X2 is a 2-level factor m = art(Y ~ X1*X2, data=df) anova(m)</td>
</tr>
<tr>
<td>Generalized Linear Model (GLM)*</td>
<td>See “Generalized Linear (Mixed) Models: Distributions and canonical links”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Within</td>
<td>Aligned Rank Transform (ART)**</td>
<td># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(ARTool) df$S = factor(df$S) # Subject id is nominal df$X1 = factor(df$X1) # X1 is a 2-level factor df$X2 = factor(df$X2) # X2 is a 2-level factor m = art(Y ~ X1*X2 + (1</td>
</tr>
<tr>
<td>Generalized Linear Mixed Model (GLMM)*</td>
<td>See “Generalized Linear (Mixed) Models: Distributions and canonical links”</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* GLMs and GLMMs are not actually nonparametric; they are generalized regression models that support tests on data not amenable to ANOVA.

**The Aligned Rank Transform within-subjects sample code uses a random intercept for subject (S). There are also random slope models, which are used when the response changes at different rates for each subject over a repeated factor. A 2-minute random slope example of county population growth over time can be seen here [https://www.youtube.com/watch?v=YDe6F7CXjWw](https://www.youtube.com/watch?v=YDe6F7CXjWw). A free webinar on the topic of random intercept and random slope models is available here [https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/](https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/).
# Nonparametric Tests

## Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Between or Within Subjects</th>
<th>Test Name</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Between</td>
<td>Aligned Rank Transform (ART)</td>
<td>“Figure 10 shows boxplots for all levels of X1×X2. A nonparametric analysis of variance based on the Aligned Rank Transform indicated no statistically significant effect on Y of X1 (F(1, 56) = 1.98, n.s.), or of X2 (F(1, 56) = 0.81, n.s.), but there was a statistically significant X1×X2 interaction (F(1, 56) = 13.65, p &lt; .001).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Generalized Linear Model (GLM)</td>
<td>See “Generalized Linear (Mixed) Models: Distributions and canonical links”</td>
</tr>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Within</td>
<td>Aligned Rank Transform (ART)</td>
<td>“Figure 11 shows boxplots for all levels of X1×X2. A nonparametric analysis of variance based on the Aligned Rank Transform indicated no statistically significant effect on Y of X1 (F(1, 42) = 1.85, n.s.) or of the X1×X2 interaction (F(1, 42) = 1.93, n.s.), but there was a statistically significant effect of X2 (F(1, 42) = 24.97, p &lt; .0001).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Generalized Linear Mixed Model</td>
<td>See “Generalized Linear (Mixed) Models: Distributions and canonical links”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(GLMM)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 10

Y by X1, X2

X2

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>

2F2LBs.csv
Figure 11

Y by X1, X2

2F2LWs.csv
Nonparametric Tests

Post hoc comparisons
## Nonparametric Tests

**Post hoc pairwise comparisons – One factor**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Kruskal-Wallis test            | Mann-Whitney U test              | Btwn  | # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and continuous response (Y)  
|         |        |                                |                                  |       | ab = wilcox.test(df[df$X == "a",]$Y, df[df$X == "b",]$Y) # a vs. b  
|         |        |                                |                                  |       | ac = wilcox.test(df[df$X == "a",]$Y, df[df$X == "c",]$Y) # a vs. c  
|         |        |                                |                                  |       | bc = wilcox.test(df[df$X == "b",]$Y, df[df$X == "c",]$Y) # b vs. c  
|         |        |                                |                                  |       | p.adjust(c(ab$p.value, ac$p.value, bc$p.value), method="holm")                                                                                                                                   |
| 1       | ≥2     | Friedman test                  | Wilcoxon signed-rank test        | Within| # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y)  
|         |        |                                |                                  |       | library(reshape2) # for dcast  
|         |        |                                |                                  |       | df2 <- dcast(df, S ~ X, value.var="Y") # make wide-format table  
|         |        |                                |                                  |       | ab = wilcox.test(df2$a, df2$b, paired=TRUE) # a vs. b  
|         |        |                                |                                  |       | ac = wilcox.test(df2$a, df2$c, paired=TRUE) # a vs. c  
|         |        |                                |                                  |       | bc = wilcox.test(df2$b, df2$c, paired=TRUE) # b vs. c  
|         |        |                                |                                  |       | p.adjust(c(ab$p.value, ac$p.value, bc$p.value), method="holm")                                                                                                                                   |
Nonparametric Tests

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Kruskal-Wallis test</td>
<td>Mann-Whitney <em>U</em> test</td>
<td>Btwn</td>
<td>“Three <em>post hoc</em> Mann-Whitney <em>U</em> tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ and ‘b’ (<em>W</em> = 309, <em>p</em> &lt; .01) and ‘a’ and ‘c’ (<em>W</em> = 343, <em>p</em> &lt; .001) were statistically significantly different, but that ‘b’ and ‘c’ were marginal (<em>W</em> = 264, <em>p</em> = .086).”</td>
</tr>
<tr>
<td>1</td>
<td>≥2</td>
<td>Friedman test</td>
<td>Wilcoxon signed-rank test</td>
<td>Within</td>
<td>“Three <em>post hoc</em> Wilcoxon signed-rank tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ and ‘c’ (<em>V</em> = 32, <em>p</em> &lt; .05) and ‘b’ and ‘c’ (<em>V</em> = 43, <em>p</em> &lt; .05) were statistically significantly different, but that ‘a’ and ‘b’ were not (<em>V</em> = 83, n.s.).”</td>
</tr>
</tbody>
</table>
## Nonparametric Tests

*Post hoc* pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Aligned Rank Transform (ART)</td>
<td>Aligned Rank Transform Contrasts (ART-C)</td>
<td>Btw, Within</td>
<td># df has subjects (S), two factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(dplyr) # for %&gt;% pipe art.con(m, ~ X1<em>X2, adjust=&quot;holm&quot;) %&gt;% # run ART-C for X1 × X2 summary() %&gt;% # add significance stars to the output mutate(sig. = symnum(p.value, corr=FALSE, na=FALSE, cutpoints = c(0, 0.001, 0.01, 0.05, 0.1, 1), symbols = c(&quot;</em><strong>&quot;, &quot;</strong>&quot;, &quot;,&quot;, &quot;,&quot;, &quot;&quot;))</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mann-Whitney U test</td>
<td></td>
<td>Btw</td>
<td># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) aa_ab = wilcox.test(df[df$X1 == &quot;a&quot; &amp; df$X2 == &quot;a&quot;,]$Y, df[df$X1 == &quot;a&quot; &amp; df$X2 == &quot;b&quot;,]$Y) aa_ba = wilcox.test(df[df$X1 == &quot;a&quot; &amp; df$X2 == &quot;a&quot;,]$Y, df[df$X1 == &quot;b&quot; &amp; df$X2 == &quot;a&quot;,]$Y) aa_bb = wilcox.test(df[df$X1 == &quot;a&quot; &amp; df$X2 == &quot;a&quot;,]$Y, df[df$X1 == &quot;b&quot; &amp; df$X2 == &quot;b&quot;,]$Y) ab_ba = wilcox.test(df[df$X1 == &quot;a&quot; &amp; df$X2 == &quot;b&quot;,]$Y, df[df$X1 == &quot;b&quot; &amp; df$X2 == &quot;a&quot;,]$Y) ab_bb = wilcox.test(df[df$X1 == &quot;a&quot; &amp; df$X2 == &quot;b&quot;,]$Y, df[df$X1 == &quot;b&quot; &amp; df$X2 == &quot;b&quot;,]$Y) ba_bb = wilcox.test(df[df$X1 == &quot;b&quot; &amp; df$X2 == &quot;a&quot;,]$Y, df[df$X1 == &quot;b&quot; &amp; df$X2 == &quot;b&quot;,]$Y) p.adjust(c(aa_ab$p.value, aa_ba$p.value, aa_bb$p.value, ab_ba$p.value, ab_bb$p.value, ba_bb$p.value), method=&quot;holm&quot;)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wilcoxon signed-rank test</td>
<td></td>
<td>Within</td>
<td># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) df2 &lt;- dcast(df, S ~ X1 + X2, value.var=&quot;Y&quot;) # make wide-format table aa_ab = wilcox.test(df2$a_a, df2$a_b, paired=TRUE) aa_ba = wilcox.test(df2$a_a, df2$b_a, paired=TRUE) aa_bb = wilcox.test(df2$a_a, df2$b_b, paired=TRUE) ab_ba = wilcox.test(df2$a_b, df2$a_a, paired=TRUE) ab_bb = wilcox.test(df2$a_b, df2$b_b, paired=TRUE) ba_bb = wilcox.test(df2$b_a, df2$b_b, paired=TRUE) p.adjust(c(aa_ab$p.value, aa_ba$p.value, aa_bb$p.value, ab_ba$p.value, ab_bb$p.value, ba_bb$p.value), method=&quot;holm&quot;)</td>
</tr>
</tbody>
</table>
### Nonparametric Tests

**Post hoc pairwise comparisons – Multiple factors**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Test</th>
<th>Test Name</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Aligned Rank Transform (ART)</td>
<td>Aligned Rank Transform Contrasts (ART-C)</td>
<td>Btwn</td>
<td>“Post hoc pairwise comparisons conducted with the ART-C procedure, and corrected with Holm’s sequential Bonferroni procedure, indicated that (a,b) vs. (b,b) showed a statistically significant difference ($t(56) = 3.42, p &lt; .01$), as did (b,a) vs. (b,b) ($t(56) = 3.21, p &lt; .05$). No other pairwise comparisons were significantly different.”</td>
</tr>
</tbody>
</table>

**Within**

“Post hoc pairwise comparisons conducted with the ART-C procedure, and corrected with Holm’s sequential Bonferroni procedure, indicated that (a,a) vs. (a,b) showed a statistically significant difference ($t(42) = -4.43, p < .001$), as did (a,b) vs. (b,a) ($t(42) = 4.38, p < .001$). The test of (a,a) vs. (b,b) was marginal ($t(42) = -2.49, p = .067$), as was the test of (b,a) vs. (b,b) ($t(42) = -2.44, p = .067$). No other pairwise comparisons were statistically significant or marginal.”

**Mann-Whitney U test**

| Btwn | “Six post hoc Mann-Whitney U tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,b) vs. (b,b) showed a statistically significant difference ($W = 179, p < .05$), and that (b,a) vs. (b,b) did as well ($W = 184, p < .05$). The other four comparisons showed no statistically significant differences.” |

| Within | “Six post hoc Wilcoxon signed-rank tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,a) vs. (a,b) showed a statistically significant difference ($V = 4, p < .01$), as did (a,b) vs. (b,a) ($V = 113, p < .01$). The test of (a,a) vs. (b,b) was marginal ($V = 19, p = .072$). The other three comparisons showed no statistically significant differences.” |
Generalized Linear (Mixed) Models

./data/GLM/
./data/GLMM/
### Terminology

<table>
<thead>
<tr>
<th>Contain Random Factors?</th>
<th>Generalize to Responses Unsuitable to ANOVA?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Random factors enable the modeling of correlated responses, i.e., within-subjects data, repeated measures data, longitudinal data, panel data, etc.*

**Between-Ss**
- Normal, binomial, multinomial, ordinal, Poisson, zero-inflated Poisson, negative binomial, zero-inflated negative binomial, gamma, exponential

**Within-Ss**
- Normal only
# GLM / GLMM

## Distributions and canonical links

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>R code for GLM (between-Ss)</th>
<th>R code for GLMM (within-Ss)</th>
</tr>
</thead>
</table>
| Normal       | identity | Linear regression: Equivalent to the linear model (LM) or linear mixed model (LMM) | library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
m = glm(Y ~ X1*X2, data=df, family=gaussian)
Anova(m, type=3, test.statistic="F") | library(lme4) # for glmer, lmer
library(lmerTest)
library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
#m = glmer(Y ~ X1*X2 + (1|S), data=df, family=gaussian) #deprecated
m = lmer(Y ~ X1*X2 + (1|S), data=df) # equivalent, use instead
Anova(m, type=3, test.statistic="F") |
| Binomial     | logit | Logistic regression: Dichotomous responses (i.e., nominal responses with two categories) | library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
df$Y = factor(df$Y) # nominal response
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
m = glm(Y ~ X1*X2, data=df, family=binomial)
Anova(m, type=3) | library(lme4) # for glmer
library(lmerTest)
library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
df$Y = factor(df$Y) # nominal response
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
m = glmer(Y ~ X1*X2 + (1|S), data=df, family=binomial)
Anova(m, type=3) |

Notes: A normal distribution is also known as a Gaussian distribution. Logistic regression is also known as binomial regression.
**GLM / GLMM**

**Distributions and canonical links**

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>Report for GLM (between-Ss)</th>
<th>Report for GLMM (within-Ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>identity</td>
<td>Linear regression: Equivalent to the linear model (LM) or linear mixed model (LMM)</td>
<td>“Figure 12a shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance based on linear regression indicated a statistically significant effect on Y of X1 (F(1, 56) = 7.06, ( p &lt; .05 )), but not of X2 (F(1, 56) = 1.02, n.s.). Also, the X1×X2 interaction was not statistically significant (F(1, 56) = 2.74, n.s.).”</td>
<td>“Figure 12b shows an interaction plot with ±1 standard deviation error bars for X1 and X2. A linear mixed model analysis of variance indicated no statistically significant effect on Y of X1 (F(1, 42) = 2.44, n.s.) or of X2 (F(1, 42) = 0.00, n.s.), but there was a statistically significant X1×X2 interaction (F(1, 42) = 5.17, ( p &lt; .05 )).”</td>
</tr>
<tr>
<td>Binomial</td>
<td>logit</td>
<td>Logistic regression: Dichotomous responses (i.e., nominal responses with two categories)</td>
<td>“Figure 13a shows the number of ‘x’ and ‘y’ outcomes for each level of X1 and X2. An analysis of variance based on logistic regression indicated a statistically significant effect on Y of X1 (( \chi^2(1, N=60) = 6.05, p &lt; .05 )) and of the X1×X2 interaction (( \chi^2(1, N=60) = 8.63, p &lt; .01 )), but not of X2 (( \chi^2(1, N=60) = 2.27, n.s. )).”</td>
<td>“Figure 13b shows the number of ‘x’ and ‘y’ outcomes for each level of X1 and X2. An analysis of variance based on mixed logistic regression indicated no statistically significant effect on Y of X1 (( \chi^2(1, N=60) = 1.02, n.s. )) or of X2 (( \chi^2(1, N=60) = 1.02, n.s. )), but there was a statistically significant X1×X2 interaction (( \chi^2(1, N=60) = 6.52, p &lt; .05 )).”</td>
</tr>
</tbody>
</table>
Figure 12b

Y by X1, X2
Figure 13a

Y by X1, X2

Count

a.a 14
a.b 7
b.a 10
b.b 8

X1.X2

Y
x
y

2F2LBs_binomial.csv
## GLM / GLMM

Distributions and canonical links

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>R code for GLM (between-Ss)</th>
<th>R code for GLMM (within-Ss)</th>
</tr>
</thead>
</table>
| Multinomial  | logit| Multinomial logistic regression: Polytomous responses (i.e., nominal responses with more than two categories) | library(nnet) # for multinom library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) df$Y = factor(df$Y) # nominal response contrasts(df$X1) <- "contr.sum" contrasts(df$X2) <- "contr.sum" m = multinom(Y ~ X1*X2, data=df) Anova(m, type=3) | library(mlogit) # for mlogit.data library(lme4) # for glmer library(merTest) library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) df$Y = factor(df$Y) # nominal response contrasts(df$X1) <- "contr.sum" contrasts(df$X2) <- "contr.sum"
# use the multinomial-Poisson transformation (Baker 1994)
df2 <- as.data.frame(mlogit.data(df, choice="Y", shape="wide", id.var="S"))
df2$S = factor(df2$S) df2$alt = factor(df2$alt)
m = glmer(Y ~ X1*alt + (1|S), data=df2, family=poisson) a = Anova(m, type=2); a[grep("X1:alt", rownames(a)),] # X1
m = glmer(Y ~ X2*alt + (1|S), data=df2, family=poisson) a = Anova(m, type=2); a[grep("X2:alt", rownames(a)),] # X2
m = glmer(Y ~ X1*X2*alt + (1|S), data=df2, family=poisson) a = Anova(m, type=2); a[grep("X1:X2:alt", rownames(a)),] # X1:X2 |

Note: Multinomial logistic regression is also known as nominal logistic regression.
### GLM / GLMM

**Distributions and canonical links**

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>Report for GLM (between-Ss)</th>
<th>Report for GLMM (within-Ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multinomial</td>
<td>logit</td>
<td><em>Multinomial logistic regression</em>: polytomous responses (i.e., nominal responses with more than two categories)</td>
<td>“Figure 14a shows the number of ‘x’, ‘y’, and ‘z’ outcomes for each level of X1 and X2. An analysis of variance based on multinomial logistic regression indicated a statistically significant effect on Y of X1 ($\chi^2(2, N=60) = 10.46, p &lt; .01$), of X2 ($\chi^2(2, N=60) = 15.21, p &lt; .001$), and a marginal X1×X2 interaction ($\chi^2(2, N=60) = 5.09, p = .078$).”</td>
<td>“Figure 14b shows the number of ‘x’, ‘y’, and ‘z’ outcomes for each level of X1 and X2. An analysis of variance based on mixed Poisson regression using the multinomial-Poisson transformation (Baker 1994) indicated no statistically significant effect on Y of X1 ($\chi^2(2, N=60) = 2.58, n.s.$) or X2 ($\chi^2(2, N=60) = 1.04, n.s.$), but there was a statistically significant X1×X2 interaction ($\chi^2(2, N=60) = 11.39, p &lt; .01$)”</td>
</tr>
<tr>
<td>Distribution</td>
<td>Link</td>
<td>Typical Uses</td>
<td>R code for GLM (between-Ss)</td>
<td>R code for GLM (within-Ss)</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>--------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Ordinal</td>
<td>cumulative logit</td>
<td><strong>Ordinal logistic regression:</strong> Ordinal responses (e.g., Likert scales)</td>
<td>library(MASS) # for polr library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) df$Y = ordered(df$Y) # ordinal response contrasts(df$X1) &lt;- &quot;contr.sum&quot; contrasts(df$X2) &lt;- &quot;contr.sum&quot; m = polr(Y ~ X1*X2, data=df, Hess=TRUE) Anova(m, type=3)</td>
<td>library(ordinal) # for clmm library(RVAideMemoire) # for Anova.clmm df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) df$Y = ordered(df$Y) # ordinal response contrasts(df$X1) &lt;- &quot;contr.sum&quot; contrasts(df$X2) &lt;- &quot;contr.sum&quot; m = clmm(Y ~ X1*X2 + (1</td>
</tr>
<tr>
<td>Poisson</td>
<td>log</td>
<td><strong>Poisson regression:</strong> Counts, rare events (e.g., gesture recognition errors, 3-pointers per quarter, number of &quot;F&quot; grades)</td>
<td>library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) contrasts(df$X1) &lt;- &quot;contr.sum&quot; contrasts(df$X2) &lt;- &quot;contr.sum&quot; m = glm(Y ~ X1*X2, data=df, family=poisson) # use family-quasipoisson if overdispersed Anova(m, type=3)</td>
<td>library(lme4) # for glmer library(lme4Test) library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) contrasts(df$X1) &lt;- &quot;contr.sum&quot; contrasts(df$X2) &lt;- &quot;contr.sum&quot; m = glmer(Y ~ X1*X2 + (1</td>
</tr>
</tbody>
</table>

Notes: Ordinal logistic regression is also known as cumulative logistic regression. When count data is overdispersed, it means the variance of the distribution is greater than its mean. For mildly overdispersed count distributions, `family=quasipoisson` can be used with `glm` but not with `glmer`. For greater overdispersion, use negative binomial regression.
## GLM / GLMM
### Distributions and canonical links

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>Report for GLM (between-Ss)</th>
<th>Report for GLMM (within-Ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinal</td>
<td>cumulative logit</td>
<td>**Ordinal logistic regression:**Ordinal responses (e.g., Likert scales)</td>
<td>“Figure 15a shows the distribution of Likert responses (1-7) for each combination of X1 and X2. An analysis of variance based on ordinal logistic regression indicated a statistically significant effect on Y of X2 ($\chi^2(1, N=60) = 6.14, p &lt; .05$), but not of X1 ($\chi^2(1, N=60) = 1.65, n.s.$) or of the X1×X2 interaction ($\chi^2(1, N=60) = 0.05, n.s.$).”</td>
<td>“Figure 15b shows the distribution of Likert responses (1-7) for each combination of X1 and X2. An analysis of variance based on mixed ordinal logistic regression indicated no statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 0.00, n.s.$) or of X2 ($\chi^2(1, N=60) = 0.00, n.s.$), but there was a statistically significant X1×X2 interaction ($\chi^2(1, N=60) = 4.22, p &lt; .05$).”</td>
</tr>
<tr>
<td>Poisson</td>
<td>log</td>
<td>**Poisson regression:**Counts, rare events (e.g., gesture recognition errors, 3-pointers per quarter, number of “F” grades)</td>
<td>“Figure 16a shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance based on Poisson regression indicated a statistically significant effect on Y of the X1×X2 interaction ($\chi^2(1, N=60) = 3.84, p &lt; .05$), but not of either X1 ($\chi^2(1, N=60) = 0.17, n.s.$) or X2 ($\chi^2(1, N=60) = 1.19, n.s.$).”</td>
<td>“Figure 16b shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance based on mixed Poisson regression indicated a statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 18.10, p &lt; .0001$), but not of X2 ($\chi^2(1, N=60) = 1.47, n.s.$) or of the X1×X2 interaction ($\chi^2(1, N=60) = 0.26, n.s.$).”</td>
</tr>
</tbody>
</table>
Figure 15b

Responses by X1, X2

Count

Likert (1-7)
Figure 16a

Y by X1, X2

Y

X1

a

b

X2

2F2LBs_poisson.csv
Figure 16b

Y by X1, X2

X2

a

b

2F2LWs_poisson.csv
## GLM / GLMM

### Distributions and canonical links

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>R code for GLM (between-Ss)</th>
<th>R code for GLMM (within-Ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-Inflated Poisson</td>
<td>log</td>
<td>Zero-inflated Poisson regression: Counts and rare events with a large proportion of zeroes</td>
<td>library(pascl) # for zeroinfl library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) contrasts(df$X1) = &quot;contr.sum&quot; contrasts(df$X2) = &quot;contr.sum&quot; m = zeroinfl(Y ~ X1*X2, data=df, dist=&quot;poisson&quot;) Anova(m, type=3)</td>
<td>library(glmmTMB) # for glmmTMB library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) contrasts(df$X1) = &quot;contr.sum&quot; contrasts(df$X2) = &quot;contr.sum&quot; m = glmmTMB(Y ~ X1*X2 + (1</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>log</td>
<td>Negative binomial regression: Same as Poisson but for use in the presence of overdispersion</td>
<td>library(MASS) # for glm.nb library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) contrasts(df$X1) = &quot;contr.sum&quot; contrasts(df$X2) = &quot;contr.sum&quot; m = glm.nb(Y ~ X1*X2, data=df) Anova(m, type=3)</td>
<td>library(lme4) # for glmer.nb library(merTest) library(car) # for Anova df$S = factor(df$S) df$X1 = factor(df$X1) df$X2 = factor(df$X2) contrasts(df$X1) = &quot;contr.sum&quot; contrasts(df$X2) = &quot;contr.sum&quot; m = glmer.nb(Y ~ X1*X2 + (1</td>
</tr>
</tbody>
</table>

Notes: Negative binomial regression can be used when count data is overdispersed, which means the variance of the distribution is greater than its mean.
## GLM / GLMM
### Distributions and canonical links

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>Report for GLM (between-Ss)</th>
<th>Report for GLM (within-Ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-Inflated Poisson</td>
<td>log</td>
<td>Zero-inflated Poisson regression: Counts and rare events with a large proportion of zeroes</td>
<td>“Figure 17a shows histograms of Y by X1 and X2. An analysis of variance based on zero-inflated Poisson regression indicated a statistically significant effect on Y of the X1×X2 interaction ($\chi^2(1, N=60) = 8.14, p &lt; .01$), and a marginal effect of X2 ($\chi^2(1, N=60) = 3.11, p = .078$). There was no statistically significant effect of X1 on Y ($\chi^2(1, N=60) = 0.29, n.s.$)”</td>
<td>“Figure 17b shows histograms of Y by X1 and X2. An analysis of variance based on zero-inflated mixed Poisson regression indicated a statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 4.58, p &lt; .05$), but neither of X2 ($\chi^2(1, N=60) = 0.05, n.s.$) nor the X1×X2 interaction ($\chi^2(1, N=60) = 2.31, n.s.$)”</td>
</tr>
<tr>
<td>Negative Binomial</td>
<td>log</td>
<td>Negative binomial regression: Same as Poisson but for use in the presence of overdispersion</td>
<td>“Figure 18a shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance based on negative binomial regression indicated a statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 13.46, p &lt; .001$), but not of X2 ($\chi^2(1, N=60) = 0.07, n.s.$) or the X1×X2 interaction ($\chi^2(1, N=60) = 0.92, n.s.$)”</td>
<td>“Figure 18b shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance based on mixed negative binomial regression indicated a statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 8.66, p &lt; .01$) and X2 ($\chi^2(1, N=60) = 6.99, p &lt; .01$), but not of the X1×X2 interaction ($\chi^2(1, N=60) = 2.23, n.s.$)”</td>
</tr>
</tbody>
</table>
Figure 17b

Y by X1, X2

Y

Count

0 1 2 3 4 5 6 7 8 9 10 11

0 1 1 1 2 2 2 0 2 2

0 1 1 1 2 3 2 0 1 0

0 3 3 2 3 1 1 0 0 0

0 1 1 1 1 1 0 1 0 0

0 1 1 1 2 2 0 2 2

2F2LWs_zipoisson.csv
Figure 18a

Y by X1, X2

2F2LBs_negbin.csv
Figure 18b

Y by X1, X2

2F2LWs_negbin.csv
# GLM / GLMM
## Distributions and canonical links

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>R code for GLM (between-Ss)</th>
<th>R code for GLMM (within-Ss)</th>
</tr>
</thead>
</table>
| Zero-Inflated Negative  | log        | Negative binomial regression: Same as Zero-Inflated Poisson but for use in the presence of overdispersion | library(pscl) # for zeroinfl
library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
m = zeroinfl(Y ~ X1*X2, data=df, dist="negbin")
Anova(m, type=3)                                                                 | library(glmmTMB) # for glmmTMB
library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
m = glmmTMB(Y ~ X1*X2 + (1|S), data=df, family=nbinom2, ziformula=~X1*X2)
Anova(m, type=3)                                                                 |
| Gamma, including Exponential | inverse, but log is often used because 1/0 is undefined | Gamma regression: Exponentially distributed responses (e.g., income, wait times) | library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
m = glm(Y ~ X1*X2, data=df, family=Gamma)
# family=Gamma(link="log") is often used
Anova(m, type=3)                                                                 | library(lme4) # for glmer
library(lmerTest)
library(car) # for Anova
df$S = factor(df$S)
df$X1 = factor(df$X1)
df$X2 = factor(df$X2)
contrasts(df$X1) <- "contr.sum"
contrasts(df$X2) <- "contr.sum"
m = glmer(Y ~ X1*X2 + (1|S), data=df, family=Gamma)
# family=Gamma(link="log") is often used
Anova(m, type=3)                                                                 |

Notes: Gamma distributions are parameterized by *shape* and *scale*. Exponential distributions are special cases of gamma distributions where *shape* equals 1.

The GLMM sample code uses a random *intercept* for *subject* (*S*). There are also random *slope* models, which are used when the response changes at different rates for each subject over the repeated factor(s). A random slope example of county population growth over time can be seen here ([https://www.youtube.com/watch?v=YDe6F7CXJWw](https://www.youtube.com/watch?v=YDe6F7CXJWw)). A free webinar on the topic of random intercept and random slope models is available here ([https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/](https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/)).
## GLM / GLMM
### Distributions and canonical links

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Typical Uses</th>
<th>Report for GLM (between-Ss)</th>
<th>Report for GLMM (within-Ss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero-Inflated Negative Binomial</td>
<td>log</td>
<td><em>Zero-inflated negative binomial regression:</em> Same as Zero-Inflated Poisson but for use in the presence of overdispersion</td>
<td>“Figure 19a shows histograms of Y by X1 and X2. An analysis of variance based on zero-inflated negative binomial regression indicated no statistically significant effects on Y of X1 ($\chi^2(1, N=60) = 0.43, \text{n.s.}$), X2 ($\chi^2(1, N=60) = 1.28, \text{n.s.}$), or the X1×X2 interaction ($\chi^2(1, N=60) = 0.10, \text{n.s.}$).”</td>
<td>“Figure 19b shows histograms of Y by X1 and X2. An analysis of variance based on zero-inflated mixed negative binomial regression indicated a statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 31.09, p &lt; .0001$) and a marginal X1×X2 interaction ($\chi^2(1, N=60) = 3.17, p = .075$), but no detectable effect of X2 ($\chi^2(1, N=60) = 0.26, \text{n.s.}$).”</td>
</tr>
<tr>
<td>Gamma, including Exponential</td>
<td>inverse, but log is often used because 1/0 is undefined</td>
<td><em>Gamma regression; Exponentially distributed responses (e.g., income, wait times)</em></td>
<td>“Figure 20a shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance based on gamma regression indicated no statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 0.40, \text{n.s.}$) or X2 ($\chi^2(1, N=60) = 0.58, \text{n.s.}$), but the X1×X2 interaction was marginal ($\chi^2(1, N=60) = 3.26, p = .071$).”</td>
<td>“Figure 20b shows an interaction plot with ±1 standard deviation error bars for X1 and X2. An analysis of variance based on mixed gamma regression indicated no statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 1.25, \text{n.s.}$) or of X2 ($\chi^2(1, N=60) = 0.29, \text{n.s.}$), but there was a statistically significant X1×X2 interaction ($\chi^2(1, N=60) = 10.84, p &lt; .001$).”</td>
</tr>
</tbody>
</table>
Figure 19a

Y by X1, X2

2F2LBs_zinegbin.csv
Figure 20a

Y by X1, X2

Y

X1

a

b

5.94 (±2.90)

4.36 (±2.36)

5.83 (±1.89)

6.10 (±2.75)

X2

a

b

2F2LBs_gamma.csv
Figure 20b

Y by X1, X2

X2

Y

4.27 (±2.36)
3.09 (±1.29)
2.27 (±1.76)
1.58 (±3.06)

X1

a
b

2F2LWs_gamma.csv
Generalized Linear (Mixed) Models

Post hoc comparisons
# GLM / GLMM

### Post hoc pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Linear regression | Z-test | Btwn, Within | # df has subjects (S), one factor (X) w/levels (a,b,c), and continuous response (Y)  
library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm")) |
|         |        |                | Z-test | Btwn | # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and continuous response (Y)  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |
|         |        |                | t-test | Within | # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y)  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="kenward-roger", type="Score")) |
| 1       | ≥2     | Logistic regression | Z-test | Btwn, Within | # df has subjects (S), one factor (X) w/levels (a,b,c), and dichotomous response (Y)  
library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm"))  
# the following also performs the equivalent contrast tests  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |

Note: Between-subjects models are from GLMs; within-subjects models are from GLMMs.
**GLM / GLMM**

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Linear regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘b’ difference was statistically significant (Z = 2.52, p &lt; .05), but those of ‘a’ vs. ‘c’ (Z = 1.16, n.s.) and ‘b’ vs. ‘c’ (Z = -1.36, n.s.) were not.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Z-test,</td>
<td>Within</td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘b’ difference was statistically significant (Z = 2.52, p &lt; .05), but those of ‘a’ vs. ‘c’ (Z = 1.16, n.s.) and ‘b’ vs. ‘c’ (Z = -1.36, n.s.) were not.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>t-test</td>
<td></td>
<td>“Pairwise comparisons using t-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ vs. ‘b’ was not statistically significantly different (t(38) = -0.27, n.s.), but ‘a’ vs. ‘c’ (t(38) = 2.38, p = .052) and ‘b’ vs. ‘c’ (Z = 2.11, p = .070) were marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that neither ‘a’ vs. ‘b’ (t(38) = -0.27, n.s.) nor ‘b’ vs. ‘c’ (t(38) = -2.11, n.s.) were detectably different, but ‘a’ vs. ‘c’ was marginal (t(38) = -2.38, p = .057).”</td>
</tr>
</tbody>
</table>
GLM / GLMM

Post hoc pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Logistic regression | Z-test | B/w | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ vs. ‘b’ was statistically significantly different (Z = 2.53, p < .05), as was ‘a’ vs. ‘c’ (Z = 3.64, p < .001), but not ‘b’ vs. ‘c’ (Z = 1.61, n.s.).”
|         |        |               |      |     | “Simultaneous pairwise comparisons using Tukey’s HSD test indicated that ‘a’ vs. ‘b’ was statistically significantly different (Z = 2.53, p < .05), as was ‘a’ vs. ‘c’ (Z = 3.64, p < .001), but not ‘b’ vs. ‘c’ (Z = 1.61, n.s.).”
|         |        |               |      |     | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ vs. ‘c’ was statistically significantly different (Z = 2.98, p < .01), as was ‘b’ vs. ‘c’ (Z = 2.53, p < .05), but not ‘a’ vs. ‘b’ (Z = 0.64, n.s.).”
|         |        |               |      |     | “Simultaneous pairwise comparisons using Tukey’s HSD test indicated that ‘a’ vs. ‘c’ was statistically significantly different (Z = 2.98, p < .01), as was ‘b’ vs. ‘c’ (Z = 2.53, p < .05), but not ‘a’ vs. ‘b’ (Z = 0.64, n.s.).”

GLM / GLMM
# GLM / GLMM

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Multinomial logistic regression</td>
<td>Pairwise multinomial logistic regression</td>
<td>Btw</td>
<td># df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and polytomous response (Y) library(nnet) # for multinom library(car) # for Anova m = multinom(Y ~ X, data=df[df$X == &quot;a&quot;</td>
</tr>
</tbody>
</table>

Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994) | Pairwise mixed Poisson regression | Within | # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and polytomous response (Y) # df2 was produced by mlogit.data and has a logical response (Y) and alt factor library(lme4) # for glmer library(lmerTest) library(car) # for Anova m = glmer(Y ~ X*alt + (1|S), data=df2[df2$X == "a" | df2$X == "b"], family=poisson) # a vs. b a = Anova(m, type=2) ab = a[grep("X:alt", rownames(a)),] m = glmer(Y ~ X*alt + (1|S), data=df2[df2$X == "a" | df2$X == "c"], family=poisson) # a vs. c a = Anova(m, type=2) ac = a[grep("X:alt", rownames(a)),] m = glmer(Y ~ X*alt + (1|S), data=df2[df2$X == "b" | df2$X == "c"], family=poisson) # b vs. c a = Anova(m, type=2) bc = a[grep("X:alt", rownames(a)),] p.adjust(c(ab$`Pr(>Chisq)``, ac$`Pr(>Chisq)``, bc$`Pr(>Chisq)``), method="holm") |
<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Multinomial logistic regression</td>
<td>Pairwise multinomial logistic regression</td>
<td>Btwn</td>
<td>“Three post hoc multinomial logistic regressions, one for each pair of X’s levels, corrected with Holm’s sequential Bonferroni procedure, indicated that the number of ‘x’, ‘y’, and ‘z’ responses was statistically significantly different for ‘a’ vs. ‘b’ ($\chi^2(4, N=40) = 11.81, p &lt; .05$) and ‘b’ vs. ‘c’ ($\chi^2(4, N=40) = 17.33, p &lt; .01$), but not for ‘a’ vs. ‘c’ ($\chi^2(4, N=40) = 6.62, n.s.$).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)</td>
<td>Pairwise mixed Poisson regression</td>
<td>Within</td>
<td>“Three post hoc mixed Poisson regressions, one for each pair of X’s levels, corrected with Holm’s sequential Bonferroni procedure, indicated that the number of ‘x’, ‘y’, and ‘z’ responses was marginally different for ‘a’ vs. ‘c’ ($\chi^2(2, N=40) = 7.82, p = .060$), but ‘a’ vs. ‘b’ ($\chi^2(2, N=40) = 4.06, n.s.$) and ‘b’ vs. ‘c’ ($\chi^2(2, N=40) = 3.48, n.s.$) were not detectably different.”</td>
</tr>
</tbody>
</table>
## GLM / GLMM

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Ordinal logistic regression  | Z-test | Btw    | # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and ordinal response (Y) library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm")) |
|         |        |                              |      |        | Btw, Within                                                                  # df has subjects (S), one factor (X) w/levels (a,b,c), and ordinal response (Y)  
library(multcomp) # for adjusted  
library(emmeans) # for as.glht, pairs, emmeans  
summary(as.glht(pairs(emmeans(m, ~ X))), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |
| 1       | ≥2     | Poisson regression          | Z-test | Btw, Within | # df has subjects (S), one factor (X) w/levels (a,b,c), and count response (Y) library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm"))  
# the following also performs the equivalent contrast tests  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |
**GLM / GLMM**

*Post hoc pairwise comparisons – One factor*

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Ordinal logistic regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that Likert scores for ‘a’ vs. ‘c’ were statistically significantly different (Z = 2.78, p &lt; .05), but not for ‘a’ vs. ‘b’ (Z = 0.85, n.s.) or ‘b’ vs. ‘c’ (Z = 1.82, n.s.).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that Likert scores for ‘a’ vs. ‘c’ were statistically significantly different (Z = -2.78, p &lt; .05), but not for ‘a’ vs. ‘b’ (Z = -0.85, n.s.) or for ‘b’ vs. ‘c’ (Z = -1.82, n.s.).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that Likert scores for ‘a’ vs. ‘c’ were statistically significantly different (Z = 2.57, p &lt; .05), and for ‘a’ vs. ‘b’ were marginal (Z = 2.20, p = .056), but Likert scores for ‘b’ vs. ‘c’ were not statistically significantly different (Z = 0.42, n.s.).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that Likert scores for ‘a’ vs. ‘c’ were statistically significantly different (Z = -2.57, p &lt; .05), for ‘a’ vs. ‘b’ were marginal (Z = -2.20, p = .071), and for ‘b’ vs. ‘c’ were not statistically significantly different (Z = -0.42, n.s.).”</td>
</tr>
</tbody>
</table>
GLM / GLMM

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Poisson regression | Z-test | Btwn | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘b’ vs. ‘c’ (Z = -2.81, p < .05) and ‘a’ vs. ‘c’ (Z = -2.33, p < .05) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = 0.50, n.s.).”

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘b’ vs. ‘c’ difference was statistically significant (Z = -2.81, p < .05), but the ‘a’ vs. ‘c’ difference was only marginal (Z = -2.33, p = .053), and the ‘a’ vs. ‘b’ difference was not statistically significant (Z = 0.50, n.s.).” |

|         |        |               |      | Within | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference was statistically significant (Z = 3.21, p < .01), but differences for ‘a’ vs. ‘b’ and ‘b’ vs. ‘c’ were not.” |

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ difference was statistically significant (Z = 3.21, p < .01), but differences for ‘a’ vs. ‘b’ and ‘b’ vs. ‘c’ were not.” |
## GLM / GLMM

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Zero-inflated Poisson regression  | Z-test | Btwn | # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and count response (Y)  
library(multcomp) # for adjusted  
library(emmeans) # for as.glht, pairs, emmeans  
summary(as.glht(pairs(emmeans(m, ~ X)))))
test=adjusted(type="holm")) |
|         |        |                                   |      |     | Within  
# df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and count response (Y)  
library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey")))
test=adjusted(type="holm"))                                                   |
|         |        |                                   |      |     | t-test  
# df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and count response (Y)  
library(multcomp) # for glht  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X)))
test=adjusted(type="holm"))   
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |
### GLM / GLMM

**Post hoc pairwise comparisons – One factor**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Zero-inflated Poisson regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘b’ (Z = 3.28, p &lt; .01) and ‘b’ vs. ‘c’ (Z = -3.14, p &lt; .01) differences were statistically significant, but not the ‘a’ vs. ‘c’ difference (Z = -0.29, n.s.).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Within pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference was statistically significant (Z = 3.29, p &lt; .01), but the ‘a’ vs. ‘b’ (Z = 1.43, n.s.) and ‘b’ vs. ‘c’ (Z = 1.75, n.s.) differences were not.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>t-test</td>
<td></td>
<td>“Pairwise comparisons using t-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference was statistically significant (t(53) = 3.29, p &lt; .01), but the ‘a’ vs. ‘b’ (t(53) = 1.43, n.s.) and ‘b’ vs. ‘c’ (t(53) = 1.75, n.s.) differences were not.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ difference was statistically significant (t(53) = 3.29, p &lt; .01), but the ‘a’ vs. ‘b’ (t(53) = 1.43, n.s.) and ‘b’ vs. ‘c’ (t(53) = 1.75, n.s.) differences were not.”</td>
</tr>
</tbody>
</table>
## GLM / GLMM

### Post hoc pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Negative binomial regression | Z-test | Btwn, Within | # df has subjects (S), one factor (X) w/levels (a,b,c), and continuous response (Y)  
library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey"))), test=adjusted(type="holm"))  
# the following also performs the equivalent contrast tests  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |

| 1       | ≥2     | Zero-inflated negative binomial regression | Z-test | Btwn | # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and count response (Y)  
library(multcomp) # for adjusted  
library(emmeans) # for as.glht, pairs, emmeans  
summary(as.glht(pairs(emmeans(m, ~ X))), test=adjusted(type="holm")) |

|        |        |                           |      |     | # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and count response (Y)  
library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey"))), test=adjusted(type="holm")) |

|        |        |                           |      |     | # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and count response (Y)  
library(multcomp) # for glht, mcp  
summary(glht(m, mcp(X="Tukey"))), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's  
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score")) |
### GLM / GLMM

**Post hoc pairwise comparisons – One factor**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Negative binomial regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ (Z = -2.81, p &lt; .01) and ‘b’ vs. ‘c’ (Z = -3.21, p &lt; .01) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = 0.40, n.s.).” Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ (Z = -2.81, p &lt; .05) and ‘b’ vs. ‘c’ (Z = -3.21, p &lt; .01) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = 0.40, n.s.).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ (Z = 4.14, p &lt; .001) and ‘b’ vs. ‘c’ (Z = 3.98, p &lt; .001) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = 0.20, n.s.).” Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ (Z = 4.14, p &lt; .001) and ‘b’ vs. ‘c’ (Z = 3.98, p &lt; .001) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = 0.20, n.s.).”</td>
</tr>
</tbody>
</table>
### GLM / GLMM

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>≥2</td>
<td>Zero-inflated negative binomial regression</td>
<td>Z-test</td>
<td>B/wn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ (Z = -2.35, p &lt; .05) and ‘b’ vs. ‘c’ (Z = -2.90, p &lt; .05) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = 0.78, n.s.).”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Within</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference was statistically significant (Z = 3.16, p &lt; .01), but not the ‘a’ vs. ‘b’ (Z = 1.53, n.s.) or the ‘b’ vs. ‘c’ (Z = 1.96, n.s.) differences.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>t-test</td>
<td></td>
<td>“Pairwise comparisons using t-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference was statistically significant (t(52) = 3.16, p &lt; .01), but not the ‘a’ vs. ‘b’ (t(52) = 1.53, n.s.) or the ‘b’ vs. ‘c’ (t(52) = 1.96, n.s.) differences.”</td>
</tr>
</tbody>
</table>

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ difference was statistically significant (t(52) = 3.16, p < .01), but not the ‘a’ vs. ‘b’ (t(52) = 1.53, n.s.) or the ‘b’ vs. ‘c’ (t(52) = 1.96, n.s.) differences.”
### GLM / GLMM

*Post hoc* pairwise comparisons – One factor

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Gamma regression, including Exponential regression | Z-test | Btwn, Within | ```
# df has subjects (S), one factor (X) w/levels (a,b,c), and continuous response (Y)
library(multcomp) # for glht, mcp
summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm"))
```

```
# the following also performs the equivalent contrast tests
library(emmeans) # for emm, emmeans
summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm"))
```

```
# or, using the Tukey HSD correction instead of Holm's
summary(emmeans(m, pairwise ~ X, adjust="tukey", mode="linear.predictor", type="Score"))
```
### GLM / GLMM

*Post hoc pairwise comparisons – One factor*

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
</table>
| 1       | ≥2     | Gamma regression, including Exponential regression | Z-test | Btwn | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘b’ difference was statistically significant (Z = 3.44, p < .01), but differences from the other two pairwise comparisons were not.”
|         |        |               |      |     | “Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘b’ difference was statistically significant (Z = 3.44, p < .01), but differences from the other two pairwise comparisons were not.” |
|         |        |               |      |     | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference (Z = -3.87, p < .001) and the ‘b’ vs. ‘c’ difference (Z = -3.63, p < .001) were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = -0.30, n.s.).”
|         |        |               |      | Within | “Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ difference (Z = -3.87, p < .001) and the ‘b’ vs. ‘c’ difference (Z = -3.63, p < .001) were statistically significant, but not the ‘a’ vs. ‘b’ difference (Z = -0.30, n.s.).” |
# GLM / GLMM

*Post hoc* pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Linear regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1<em>X2)), test=adjusted(type=&quot;holm&quot;)) # or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1</em>X2, adjust=&quot;tukey&quot;, mode=&quot;linear.predictor&quot;, type=&quot;Score&quot;))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Logistic regression</td>
<td>Z-test</td>
<td>Btwn, Within</td>
<td># df has subjects (S), two factors (X1,X2) each w/levels (a,b), and dichotomous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1<em>X2)), test=adjusted(type=&quot;holm&quot;)) # or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1</em>X2, adjust=&quot;tukey&quot;, mode=&quot;linear.predictor&quot;, type=&quot;Score&quot;))</td>
</tr>
</tbody>
</table>

**t-test** | Within | # df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1*X2, adjust="tukey", mode="kenward-roger", type="Score")) |
### GLM / GLMM

*Post hoc pairwise comparisons – Multiple factors*

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
</table>
| ≥2      | ≥2     | Linear regression | Z-test | Btwn | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant (Z = -3.05, p < .05), as was the (b,a) vs. (a,b) difference (Z = 2.59, p < .05). The other pairwise comparisons were not statistically significantly different.”

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant (Z = -3.05, p < .05), as was the (b,a) vs. (a,b) difference (Z = 2.59, p < .05). The other pairwise comparisons were not statistically significantly different.” |

| t-test  | Within | “Pairwise comparisons using t-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was marginal (t(42) = -2.71, p = .058). No other pairwise comparisons were statistically significant or marginal.”

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant (t(42) = -2.71, p < .05). The other pairwise comparisons were not statistically significant.” |
### GLM / GLMM

**Post hoc pairwise comparisons – Multiple factors**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>Report</th>
</tr>
</thead>
</table>
| ≥2      | ≥2     | Logistic regression | Z-test | Btwn | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = -2.85, p < .05$). Furthermore, the (a,a) vs. (a,b) difference was marginal ($Z = -2.40, p = .083$). No other pairwise comparisons were detectably different or marginal.”  

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = -2.85, p < .05$). Furthermore, the (a,a) vs. (a,b) difference was marginal ($Z = -2.40, p = .078$). No other pairwise comparisons were detectably different or marginal.” |
|         |        |               |        |     |        |
| Within  |        |               |        |     | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) and (a,a) vs. (a,b) differences were both marginal (both: $Z = 2.45, p = .086$). No other pairwise comparisons were detectably different or marginal.”  

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) and (a,a) vs. (a,b) differences were both marginal (both: $Z = 2.45, p = .068$). No other pairwise comparisons were detectably different or marginal.” |
<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Multinomial logistic regression</td>
<td>Pairwise multinomial logistic regression</td>
<td>Btwn</td>
<td></td>
</tr>
</tbody>
</table>

# df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and polytomous response (Y)
library(nnet) # for multinom
library(car) # for Anova

df2 <- df[(df$X1 == "a" & df$X2 == "a") | (df$X1 == "a" & df$X2 == "b"),] # aa vs. ab
df2$X12 = with(df2, interaction(X1, X2))
m = multinom(Y ~ X12, data=df2)
aa_ab = Anova(m, type=3)

df2 <- df[(df$X1 == "a" & df$X2 == "a") | (df$X1 == "b" & df$X2 == "a"),] # aa vs. ba
df2$X12 = with(df2, interaction(X1, X2))
m = multinom(Y ~ X12, data=df2)
aa_ba = Anova(m, type=3)

df2 <- df[(df$X1 == "a" & df$X2 == "a") | (df$X1 == "b" & df$X2 == "b"),] # aa vs. bb
df2$X12 = with(df2, interaction(X1, X2))
m = multinom(Y ~ X12, data=df2)
aa_bb = Anova(m, type=3)

df2 <- df[(df$X1 == "a" & df$X2 == "b") | (df$X1 == "b" & df$X2 == "a"),] # ab vs. ba
df2$X12 = with(df2, interaction(X1, X2))
m = multinom(Y ~ X12, data=df2)
ab_ba = Anova(m, type=3)

df2 <- df[(df$X1 == "a" & df$X2 == "b") | (df$X1 == "b" & df$X2 == "b"),] # ab vs. bb
df2$X12 = with(df2, interaction(X1, X2))
m = multinom(Y ~ X12, data=df2)
ab_bb = Anova(m, type=3)

# ... Continued on next slide ... #
## GLM / GLMM

*Post hoc* pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| ≥2      | ≥2     | Multinomial logistic regression | Pairwise multinomial logistic regression | Btwn | # ... Continued from previous slide ... #

```r
df2 <- df[(df$X1 == "b" & df$X2 == "a") | (df$X1 == "b" & df$X2 == "b"),] # ba vs. bb
df2$X12 = with(df2, interaction(X1, X2))
m = multinom(Y ~ X12, data=df2)
ba_bb = Anova(m, type=3)
p.adjust(c(
  aa_ab$`Pr(>Chisq)`,
  aa_ba$`Pr(>Chisq)`,
  aa_bb$`Pr(>Chisq)`,
  ab_ba$`Pr(>Chisq)`,
  ab_bb$`Pr(>Chisq)`,
  ba_bb$`Pr(>Chisq)`), method="holm")
```

- GLM / GLMM
- Post hoc pairwise comparisons – Multiple factors
<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)</td>
<td>Pairwise mixed Poisson regression</td>
</tr>
</tbody>
</table>

**GLM / GLMM**

*Post hoc* pairwise comparisons – Multiple factors

### Factors
- **df** has subjects ($S$), two within-$S$s factors ($X_1, X_2$) each w/levels (a,b), and polytomous response ($Y$)
- df2 was produced by mlogit.data and has a logical response ($Y$) and alt factor

```r
library(lme4) # for glmer
library(lmerTest)
library(car) # for Anova
```

```r
df3 <- df2[(df2$X1 == "a" & df2$X2 == "a") | (df2$X1 == "a" & df2$X2 == "b"),] # aa vs. ab
df3$X12 = with(df3, interaction(X1, X2))
m = glmer(Y ~ X12*alt + (1|S), data=df3, family=poisson)
a = Anova(m, type=2)
aa_ab = a[grep("X12:alt", rownames(a)),]
```

```r
df3 <- df2[(df2$X1 == "a" & df2$X2 == "a") | (df2$X1 == "b" & df2$X2 == "a"),] # aa vs. ba
df3$X12 = with(df3, interaction(X1, X2))
m = glmer(Y ~ X12*alt + (1|S), data=df3, family=poisson)
a = Anova(m, type=2)
aa_ba = a[grep("X12:alt", rownames(a)),]
```

```r
df3 <- df2[(df2$X1 == "a" & df2$X2 == "a") | (df2$X1 == "b" & df2$X2 == "b"),] # aa vs. bb
df3$X12 = with(df3, interaction(X1, X2))
m = glmer(Y ~ X12*alt + (1|S), data=df3, family=poisson)
a = Anova(m, type=2)
aa_bb = a[grep("X12:alt", rownames(a)),]
```

```r
df3 <- df2[(df2$X1 == "a" & df2$X2 == "b") | (df2$X1 == "b" & df2$X2 == "a"),] # ab vs. ba
df3$X12 = with(df3, interaction(X1, X2))
m = glmer(Y ~ X12*alt + (1|S), data=df3, family=poisson)
a = Anova(m, type=2)
ab_ba = a[grep("X12:alt", rownames(a)),]
```

# ... Continued on next slide ... #
## GLM / GLMM

*Post hoc* pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| ≥2      | ≥2     | Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994) | Pairwise mixed Poisson regression | Within | # ... Continued from previous slide ... #
|         |        |                                                                   |                       |     | df3 <- df2[(df2$X1 == "a" & df2$X2 == "b") | (df2$X1 == "b" & df2$X2 == "b"),] # ab vs. bb
|         |        |                                                                   |                       |     | df3$X12 = with(df3, interaction(X1, X2))
|         |        |                                                                   |                       |     | m = glmer(Y ~ X12*alt + (1|S), data=df3, family=poisson)
|         |        |                                                                   |                       |     | a = Anova(m, type=2)
|         |        |                                                                   |                       |     | ab_bb = a[grep("X12:alt", rownames(a)),]
|         |        |                                                                   |                       |     | df3 <- df2[(df2$X1 == "b" & df2$X2 == "a") | (df2$X1 == "b" & df2$X2 == "b"),] # ba vs. bb
|         |        |                                                                   |                       |     | df3$X12 = with(df3, interaction(X1, X2))
|         |        |                                                                   |                       |     | m = glmer(Y ~ X12*alt + (1|S), data=df3, family=poisson)
|         |        |                                                                   |                       |     | a = Anova(m, type=2)
|         |        |                                                                   |                       |     | ba_bb = a[grep("X12:alt", rownames(a)),]
|         |        |                                                                   |                       |     | p.adjust(c(
|         |        |                                                                   |                       |     |   aa_ab$`Pr(>Chisq)`",
|         |        |                                                                   |                       |     |   aa_ba$`Pr(>Chisq)`",
|         |        |                                                                   |                       |     |   aa_bb$`Pr(>Chisq)`",
|         |        |                                                                   |                       |     |   ab_ba$`Pr(>Chisq)`",
|         |        |                                                                   |                       |     |   ab_bb$`Pr(>Chisq)`",
|         |        |                                                                   |                       |     |   ba_bb$`Pr(>Chisq)`"), method="holm")
### GLM / GLMM

*Post hoc pairwise comparisons – Multiple factors*

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Multinomial logistic regression</td>
<td>Pairwise multinomial logistic regression</td>
<td>Btwn</td>
<td>“Six post hoc multinomial logistic regressions, one for each combination of the levels of X1 and X2, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,b) vs. (b,a) difference was statistically significant ($\chi^2(6, N=30) = 20.00, p &lt; .05$). No other pairwise comparisons were statistically significant.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)</td>
<td>Pairwise mixed Poisson regression</td>
<td>Within</td>
<td>“Six post hoc mixed Poisson regressions, one for each combination of the levels of X1 and X2, corrected with Holm’s sequential Bonferroni procedure, indicated that the difference in the number of ‘x’, ‘y’, and ‘z’ responses was marginal for (a,b) vs. (b,b) ($\chi^2(2, N=30) = 8.68, p = .078$). No other pairwise comparisons were detectably different or marginal.”</td>
</tr>
</tbody>
</table>
### GLM / GLMM

*Post hoc* pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| ≥2     | ≥2     | Ordinal logistic regression | Z-test | Btwn, Within | # df has subjects (S), two factors (X1,X2) each w/levels (a,b), and ordinal response (Y)  
library(multcomp) # for adjusted  
library(emmeans) # for as.glht, pairs, emmeans  
summary(as.glht(pairs(emmeans(m, ~ X1*X2)))), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1*X2, adjust="tukey", mode="linear.predictor", type="Score")) |
| ≥2     | ≥2     | Poisson regression | Z-test | Btwn, Within | # df has subjects (S), two factors (X1,X2) each w/levels (a,b), and count response (Y)  
library(multcomp) # for glht  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1*X2, adjust="tukey", mode="linear.predictor", type="Score")) |
| ≥2     | ≥2     | Zero-inflated Poisson regression | Z-test | Btwn | # df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and count response (Y)  
library(multcomp) # for adjusted  
library(emmeans) # for as.glht, pairs, emmeans  
summary(as.glht(pairs(emmeans(m, ~ X1*X2)))), test=adjusted(type="holm")) |
|         |        |               |      |     | t-test | # df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and count response (Y)  
library(multcomp) # for glht  
library(emmeans) # for emm, emmeans  
summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm"))  
# or, using the Tukey HSD correction instead of Holm's summary(emmeans(m, pairwise ~ X1*X2, adjust="tukey", mode="linear.predictor", type="Score")) |
### GLM / GLMM

*Post hoc* pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Ordinal logistic</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,b) difference was marginal (Z = 2.55, p = .065). No other pairwise comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>regression</td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,b) difference was marginal (Z = -2.55, p = .053). No other pairwise comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Poisson regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Within</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the following differences were statistically significant: (a,a) vs. (b,a) (Z = 3.25, p &lt; .01); (b,a) vs. (a,b) (Z = -3.78, p &lt; .001); and (a,b) vs. (b,b) (Z = 2.76, p &lt; .05). Furthermore, the (a,a) vs. (b,b) difference was marginal (Z = 2.20, p = .083). No other comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the following differences were statistically significant: (a,a) vs. (b,a) (Z = 3.25, p &lt; .01); (b,a) vs. (a,b) (Z = -3.78, p &lt; .001); and (a,b) vs. (b,b) (Z = 2.76, p &lt; .05). No other comparisons were statistically significant or marginal.”</td>
</tr>
</tbody>
</table>
**GLM / GLMM**

*Post hoc pairwise comparisons – Multiple factors*

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Zero-inflated Poisson regression</td>
<td><em>Z</em>-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using <em>Z</em>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>t</em>-test</td>
<td>Within</td>
<td>“Pairwise comparisons using <em>t</em>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant (<em>t</em>(51) = 2.89, <em>p</em> &lt; .05). No other comparisons were detectably different.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant (<em>t</em>(51) = 2.89, <em>p</em> &lt; .05). No other comparisons were detectably different.”</td>
</tr>
</tbody>
</table>
## GLM / GLMM

### Post hoc pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2 ≥2</td>
<td>Negative binomial regression</td>
<td>Z-test, Btwn, Within</td>
<td>library(multcomp) # for <code>glt</code>&lt;br&gt;library(emmeans) # for <code>emm</code>, <code>emmeans</code>&lt;br&gt;summary(glt(m, emm(pairwise ~ X1<em>X2)), test=adjusted(type=&quot;holm&quot;)) # or, using the Tukey HSD correction instead of Holm's&lt;br&gt;summary(emmeans(m, pairwise ~ X1</em>X2, adjust=&quot;tukey&quot;, mode=&quot;linear.predictor&quot;, type=&quot;Score&quot;))</td>
<td></td>
</tr>
<tr>
<td>≥2 ≥2</td>
<td>Zero-inflated negative binomial regression</td>
<td>Z-test, Btwn</td>
<td>library(multcomp) # for adjusted&lt;br&gt;library(emmeans) # for as.glht, <code>pairs</code>, <code>emmeans</code>&lt;br&gt;summary(as.glht(pairs(emmeans(m, ~ X1*X2))), test=adjusted(type=&quot;holm&quot;))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>t-test, Within</td>
<td>library(multcomp) # for <code>glt</code>&lt;br&gt;library(emmeans) # for <code>emm</code>, <code>emmeans</code>&lt;br&gt;summary(glt(m, emm(pairwise ~ X1<em>X2)), test=adjusted(type=&quot;holm&quot;)) # or, using the Tukey HSD correction instead of Holm's&lt;br&gt;summary(emmeans(m, pairwise ~ X1</em>X2, adjust=&quot;tukey&quot;, mode=&quot;linear.predictor&quot;, type=&quot;Score&quot;))</td>
<td></td>
</tr>
<tr>
<td>≥2 ≥2</td>
<td>Gamma regression, including Exponential regression</td>
<td>Z-test, Btwn, Within</td>
<td>library(multcomp) # for <code>glt</code>&lt;br&gt;library(emmeans) # for <code>emm</code>, <code>emmeans</code>&lt;br&gt;summary(glt(m, emm(pairwise ~ X1<em>X2)), test=adjusted(type=&quot;holm&quot;)) # or, using the Tukey HSD correction instead of Holm's&lt;br&gt;summary(emmeans(m, pairwise ~ X1</em>X2, adjust=&quot;tukey&quot;, mode=&quot;linear.predictor&quot;, type=&quot;Score&quot;))</td>
<td></td>
</tr>
</tbody>
</table>
### GLM / GLMM

**Post hoc pairwise comparisons – Multiple factors**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥2</td>
<td>≥2</td>
<td>Negative binomial regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the differences between (a,a) vs. (b,b) ((Z = -2.82, p &lt; .05)) and (a,b) vs. (b,b) ((Z = -3.25, p &lt; .01)) were statistically significant. In addition, the difference between (b,a) vs. (a,b) was marginal ((Z = 2.34, p = .078)). No other differences were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the differences between (a,a) vs. (b,b) ((Z = -2.82, p &lt; .05)) and (a,b) vs. (b,b) ((Z = -3.25, p &lt; .01)) were statistically significant. In addition, the difference between (b,a) vs. (a,b) was marginal ((Z = 2.34, p = .090)). No other differences were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zero-inflated negative binomial regression</td>
<td>Z-test</td>
<td>Btwn</td>
<td>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant or marginal.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that these differences were statistically significant: (a,a) vs. (b,a) ((t(50) = -2.95, p &lt; .05)); (a,a) vs. (b,b) ((t(50) = -3.87, p &lt; .01)); (b,a) vs. (a,b) ((t(50) = 4.04, p &lt; .001)); and (a,b) vs. (b,b) ((t(50) = -4.81, p &lt; .001)). No other comparisons were detectably different.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that these differences were statistically significant: (a,a) vs. (b,a) ((t(50) = -2.95, p &lt; .05)); (a,a) vs. (b,b) ((t(50) = -3.87, p &lt; .01)); (b,a) vs. (a,b) ((t(50) = 4.04, p &lt; .01)); and (a,b) vs. (b,b) ((t(50) = -4.81, p &lt; .001)). No other comparisons were detectably different.”</td>
</tr>
</tbody>
</table>
### GLM / GLMM

*Post hoc* pairwise comparisons – Multiple factors

<table>
<thead>
<tr>
<th>Factors</th>
<th>Levels</th>
<th>Omnibus Model</th>
<th>Test</th>
<th>B/W</th>
<th>R Code</th>
</tr>
</thead>
</table>
| ≥2      | ≥2     | Gamma regression, including Exponential regression | Z-test | Btwn | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant.”

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no comparisons were statistically significant.”

|   |   |   |   | Within | “Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant (Z = 3.18, p < .01). Also, the (b,a) vs. (b,b) difference was statistically significant (Z = -2.89, p < .05). No other pairwise comparisons were statistically significant.”

“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant (Z = 3.18, p < .01). Also, the (b,a) vs. (b,b) difference was significant (Z = -2.89, p < .05). No other pairwise comparisons were statistically significant.”

---
Tests of Proportion & Association:


- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Series 5*, vol. 50, no. 302, pp. 157-175. [https://doi.org/10.1080/14786440009463897](https://doi.org/10.1080/14786440009463897)


Distribution Tests:


ANOVA Assumptions:


Linear Models (LM):

Linear Mixed Models (LMM):


Classic Nonparametric Tests:


Aligned Rank Transform (ART):


Generalized Linear Models (GLM):


Generalized Linear Mixed Models (GLMM):

Binomial Regression:


Multinomial Regression:


Ordinal Regression:


Bibliography
Bibliography

Poisson Regression:


Negative Binomial Regression:


Post Hoc Comparisons:


