

#Rstats

Statistical Analysis & Reporting in R

Jacob O. Wobbrock, Ph.D.

The Information School | DUB Group

University of Washington

wobbrock@uw.edu

Table of Analyses

Proportion & Association

Samples	Categories	Tests
1	2	One-sample χ^2 test, binomial test
1	≥ 2	One-sample χ^2 test, multinomial test
2	≥ 2	Two-sample χ^2 test, <i>G</i> -test, Fisher's exact test
2 (dep.)	≥ 2	Symmetry test

Assumptions

- Normality:
Shapiro-Wilk test
- Homoscedasticity:
Levene's test
- Sphericity:
Mauchly's test

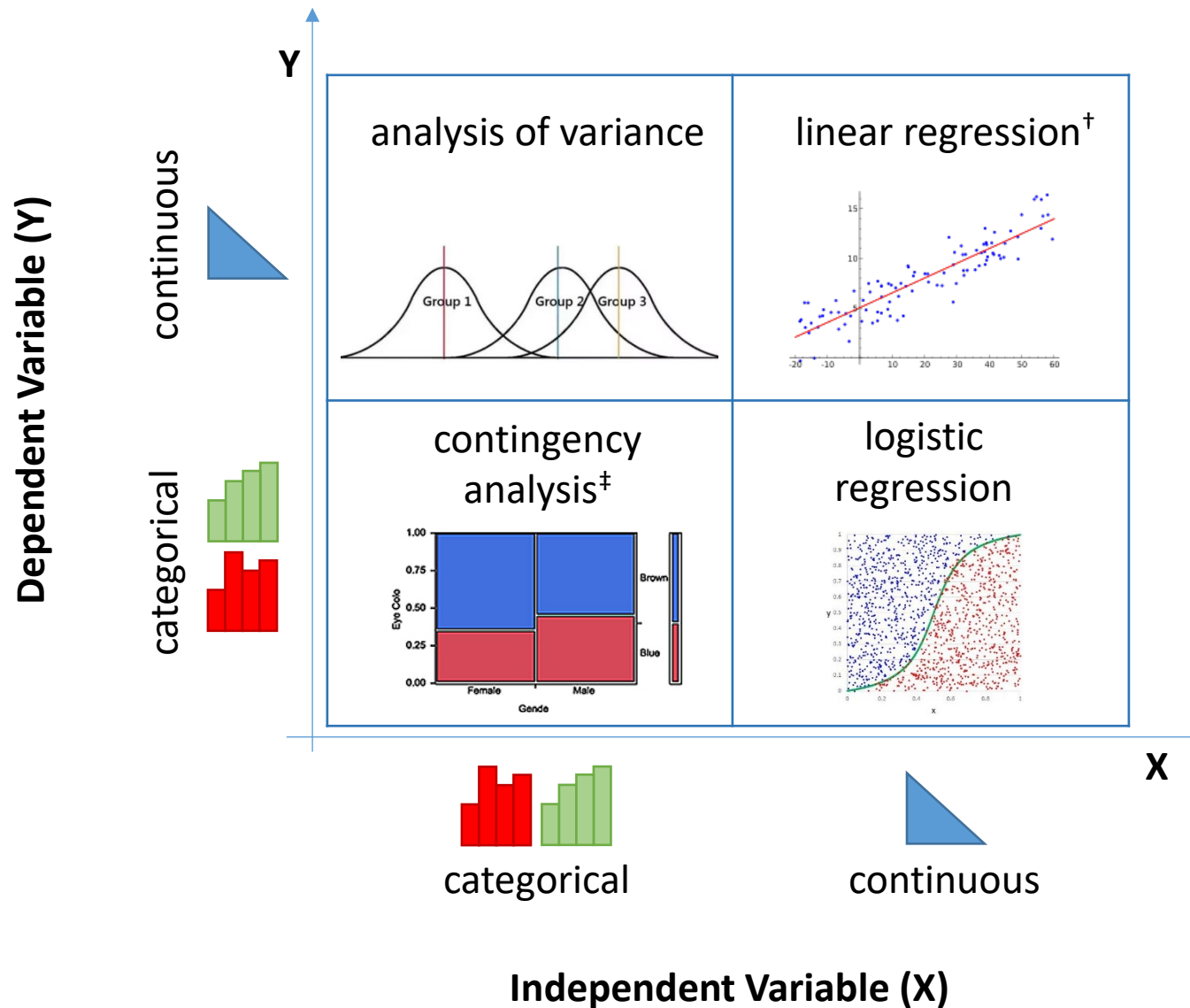
Distributions

- Shapiro-Wilk test
- χ^2 goodness-of-fit test

Analyses of Variance

Factors	Levels	<u>B</u> etween or <u>W</u> ithin	Parametric Tests	Nonparametric Tests	
			Linear Models	Generalized Models	
1	2	B	Independent-samples <i>t</i> -test	Median test	Mann-Whitney <i>U</i> test
1	2	W	Paired-samples <i>t</i> -test	Sign test	Wilcoxon signed-rank test
1	≥ 2	B	One-way ANOVA	Kruskal-Wallis test	
1	≥ 2	W	One-way repeated measures ANOVA	Friedman test	
≥ 2	≥ 2	B	Factorial ANOVA Linear Model (LM)	Aligned Rank Transform (ART)	
				Generalized Linear Model (GLM)	
≥ 2	≥ 2	W	Factorial repeated measures ANOVA Linear Mixed Model (LMM)	Aligned Rank Transform (ART)	
				Generalized Linear Mixed Model (GLMM)	

Analysis Categories



Credit: This table is adapted from the “Fit Y by X” dialog in SAS® JMP™. <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.

Proportion & Association

Proportion & Association

One independent sample

Samples	Response Categories	Test Name	Exact Test?	R Code
1	2	Binomial test	Yes, use with $N \leq 200$	<pre># df is a long-format data table w/columns for participant (PId) and 2-category outcome (Y) df\$PId = factor(df\$PId) # participant 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")</pre>
1	≥ 2	Multinomial test	Yes, use with $N \leq 200$	<pre># df is a long-format data table w/columns for participant (PId) and N-category outcome (Y) library(XNomial) # for xmulti df\$PId = factor(df\$PId) # participant 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)</pre>
		One-sample Pearson chi-squared test	No, use with $N > 200$	<pre># df is a long-format data table w/columns for participant (PId) and N-category outcome (Y) df\$PId = factor(df\$PId) # participant 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)</pre>

Proportion & Association

One independent sample

Samples	Response Categories	Test Name	Exact Test?	Report
1	2	Binomial test	Yes, use with $N \leq 200$	“Out of 60 responses, 42 were ‘yes’ and 18 were ‘no’. A two-sided exact binomial test indicated that these proportions were significantly different from chance ($p = .003$).”
1	≥ 2	Multinomial test	Yes, use with $N \leq 200$	“Out of 60 outcomes, 17 were ‘yes’, 12 were ‘no’, and 31 were ‘maybe’. An exact multinomial test indicated that these proportions were statistically significantly different from chance ($p = .009$).”
		One-sample Pearson chi-squared test	No, use with $N > 200$	“Out of 60 outcomes, 17 were ‘yes’, 12 were ‘no’, and 31 were ‘maybe’. A one-sample Pearson chi-squared test indicated that these proportions were significantly different from chance ($\chi^2(2, N=60) = 9.70, p = .008$).”

Proportion & Association

Two independent samples

Samples	Response Categories	Test Name	Exact Test?	R Code
2	≥ 2	Fisher's exact test	Yes, use with $N \leq 200$	<pre># df is a long-format data table w/participant (PIId), factor (X), and N-category outcome (Y) df\$PIId = factor(df\$PIId) # participant is nominal (unused) df\$X = factor(df\$X) # X is a factor of m ≥ 2 levels df\$Y = factor(df\$Y) # Y is an outcome of n ≥ 2 categories xt = xtabs(~ X + Y, data=df) # make m × n crosstabs fisher.test(xt)</pre>
		G-test	No, use with $N > 200$	<pre># df is a long-format data table w/participant (PIId), factor (X), and N-category outcome (Y) library(RVAideMemoire) # for G.test df\$PIId = factor(df\$PIId) # participant is nominal (unused) df\$X = factor(df\$X) # X is a factor of m ≥ 2 levels df\$Y = factor(df\$Y) # Y is an outcome of n ≥ 2 categories xt = xtabs(~ X + Y, data=df) # make m × n crosstabs G.test(xt)</pre>
		Two-sample Pearson chi-squared test	No, use with $N > 200$	<pre># df is a long-format data table w/participant (PIId), factor (X), and N-category outcome (Y) df\$PIId = factor(df\$PIId) # participant is nominal (unused) df\$X = factor(df\$X) # X is a factor of m ≥ 2 levels df\$Y = factor(df\$Y) # Y is an outcome of n ≥ 2 categories xt = xtabs(~ X + Y, data=df) # make m × n crosstabs chisq.test(xt)</pre>

Proportion & Association

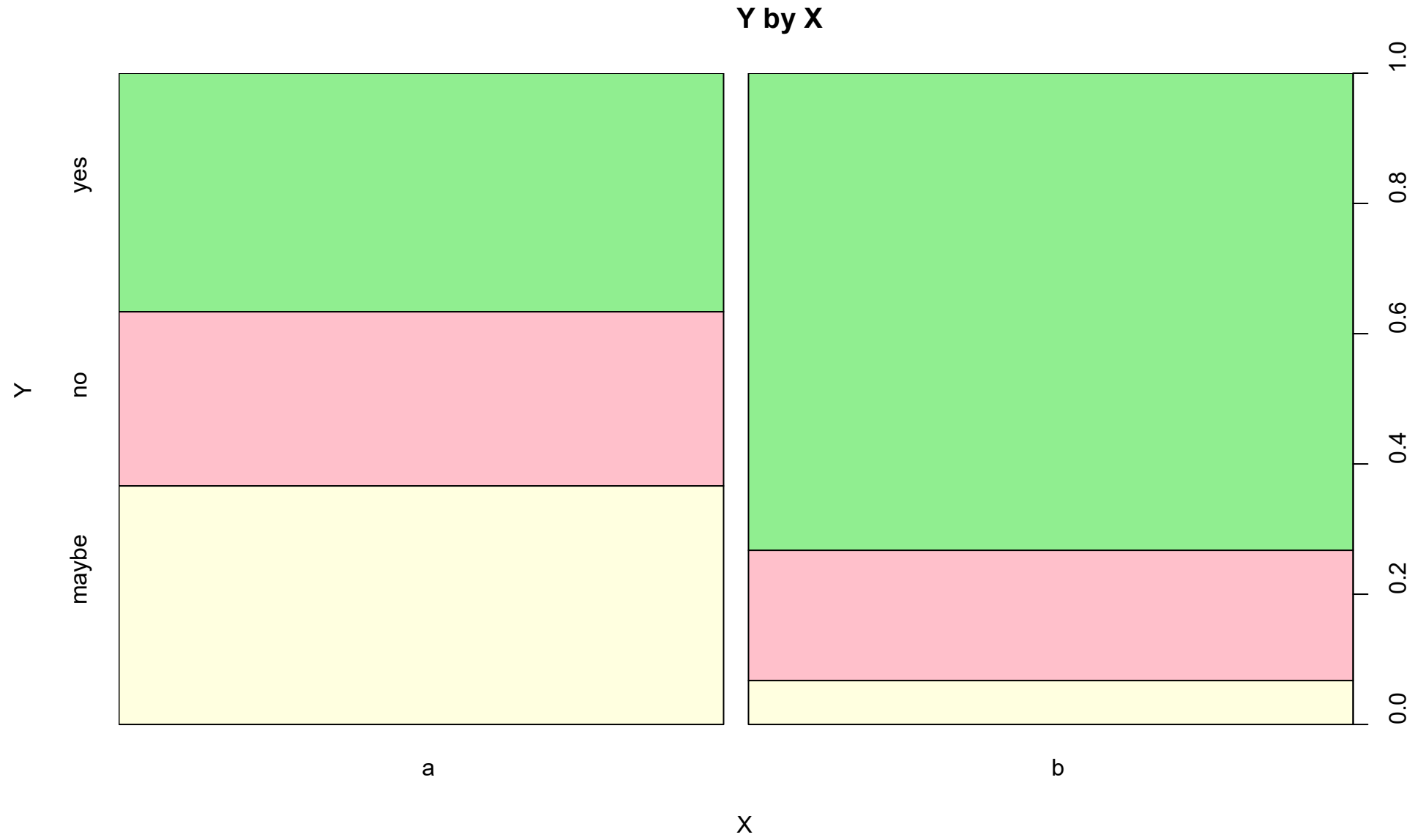
Two independent samples

Samples	Response Categories	Test Name	Exact Test?	Report
2	≥ 2	Fisher's exact test	Yes, use with $N \leq 200$	"Table 1 shows the counts of 'yes', 'no', and 'maybe' responses (Y) for each of 'a' and 'b' (X). Figure 1 plots these proportions. Fisher's exact test indicated a statistically significant association between X and Y ($p = .005$)."
		G-test	No, use with $N > 200$	"Table 1 shows the counts of 'yes', 'no', and 'maybe' responses (Y) for each of 'a' and 'b' (X). Figure 1 plots these proportions. A G-test indicated a statistically significant association between X and Y ($G(2) = 10.88, p = .004$)."
		Two-sample Pearson chi-squared test	No, use with $N > 200$	"Table 1 shows the counts of 'yes', 'no', and 'maybe' responses (Y) for each of 'a' and 'b' (X). Figure 1 plots these proportions. A two-sample Pearson chi-squared test indicated a statistically significant association between X and Y ($\chi^2(2, N=60) = 10.18, p = .006$)."

Table 1

		Y		
		yes	no	maybe
X	a	11	8	11
	b	22	6	2

Figure 1



Proportion & Association

Two dependent samples

Samples	Response Categories	Test Name	Exact Test?	R Code
2	≥ 2	Symmetry test	No, asymptotic	<pre># df is a long-format data table w/columns for participant (PIId), a within-Ss. factor (X), # and N-category outcome (Y) library(coin) # for symmetry_test df\$PIId = factor(df\$PIId) # participant is nominal df\$X = factor(df\$X) # X is a within-Ss. factor of ≥ 2 levels df\$Y = factor(df\$Y) # Y is an outcome of ≥ 2 categories symmetry_test(Y ~ X PIId, data=df)</pre>

Proportion & Association

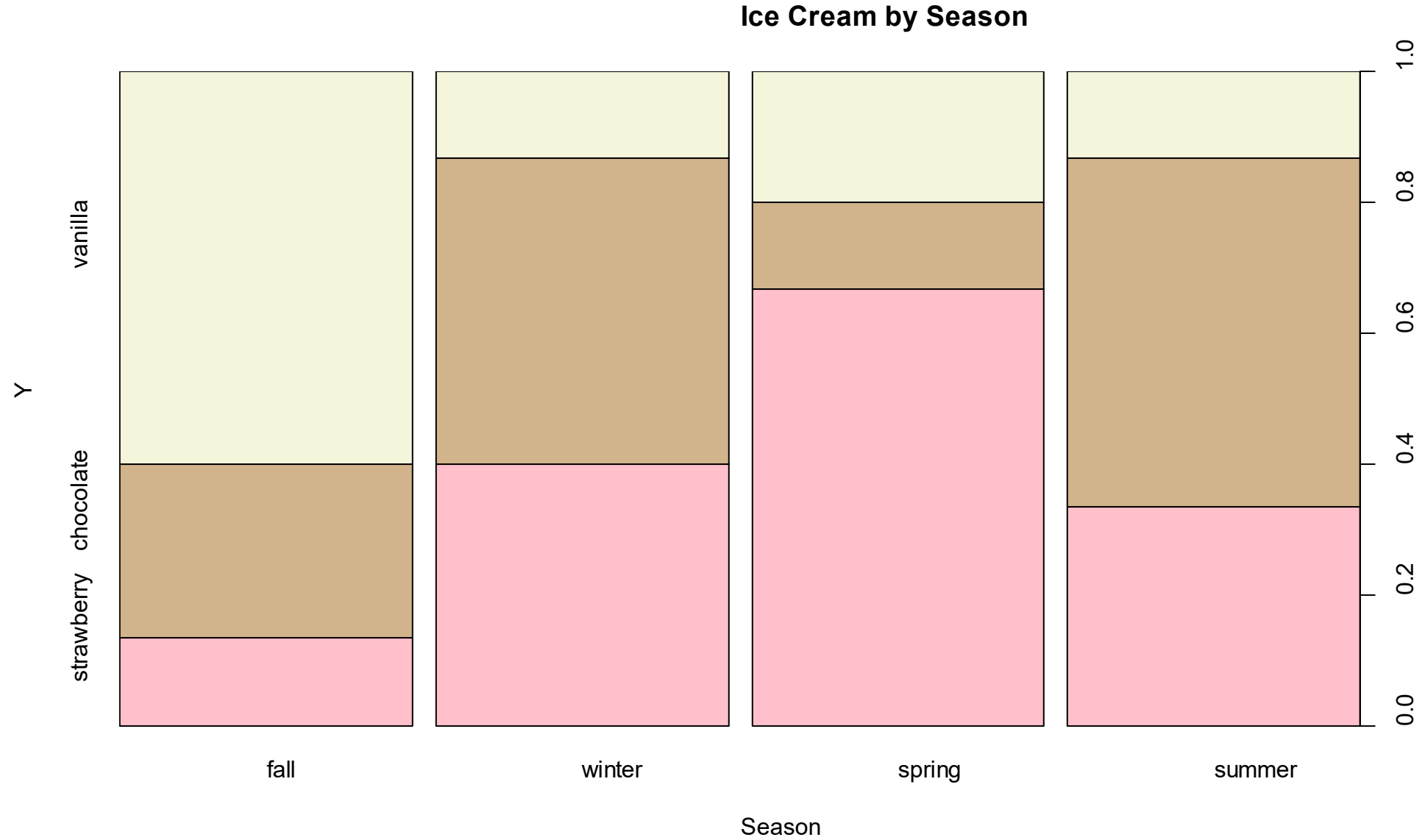
Two dependent samples

Samples	Response Categories	Test Name	Exact Test?	Report
2	≥ 2	Symmetry test	No, asymptotic	“Fifteen participants each provided four responses indicating their favorite ice cream flavor in each season (‘fall’, ‘winter’, ‘spring’, and ‘summer’). Table 2 shows the counts, and Figure 2 plots the proportions. Out of the 60 responses, 16 were ‘vanilla’, 21 were ‘chocolate’, and 23 were ‘strawberry’. A symmetry test shows that the ice cream proportions across seasons were significantly different ($p = .013$).”

Table 2

		Y		
		vanilla	chocolate	strawberry
X	fall	9	4	2
	winter	2	7	6
	spring	3	2	10
	summer	2	8	5

Figure 2



Proportion & Association

Post hoc comparisons

Proportion & Association

Post hoc tests – One independent sample

Samples	Response Categories	Omnibus Test	Contrast Test	R Code
1	≥2	Multinomial test	Pairwise binomial tests	<pre># For Y's response categories, test each in pairwise fashion against 50/50. yn = binom.test(c(sum(df\$Y == "yes"), sum(df\$Y == "no")), p=1/2) # yes vs. no ym = binom.test(c(sum(df\$Y == "yes"), sum(df\$Y == "maybe")), p=1/2) # yes vs. maybe nm = binom.test(c(sum(df\$Y == "no"), sum(df\$Y == "maybe")), p=1/2) # no vs. maybe p.adjust(c(yn\$p.value, ym\$p.value, nm\$p.value), method="holm") # or, equivalently, if xt is a table of counts for each category in Y library(RVAideMemoire) # for multinomial.multcomp multinomial.multcomp(xt, p.method="holm") # same results as above</pre>
1	≥2	One-sample Pearson chi-squared test	Pairwise chi-squared tests	<pre># 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: qchisq(1 - .011, df=1) # 6.465</pre>
1	≥2	Multinomial test, one-sample Pearson chi-squared test	Individual binomial tests against chance	<pre># A different kind of post hoc test for one sample. For Y's response categories, # test each proportion against chance. # df is a long-format data table w/participant (Pid) and N-category outcome (Y) y = binom.test(sum(df\$Y == "yes"), nrow(df), p=1/3) # yes n = binom.test(sum(df\$Y == "no"), nrow(df), p=1/3) # no m = binom.test(sum(df\$Y == "maybe"), nrow(df), p=1/3) # maybe p.adjust(c(y\$p.value, n\$p.value, m\$p.value), method="holm")</pre>

Proportion & Association

Post hoc tests – One independent sample

Samples	Response Categories	Omnibus Test	Contrast Test	Report
1	≥2	Multinomial test	Pairwise binomial tests	“Three <i>post hoc</i> pairwise comparisons using exact binomial tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that the proportions of ‘yes’ vs. ‘no’ and ‘yes’ vs. ‘maybe’ were not significantly different, but that the proportions of ‘no’ vs. ‘maybe’ were ($p = .016$).”
1	≥2	One-sample Pearson chi-squared test	Pairwise chi-squared tests	“Three <i>post hoc</i> pairwise comparisons using Pearson chi-squared tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that only the proportions of ‘no’ vs. ‘maybe’ were significantly different ($\chi^2(1, N=43) = 6.47, p = .011$).”
1	≥2	Multinomial test, one-sample Pearson chi-squared test	Individual binomial tests against chance	“Three <i>post hoc</i> binomial tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that only the number of ‘maybe’ responses was significantly different from chance ($p = .011$).”

Proportion & Association

Post hoc tests – Two independent samples

Samples	Response Categories	Omnibus Test	Contrast Test	R Code
2	≥2	Fisher's exact test	Pairwise Fisher's exact tests	<pre># xt is an m × n crosstabs with categories X and Y library(RVAideMemoire) # for fisher.multcomp fisher.multcomp(xt, p.method="holm") # xt shows levels</pre>
2	≥2	G-test	Pairwise G-tests	<pre># xt is an m × n crosstabs with categories X and Y library(RVAideMemoire) # for G.multcomp G.multcomp(xt, p.method="holm") # xt shows levels</pre>
2	≥2	Two-sample Pearson chi-squared test	Pairwise chi-squared tests	<pre># xt is an m × n crosstabs with categories X and 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: qchisq(1 - .00067, df=1) # 11.571</pre>
2	≥2	Fisher's exact test, G-test, two-sample Pearson chi-squared test	Individual binomial tests against chance	<pre># A different kind of post hoc test for two samples. For X's categories (a,b) and Y's # response categories (yes, no, maybe), test each proportion of Y within each level of X # against chance. # df is a long-format data table w/participant (PIId), factor (X), and N-category outcome (Y) ay = binom.test(sum(df[df\$X == "a",]\$Y == "yes"), nrow(df[df\$X == "a",]), p=1/3) an = binom.test(sum(df[df\$X == "a",]\$Y == "no"), nrow(df[df\$X == "a",]), p=1/3) am = binom.test(sum(df[df\$X == "a",]\$Y == "maybe"), nrow(df[df\$X == "a",]), p=1/3) p.adjust(c(ay\$p.value, an\$p.value, am\$p.value), method="holm") by = binom.test(sum(df[df\$X == "b",]\$Y == "yes"), nrow(df[df\$X == "b",]), p=1/3) bn = binom.test(sum(df[df\$X == "b",]\$Y == "no"), nrow(df[df\$X == "b",]), p=1/3) bm = binom.test(sum(df[df\$X == "b",]\$Y == "maybe"), nrow(df[df\$X == "b",]), p=1/3) p.adjust(c(by\$p.value, bn\$p.value, bm\$p.value), method="holm")</pre>

Proportion & Association

Post hoc tests – Two independent samples

Samples	Response Categories	Omnibus Test	Contrast Test	Report
2	≥2	Fisher's exact test	Pairwise Fisher's exact tests	"Three <i>post hoc</i> Fisher's exact tests of association, corrected with Holm's sequential Bonferroni procedure (Holm 1979), indicated that there was a significant association for {a, b}: {yes, maybe} ($p = .008$), but neither for {a, b}: {no, maybe} nor {a, b}: {yes, no}."
2	≥2	G-test	Pairwise G-tests	"Fifteen <i>post hoc</i> pairwise comparisons using G-tests, corrected with Holm's sequential Bonferroni procedure (Holm 1979), indicated that {b, yes} vs. {b, maybe} was statistically significant ($p < .001$), as was {b, no} vs. {b, maybe} ($p = .026$). The other pairwise comparisons were not statistically significant."
2	≥2	Two-sample Pearson chi-squared test	Pairwise chi-squared tests	"Fifteen <i>post hoc</i> pairwise comparisons using Pearson chi-squared tests, corrected with Holm's sequential Bonferroni procedure (Holm 1979), indicated that {b, yes} vs. {b, maybe} was statistically significant ($p = .001$), as was {b, no} vs. {b, maybe} ($p = .035$). The other pairwise comparisons were not statistically significant."
2	≥2	Fisher's exact test, G-test, two-sample Pearson chi-squared test	Individual binomial tests against chance	"Three <i>post hoc</i> binomial tests, corrected with Holm's sequential Bonferroni procedure (Holm 1979), indicated that, for X='a', the proportions of 'yes', 'no', and 'maybe' were not significantly different from chance (<i>i.e.</i> , 10 each). Three similar <i>post hoc</i> binomial tests for X='b' indicated that the proportion of 'yes' responses was significantly different from chance ($p < .0001$), as was the proportion of 'maybe' responses ($p = .003$), but not the proportion of 'no' responses ($p = .174$)."

Proportion & Association

Post hoc tests – Two dependent samples

Samples	Response Categories	Omnibus Test	Contrast Test	R Code
2	≥ 2	Symmetry test	Sign test	<pre> # df is a long-format data table w/columns for participant (PIId), a within-Ss. factor (X), # and N-category outcome (Y) library(coin) # for sign_test # Post hoc sign tests for the "vanilla" flavor, with obvious replication to other flavors. # Add new columns to the data table: df\$chose.vanilla.in.fall = ifelse(df\$Y == "vanilla" & df\$X == "fall", 1, 0) df\$chose.vanilla.in.winter = ifelse(df\$Y == "vanilla" & df\$X == "winter", 1, 0) df\$chose.vanilla.in.spring = ifelse(df\$Y == "vanilla" & df\$X == "spring", 1, 0) df\$chose.vanilla.in.summer = ifelse(df\$Y == "vanilla" & df\$X == "summer", 1, 0) fa.wi = pvalue(sign_test(chose.vanilla.in.fall ~ chose.vanilla.in.winter, data=df)) fa.sp = pvalue(sign_test(chose.vanilla.in.fall ~ chose.vanilla.in.spring, data=df)) fa.su = pvalue(sign_test(chose.vanilla.in.fall ~ chose.vanilla.in.summer, data=df)) wi.sp = pvalue(sign_test(chose.vanilla.in.winter ~ chose.vanilla.in.spring, data=df)) wi.su = pvalue(sign_test(chose.vanilla.in.winter ~ chose.vanilla.in.summer, data=df)) sp.su = pvalue(sign_test(chose.vanilla.in.spring ~ chose.vanilla.in.summer, data=df)) p.adjust(c(fa.wi, fa.sp, fa.su, wi.sp, wi.su, sp.su), method="holm") </pre>

Proportion & Association

Post hoc tests – Two dependent samples

Samples	Response Categories	Omnibus Test	Contrast Test	Report
2	≥ 2	Symmetry test	Sign test	“The proportions of vanilla ice cream were compared across the four seasons using six <i>post hoc</i> sign tests corrected with Holm’s sequential Bonferroni procedure (Holm 1979). The proportions of vanilla preferred in fall, winter, spring, and summer did not significantly differ.”

Bibliography (contingency tables, tests of proportion and association)

- Conover, W.J. (1999). The sign test. In *Practical Nonparametric Statistics* (3rd ed.). New York, NY: John Wiley & Sons, pp. 157-165. <https://www.wiley.com/en-us/Practical+Nonparametric+Statistics%2C+3rd+Edition-p-9780471160687>
- Fisher, R.A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society* 85 (1), pp. 87-94. <https://doi.org/10.2307/2340521>
- Mehta, C.R. and Patel, N.R. (1983). A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *Journal of the American Statistical Association* 78 (382), pp. 427-434. <https://doi.org/10.2307/2288652>
- 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* (302), pp. 157-175. <https://doi.org/10.1080/14786440009463897>
- Sokal, R.R. and Rohlf, F.J. (1969). *Biometry: The Principles and Practice of Statistics in Biological Research*. San Francisco, CA: W. H. Freeman & Co. <https://books.google.com/books?id=C-OTQgAACAAJ>
- Strasser, H. and Weber, C. (1999). On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics* 8 (2), pp. 220-250. <https://doi.org/10.57938/ff565ba0-aa64-4fe0-a158-86fd331bee78>

Assumptions

ANOVA Assumptions

Normality

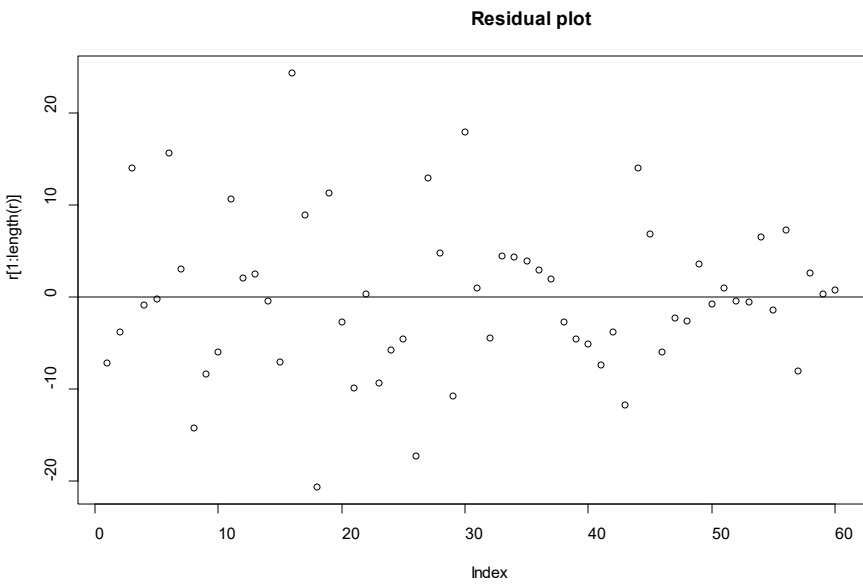
Assumption	Test Name	Context of Use	R Code
Normality	Shapiro-Wilk test (on the response in each condition)	<i>t</i> -test, ANOVA	<pre># df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y shapiro.test(df[df\$X1 == "a" & df\$X2 == "a",]\$Y) # condition a,a shapiro.test(df[df\$X1 == "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</pre>
Normality	Shapiro-Wilk test (on residuals)	<i>t</i> -test, ANOVA	<pre># df has two factors (X1,X2) each w/two levels (a,b) and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_normality m = aov_ez(dv="Y", between=c("X1","X2"), id="PIId", type=3, data=df) # make model r = residuals(m\$lm) # extract residuals mean(r); sum(r) # both should be ~0 plot(r[1:length(r)], main="Residual plot"); abline(h=0) # should look random hist(r, main="Histogram of residuals") # should look normal qqnorm(r); qqline(r) # Q-Q plot shapiro.test(r) # Shapiro-Wilk test print(check_normality(m)) # same</pre>
Normality	Shapiro-Wilk test (on residuals)	Linear mixed model (LMM)	<pre># df has two factors (X1,X2) each w/two levels (a,b) and continuous response (Y) library(lme4) # for lmer library(lmerTest) library(performance) # for check_normality m = lmer(Y ~ X1*X2 + (1 PIId), data=df) # make linear mixed model r = residuals(m) # extract model residuals mean(r); sum(r) # both should be ~0 plot(r[1:length(r)], main="Residual plot"); abline(h=0) # should look random hist(r, main="Histogram of residuals") # should look normal qqnorm(r); qqline(r) # Q-Q plot shapiro.test(r) # Shapiro-Wilk test print(check_normality(m)) # same</pre>

ANOVA Assumptions

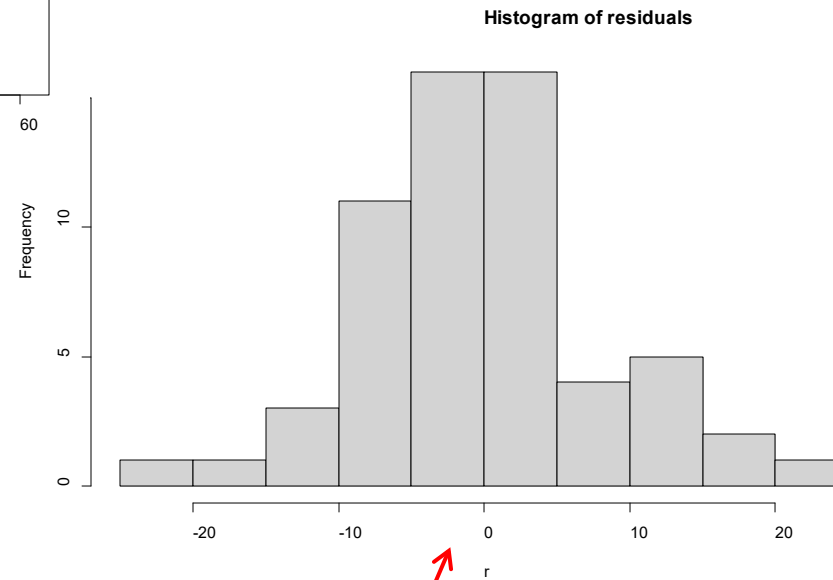
Normality

Assumption	Test Name	Context of Use	Report
Normality	Shapiro-Wilk test (on the response in each condition)	<i>t</i> -test, ANOVA	“To test the normality assumption, 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, indicating compliance with the normality assumption.”
Normality	Shapiro-Wilk test (on residuals)	<i>t</i> -test, ANOVA	“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 ($W = .984$, $p = .627$), indicating compliance with the normality assumption. A plot of residuals, histogram of residuals, and Q-Q plot all visually confirm the same (Figure 3).”
Normality	Shapiro-Wilk test (on residuals)	Linear mixed model (LMM)	“To test the normality assumption, a Shapiro-Wilk test was run on the residuals of a within-subjects linear mixed model (LMM). The test was statistically non-significant ($W = .984$, $p = .627$), indicating compliance with the normality assumption. A plot of residuals, histogram of residuals, and Q-Q plot all visually confirm the same (Figure 3).”

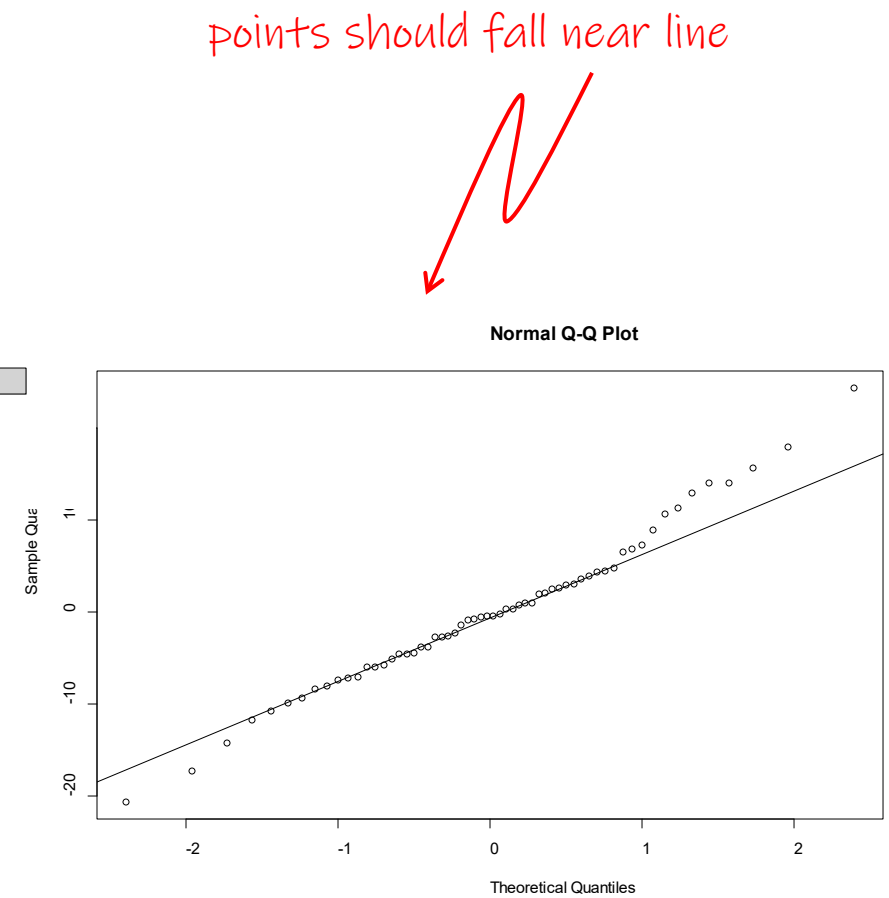
Figure 3



should look random



should look normal



ANOVA Assumptions

Homoscedasticity & sphericity

Assumption	Test Name	Context of Use	R Code
Homogeneity of variance	Levene's test	Any ANOVA model with at least one between-subjects factor	<pre># df has one between-Ss factor (X1), one within-Ss factor (X2), and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_homogeneity library(car) # for Anova m = aov_ez(dv="Y", between="X1", within="X2", id="PIId", type=3, data=df) print(check_homogeneity(m)) # Levene's test # if a violation occurs (p<.05), use a Welch t-test for one factor of two levels... t.test(Y ~ X1, data=df, var.equal=FALSE) # ...or a White-adjusted ANOVA for >1 factor or >2 levels Anova(m\$lm, type=3, white.adjust=TRUE) # shows between-Ss. factors only</pre>
Sphericity	Mauchly's test	Any ANOVA model with at least one within-subjects factor	<pre># df has one between-Ss factor (X1), one within-Ss factor (X2), and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_sphericity m = aov_ez(dv="Y", between="X1", within="X2", id="PIId", type=3, data=df) print(check_sphericity(m)) anova(m, correction="none") # use if p≥.05, no violation of sphericity anova(m, correction="GG") # use if p<.05, sphericity violation</pre>

ANOVA Assumptions

Homoscedasticity & sphericity

Assumption	Test Name	Context of Use	Report
Homogeneity of variance	Levene's test	Any ANOVA model with at least one between-subjects factor	"To test the homogeneity of variance assumption, Levene's test was run on a mixed factorial ANOVA with a between-subjects factor X1 and a within-subjects factor X2. The test was statistically non-significant ($p = .753$), indicating no violation."
Sphericity	Mauchly's test	Any ANOVA model with at least one within-subjects factor	"To test the sphericity assumption, Mauchly's test of sphericity was run on a mixed factorial ANOVA with a between-subjects factor X1 and a within-subjects factor X2. The test was statistically non-significant ($p = .999$), indicating no sphericity violation."

Bibliography (ANOVA assumptions of normality, homogeneity of variance, sphericity)

- Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London, Series A* 160 (901), pp. 268-282.
<https://www.jstor.org/stable/96803>
- Brown, M.B. and Forsythe, A.B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association* 69 (346), pp. 364-367. <https://doi.org/10.2307/2285659>
- Fligner, M.A. and Killeen, T.J. (1976). Distribution-free two-sample tests for scale. *Journal of the American Statistical Association* 71 (353), pp. 210-213. <https://www.jstor.org/stable/2285771>
- Greenhouse, S.W. and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24 (2), pp. 95-112.
<https://doi.org/10.1007/BF02289823>
- Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, S.G. Ghurye, H. Hoeffding, W.G. Madow and H.B. Mann (eds.). Palo Alto, CA: Stanford University Press, pp. 278-292.
- Mauchly, J.W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics* 11 (2), pp. 204-209.
<https://www.jstor.org/stable/2235878>
- Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52 (3/4), pp. 591-611.
<https://doi.org/10.2307/2333709>
- Welch, B.L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika* 38 (3/4), pp. 330-336.
<https://doi.org/10.2307/2332579>
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4), pp. 817-838. <https://doi.org/10.2307/1912934>

Distributions

Distributions

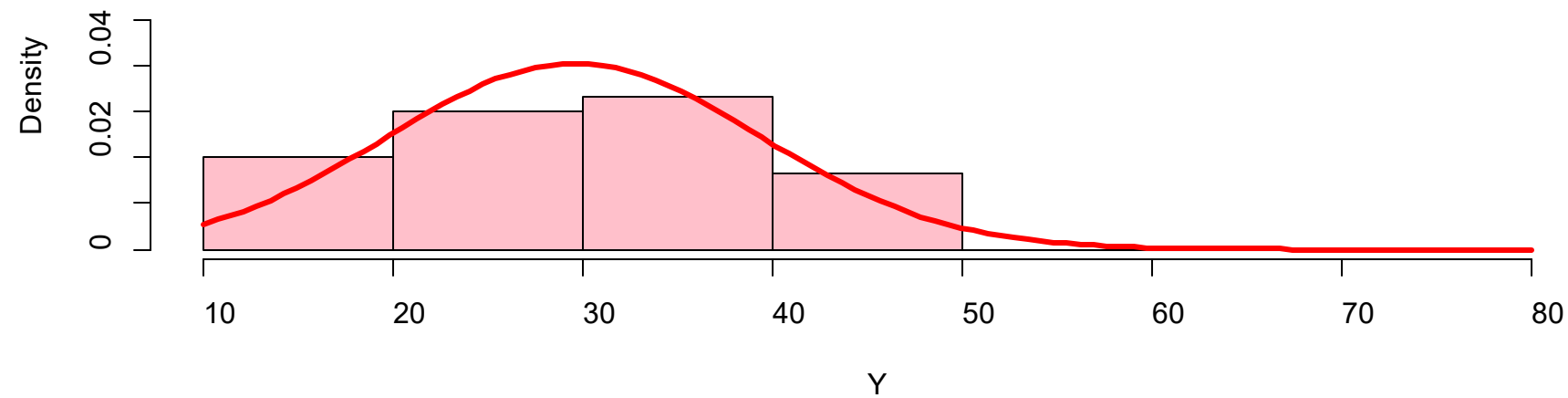
Distribution	Parameterization	R Distribution Fns	R Code
Normal	mean (μ): mean standard deviation (σ): sd	<code>_norm:</code> <code>dnorm</code> <code>pnorm</code> <code>qnorm</code> <code>rnorm</code>	<pre># df has one factor (X) w/two levels (a,b) and continuous response Y library(EnvStats) # for gofTest fa = gofTest(df[df\$X == "a",]\$Y, distribution="norm") # create fit for X=a print(fa) # Shapiro-Wilk GOF fb = gofTest(df[df\$X == "b",]\$Y, distribution="norm") # create fit for X=b print(fb) # Shapiro-Wilk GOF</pre>
Lognormal	mean (μ): meanlog standard deviation (σ): sdlog	<code>_lnorm:</code> <code>dlnorm</code> <code>plnorm</code> <code>qlnorm</code> <code>rlnorm</code>	<pre># df has one factor (X) w/two levels (a,b) and positively skewed response Y library(EnvStats) # for gofTest fa = gofTest(df[df\$X == "a",]\$Y, distribution="lnorm") # create fit for X=a print(fa) # Shapiro-Wilk GOF fb = gofTest(df[df\$X == "b",]\$Y, distribution="lnorm") # create fit for X=b print(fb) # Shapiro-Wilk GOF</pre>
Poisson	lambda (λ): lambda	<code>_pois:</code> <code>dpois</code> <code>ppois</code> <code>qpois</code> <code>rpois</code>	<pre># df has one factor (X) w/two levels (a,b) and integer count response Y library(fitdistrplus) # for fitdist, gofstat fa = fitdist(df[df\$X == "a",]\$Y, distr="pois") # create fit for X=a gofstat(fa) # chi-square GOF fb = fitdist(df[df\$X == "b",]\$Y, distr="pois") # create fit for X=b gofstat(fb) # chi-square GOF # if var/mean > 1.15, we have overdispersion; if so, use quasipoisson or an nbinom GL(M)M abs(var(df[df\$X == "a",]\$Y) / mean(df[df\$X == "a",]\$Y)) > 1.15 abs(var(df[df\$X == "b",]\$Y) / mean(df[df\$X == "b",]\$Y)) > 1.15</pre>
Negative Binomial	theta (θ): size mu (μ): mu	<code>_nbinom:</code> <code>dnbinom</code> <code>pnbinom</code> <code>qnbinom</code> <code>rnbinom</code>	<pre># df has one factor (X) w/two levels (a,b) and integer count response Y library(fitdistrplus) # for fitdist, gofstat fa = fitdist(df[df\$X == "a",]\$Y, distr="nbinom") # create fit for X=a gofstat(fa) # chi-square GOF fb = fitdist(df[df\$X == "b",]\$Y, distr="nbinom") # create fit for X=b gofstat(fb) # chi-square GOF</pre>

Distributions

Distribution	Parameterization	R Distribution Fns	Report
Normal	mean (μ): <code>mean</code> standard deviation (σ): <code>sd</code>	<code>_norm:</code> <code>dnorm</code> <code>pnorm</code> <code>qnorm</code> <code>rnorm</code>	“Figure 4 shows the distributions of response Y for both levels of factor X. To test whether these distributions were normally distributed, a Shapiro-Wilk goodness-of-fit test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($W = .979, p = .796$), as was the test for level ‘b’ ($W = .987, p = .961$), indicating no detectable departure from a normal distribution for either level of X.”
Lognormal	mean (μ): <code>meanlog</code> standard deviation (σ): <code>sdlog</code>	<code>_lnorm:</code> <code>dlnorm</code> <code>plnorm</code> <code>qlnorm</code> <code>rlnorm</code>	“Figure 5 shows the distributions of response Y for both levels of factor X. To test whether these distributions were lognormally distributed, a Shapiro-Wilk goodness-of-fit test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($W = .979, p = .796$), as was the test for level ‘b’ ($W = .987, p = .961$), indicating no detectable departure from a lognormal distribution for either level of X.”
Poisson	lambda (λ): <code>lambda</code>	<code>_pois:</code> <code>dpois</code> <code>ppois</code> <code>qpois</code> <code>rpois</code>	“Figure 6 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 ($\chi^2(4, N=30) = 0.56, p = .967$), as was the test for level ‘b’ ($\chi^2(4, N=30) = 3.89, p = .422$), indicating no detectable departure from a Poisson distribution for either level of X.”
Negative Binomial	theta (θ): <code>size</code> mu (μ): <code>mu</code>	<code>_nbinom:</code> <code>dnbinom</code> <code>pnbinom</code> <code>qnbinom</code> <code>rnbinom</code>	“Figure 7 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 ($\chi^2(4, N=30) = 0.88, p = .927$), as was the test for level ‘b’ ($\chi^2(2, N=30) = 0.81, p = .666$), indicating no detectable departure from a negative binomial distribution for either level of X.”

Figure 4

Histogram of Y for X=a



Histogram of Y for X=b

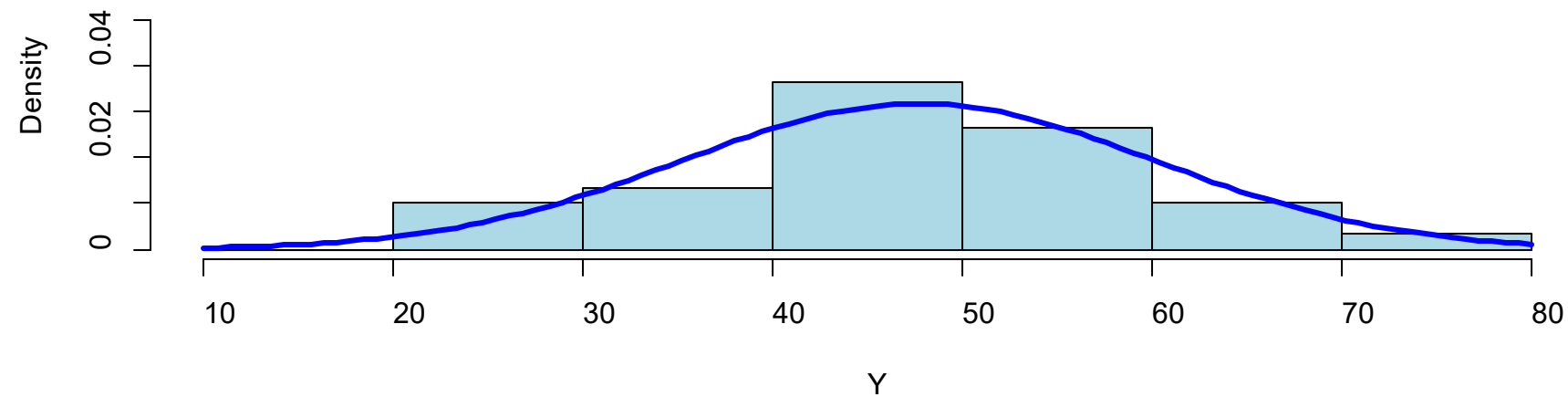
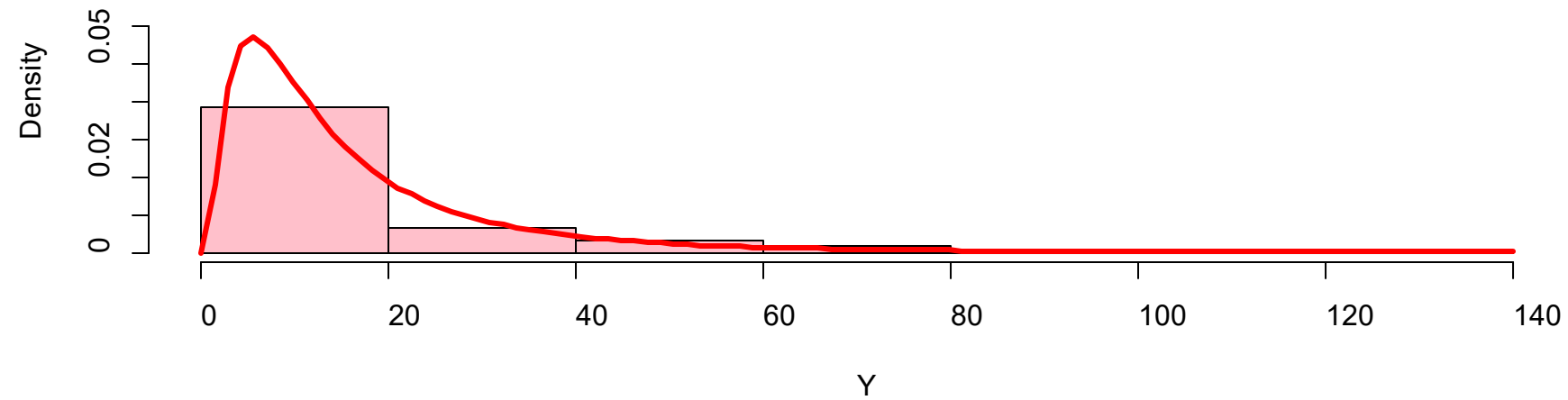


Figure 5

Histogram of Y for X=a



Histogram of Y for X=b

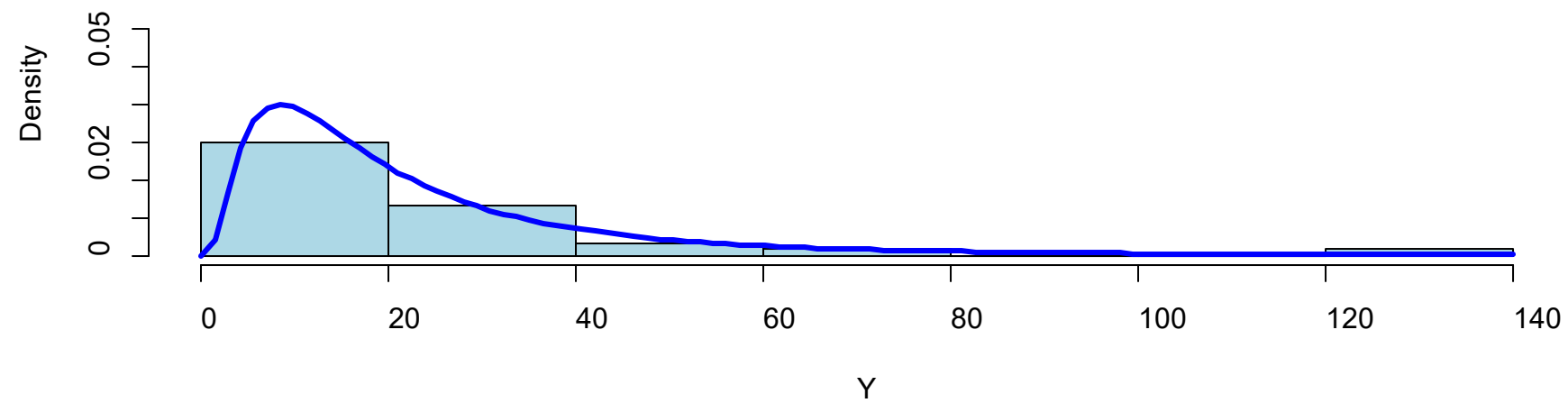


Figure 6

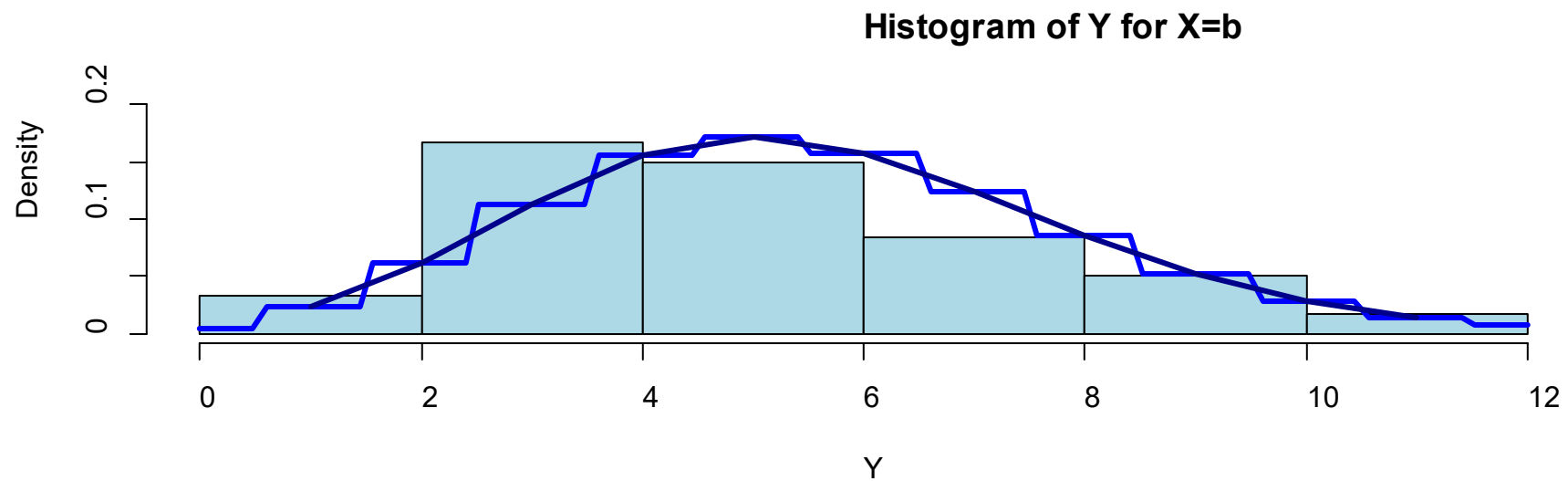
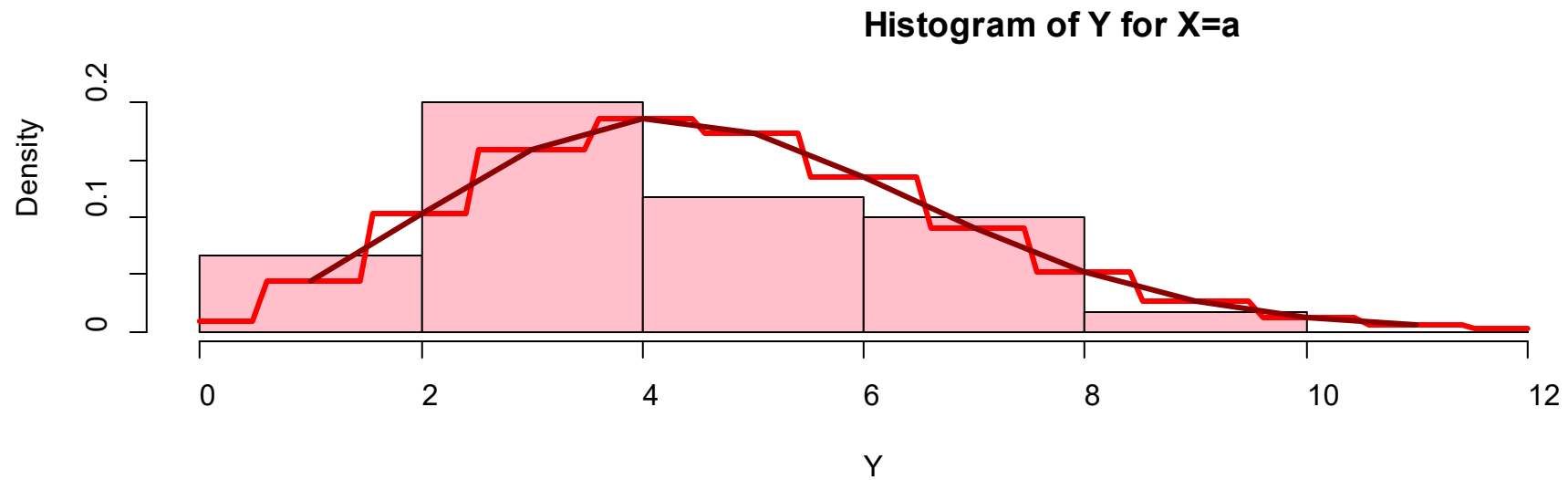
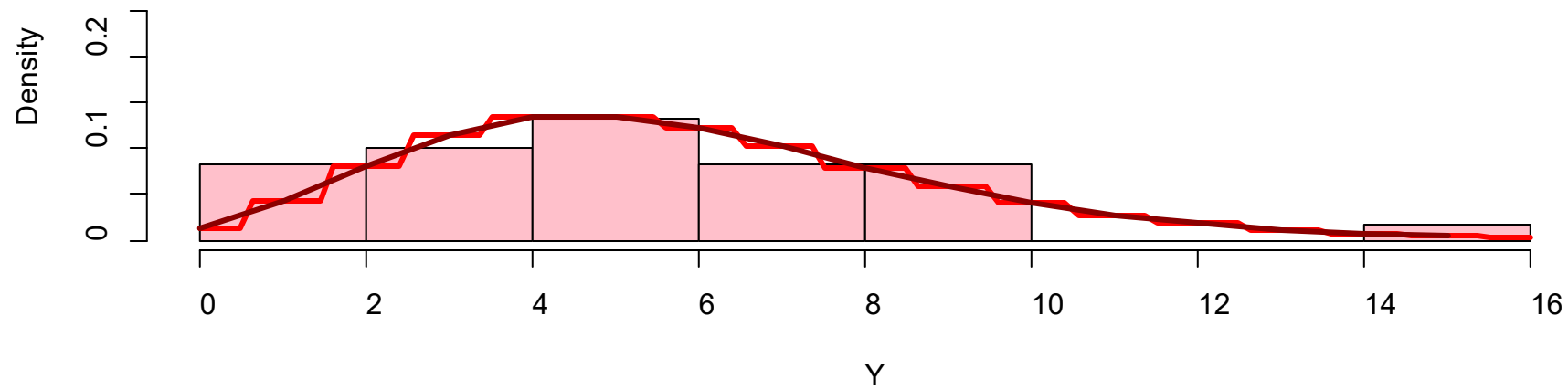
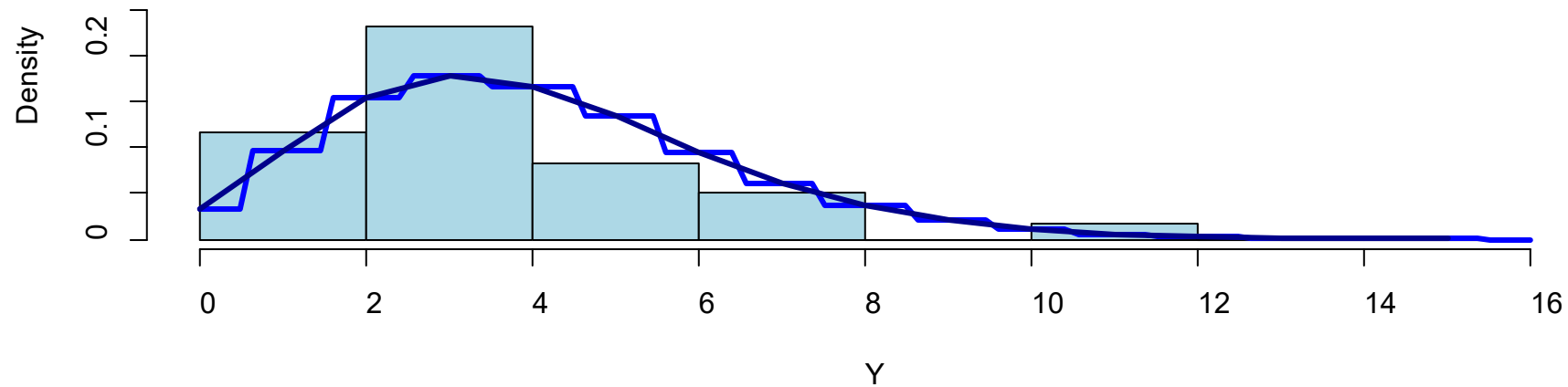


Figure 7

Histogram of Y for X=a



Histogram of Y for X=b



Distribution Tests

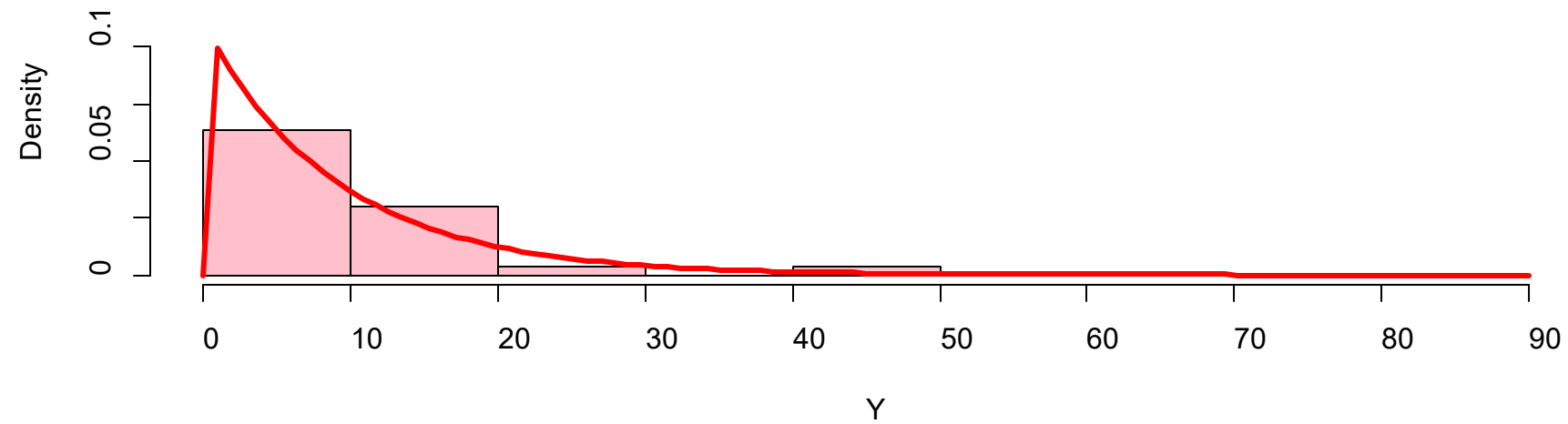
Distribution	Parameterization	R Distribution Fns	R Code
Exponential	rate (λ): rate	_exp: dexp pexp qexp rexp	<pre># df has one factor (X) w/two levels (a,b) and exponential response Y library(EnvStats) # for gofTest fa = gofTest(df[df\$X == "a",]\$Y, distribution="exp") # create fit for X=a print(fa) # Shapiro-Wilk GOF fb = gofTest(df[df\$X == "b",]\$Y, distribution="exp") # create fit for X=b print(fb) # Shapiro-Wilk GOF</pre>
Gamma	shape (α): shape scale (β): scale	_gamma: dgamma pgamma qgamma rgamma	<pre># df has one factor (X) w/two levels (a,b) and positively skewed response Y library(EnvStats) # for gofTest fa = gofTest(df[df\$X == "a",]\$Y, distribution="gamma") # create fit for X=a print(fa) # Shapiro-Wilk GOF fb = gofTest(df[df\$X == "b",]\$Y, distribution="gamma") # create fit for X=b print(fb) # Shapiro-Wilk GOF</pre>

Distribution Tests

Distribution	Parameterization	R Distribution Fns	Report
Exponential	rate (λ): <code>rate</code>	<code>_exp:</code> <code>dexp</code> <code>pexp</code> <code>qexp</code> <code>rexp</code>	“Figure 8 shows the distributions of response Y for both levels of factor X. To test whether these distributions were exponentially distributed, a Shapiro-Wilk goodness-of-fit test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($W = .966, p = .428$), as was the test for level ‘b’ ($W = .952, p = .190$), indicating no detectable departure from an exponential distribution for either level of X.”
Gamma	shape (α): <code>shape</code> scale (β): <code>scale</code>	<code>_gamma:</code> <code>dgamma</code> <code>pgamma</code> <code>qgamma</code> <code>rgamma</code>	“Figure 9 shows the distributions of response Y for both levels of factor X. To test whether these distributions were gamma distributed, a Shapiro-Wilk goodness-of-fit test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($W = .968, p = .483$), as was the test for level ‘b’ ($W = .967, p = .459$), indicating no detectable departure from a gamma distribution for either level of X.”

Figure 8

Histogram of Y for X=a



Histogram of Y for X=b

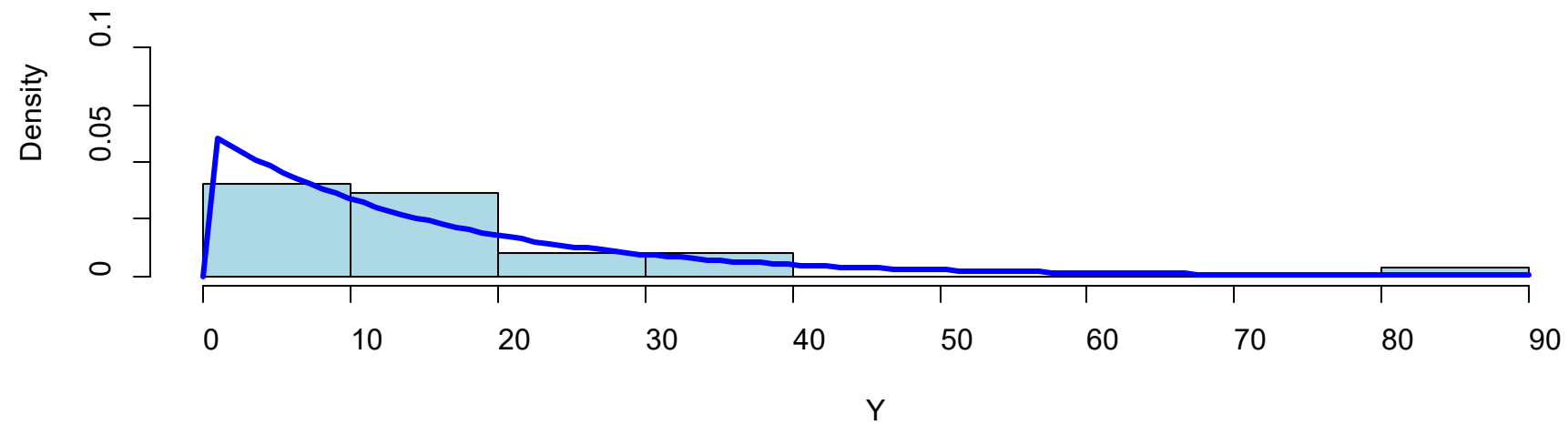
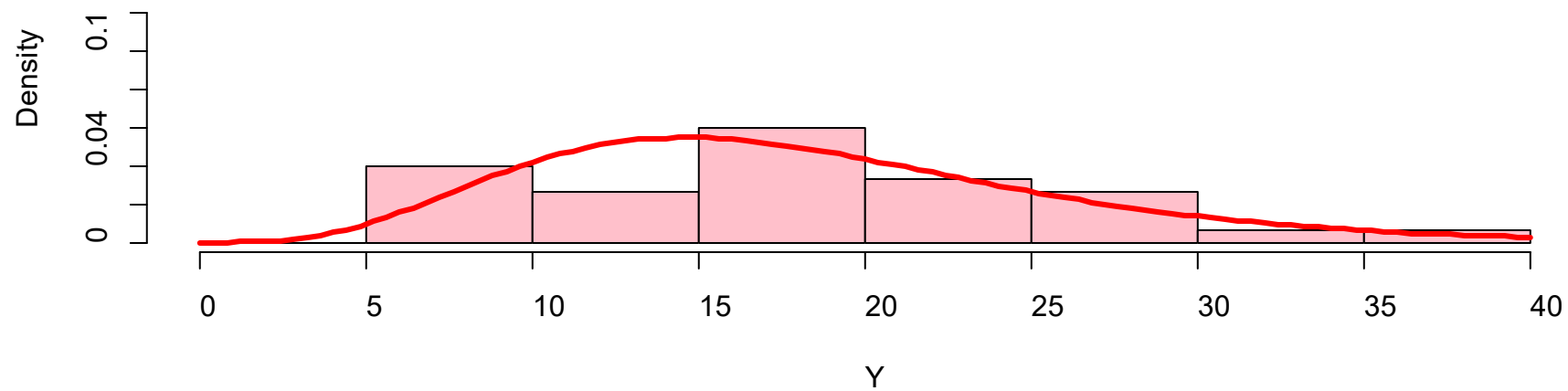
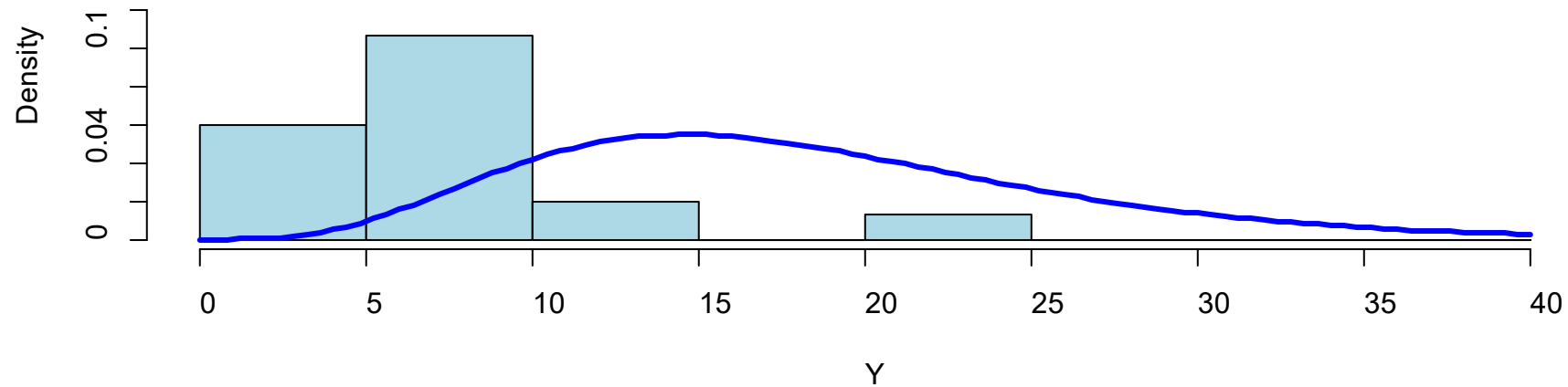


Figure 9

Histogram of Y for X=a



Histogram of Y for X=b



Bibliography (data distributions)

- Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution*. Cambridge, UK: Cambridge University Press.
<http://www.jstor.org/stable/2227716>
- Hilbe, J.M. (2011). *Negative Binomial Regression*. Cambridge, UK: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511973420>
- Hilbe, J.M. (2014). *Modeling Count Data*. New York, NY: Cambridge University Press.
<https://books.google.com/books?id=aZLfAwAAQBAJ>
- Lawless, J.F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 15 (3), pp. 209-225.
<https://doi.org/10.2307/3314912>
- Ng, V.K.Y. and Cribbie, R.A. (2017). Using the gamma generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. *Current Psychology* 36 (2), pp. 225-235. <https://doi.org/10.1007/s12144-015-9404-0>
- 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* (302), pp. 157-175. <https://doi.org/10.1080/14786440009463897>
- Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52 (3/4), pp. 591-611.
<https://doi.org/10.2307/2333709>
- Vermunt, J.K. (1997). *Log-linear Models for Event Histories*. Thousand Oaks, CA: Sage Publications.
https://books.google.com/books?id=apI_AQAAIAAJ
- von Bortkiewicz, L. (1898). *Das Gesetz der Kleinen Zahlen (The Law of Small Numbers)*. Leipzig, Germany: Druck und Verlag von B.G. Teubner. https://books.google.com/books?id=o_k3AAAAMAAJ

Parametric Tests

Parametric Tests

One between-Ss. factor

Factors	Levels	Between or Within Subjects	Test Name	R Code
1	2	Between	Independent-samples <i>t</i> -test	<pre># df has one between-Ss. factor (X) w/levels (a,b) and continuous response (Y) library(car) # for leveneTest df\$PId = factor(df\$PId) # participant is nominal (unused) df\$X = factor(df\$X) # X is a 2-level factor contrasts(df\$X) <- "contr.sum" leveneTest(Y ~ X, data=df, center=mean) # check homogeneity of variance t.test(Y ~ X, data=df, var.equal=TRUE) # if p≥.05, no violation of homogeneity t.test(Y ~ X, data=df, var.equal=FALSE) # if p<.05, violation of homogeneity</pre>
1	≥2	Between	One-way ANOVA	<pre># df has one between-Ss. factor (X) w/levels (a,b,c) and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_homogeneity df\$PId = factor(df\$PId) # participant is nominal df\$X = factor(df\$X) # X is a 3-level factor contrasts(df\$X) <- "contr.sum" m = aov_ez(dv="Y", between="X", id="PId", type=3, data=df) # fit model print(check_homogeneity(m)) # Levene's test leveneTest(Y ~ X, data=df, center=mean) # same anova(m) # if p≥.05, no violation of homogeneity Anova(m\$lm, type=3, white.adjust=TRUE) # if p<.05, violation of homogeneity</pre>

Parametric Tests

One between-Ss. factor

Factors	Levels	Between or Within Subjects	Test Name	Report
1	2	Between	Independent-samples <i>t</i> -test	“The mean of ‘a’ was 29.29 (<i>SD</i> = 14.72) and of ‘b’ was 47.68 (<i>SD</i> = 12.53). This difference was statistically significant according to an independent-samples <i>t</i> -test ($t(58) = -5.21, p < .0001$).”
1	≥2	Between	One-way ANOVA	“The mean of ‘a’ was 32.12 (<i>SD</i> = 14.59), of ‘b’ was 44.23 (<i>SD</i> = 12.45), and of ‘c’ was 41.60 (<i>SD</i> = 14.36). These differences were statistically significant according to a one-way ANOVA ($F(2, 57) = 4.24, p = .019$).”

Note: “*SD*” stands for “standard deviation,” i.e., the spread of values around the mean.

Parametric Tests

One within-Ss. factor

Factors	Levels	Between or Within Subjects	Test Name	R Code
1	2	Within	Paired-samples t-test	<pre># df has one within-Ss. factor (X) w/levels (a,b) and continuous response (Y) library(reshape2) # for dcast df\$PId = factor(df\$PId) # participant is nominal df\$X = factor(df\$X) # X is a 2-level factor contrasts(df\$X) <- "contr.sum" df2 <- dcast(df, PId ~ X, value.var="Y") # make wide-format table t.test(df2\$a, df2\$b, paired=TRUE) # neither homogeneity nor sphericity applies to a paired t-test</pre>
1	≥2	Within	One-way repeated measures ANOVA	<pre># df has one within-Ss. factor (X) w/levels (a,b,c) and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_sphericity df\$PId = factor(df\$PId) # participant is nominal df\$X = factor(df\$X) # X is a 3-level factor contrasts(df\$X) <- "contr.sum" m = aov_ez(dv="Y", within="X", id="PId", type=3, data=df) # fit model print(check_sphericity(m)) # Mauchly's test anova(m, correction="none") # if p≥.05, no violation of sphericity anova(m, correction="GG") # if p<.05, violation of sphericity</pre>

Parametric Tests

One within-Ss. factor

Factors	Levels	Between or Within Subjects	Test Name	Report
1	2	Within	Paired-samples <i>t</i> -test	“The mean of ‘a’ was 29.29 (<i>SD</i> = 14.72) and of ‘b’ was 47.68 (<i>SD</i> = 12.53). This difference was statistically significant according to a paired-samples <i>t</i> -test ($t(29) = -4.85, p < .0001$).”
1	≥2	Within	One-way repeated measures ANOVA	“The mean of ‘a’ was 32.12 (<i>SD</i> = 14.59), of ‘b’ was 44.23 (<i>SD</i> = 12.45), and of ‘c’ was 41.60 (<i>SD</i> = 14.36). These differences were statistically significant ($F(2, 38) = 4.06, p = .025$).”

Note: “*SD*” stands for “standard deviation,” i.e., the spread of values around the mean.

Parametric Tests

Multiple between-Ss. factors

Factors	Levels	Between or Within Subjects	Test Name	R Code
≥2	≥2	Between	Factorial ANOVA	<pre># df has two between-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_homogeneity library(car) # for Anova df\$PId = factor(df\$PId) # participant is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = aov_ez(dv="Y", between=c("X1","X2"), id="PId", type=3, data=df) # fit model print(check_homogeneity(m)) # Levene's test leveneTest(Y ~ X1*X2, data=df, center=mean) # same anova(m) # if p≥.05, no violation of homogeneity Anova(m\$lm, type=3, white.adjust=TRUE) # if p<.05, violation of homogeneity</pre>
			Linear model (LM)	<pre># df has two between-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(car) # for Anova df\$PId = factor(df\$PId) # participant 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 contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = lm(Y ~ X1*X2, data=df) # fit model m = aov(Y ~ X1*X2, data=df) # equivalent print(check_homogeneity(m)) # Levene's test leveneTest(Y ~ X1*X2, data=df, center=mean) # same anova(m) # if p≥.05, no violation of homogeneity Anova(m, type=3, white.adjust=TRUE) # if p<.05, violation of homogeneity</pre>

Parametric Tests

Multiple between-Ss. factors

Factors	Levels	Between or Within Subjects	Test Name	Report
≥2	≥2	Between	Factorial ANOVA	“Figure 10 shows an interaction plot with ± 1 standard deviation error bars for X1 and X2. Levene’s test indicated a violation of the assumption of homogeneity of variance ($F(3, 56) = 6.86, p = .001$). Therefore, a White-adjusted factorial ANOVA was used. It indicated a significant effect on Y of X1 ($F(1, 56) = 20.51, p < .0001$), no significant effect of X2 ($F(1, 56) = 0.09, p = .766$), and a significant X1×X2 interaction ($F(1, 56) = 7.15, p = .010$).”
			Linear Model (LM)	“Figure 10 shows an interaction plot with ± 1 standard deviation error bars for X1 and X2. Levene’s test indicated a violation of the assumption of homogeneity of variance ($F(3, 56) = 6.86, p = .001$). Therefore, a White-adjusted factorial ANOVA was used. It indicated a significant effect on Y of X1 ($F(1, 56) = 20.51, p < .0001$), no significant effect of X2 ($F(1, 56) = 0.09, p = .766$), and a significant X1×X2 interaction ($F(1, 56) = 7.15, p = .010$).”

Parametric Tests

Multiple within-Ss. factors

Factors	Levels	Between or Within Subjects	Test Name	R Code
≥2	≥2	Within	Factorial repeated measures ANOVA	<pre># df has two within-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_sphericity df\$PId = factor(df\$PId) # participant is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = aov_ez(dv="Y", within=c("X1","X2"), id="PId", type=3, data=df) # fit model print(check_sphericity(m)) # Mauchly's test anova(m, correction="none") # use if p≥.05, no violation of sphericity anova(m, correction="GG") # use if p<.05, sphericity violation</pre>
≥2	≥2	Within	Linear mixed model (LMM)*	<pre># df has two within-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(lme4) # for lmer library(lmerTest) library(car) # for Anova df\$PId = factor(df\$PId) # participant is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = lmer(Y ~ X1*X2 + (1 PId), data=df) # sphericity is N/A for LMMs Anova(m, type=3, test.statistic="F")</pre>

*The LMM sample code uses a random intercept for *participant (PId)*. There are also random slope models, which are used when the response changes at different rates for each *PId* 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>). A free webinar on the topic of random intercept and random slope models is available here (<https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/>).

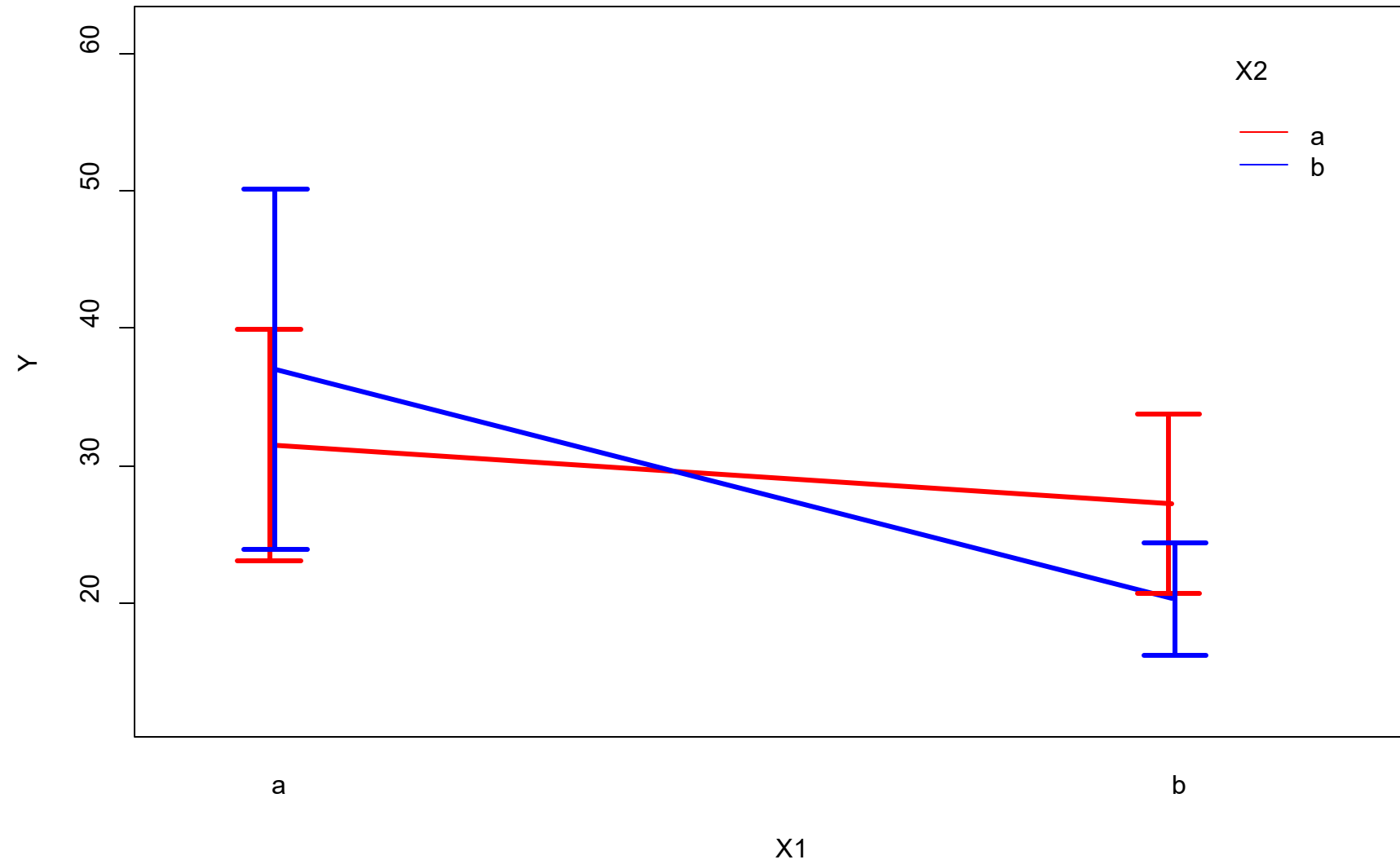
Parametric Tests

Multiple within-Ss. factors

Factors	Levels	Between or Within Subjects	Test Name	Report
≥2	≥2	Within	Factorial repeated measures ANOVA	“Figure 10 shows an interaction plot with ± 1 standard deviation error bars for X1 and X2. A factorial repeated measures ANOVA indicated a significant effect on Y of X1 ($F(1, 14) = 29.58, p < .0001$), no significant effect of X2 ($F(1, 14) = 0.08, p = .785$), and a significant X1×X2 interaction ($F(1, 14) = 5.11, p = .040$).”
≥2	≥2	Within	Linear Mixed Model (LMM)	“Figure 10 shows an interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance based on a linear mixed model (LMM) indicated a significant effect on Y of X1 ($F(1, 42) = 21.98, p < .0001$), no significant effect of X2 ($F(1, 42) = 0.10, p = .758$), and a significant X1×X2 interaction ($F(1, 42) = 7.66, p = .008$).”

Figure 10

Y by X1, X2



Parametric Tests

Post hoc pairwise comparisons

Parametric Tests

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Test	Test Name	B/W	R Code
1	>2	One-way ANOVA	Independent samples <i>t</i> -test	Btwn	<pre># df has one between-Ss. factor (X) w/levels (a,b,c) and continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from afex::aov_ez</pre>
1	>2	One-way repeated measures ANOVA	Paired samples <i>t</i> -test	Within	<pre># df has one within-Ss. factor (X) w/levels (a,b,c) and continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from afex::aov_ez</pre>

Parametric Tests

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Test	Test Name	B/W	Report
1	≥ 2	One-way ANOVA	Independent samples <i>t</i> -test	Btwn	“Three <i>post hoc</i> pairwise comparisons using independent-samples <i>t</i> -tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ was significantly different ($t(57) = -2.77, p = .023$), but ‘a’ vs. ‘c’ and ‘b’ vs. ‘c’ were not.”
1	≥ 2	One-way repeated measures ANOVA	Paired samples <i>t</i> -test	Within	“Three <i>post hoc</i> pairwise comparisons using independent-samples <i>t</i> -tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ was significantly different ($t(19) = -2.70, p = .042$), but ‘a’ vs. ‘c’ and ‘b’ vs. ‘c’ were not.”

Parametric Tests

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Test	Test Name	B/W	R Code
≥2	≥2	Factorial ANOVA, Linear model (LM)	Independent samples <i>t</i> -test	Btwn	<pre># df has two between-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from lm, aov, or afex::aov_ez</pre>
≥2	≥2	Factorial repeated measures ANOVA	Paired samples <i>t</i> -test	Within	<pre># df has two within-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from afex::aov_ez</pre>
≥2	≥2	Linear mixed model (LMM)	Paired samples <i>t</i> -test	Within	<pre># df has two within-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from lme4::lmer</pre>

Parametric Tests

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Test	Test Name	B/W	Report
≥2	≥2	Factorial ANOVA, Linear model (LM)	Independent samples <i>t</i> -test	Btwn	“Six <i>post hoc</i> pairwise comparisons using independent-samples <i>t</i> -tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,b} ($t(56) = 3.53, p = .004$), {b,a} vs. {a,b} ($t(56) = -3.10, p = .012$), and {a,b} vs. {b,b} ($t(56) = 5.27, p < .0001$) were significantly different. The other three pairwise comparisons were not detectably different.”
≥2	≥2	Factorial repeated measures ANOVA	Paired samples <i>t</i> -test	Within	“Six <i>post hoc</i> pairwise comparisons using paired-samples <i>t</i> -tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,b} ($t(14) = 4.82, p = .002$), {b,a} vs. {b,b} ($t(14) = 3.27, p = .023$), {a,b} vs. {b,b} ($t(14) = 4.25, p = .004$) were significantly different. The other three pairwise comparisons were not detectably different.”
≥2	≥2	Linear mixed model (LMM)	Paired samples <i>t</i> -test	Within	“Six <i>post hoc</i> pairwise comparisons using paired-samples <i>t</i> -tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,b} ($t(42) = 3.53, p = .005$), {b,a} vs. {a,b} ($t(42) = -3.10, p = .014$), and {a,b} vs. {b,b} ($t(42) = 5.27, p < .0001$) were significantly different. The other three pairwise comparisons were not detectably different.”

Bibliography (parametric t -tests, F-tests, and effect sizes)

- Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement* 33 (1), pp. 107-112. <https://doi.org/10.1177/001316447303300111>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 20-27.
- Fisher, R.A. (1921). On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron* 1 (4), pp. 3-32.
- Fisher, R.A. (1925). *Statistical methods for research workers*. Edinburgh, Scotland: Oliver and Boyd.
https://en.wikipedia.org/wiki/Statistical_Methods_for_Research_Workers
- Greenhouse, S.W. and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24 (2), pp. 95-112.
<https://doi.org/10.1007/BF02289823>
- Kenward, M.G. and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53 (3), pp. 983-997. <https://doi.org/10.2307/2533558>
- Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, S.G. Ghurye, H. Hoeffding, W.G. Madow and H.B. Mann (eds.). Palo Alto, CA: Stanford University Press, pp. 278-292.
- Mauchly, J.W. (1940). Significance test for sphericity of a normal n -variate distribution. *The Annals of Mathematical Statistics* 11 (2), pp. 204-209. <https://www.jstor.org/stable/2235878>
- Olejnik, S. and Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods* 8 (4), pp. 434-447. <https://doi.org/10.1037/1082-989X.8.4.434>

Bibliography (parametric t -tests, F-tests, and effect sizes)

- Student. (1908). The probable error of a mean. *Biometrika* 6 (1), pp. 1-25. <https://doi.org/10.2307/2331554>
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54 (3), pp. 426-482. <https://doi.org/10.2307/1990256>
- Welch, B.L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika* 38 (3/4), pp. 330-336. <https://doi.org/10.2307/2332579>
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4), pp. 817-838. <https://doi.org/10.2307/1912934>

Variance-Covariance Structures

(for use with `nlme::lme` instead of `lme4::lmer`)

Covariance Structures

Optional when fitting linear mixed models (LMMs)

Abbreviation	Name	Description	R Code
ID	Scaled identity	All variances are equal, and all covariances are zero.	<pre># df has 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 df\$PId = factor(df\$PId) # participant is nominal df\$X = factor(df\$X) # X is a 3-level factor contrasts(df\$X) <- "contr.sum" m = lme(Y ~ X, random=~1 PId, data=df) # ID getVarCov(m, type="marginal") # get VCV matrix anova(m, type="marginal") # for F-test Anova(m, type=3, test.statistic="Chisq") # for chisq test emmeans(m, pairwise ~ X, adjust="holm", mode="containment") # post hoc tests</pre>
DIAG	Diagonal	All variances can differ; otherwise, like ID.	<pre># See R Code for ID. Only the blue model-building line changes to: m = lme(Y ~ X, random=~1 PId, data=df, weights=varIdent(form=~1 X)) # DIAG</pre>
CS	Compound symmetry	All variances are equal, and all covariances are equal.	<pre># See R Code for ID. Only the blue model-building line changes to: m = lme(Y ~ X, random=~1 PId, data=df, correlation=corCompSymm(form=~1 PId)) # CS</pre>
CSH	Heterogeneous compound symmetry	All variances can differ; otherwise, like CS.	<pre># See R Code for ID. Only the blue model-building line changes to: m = lme(Y ~ X, random=~1 PId, data=df, correlation=corCompSymm(form=~1 PId), weights=varIdent(form=~1 X)) # CSH</pre>

Note: The `lme4::lmer` function does not allow specifying common variance-covariance (VCV) structures for repeated factors or residuals. Therefore, we must use `nlme::lme` for this. For a list of common VCV structures, see <https://www.ibm.com/docs/en/spss-statistics/26.0.0?topic=mixed-covariance-structure-list-command>. For their matrix formulations, see <https://www.ibm.com/docs/en/spss-statistics/26.0.0?topic=statistics-covariance-structures>. For a treatment in R, see <https://rpubs.com/samuelkn/CovarianceStructuresInR>.

Note: The `correlation` parameter sets covariances (matrix off-diagonal) and the `weights` parameter sets variances (matrix on-diagonal). When `correlation=NULL` or is unspecified, the off-diagonal values are zero. When `weights=NULL` or is unspecified, the on-diagonal variances are equal. The R help pages called up with `?corClasses` and `?varClasses` explain these parameters.

Covariance Structures

Optional when fitting linear mixed models (LMMs)

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

Note: See ?corClasses and ?varClasses for additional variance-covariance structures. Or see <https://rdrr.io/cran/nlme/man/corClasses.html> and <https://rdrr.io/cran/nlme/man/varClasses.html>.

Bibliography (variance-covariance structures, linear mixed models, random effects)

- Frederick, B.N. (1999). Fixed-, random-, and mixed-effects ANOVA models: A user-friendly guide for increasing the generalizability of ANOVA results. In *Advances in Social Science Methodology*, B. Thompson (ed.). Stamford, CT: JAI Press, pp. 111-122. <http://eric.ed.gov/?id=ED426098>
- Krueger, C. and Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing* 6 (2), pp. 151-157. <https://doi.org/10.1177/1099800404267682>
- Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* 38 (4), pp. 963-974. <https://www.jstor.org/stable/2529876>
- Littell, R.C., Henry, P.R. and Ammerman, C.B. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science* 76 (4), pp. 1216-1231. <https://doi.org/10.2527/1998.7641216x>
- Littell, R.C., Pendergast, J. and Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* 19 (13), pp. 1793-1819. [https://doi.org/10.1002/1097-0258\(20000715\)19:13<1793::AID-SIM482>3.0.CO;2-Q](https://doi.org/10.1002/1097-0258(20000715)19:13<1793::AID-SIM482>3.0.CO;2-Q)
- Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer. <https://link.springer.com/book/10.1007/b98882>
- Schuster, C. and von Eye, A. (2001). The relationship of ANOVA models with random effects and repeated measurement designs. *Journal of Adolescent Research* 16 (2), pp. 205-220. <https://doi.org/10.1177/0743558401162006>
- Ware, J.H. (1985). Linear models for the analysis of serial measurements in longitudinal studies. *American Statistician* 39 (2), pp. 95-101. <https://doi.org/10.2307/2682803>
- West, B.T., Welch, K.B. and Galecki, A.T. (2015). *Linear Mixed Models*. Boca Raton, FL: CRC Press.

Nonparametric Tests

Nonparametric Tests

One factor with 2 levels

Factors	Levels	Between or Within Subjects	Test Name	R Code
1	2	Between	Median test	<pre># df has one between-Ss. factor (X) w/levels (a,b) and a (1,0) response library(coin) # for median_test df\$PId = factor(df\$PId) # participant is nominal (unused) df\$X = factor(df\$X) # X is a 2-level factor contrasts(df\$X) <- "contr.sum" median_test(Y ~ X, data=df)</pre>
1	2	Between	Mann-Whitney U test	<pre># df has one between-Ss. factor (X) w/levels (a,b) and continuous response (Y) library(coin) # for wilcox_test df\$PId = factor(df\$PId) # participant is nominal (unused) df\$X = factor(df\$X) # X is a 2-level factor contrasts(df\$X) <- "contr.sum" wilcox_test(Y ~ X, data=df, distribution="exact")</pre>
1	2	Within	Sign test	<pre># df has one within-Ss. factor (X) w/levels (a,b) and a (1,0) response library(coin) # for sign_test df\$PId = factor(df\$PId) # participant is nominal df\$X = factor(df\$X) # X is a 2-level factor contrasts(df\$X) <- "contr.sum" sign_test(Y ~ X PId, data=df)</pre>
1	2	Within	Wilcoxon signed-rank test	<pre># df has one within-Ss. factor (X) w/levels (a,b) and continuous response (Y) library(coin) # for wilcoxsign_test df\$PId = factor(df\$PId) # participant is nominal df\$X = factor(df\$X) # X is a 2-level factor contrasts(df\$X) <- "contr.sum" wilcoxsign_test(Y ~ X PId, data=df, distribution="exact")</pre>

Note: The Mann-Whitney U test is also known as the Wilcoxon-Mann-Whitney test and the Wilcoxon rank-sum test, neither to be confused with the Wilcoxon signed-rank test.

Nonparametric Tests

One factor with 2 levels

Factors	Levels	Between or Within Subjects	Test Name	Report
1	2	Between	Median test	“The sum of ‘a’ was 20 and of ‘b’ was 8. This difference was statistically significant according to a median test ($Z = 3.08, p = .002$).”
1	2	Between	Mann-Whitney U test	“The median of ‘a’ was 29.12 ($IQR = 13.92$) and of ‘b’ was 45.72 ($IQR = 15.91$). This difference was statistically significant according to a Mann-Whitney U test ($Z = -4.70, p < .0001$).”
1	2	Within	Sign test	“The sum of ‘a’ was 20 and of ‘b’ was 8. This difference was statistically significant according to a sign test ($Z = 2.83, p = .005$).”
1	2	Within	Wilcoxon signed-rank test	“The median of ‘a’ was 29.12 ($IQR = 13.92$) and of ‘b’ was 45.72 ($IQR = 15.91$). This difference was statistically significant according to a Wilcoxon signed-rank test ($Z = -3.92, p < .0001$).”

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

One factor with ≥ 2 levels

Factors	Levels	Between or Within Subjects	Test Name	R Code
1	≥ 2	Between	Kruskal-Wallis test	<pre># df has one between-Ss. factor (X) w/levels (a,b,c) and continuous response (Y) library(coin) # for kruskal_test df\$PId = factor(df\$PId) # participant is nominal (unused) df\$X = factor(df\$X) # X is a 3-level factor contrasts(df\$X) <- "contr.sum" kruskal_test(Y ~ X, data=df, distribution="asymptotic")</pre>
1	≥ 2	Within	Friedman test	<pre># df has one within-Ss. factor (X) w/levels (a,b,c), and continuous response (Y) library(coin) df\$PId = factor(df\$PId) # participant is nominal df\$X = factor(df\$X) # X is a 3-level factor friedman_test(Y ~ X PId, data=df, distribution="asymptotic")</pre>

Nonparametric Tests

One factor with ≥ 2 levels

Factors	Levels	Between or Within Subjects	Test Name	Report
1	≥ 2	Between	Kruskal-Wallis test	“The median of ‘a’ was 31.44 ($IQR = 12.50$), of ‘b’ was 42.90 ($IQR = 20.60$), and of ‘c’ was 39.50 ($IQR = 14.73$). These differences were statistically significant according to a Kruskal-Wallis test ($\chi^2(2, N=60) = 9.36, p = .009$).”
1	≥ 2	Within	Friedman test	“The median of ‘a’ was 31.44 ($IQR = 12.50$), of ‘b’ was 42.90 ($IQR = 20.60$), and of ‘c’ was 39.50 ($IQR = 14.73$). These differences were statistically significant according to a Friedman test ($\chi^2(2, N=60) = 12.90, p = .002$).”

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

Factors	Levels	Between or Within Subjects	Test Name	R Code
≥2	≥2	Between	Aligned Rank Transform (ART)	<pre># df has two between-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(ARTool) df\$PId = factor(df\$PId) # participant 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 contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = art(Y ~ X1*X2, data=df) anova(m)</pre>
≥2	≥2	Within	Aligned Rank Transform (ART)*	<pre># df has two within-Ss. factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(ARTool) df\$PId = factor(df\$PId) # participant is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = art(Y ~ X1*X2 + (1 PId), data=df) # PId is a random factor anova(m)</pre>

*The Aligned Rank Transform within-subjects sample code uses a random intercept for *participant (PId)*. There are also random slope models, which are used when the response changes at different rates for each participant 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>). A free webinar on the topic of random intercept and random slope models is available here (<https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/>).

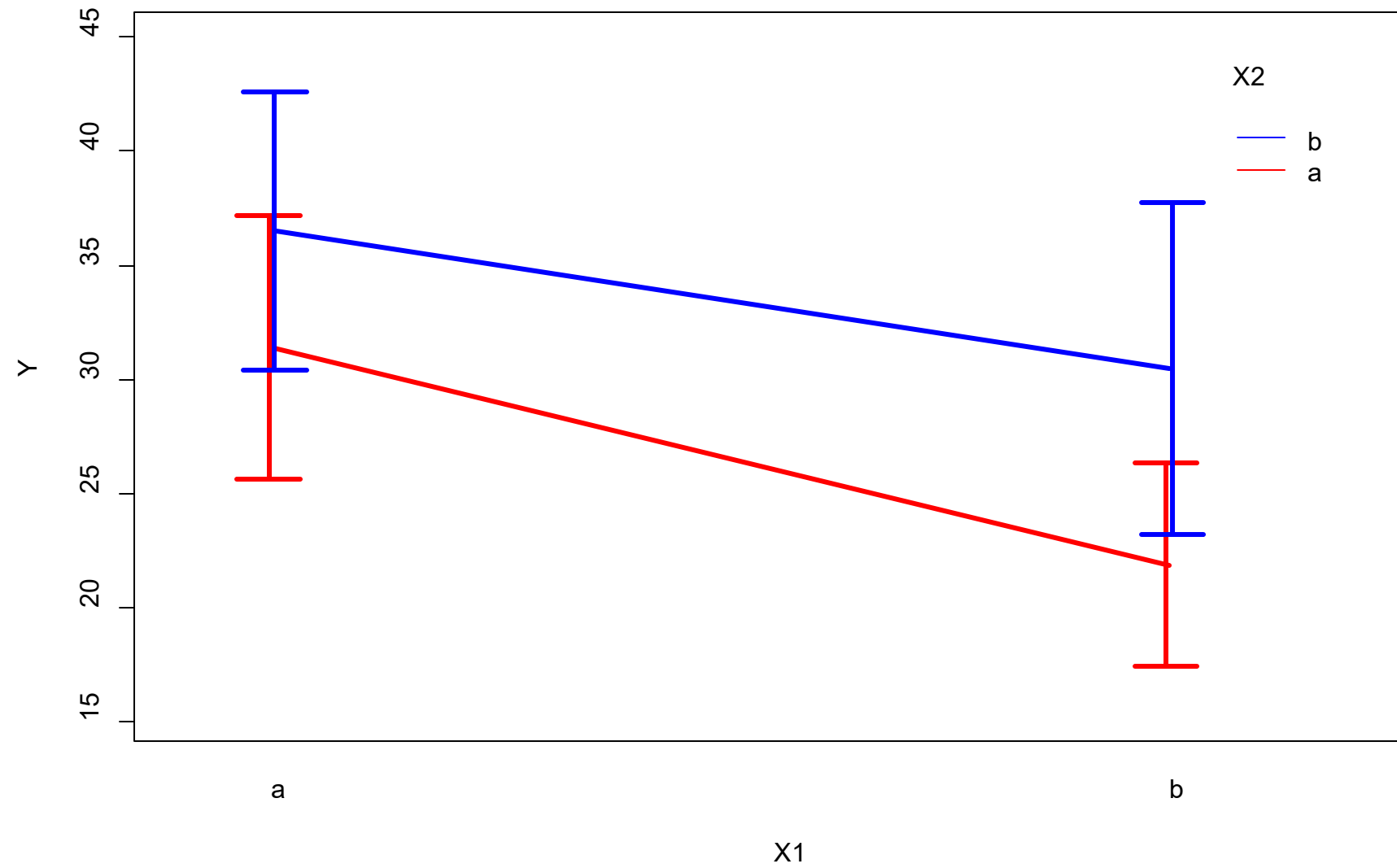
Nonparametric Tests

Multiple factors

Factors	Levels	Between or Within Subjects	Test Name	Report
≥ 2	≥ 2	Between	Aligned Rank Transform (ART)	“Figure 11 shows an interaction plot for Y by X1, X2 with ± 1 SD error bars. A nonparametric analysis of variance based on the Aligned Rank Transform indicated a significant effect on Y of X1 ($F(1, 56) = 24.85, p < .0001$) and of X2 ($F(1, 56) = 19.54, p < .0001$), but no significant X1×X2 interaction ($F(1, 56) = 0.98, p = .327$).”
≥ 2	≥ 2	Within	Aligned Rank Transform (ART)	“Figure 11 shows an interaction plot for Y by X1, X2 with ± 1 SD error bars. A nonparametric analysis of variance based on the Aligned Rank Transform indicated a significant effect on Y of X1 ($F(1, 42) = 24.85, p < .0001$) and of X2 ($F(1, 42) = 19.54, p < .0001$), but no significant X1×X2 interaction ($F(1, 42) = 0.98, p = .328$).”

Figure 11

Y by X1, X2



Nonparametric Tests

Post hoc pairwise comparisons

Nonparametric Tests

Post hoc pairwise comparisons – One factor

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

Nonparametric Tests

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Test	Test Name	B/W	Report
1	≥ 2	Kruskal-Wallis test	Mann-Whitney <i>U</i> test	Btwn	“Three <i>post hoc</i> Mann-Whitney <i>U</i> tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -2.89, p = .012$) was significantly different, but that ‘a’ vs. ‘c’ and ‘b’ vs. ‘c’ were not.”
1	≥ 2	Friedman test	Wilcoxon signed-rank test	Within	“Three <i>post hoc</i> Wilcoxon signed-rank tests, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -2.87, p = .013$) was significantly different, but that ‘a’ vs. ‘c’ and ‘b’ vs. ‘c’ were not.”

Nonparametric Tests

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Test	Test Name	B/W	R Code
≥2	≥2	Aligned Rank Transform (ART)	Aligned Rank Transform Contrasts (ART-C)	Btwn, Within	<pre> # df has two factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(plyr) # for mutate library(dplyr) # for %>% art.con(m, ~ X1*X2, adjust="holm") %>% # run ART-C for X1×X2 summary() %>% # (optional) add significance stars to the output plyr::mutate(sig. = symnum(p.value, corr=FALSE, na=FALSE, cutpoints = c(0, .001, .01, .05, .10, 1), symbols = c("****", "***", "**", ".", " "))) </pre>

Nonparametric Tests

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Test	Test Name	B/W	Report
≥2	≥2	Aligned Rank Transform (ART)	Aligned Rank Transform Contrasts (ART-C)	Btwn	“Six <i>post hoc</i> pairwise comparisons conducted with the ART-C procedure (Elkin <i>et al.</i> 2021), and corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,a} ($t(56) = 4.37, p < .001$), {a,b} vs. {b,a} ($t(56) = 6.67, p < .0001$), {a,b} vs. {b,b} ($t(56) = 2.67, p = .030$), and {b,a} vs. {b,b} ($t(56) = -4.00, p = .001$) were significantly different. The two other pairwise comparisons were not detectably different.”
				Within	“Six <i>post hoc</i> pairwise comparisons conducted with the ART-C procedure (Elkin <i>et al.</i> 2021), and corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,a} ($t(42) = 4.37, p < .001$), {a,b} vs. {b,a} ($t(42) = 6.67, p < .0001$), {a,b} vs. {b,b} ($t(42) = 2.67, p = .032$), and {b,a} vs. {b,b} ($t(42) = -4.00, p = .001$). The two other pairwise comparisons were not detectably different.”

Bibliography (one-way nonparametric tests, aligned rank transform)

- Brown, G.W. and Mood, A.M. (1948). Homogeneity of several samples. *The American Statistician* 2 (3), p. 22.
<https://doi.org/10.2307/2682087>
- Brown, G.W. and Mood, A.M. (1951). On median tests for linear hypotheses. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press, pp. 159-166.
<https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Second-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/On-Median-Tests-for-Linear-Hypotheses/bsmsp/1200500226>
- Conover, W.J. (1999). The sign test. In *Practical Nonparametric Statistics (3rd ed.)*. New York, NY: John Wiley & Sons, pp. 157-165.
<https://www.wiley.com/en-us/Practical+Nonparametric+Statistics%2C+3rd+Edition-p-9780471160687>
- Elkin, L.A., Kay, M., Higgins, J. and Wobbrock, J.O. (2021). An aligned rank transform procedure for multifactor contrast tests. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '21)*. New York, NY: ACM Press, pp. 754-768.
<https://doi.org/10.1145/3472749.3474784>
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association* 32 (200), pp. 675-701. <https://doi.org/10.2307/2279372>
- Higgins, J.J., Blair, R.C. and Tashtoush, S. (1990). The aligned rank transform procedure. *Proceedings of the Conference on Applied Statistics in Agriculture*. Manhattan, Kansas: New Prairie Press, pp. 185-195.
<http://newprairiepress.org/agstatconference/1990/proceedings/18/>
- Higgins, J.J. and Tashtoush, S. (1994). An aligned rank transform test for interaction. *Nonlinear World* 1 (2), pp. 201-211.
- Hodges, J.L. and Lehmann, E.L. (1962). Rank methods for combination of independent experiments in the analysis of variance. *Annals of Mathematical Statistics* 33 (2), pp. 482-497. <https://www.jstor.org/stable/2237528>

Bibliography (one-way nonparametric tests, aligned rank transform)

- Kruskal, W.H. and Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47 (260), pp. 583-621. <https://doi.org/10.2307/2280779>
- Mann, H.B. and Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18 (1), pp. 50-60. <https://www.jstor.org/stable/2236101>
- Salter, K.C. and Fawcett, R.F. (1985). A robust and powerful rank test of treatment effects in balanced incomplete block designs. *Communications in Statistics: Simulation and Computation* 14 (4), pp. 807-828. <https://doi.org/10.1080/03610918508812475>
- Salter, K.C. and Fawcett, R.F. (1993). The ART test of interaction: A robust and powerful rank test of interaction in factorial models. *Communications in Statistics: Simulation and Computation* 22 (1), pp. 137-153. <https://doi.org/10.1080/03610919308813085>
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* 1 (6), pp. 80-83. <https://doi.org/10.2307/3001968>
- Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)*. New York, NY: ACM Press, pp. 143-146. <https://doi.org/10.1145/1978942.1978963>

Generalized Linear (Mixed) Models

Terminology

		Generalize to Responses Unsuitable to ANOVA?			
		No	Yes		
Contain Random Factors?*	No	<p>(General) Linear Model (LM)</p> <p>(Sometimes abbreviated GLM)</p> <p><code>lm</code> in R</p> <p>(see also <code>aov</code>)</p>	<p><u>Generalized</u> Linear Model (GLM)</p> <p>(Sometimes abbreviated GZLM or GLIM)</p> <p><code>glm</code> in R</p>	Between-Ss.	
	Yes	<p>(General) Linear Mixed Model (LMM)</p> <p><code>lme4::lmer</code> in R</p> <p>(see also <code>nlme::lme</code>)</p>	<p><u>Generalized</u> Linear Mixed Model (GLMM)</p> <p><code>lme4::glmer</code> in R</p>	Within-Ss.	
		Normal only	Normal, binomial, multinomial, ordinal, Poisson, zero-inflated Poisson, negative binomial, zero-inflated negative binomial, exponential, gamma		

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

Distribution	GLM	GLMM
Normal	<code>lm()</code> ^[1]	<code>lme4::lmer()</code> ^[2]
Lognormal ^[3]	<code>lm(log(Y) ~ ...)</code>	<code>lme4::lmer(log(Y) ~ ...)</code>
Binomial	<code>glm() family=binomial</code>	<code>lme4::glmer() family=binomial</code>
Multinomial ^[4]	<code>multpois::glm.mp()</code> <code>multpois::Anova.mp()</code>	<code>multpois::glmer.mp()</code> <code>multpois::Anova.mp()</code>
Ordinal	<code>MASS::polr()</code>	<code>ordinal::clmm()</code> <code>RVAideMemoire::Anova.clmm()</code>
Poisson	<code>glm() family=poisson</code> <code>#or family=quasipoisson</code> ^[5]	<code>lme4::glmer() family=poisson</code>
Zero-inflated Poisson ^[6]	<code>glmmTMB::glmmTMB() family=poisson</code> <code>ziformula=~1</code>	<code>glmmTMB::glmmTMB() family=poisson</code> <code>ziformula=~1 REML=TRUE</code> ^[7]
Negative binomial	<code>MASS::glm.nb()</code>	<code>lme4::glmer.nb()</code>
Zero-inflated negative binomial ^[6]	<code>glmmTMB::glmmTMB() family=nbinom2</code> <code>ziformula=~1</code>	<code>glmmTMB::glmmTMB() family=nbinom2</code> <code>ziformula=~1 REML=TRUE</code> ^[7]
Exponential	<code>glm() family=Gamma(link="log")</code>	<code>lme4::glmer() family=Gamma(link="log")</code>
Gamma ^[8]	<code>glm() family=Gamma</code>	<code>lme4::glmer() family=Gamma</code>

Footnotes [1] – [8] appear on next slide.

Footnotes

- [1] `lm` and `glm` are from the base `stats` package in R. A call to `glm` with `family=gaussian` is the same as a call to `lm`.
- [2] An `lme4::glmer` call with `family=gaussian` will produce a message that one should just use `lme4::lmer`, which is equivalent.
- [3] A lognormal distribution is just a log-transform of the response Y and otherwise a linear (mixed) model.
- [4] The `multpois` functions use the multinomial-Poisson transformation (Baker 1994). See **NB**, below.
- [5] Use `family=quasipoisson` when mild overdispersion is present. If overdispersion is large, use negative binomial regression. Note that `family=quasipoisson` cannot be used with `lme4::glmer`.
- [6] Use zero-inflated variants of Poisson regression or negative binomial regression when distributions show large numbers of zeros relative to other counts. Use `print(performance::check_zeroinflation(m))` to test whether a regular Poisson or negative binomial model `m` is zero-inflated. The `glmmTMB` function takes a `ziformula` parameter that can be set to `~1` to fit a single zero-inflation parameter applying to all observations. (The `glmmTMB` [vignette](#) offers more detail.)
- [7] `glmmTMB` models with random factors should set `REML` to `TRUE` to use restricted maximum likelihood estimation.
- [8] Gamma models use the inverse function as their canonical link, $f(x) = x^{-1}$. Sometimes this fails; in such cases, it is reasonable to try the log link function, $f(x) = \log(x)$. Set the `family` parameter to `Gamma(link="log")`.

NB: Unfortunately, there is no `family=multinomial` option for `glm` or `lme4::glmer`. Also, the `nnet::multinom` function produces models that are difficult to use with `emmeans`. Moreover, `multinom` cannot accept random factors to handle repeated measures. Some other functions *do* offer multinomial regression modeling, such as `mclogit::mblogit`, but do not enable ANOVA-style output. Markov chain Monte Carlo (MCMC) methods in the `MCMCglmm` package also exist, but these methods utilize Bayesian approaches distinct from the other models presented here. Fortunately, Baker (1994) illustrated how multinomial models can be treated as Poisson models using the multinomial-Poisson transformation. By analyzing our data as counts for each categorical alternative, we can use `glm` or `lme4::glmer` with `family=poisson` to obtain the equivalent results as if they had a `family=multinomial` option. The `multpois` package offers such models, and the `multpois::Anova.mp` function produces ANOVA-style output. The `*.mp.con` functions carry out *post hoc* pairwise comparisons.

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss.)	R code for GLMM (within-Ss.)
Normal	identity	<i>Linear regression:</i> Normally distributed responses; equivalent to the linear model (LM) or linear mixed model (LMM)	<pre>library(car) # for Anova library(performance) # for check_* df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = lm(Y ~ X1*X2, data=df) print(check_normality(m)) print(check_homogeneity(m)) Anova(m, type=3, test.statistic="F")</pre>	<pre>library(lme4) # for lmer library(lmerTest) library(car) # for Anova library(performance) # for check_* df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = lmer(Y ~ X1*X2 + (1 PId), data=df) print(check_normality(m)) print(check_homogeneity(m)) Anova(m, type=3, test.statistic="F")</pre>
Lognormal	identity	<i>Linear regression:</i> Lognormally distributed responses (e.g., time measurements)	<pre>library(car) # for Anova library(performance) # for check_* df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = lm(log(Y) ~ X1*X2, data=df) print(check_normality(m)) print(check_homogeneity(m)) Anova(m, type=3, test.statistic="F")</pre>	<pre>library(lme4) # for lmer library(lmerTest) library(car) # for Anova library(performance) # for check_* df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = lmer(log(Y) ~ X1*X2 + (1 PId), data=df) print(check_normality(m)) print(check_homogeneity(m)) Anova(m, type=3, test.statistic="F")</pre>

Note: A normal distribution is also known as a Gaussian distribution. The GLMM sample code uses a random intercept for *participant (PId)*. 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>). A free webinar on the topic of random intercept and random slope models is available here (<https://thecraftofstatisticalanalysis.com/random-intercept-random-slope-models/>).

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	Report for GLM (between-Ss.)	Report for GLMM (within-Ss.)
Normal	identity	<i>Linear regression:</i> Normally distributed responses; equivalent to the linear model (LM) or linear mixed model (LMM)	“Figure 12 shows an interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance based on a linear model indicated a statistically significant X1×X2 interaction ($F(1, 56) = 8.05, p = .006$).”	“Figure 12 shows an interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance based on a linear mixed model indicated a statistically significant X1×X2 interaction ($F(1, 42) = 8.05, p = .007$).”
Lognormal	identity	<i>Linear regression:</i> Lognormally distributed responses (e.g., time measurements)	“Figure 13 shows four lognormal histograms for each X1×X2 condition. An analysis of variance based on a linear model indicated no statistically significant effects on $\log(Y)$ of X1, X2, or the X1×X2 interaction.”	“Figure 13 shows four lognormal histograms for each X1×X2 condition. An analysis of variance based on a linear mixed model indicated no statistically significant effects on $\log(Y)$ of X1, X2, or the X1×X2 interaction.”

Figure 12

Y by X1, X2

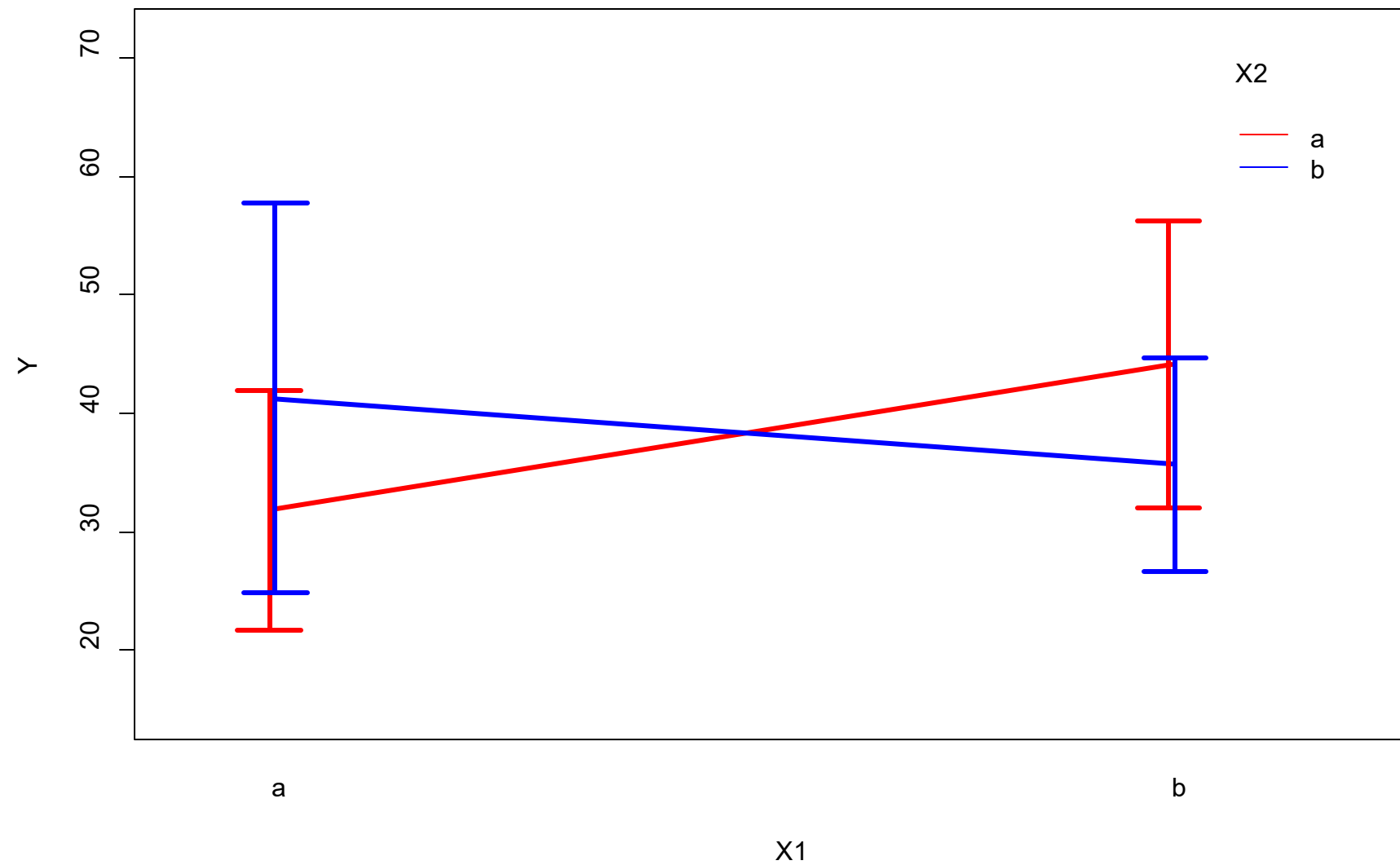
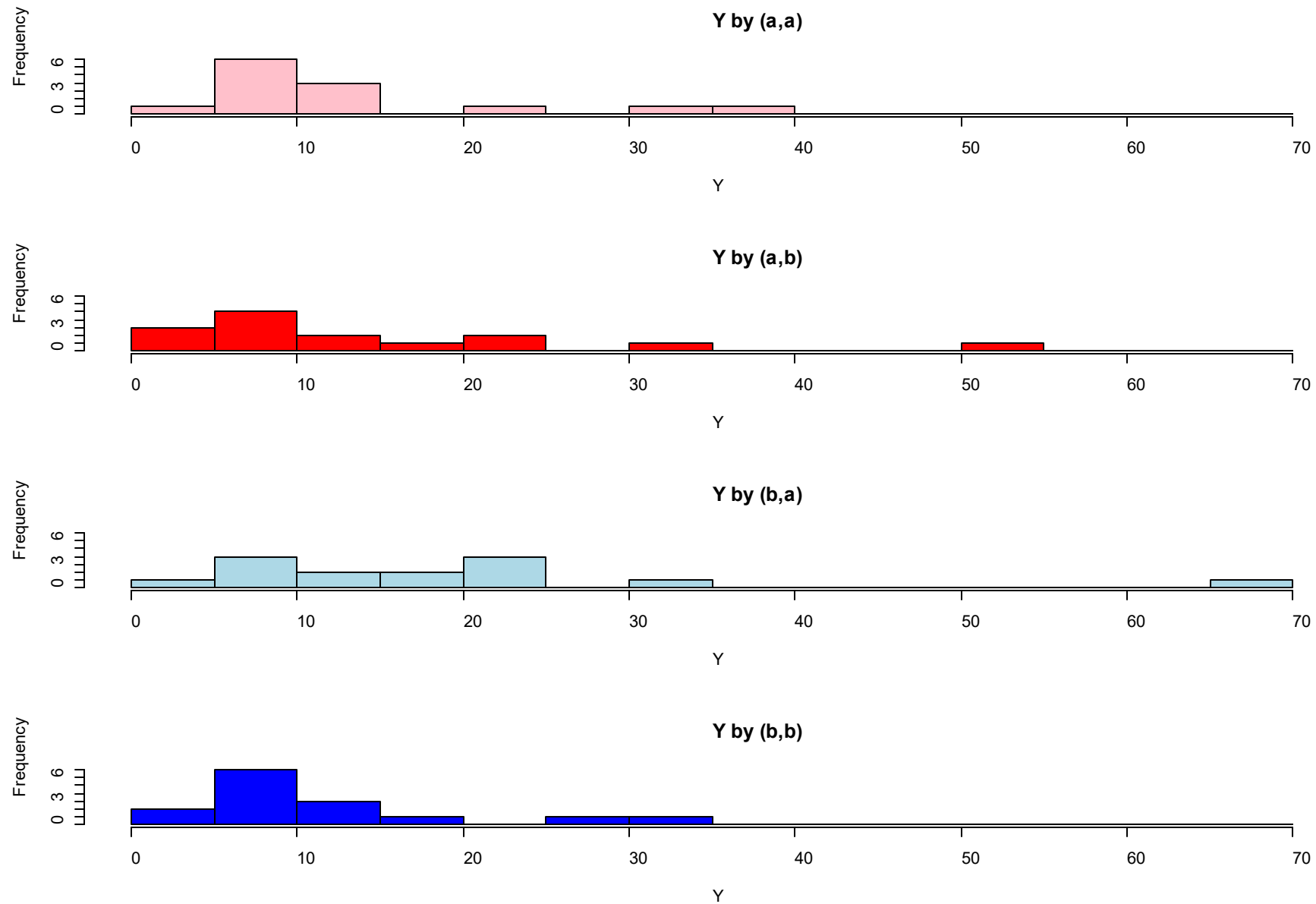


Figure 13



GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss.)	R code for GLMM (within-Ss.)
Binomial	logit	<i>Logistic regression:</i> Dichotomous responses (i.e., nominal responses with two categories)	<pre>library(car) # for Anova df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = factor(df\$Y) # dichotomous response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glm(Y ~ X1*X2, data=df, family=binomial) Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(car) # for Anova df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = factor(df\$Y) # dichotomous response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmer(Y ~ X1*X2 + (1 PId), data=df, family=binomial) Anova(m, type=3)</pre>
Multinomial	logit	<i>Multinomial logistic regression:</i> Polytomous responses (i.e., nominal responses with more than two categories)	<pre>library(multpois) # for glm.mp, Anova.mp df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = factor(df\$Y) # polytomous response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glm.mp(Y ~ X1*X2, data=df) Anova.mp(m, type=3)</pre>	<pre>library(multpois) # for glmer.mp, Anova.mp df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = factor(df\$Y) # polytomous response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmer.mp(Y ~ X1*X2 + (1 PId), data=df) Anova.mp(m, type=3)</pre>

Notes: Logistic regression is also known as binomial regression. Multinomial logistic regression is also known as nominal logistic regression. It is carried out here using the multinomial-Poisson transformation (Baker 1994).

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	Report for GLM (between-Ss.)	Report for GLMM (within-Ss.)
Binomial	logit	<i>Logistic regression:</i> Dichotomous responses (i.e., nominal responses with two categories)	“Figure 14 shows the number of ‘yes’ and ‘no’ responses for each X1×X2 condition. An analysis of variance based on logistic regression indicated a statistically significant effect of X2 on Y ($\chi^2(1, N=60) = 11.69, p = .001$).”	“Figure 14 shows the number of ‘yes’ and ‘no’ responses for each X1×X2 condition. An analysis of variance based on mixed logistic regression indicated a statistically significant effect of X2 on Y ($\chi^2(1, N=60) = 9.13, p = .003$).”
Multinomial	logit	<i>Multinomial logistic regression:</i> Polytomous responses (i.e., nominal responses with more than two categories)	“Figure 15 shows the number of ‘yes’, ‘no’, and ‘maybe’ responses for each X1×X2 condition. An analysis of variance based on multinomial logistic regression, implemented using the multinomial-Poisson transformation (Baker 1994), indicated a statistically significant effect of X1 on Y ($\chi^2(2, N=60) = 10.33, p = .006$).”	“Figure 15 shows the number of ‘yes’, ‘no’, and ‘maybe’ responses for each X1×X2 condition. An analysis of variance based on mixed multinomial logistic regression, implemented using the multinomial-Poisson transformation (Baker 1994), indicated a statistically significant effect of X1 on Y ($\chi^2(2, N=60) = 9.28, p = .010$).”

Figure 14

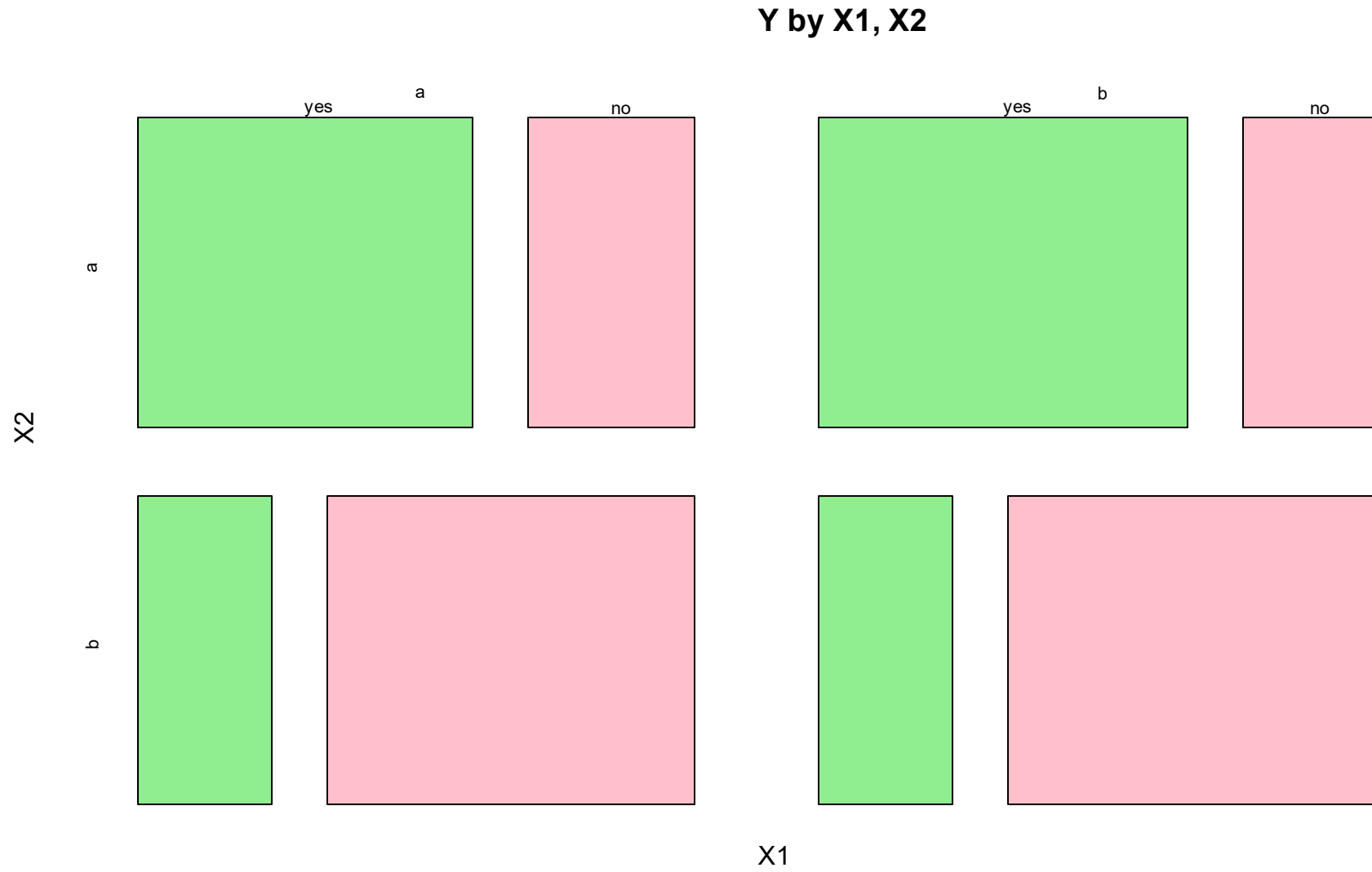
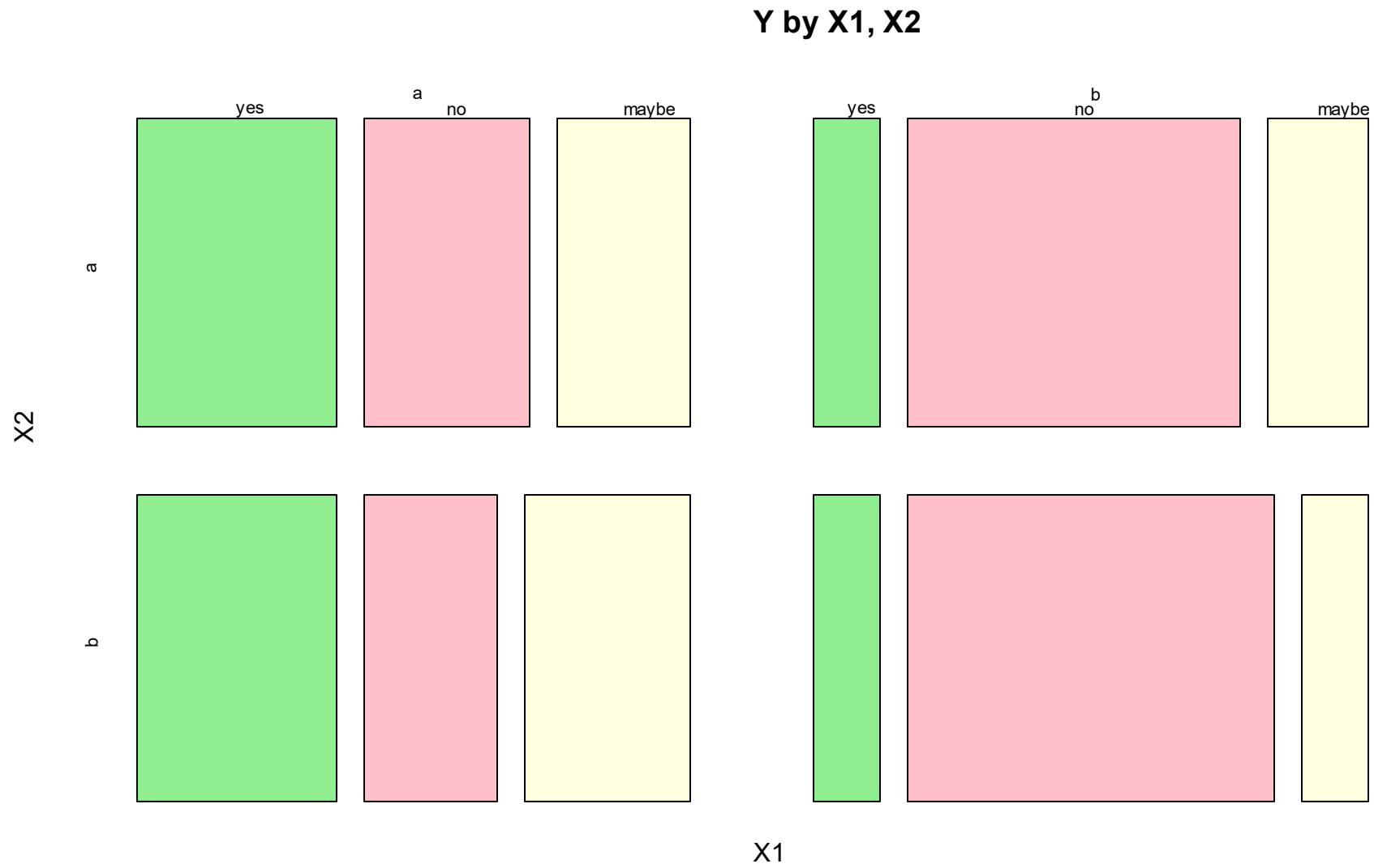


Figure 15



GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss.)	R code for GLMM (within-Ss.)
Ordinal	cumulative logit	<i>Ordinal logistic regression:</i> Ordinal responses (e.g., Likert-type scales)	<pre>library(MASS) # for polr library(car) # for Anova df\$PIId = factor(df\$PIId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = ordered(df\$Y) # ordinal response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = polr(Y ~ X1*X2, data=df, Hess=TRUE) Anova(m, type=3)</pre>	<pre>library(ordinal) # for clmm library(RVAideMemoire) # for Anova.clmm df\$PIId = factor(df\$PIId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = ordered(df\$Y) # ordinal response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" df2 <- as.data.frame(df) # Anova.clmm fails without this m = clmm(Y ~ X1*X2 + (1 PIId), data=df2, Hess=TRUE, link="logit") # logit, probit, cloglog, loglog, cauchit links are valid Anova.clmm(m)</pre>

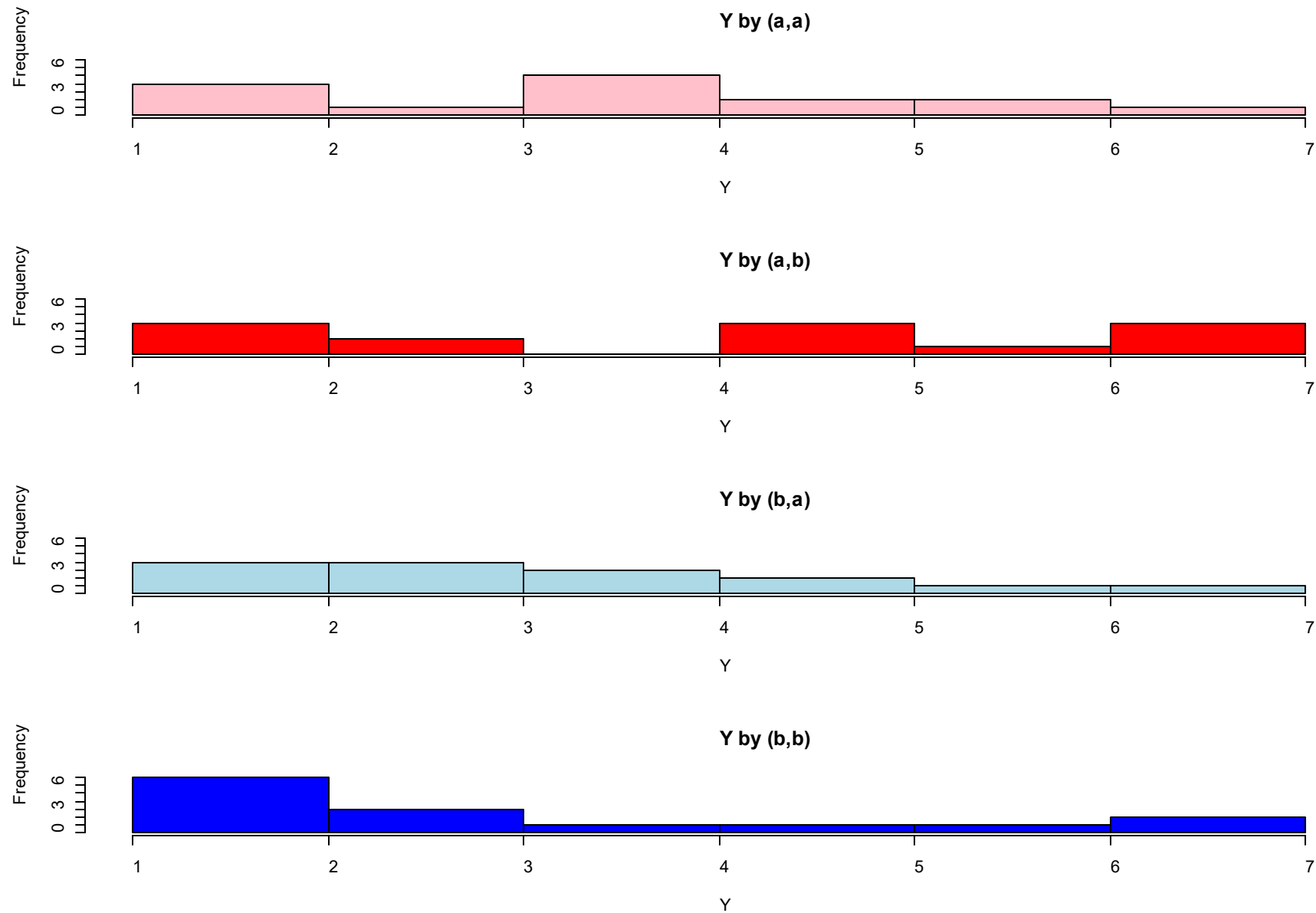
Note: Ordinal logistic regression is also known as cumulative logistic regression.

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	Report for GLM (between-Ss)	Report for GLMM (within-Ss)
Ordinal	cumulative logit	<i>Ordinal logistic regression:</i> Ordinal responses (e.g., Likert-type scales)	“Figure 16 shows the distribution of Likert responses (1-7) for each $X1 \times X2$ condition. An analysis of variance based on ordinal logistic regression indicated no statistically significant effects on Y of X1, X2, or the $X1 \times X2$ interaction.”	“Figure 16 shows the distribution of Likert responses (1-7) for each $X1 \times X2$ condition. An analysis of variance based on mixed ordinal logistic regression indicated no statistically significant effects on Y of X1, X2, or the $X1 \times X2$ interaction.”

Figure 16



GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss.)	R code for GLMM (within-Ss.)
Poisson	log	<i>Poisson regression:</i> Count responses	<pre>library(car) # for Anova library(performance) # for check_overdispersion df\$PId = factor(df\$PId) 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=poisson) print(check_overdispersion(m)) # use family=quasipoisson or negative binomial # regression if overdispersed Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(car) # for Anova library(performance) # for check_overdispersion df\$PId = factor(df\$PId) 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 PId), data=df, family=poisson) print(check_overdispersion(m)) # use mixed negative binomial regression if overdispersed Anova(m, type=3)</pre>
Zero-Inflated Poisson	log	<i>Zero-inflated Poisson regression:</i> Zero-inflated count responses	<pre>library(glmmTMB) # for glmmTMB library(car) # for Anova library(performance) # for check_zeroinflation df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m0 = glm(Y ~ X1*X2, data=df, family=poisson) print(check_zeroinflation(m0)) m = glmmTMB(Y ~ X1*X2, data=df, family=poisson, ziformula=~1) Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(glmmTMB) # for glmmTMB library(car) # for Anova library(performance) # for check_zeroinflation df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m0 = glmer(Y ~ X1*X2 + (1 PId), data=df, family=poisson) print(check_zeroinflation(m0)) m = glmmTMB(Y ~ X1*X2 + (1 PId), data=df, family=poisson, ziformula=~1, REML=TRUE) Anova(m, type=3)</pre>

Notes: When count data is overdispersed, it means the variance of the response in each condition is greater than its mean. To test for overdispersion, one can use `performance::check_overdispersion(m)`, where “m” is a fitted model with `family=poisson`. One can also use `abs(var(Y)/mean(Y)) > 1.15`, where “Y” is the response in each condition. For mildly overdispersed count data, `family=quasipoisson` can be used with `glm` but not with `lme4::glmer`. With high overdispersion, use negative binomial regression.

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	Report for GLM (between-Ss.)	Report for GLMM (within-Ss.)
Poisson	log	<i>Poisson regression:</i> Count responses	“Figure 17 shows histograms of count responses in each $X1 \times X2$ condition. An analysis of variance based on Poisson regression indicated a statistically significant effect of X1 on Y ($\chi^2(1, N=60) = 4.47, p = .034$) and a significant $X1 \times X2$ interaction ($\chi^2(1, N=60) = 5.59, p = .018$).”	“Figure 17 shows histograms of count responses in each $X1 \times X2$ condition. An analysis of variance based on mixed Poisson regression indicated a statistically significant effect of X1 on Y ($\chi^2(1, N=60) = 4.44, p = .035$) and a significant $X1 \times X2$ interaction ($\chi^2(1, N=60) = 5.54, p = .019$).”
Zero-Inflated Poisson	log	<i>Zero-inflated Poisson regression:</i> Zero-inflated count responses	“Figure 18 shows histograms of zero-inflated count responses in each $X1 \times X2$ condition. An analysis of variance based on zero-inflated Poisson regression indicated a statistically significant $X1 \times X2$ interaction ($\chi^2(1, N=60) = 7.34, p = .007$).”	“Figure 18 shows histograms of zero-inflated count responses in each $X1 \times X2$ condition. An analysis of variance based on mixed zero-inflated Poisson regression indicated a statistically significant $X1 \times X2$ interaction ($\chi^2(1, N=60) = 7.34, p = .007$).”

Figure 17

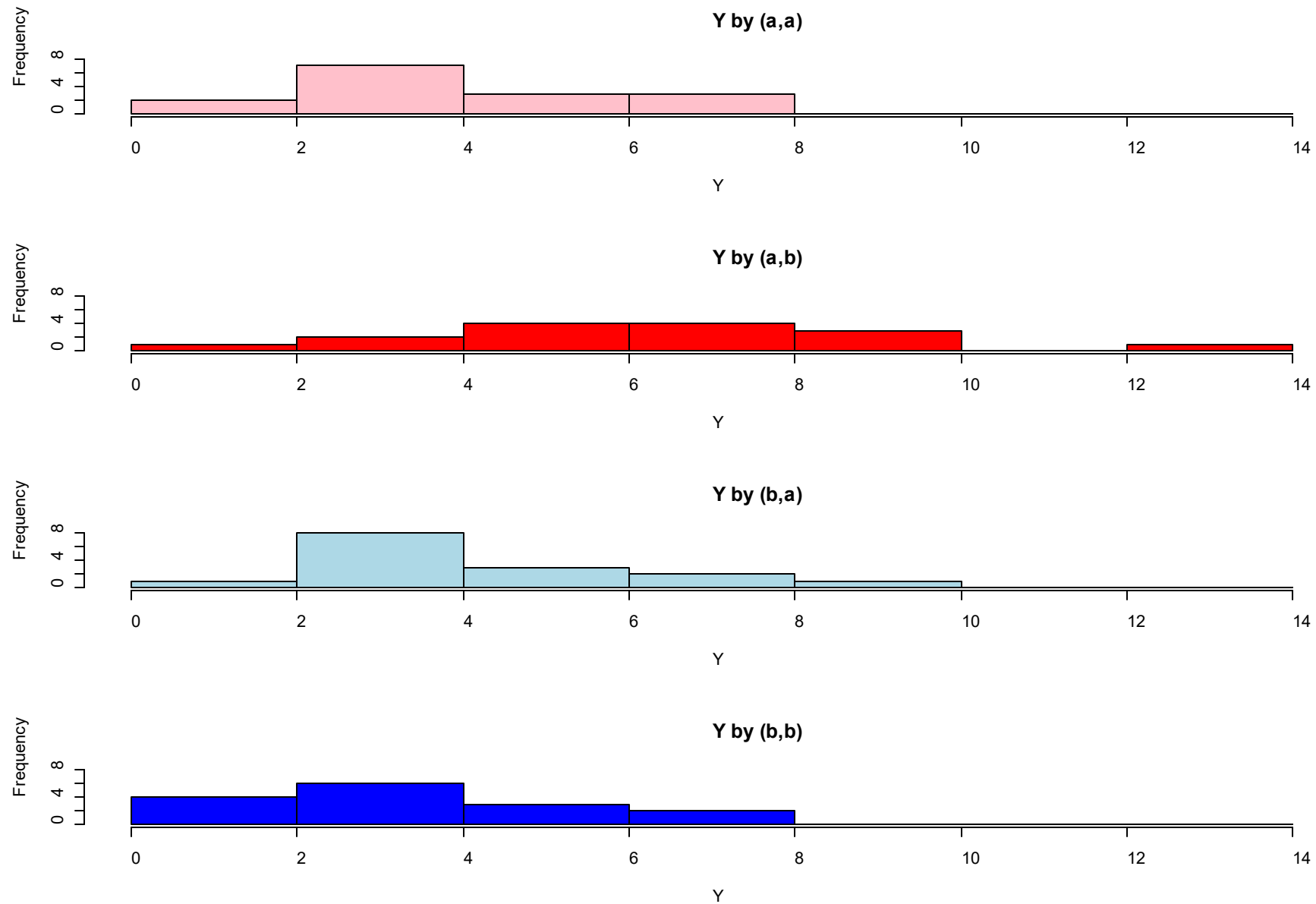
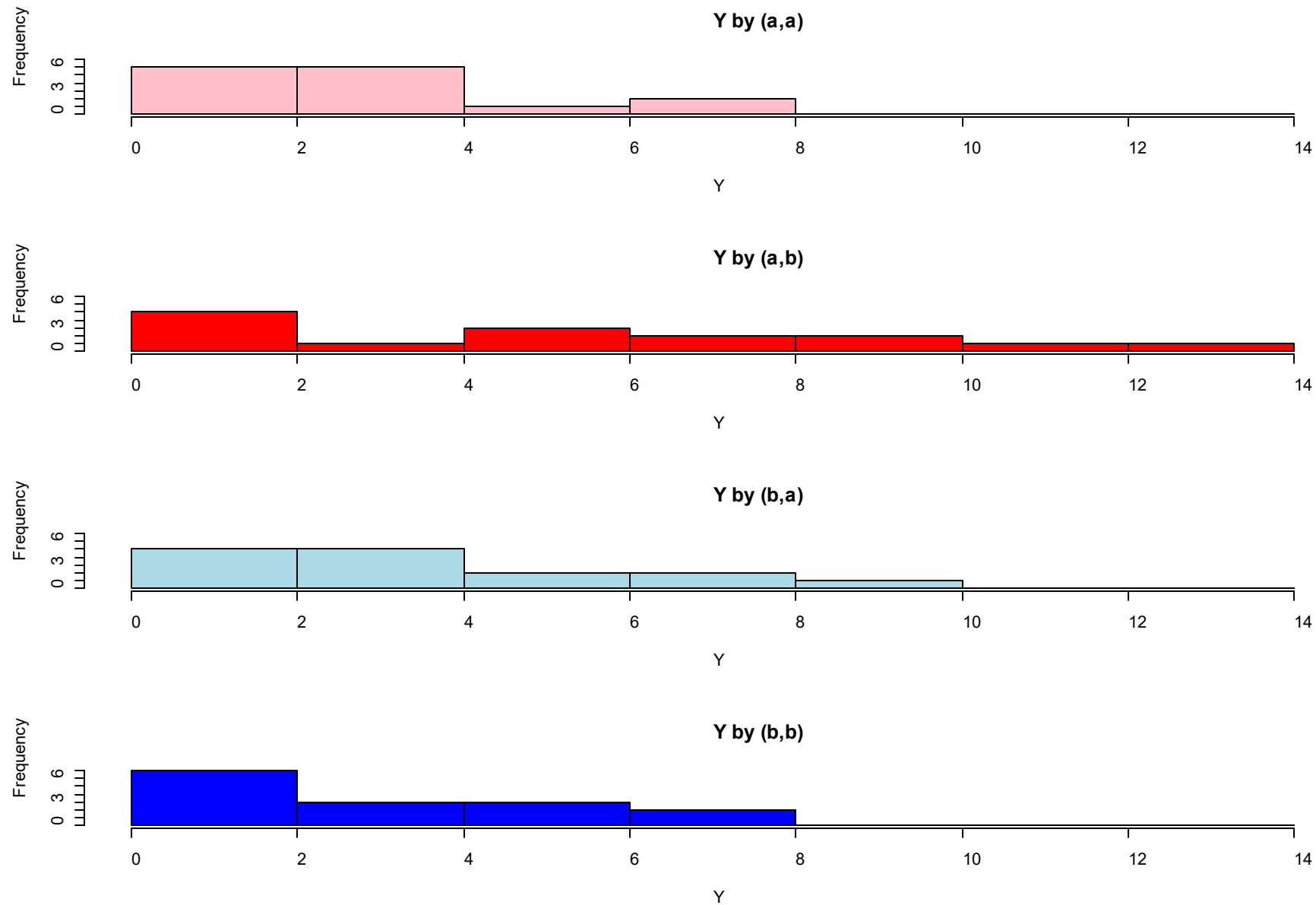


Figure 18



GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss.)	R code for GLMM (within-Ss.)
Negative Binomial	log	<i>Negative binomial regression:</i> Overdispersed count responses	<pre>library(MASS) # for glm.nb library(car) # for Anova df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glm.nb(Y ~ X1*X2, data=df) Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer.nb library(lmerTest) library(car) # for Anova df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmer.nb(Y ~ X1*X2 + (1 PId), data=df) Anova(m, type=3)</pre>
Zero-Inflated Negative Binomial	log	<i>Zero-inflated negative binomial regression:</i> Overdispersed zero-inflated count responses	<pre>library(MASS) # for glm.nb library(glmmTMB) # for glmmTMB library(car) # for Anova library(performance) # for check_zeroinflation df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m0 = glm.nb(Y ~ X1*X2, data=df) print(check_zeroinflation(m0)) m = glmmTMB(Y ~ X1*X2, data=df, family=nbinom2, ziformula=~1) Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer.nb library(lmerTest) library(glmmTMB) # for glmmTMB library(car) # for Anova library(performance) # for check_zeroinflation df\$PId = factor(df\$PId) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m0 = glmer.nb(Y ~ X1*X2 + (1 PId), data=df) print(check_zeroinflation(m0)) m = glmmTMB(Y ~ X1*X2 + (1 PId), data=df, family=nbinom2, ziformula=~1, REML=TRUE) Anova(m, type=3)</pre>

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	Report for GLM (between-Ss.)	Report for GLMM (within-Ss.)
Negative Binomial	log	<i>Negative binomial regression:</i> Overdispersed count responses	“Figure 19 shows histograms of count responses in each X1×X2 condition. An analysis of variance based on negative binomial regression indicated a statistically significant effect of X1 on Y ($\chi^2(1, N=60) = 4.20, p = .040$).”	“Figure 19 shows histograms of count responses in each X1×X2 condition. An analysis of variance based on mixed negative binomial regression indicated a statistically significant effect of X1 on Y ($\chi^2(1, N=60) = 4.21, p = .040$).”
Zero-Inflated Negative Binomial	log	<i>Zero-inflated negative binomial regression:</i> Overdispersed zero-inflated count responses	“Figure 20 shows histograms of zero-inflated count responses in each X1×X2 condition. An analysis of variance based on zero-inflated negative binomial regression indicated a statistically significant effect of X1 on Y ($\chi^2(1, N=60) = 6.25, p = .012$).”	“Figure 20 shows histograms of zero-inflated count responses in each X1×X2 condition. An analysis of variance based on mixed zero-inflated negative binomial regression indicated a statistically significant effect of X1 on Y ($\chi^2(1, N=60) = 5.52, p = .019$).”

Figure 19

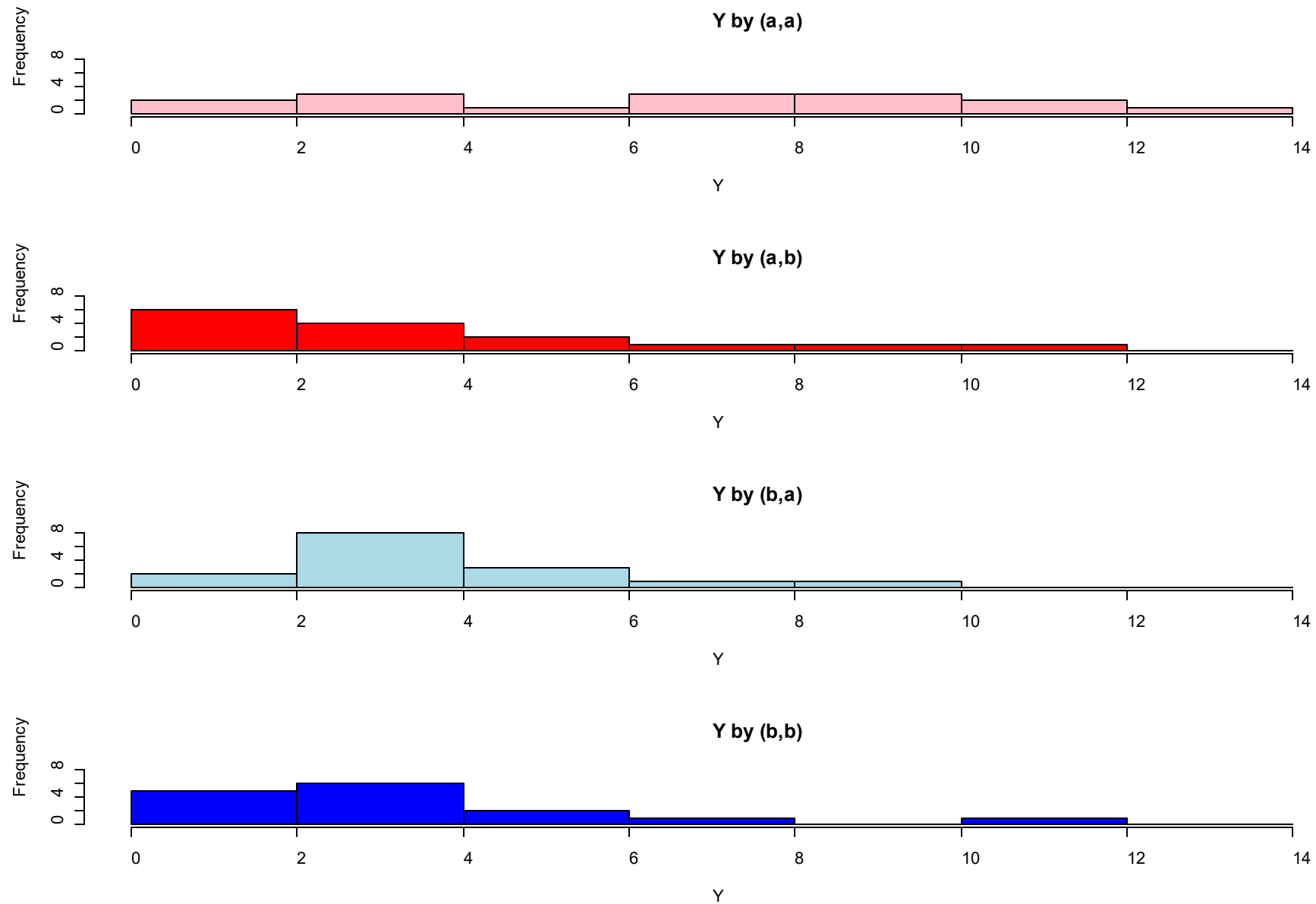
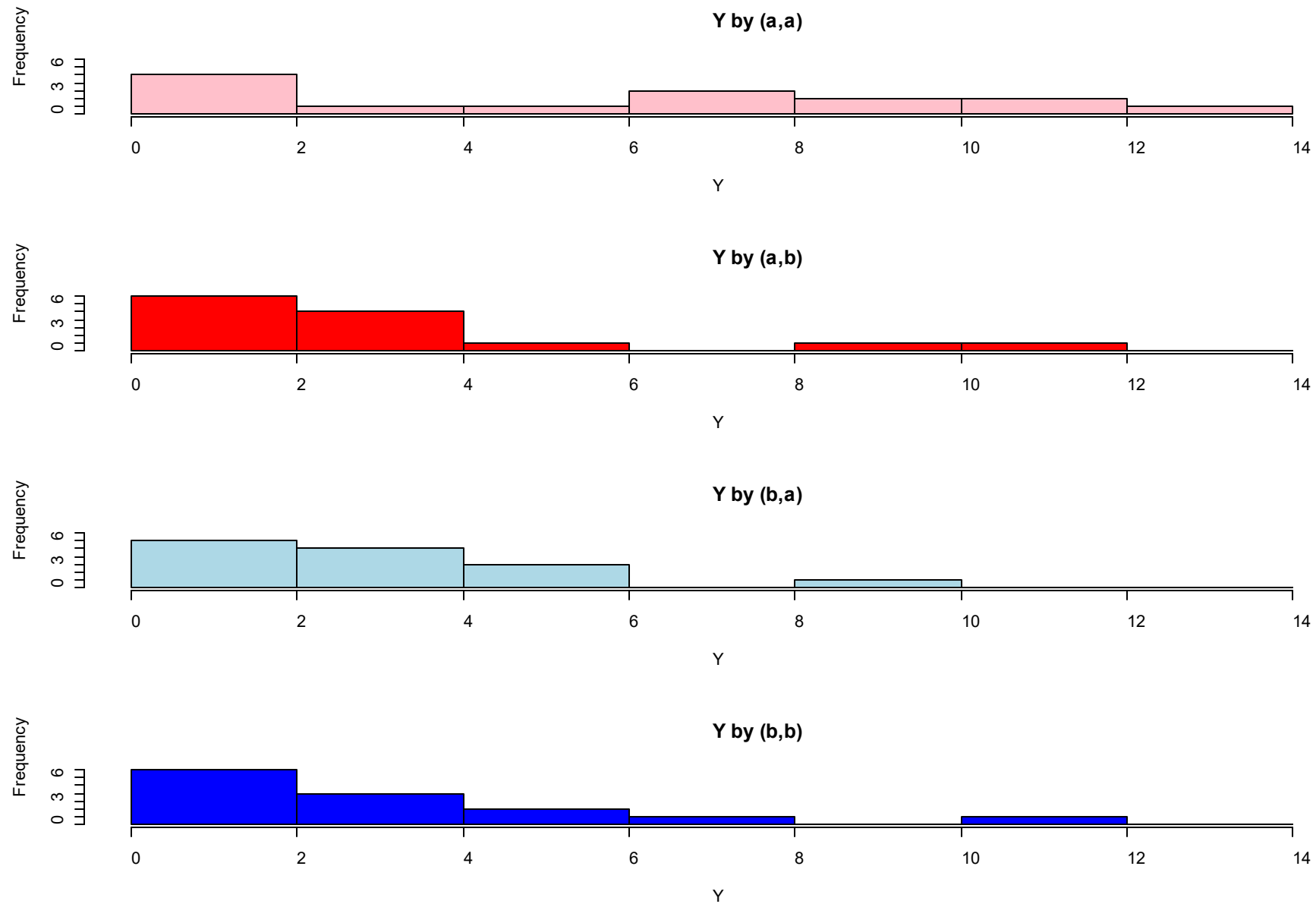


Figure 20



GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss.)	R code for GLMM (within-Ss.)
Exponential	log	<i>Exponential regression:</i> Exponentially distributed responses (e.g., income)	<pre>library(car) # for Anova df\$PId = factor(df\$PId) 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(link="log")) Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(car) # for Anova df\$PId = factor(df\$PId) 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 PId), data=df, family=Gamma(link="log")) Anova(m, type=3)</pre>
Gamma	inverse	<i>Gamma regression:</i> Skewed continuous responses (e.g., time measurements)	<pre>library(car) # for Anova df\$PId = factor(df\$PId) 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) Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(car) # for Anova df\$PId = factor(df\$PId) 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 PId), data=df, family=Gamma) Anova(m, type=3)</pre>

Notes: Gamma distributions are parameterized by *shape* and *scale* (or sometimes *rate*, which is $1/\text{scale}$). Exponential distributions are special cases of Gamma distributions where *shape* always equals 1. The log link function is used with the exponential distribution to fix the dispersion parameter at 1.0.

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	Report for GLM (between-Ss.)	Report for GLMM (within-Ss.)
Exponential	log	<i>Exponential regression:</i> Exponentially distributed responses (e.g., income)	“Figure 21 shows exponential histograms for each X1×X2 condition. An analysis of variance based on exponential regression indicated a statistically significant effect of X2 on Y ($\chi^2(1, N=60) = 20.96, p < .0001$).”	“Figure 21 shows exponential histograms for each X1×X2 condition. An analysis of variance based on mixed exponential regression indicated a statistically significant effect of X2 on Y ($\chi^2(1, N=60) = 31.00, p < .0001$).”
Gamma	inverse	<i>Gamma regression:</i> Skewed continuous responses (e.g., time measurements)	“Figure 22 shows gamma histograms for each X1×X2 condition. An analysis of variance based on gamma regression indicated a statistically significant effect of X2 on Y ($\chi^2(1, N=60) = 24.22, p < .0001$) and an X1×X2 interaction ($\chi^2(1, N=60) = 7.97, p = .005$).”	“Figure 22 shows gamma histograms for each X1×X2 condition. An analysis of variance based on mixed gamma regression indicated a statistically significant effect of X2 on Y ($\chi^2(1, N=60) = 25.36, p < .0001$) and an X1×X2 interaction ($\chi^2(1, N=60) = 8.86, p = .003$).”

Figure 21

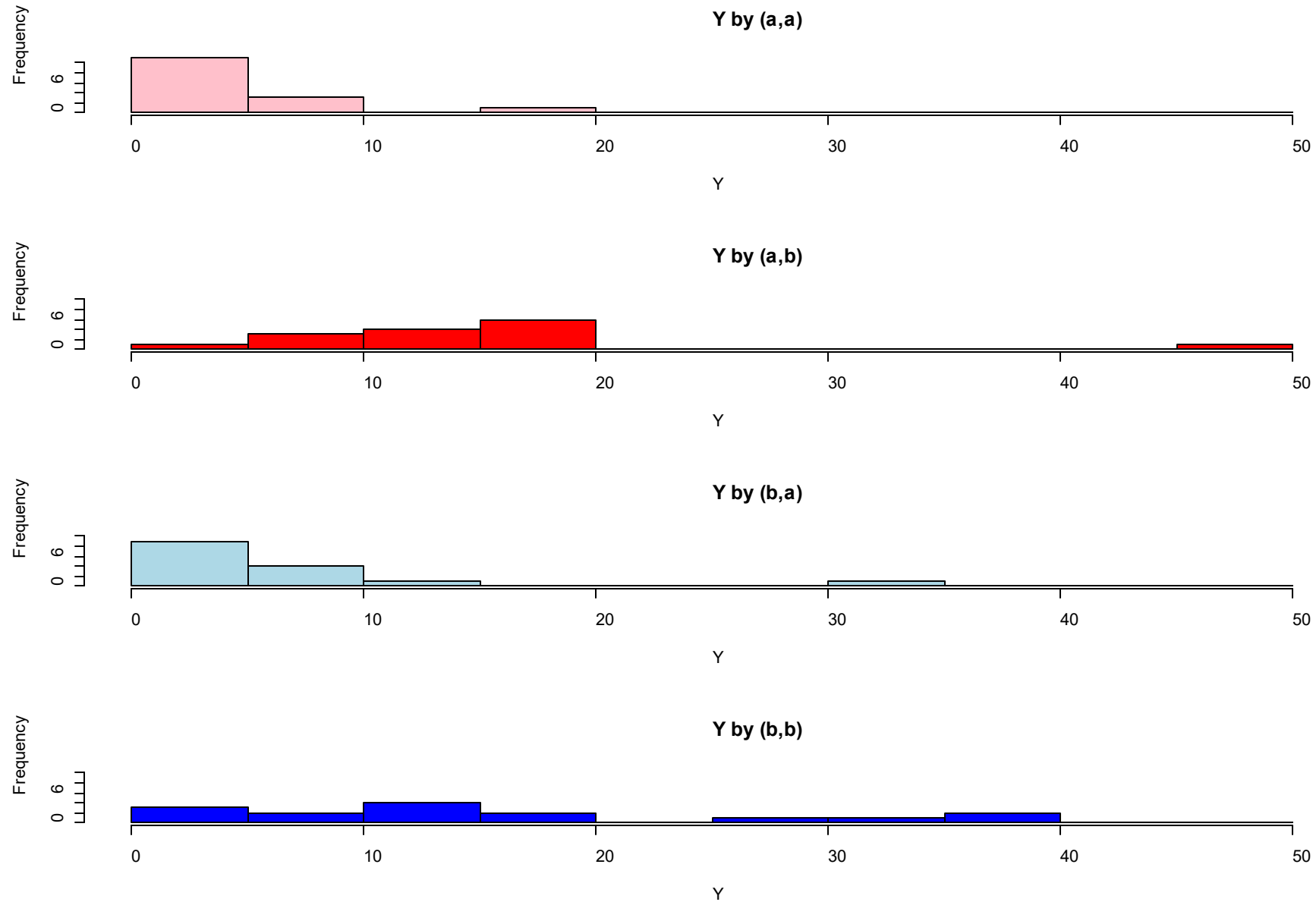
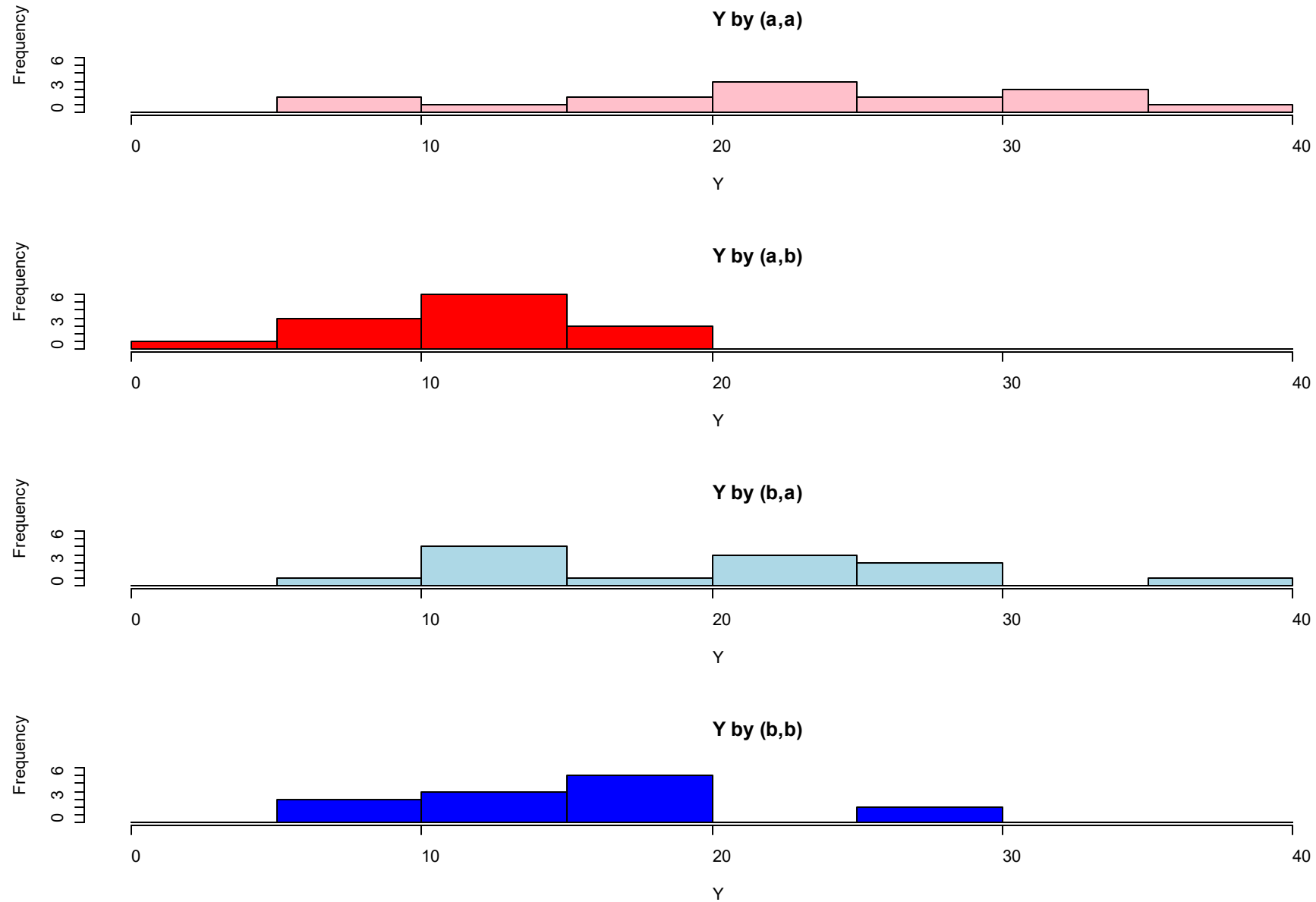


Figure 22



Generalized Linear (Mixed) Models

Post hoc pairwise comparisons

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	R Code
1	≥2	Linear regression	t-test	Btwn, Within	# df has one factor (X) w/levels (a,b,c) and continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from lm or lme4::lmer
1	≥2	Logistic regression	Z-test	Btwn, Within	# df has one factor (X) w/levels (a,b,c) and dichotomous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from glm or lme4::glmer with family=binomial
1	≥2	Multinomial logistic regression	Chi-squared test	Btwn	# df has one between-Ss. factor (X) w/levels (a,b,c) and polytomous response (Y) library(multpois) # for glm.mp.con glm.mp.con(m, pairwise ~ X, adjust="holm") # m is from multpois::glm.mp
				Within	# df has one within-Ss. factor (X) w/levels (a,b,c) and polytomous response (Y) library(multpois) # for glmer.mp.con glmer.mp.con(m, pairwise ~ X, adjust="holm") # m is from multpois::glmer.mp
1	≥2	Ordinal logistic regression	Z-test	Btwn, Within	# df has one factor (X) w/levels (a,b,c) and ordinal response (1-7) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from MASS::polr or ordinal::clmm

Notes: Between-subjects models are from GLMs, within-subjects models are from GLMMs. Multinomial logistic regression is implemented via the multinomial-Poisson transformation (Baker 1994).

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	Report
1	≥2	Linear regression	t-test	Btwn	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($t(57) = -3.16, p = .008$) and ‘a’ vs. ‘c’ ($t(57) = -2.47, p = .033$) were significantly different, but not ‘b’ vs. ‘c’.”
				Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($t(38) = -3.16, p = .009$) and ‘a’ vs. ‘c’ ($t(38) = -2.47, p = .036$) were significantly different, but not ‘b’ vs. ‘c’.”
1	≥2	Logistic regression	Z-test	Btwn	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘c’ was significantly different ($Z = -2.46, p = .041$), but not ‘a’ vs. ‘b’ or ‘b’ vs. ‘c’.”
				Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that no two levels were significantly different.”
1	≥2	Multinomial logistic regression	Chi-squared test	Btwn	“Three <i>post hoc</i> pairwise comparisons, conducted using the multinomial-Poisson transformation (Baker 1994), and corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘c’ ($\chi^2(2, N=40) = 14.71, p = .001$) and ‘b’ vs. ‘c’ ($\chi^2(2, N=40) = 17.65, p < .001$) were significantly different, but not ‘a’ vs. ‘b’.”
				Within	“Three <i>post hoc</i> pairwise comparisons, conducted using the multinomial-Poisson transformation (Baker 1994), and corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘c’ ($\chi^2(2, N=40) = 12.22, p = .007$) and ‘b’ vs. ‘c’ ($\chi^2(2, N=40) = 10.73, p = .009$) were significantly different, but not ‘a’ vs. ‘b’.”
1	≥2	Ordinal logistic regression	Z-test	Btwn, Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘b’ vs. ‘c’ was significantly different ($Z = -2.72, p = .020$), but not ‘a’ vs. ‘b’ or ‘a’ vs. ‘c’.”

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	R Code
1	≥2	Poisson regression	Z-test	Btwn, Within	<pre># df has one factor (X) w/levels (a,b,c) and count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from glm or lme4::glmer with family=poisson</pre>
1	≥2	Zero-inflated Poisson regression	Z-test	Btwn, Within	<pre># df has one factor (X) w/levels (a,b,c) and zero-inflated count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from glmmTMB::glmmTMB with family=poisson</pre>
1	≥2	Negative binomial regression	Z-test	Btwn, Within	<pre># df has one factor (X) w/levels (a,b,c) and overdispersed count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from MASS::glm.nb or lme4::glmer.nb</pre>
1	≥2	Zero-inflated negative binomial regression	Z-test	Btwn, Within	<pre># df has one factor (X) w/levels (a,b,c) and zero-inflated overdispersed count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from glmmTMB::glmmTMB with family=nbinom2</pre>

Notes: Between-subjects models are from GLMs, within-subjects models are from GLMMs.

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	Report
1	≥2	Poisson regression	Z-test	Btwn, Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -2.92, p = .007$) and ‘b’ vs. ‘c’ ($Z = 3.28, p = .003$) were significantly different, but not ‘a’ vs. ‘c’.”
1	≥2	Zero-inflated Poisson regression	Z-test	Btwn	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -2.81, p = .010$) and ‘b’ vs. ‘c’ ($Z = 2.98, p = .009$) were significantly different, but not ‘a’ vs. ‘c’.”
				Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -2.81, p = .010$) and ‘b’ vs. ‘c’ ($Z = 2.99, p = .009$) were significantly different, but not ‘a’ vs. ‘c’.”
1	≥2	Negative binomial regression	Z-test	Btwn, Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ was significantly different ($Z = -2.50, p = .038$), but not ‘a’ vs. ‘c’ or ‘b’ vs. ‘c’.”
1	≥2	Zero-inflated negative binomial regression	Z-test	Btwn	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -4.37, p < .0001$) and ‘b’ vs. ‘c’ ($Z = 4.47, p < .0001$) were significantly different, but not ‘a’ vs. ‘c’.”
				Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -4.24, p < .0001$) and ‘b’ vs. ‘c’ ($Z = 4.32, p < .0001$) were significantly different, but not ‘a’ vs. ‘c’.”

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	R Code
1	≥2	Exponential regression	Z-test	Btwn, Within	<pre># df has one factor (X) w/levels (a,b,c) and exponential response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from glm or lme4::glmer with # family=Gamma(link="log")</pre>
1	≥2	Gamma regression	Z-test	Btwn, Within	<pre># df has one factor (X) w/levels (a,b,c) and skewed continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X, adjust="holm") # m is from glm or lme4::glmer with family=Gamma</pre>

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	Report
1	≥2	Exponential regression	Z-test	Btwn	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -3.15, p = .008$) and ‘b’ vs. ‘c’ ($Z = 2.36, p = .044$) were significantly different, but not ‘a’ vs. ‘c’.”
				Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -3.64, p = .001$) and ‘b’ vs. ‘c’ ($Z = 2.86, p = .008$) were significantly different, but not ‘a’ vs. ‘c’.”
1	≥2	Gamma regression	Z-test	Btwn	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -5.44, p < .0001$), ‘a’ vs. ‘c’ ($Z = -3.20, p = .005$), and ‘b’ vs. ‘c’ ($Z = 3.00, p = .005$) were all significantly different.”
				Within	“Three <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that ‘a’ vs. ‘b’ ($Z = -6.15, p < .0001$), ‘a’ vs. ‘c’ ($Z = -3.58, p = .001$), and ‘b’ vs. ‘c’ ($Z = 3.48, p = .001$) were all significantly different.”

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Linear regression	t-test	Btwn, Within	# df has two factors (X1,X2) each w/levels (a,b) and continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from lm or lme4::lmer
≥2	≥2	Logistic regression	Z-test	Btwn, Within	# df has two factors (X1,X2) each w/levels (a,b) and dichotomous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from glm or lme4::glmer with family=binomial
≥2	≥2	Multinomial logistic regression	Chi-squared test	Btwn	# df has two between-Ss. factors (X1,X2) each w/levels (a,b) and polytomous response (Y) library(multpois) # for glm.mp.con glm.mp.con(m, pairwise ~ X1*X2, adjust="holm") # m is from multpois::glm.mp
				Within	# df has two within-Ss. factors (X1,X2) each w/levels (a,b) and polytomous response (Y) library(multpois) # for glmer.mp.con glmer.mp.con(m, pairwise ~ X1*X2, adjust="holm") # m is from multpois::glmer.mp
≥2	≥2	Ordinal logistic regression	Z-test	Btwn, Within	# df has two factors (X1,X2) each w/levels (a,b) and ordinal response (1-7) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from MASS::polr or ordinal::clmm

Notes: Between-subjects models are from GLMs, within-subjects models are from GLMMs. Multinomial logistic regression is implemented via the multinomial-Poisson transformation (Baker 1994).

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Linear regression	t-test	Btwn	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,a} was significantly different ($t(56) = -2.75, p = .048$). All other pairwise comparisons were not detectably different.”
				Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that no pairwise comparison was detectably different.”
≥2	≥2	Logistic regression	Z-test	Btwn, Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that no pairwise comparisons were detectably different.”
≥2	≥2	Multinomial logistic regression	Chi-squared test	Btwn, Within	“Six <i>post hoc</i> pairwise comparisons, conducted using the multinomial-Poisson transformation (Baker 1994), and corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that no pairwise comparisons were detectably different.”
≥2	≥2	Ordinal logistic regression	Z-test	Btwn, Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that no individual pairwise comparison was detectably different.”

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Poisson regression	Z-test	Btwn, Within	# df has two factors (X1,X2) each w/levels (a,b) and count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from glm or lme4::glmer with family=poisson
≥2	≥2	Zero-inflated Poisson regression	Z-test	Btwn, Within	# df has two factors (X1,X2) each w/levels (a,b) and zero-inflated count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from glmmTMB::glmmTMB with family=poisson
≥2	≥2	Negative binomial regression	Z-test	Btwn, Within	# df has two factors (X1,X2) each w/levels (a,b) and overdispersed count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from MASS::glm.nb or lme4::glmer.nb
≥2	≥2	Zero-inflated negative binomial regression	Z-test	Btwn, Within	# df has two factors (X1,X2) each w/levels (a,b) and zero-inflated overdispersed count response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from glmmTMB::glmmTMB with family=nbinom2

Notes: Between-subjects models are from GLMs, within-subjects models are from GLMMs.

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Poisson regression	Z-test	Btwn, Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,b} vs. {b,b} was significantly different ($Z = 3.21, p = .008$). All other pairwise comparisons were not detectably different.”
≥2	≥2	Zero-inflated Poisson regression	Z-test	Btwn	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {a,b} ($Z = -3.10, p = .012$) and {a,b} vs. {b,b} ($Z = 3.01, p = .013$) were significantly different. Other pairwise comparisons were not detectably different.”
				Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {a,b} ($Z = -3.10, p = .011$) and {a,b} vs. {b,b} ($Z = 3.01, p = .013$) were significantly different. Other pairwise comparisons were not detectably different.”
≥2	≥2	Negative binomial regression	Z-test	Btwn	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,b} was significantly different ($Z = 2.82, p = .029$). All other pairwise comparisons were not detectably different.”
				Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,b} was significantly different ($Z = 2.83, p = .028$). All other pairwise comparisons were not detectably different.”
≥2	≥2	Zero-inflated negative binomial regression	Z-test	Btwn	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,a} ($Z = 2.99, p = .017$) and {a,a} vs. {b,b} ($Z = 2.66, p = .040$) were significantly different. Other pairwise comparisons were not detectably different.”
				Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {b,a} was significantly different ($Z = 2.82, p = .029$). All other pairwise comparisons were not detectably different.”

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Exponential regression	Z-test	Btwn, Within	<pre># df has two factors (X1,X2) each w/levels (a,b) and exponential response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from glm or lme4::glmer with # family=Gamma(link="log")</pre>
≥2	≥2	Gamma regression	Z-test	Btwn, Within	<pre># df has two factors (X1,X2) each w/levels (a,b) and skewed continuous response (Y) library(emmeans) # for emmeans emmeans(m, pairwise ~ X1*X2, adjust="holm") # m is from glm or lme4::glmer with family=Gamma</pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Exponential regression	Z-test	Btwn	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {a,b} ($Z = -3.97, p = .001$), {a,a} vs. {b,b} ($Z = -4.31, p < .001$), and {b,a} vs. {b,b} ($Z = -2.70, p = .037$) were significantly different. Other pairwise comparisons were not detectably different.”
				Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {a,b} ($Z = -4.66, p < .0001$), {a,a} vs. {b,b} ($Z = -5.06, p < .0001$), {b,a} vs. {a,b} ($Z = -2.86, p = .013$), and {b,a} vs. {b,b} ($Z = -3.27, p = .004$) were significantly different. The other two pairwise comparisons were not detectably different.”
≥2	≥2	Gamma regression	Z-test	Btwn	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {a,b} ($Z = -4.77, p < .001$), {a,a} vs. {b,b} ($Z = -2.65, p = .042$), {b,a} vs. {a,b} ($Z = -3.87, p = .001$), and {a,b} vs. {b,b} ($Z = 2.52, p = .044$) were significantly different. The other two pairwise comparisons were not detectably different.”
				Within	“Six <i>post hoc</i> pairwise comparisons, corrected with Holm’s sequential Bonferroni procedure (Holm 1979), indicated that {a,a} vs. {a,b} ($Z = -5.13, p < .0001$), {a,a} vs. {b,b} ($Z = -2.85, p = .018$), {b,a} vs. {a,b} ($Z = -4.16, p < .001$), and {a,b} vs. {b,b} ($Z = 2.71, p = .020$) were significantly different. The other two pairwise comparisons were not detectably different.”

Bibliography

Generalized linear (mixed) models

- Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88 (421), pp. 9-25. <https://doi.org/10.2307/2290687>
- Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* 135 (3), pp. 370-384. <https://doi.org/10.2307/2344614>
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54 (3), pp. 426-482. <https://doi.org/10.2307/1990256>

Lognormal models

- Aitchison, J. and Brown, J.A.C. (1957). *The Lognormal Distribution*. Cambridge, UK: Cambridge University Press. <http://www.jstor.org/stable/2227716>
- Berry, D.A. (1987). Logarithmic transformations in ANOVA. *Biometrics* 43 (2), pp. 439-456. <https://doi.org/10.2307/2531826>
- Lawrence, R.J. (1988). The log-normal as event-time distribution. In *Log-normal Distributions: Theory and Application*, E.L. Crow and K. Shimizu (eds.). New York, NY: Dekker, pp. 211-228.
- Limpert, E., Stahel, W.A. and Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience* 51 (5), pp. 341-352. [https://doi.org/10.1641/0006-3568\(2001\)051\[0341:LNDATS\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2)

Binomial models

- Berkson, J. (1944). Application of the logistic function to bio-assay. *Journal of the American Statistical Association* 39 (227), pp. 357-365. <https://doi.org/10.2307/2280041>
- Gilmour, A.R., Anderson, R.D. and Rae, A.L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72 (3), pp. 593-599. <https://doi.org/10.2307/2336731>
- Stiratelli, R., Laird, N. and Ware, J.H. (1984). Random-effects models for serial observations with binary response. *Biometrics* 40 (4), pp. 961-971. <https://doi.org/10.2307/2531147>

Bibliography

Multinomial models

- Begg, C.B. and Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika* 71 (1), pp. 11-18. <https://www.jstor.org/stable/2336391>
- Cox, D.R. (1966). Some procedures connected with the logistic qualitative response curve. In *Research Papers in Probability and Statistics (Festschrift for J. Neyman)*, F.N. David (ed.). London, UK: John Wiley & Sons, pp. 55-71.
- Daniels, M.J. and Gatsonis, C. (1997). Hierarchical polytomous regression models with applications to health services research. *Statistics in Medicine* 16 (20), pp. 2311-2325. [https://doi.org/10.1002/\(SICI\)1097-0258\(19971030\)16:20%3C2311::AID-SIM654%3E3.0.CO;2-E](https://doi.org/10.1002/(SICI)1097-0258(19971030)16:20%3C2311::AID-SIM654%3E3.0.CO;2-E)
- Hartzel, J., Agresti, A. and Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling* 1 (2), pp. 81-102. <https://doi.org/10.1177/1471082X0100100201>
- Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine* 22 (9), pp. 1433-1446. <https://doi.org/10.1002/sim.1522>
- Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review* 10 (3), pp. 251-259. <https://doi.org/10.2307/2525642>

Bibliography

Multinomial-Poisson trick

- Baker, S.G. (1994). The multinomial-Poisson transformation. *The Statistician* 43 (4), pp. 495-504. <https://doi.org/10.2307/2348134>
- Chen, Z. and Kuo, L. (2001). A note on the estimation of the multinomial logit model with random effects. *The American Statistician* 55 (2), pp. 89-95. <https://www.jstor.org/stable/2685993>
- Guimaraes, P. (2004). Understanding the multinomial-Poisson transformation. *The Stata Journal* 4 (3), pp. 265-273. <https://www.stata-journal.com/article.html?article=st0069>
- Lee, J.Y.L., Green, P.J. and Ryan, L.M. (2017). On the “Poisson trick” and its extensions for fitting multinomial regression models. arXiv:1707.08538v1 [stat.ME], 25 pages. <https://arxiv.org/abs/1707.08538>

Ordinal models

- Agresti, A. (2010). *Analysis of Ordinal Categorical Data (2nd Ed)*. Hoboken, NJ : John Wiley & Sons, Inc. <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470594001>
- Hedeker, D. and Gibbons, R.D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* 50 (4), pp. 933-944. <https://www.jstor.org/stable/2533433>
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B* 42 (2), pp. 109-142. <https://www.jstor.org/stable/2984952>
- McKelvey, R.D. and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* 4 (1), pp. 103-120. <https://doi.org/10.1080/0022250X.1975.9989847>
- Winship, C. and Mare, R.D. (1984). Regression models with ordinal variables. *American Sociological Review* 49 (4), pp. 512-525. <https://doi.org/10.2307/2095465>

Bibliography

Count models – Poisson and negative binomial

- Cox, D.R. (1983). Some remarks on overdispersion. *Biometrika* 70 (1), pp. 269-274. <https://doi.org/10.2307/2335966>
- Gardner, W., Mulvey, E.P. and Shaw, E.C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin* 118 (3), pp. 392-404. <https://doi.org/10.1037/0033-2909.118.3.392>
- Hilbe, J.M. (2011). *Negative Binomial Regression*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9780511973420>
- Hilbe, J.M. (2014). *Modeling Count Data*. New York, NY: Cambridge University Press. <https://books.google.com/books?id=aZLfAwAAQBAJ>
- Lawless, J.F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* 15 (3), pp. 209-225. <https://doi.org/10.2307/3314912>
- Ver Hoef, J.M. and Boveng, P.L. (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology* 88 (11), pp. 2766-2772. <https://doi.org/10.1890/07-0043.1>
- Vermunt, J.K. (1997). *Log-linear Models for Event Histories*. Thousand Oaks, CA: Sage Publications. Retrieved from https://books.google.com/books?id=apI_AQAAIAAJ
- von Bortkiewicz, L. (1898). *Das Gesetz der Kleinen Zahlen (The Law of Small Numbers)*. Leipzig, Germany: Druck und Verlag von B.G. Teubner. https://books.google.com/books?id=o_k3AAAAMAAJ

Bibliography

Zero-inflated count models – Poisson and negative binomial

- Hall, D.B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics* 56 (4), pp. 1030-1039. <https://doi.org/10.1111/j.0006-341X.2000.01030.x>
- Heilbron, D.C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal* 36 (5), pp. 531-547. <https://doi.org/10.1002/bimj.4710360505>
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34 (1), pp. 1-14. <https://doi.org/10.2307/1269547>

Gamma models

- Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A* 135 (3), pp. 370-384. <https://doi.org/10.2307/2344614>
- Ng, V.K.Y. and Cribbie, R.A. (2017). Using the gamma generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. *Current Psychology* 36 (2), pp. 225-235. <https://doi.org/10.1007/s12144-015-9404-0>
- Stroup, W.W., Ptukhina, M. and Garai, J. (2024). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Boca Raton, FL: CRC Press, pp. 453-472. <https://doi.org/10.1201/9780429092060>

Bibliography

Post hoc comparisons

- Boik, R.J. (1979). Interactions, partial interactions, and interaction contrasts in the analysis of variance. *Psychological Bulletin* 86 (5), pp. 1084-1089. <https://doi.org/10.1037/0033-2909.86.5.1084>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6 (2), pp. 65-70. <http://www.jstor.org/stable/4615733>
- Kenward, M.G. and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53 (3), pp. 983-997. <https://doi.org/10.2307/2533558>
- Kramer, C.Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics* 12 (3), pp. 307-310. <https://doi.org/10.2307/3001469>
- Marascuilo, L.A. and Levin, J.R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. *American Educational Research Journal* 7 (3), pp. 397-421. <https://doi.org/10.3102/00028312007003397>
- Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2 (6), pp. 110-114. <https://doi.org/10.2307/3002019>
- Tukey, J.W. (1949). Comparing individual means in the analysis of variance. *Biometrics* 5 (2), pp. 99-114. <https://doi.org/10.2307/3001913>
- Tukey, J.W. (1953). *The Problem of Multiple Comparisons*. Princeton, NJ: Princeton University.