

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

Assumptions

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

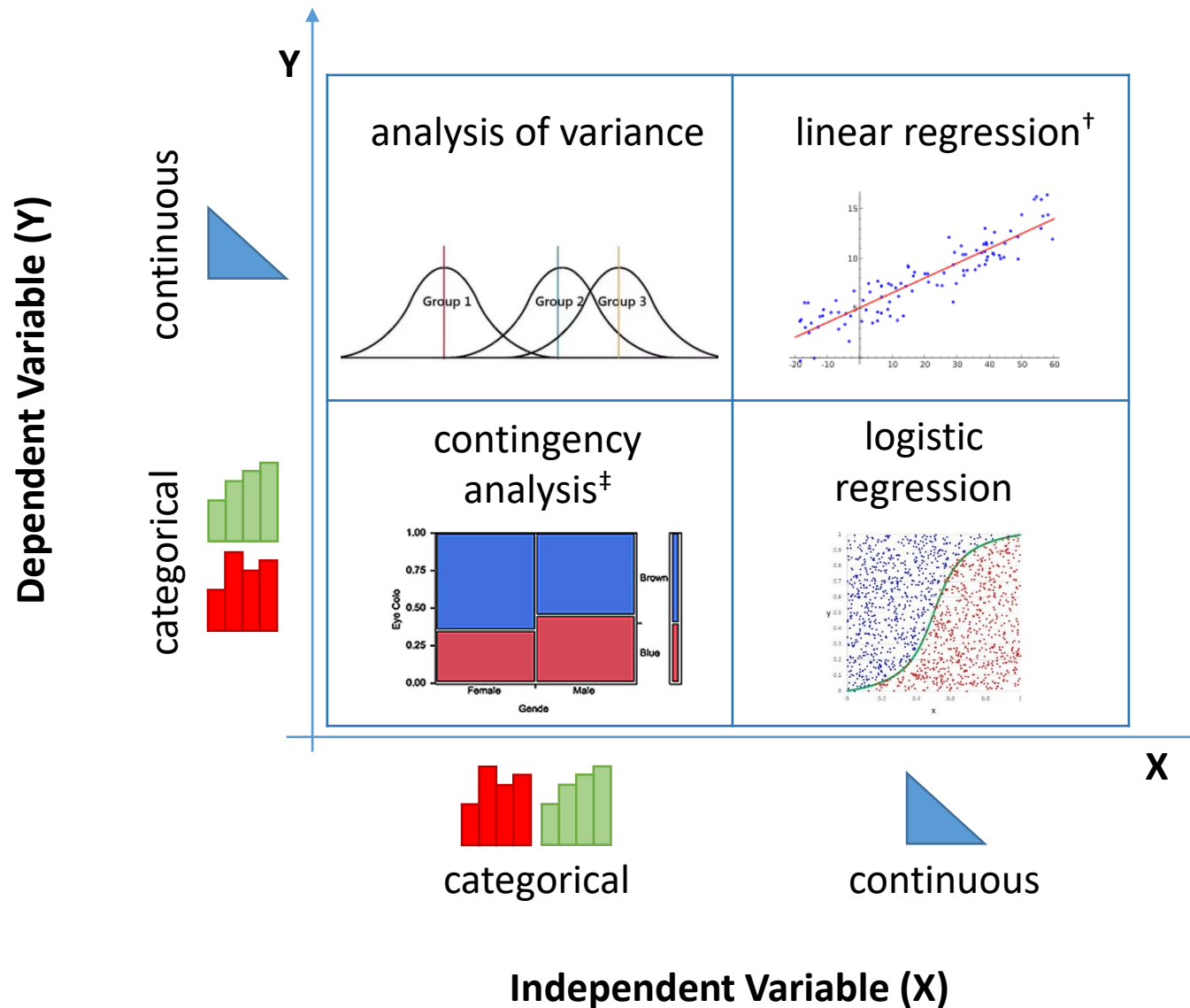
Distributions

- Kolmogorov-Smirnov test
- χ^2 goodness-of-fit test

Analyses of Variance

Factors	Levels	<u>B</u> etween or <u>W</u> ithin	Parametric Tests & Linear (Mixed) Models	Nonparametric Tests & Generalized Linear (Mixed) Models
1	2	B	Independent-samples <i>t</i> -test	Mann-Whitney <i>U</i> test
1	2	W	Paired-samples <i>t</i> -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

./data/Proportions/

Proportion & Association

One 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 subject (S) and 2-category outcome (Y) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$Y = factor(df\$Y) # Y is an outcome of 2 categories xt = xtabs(~ Y, data=df) # make counts binom.test(xt, p=0.5, alternative="two.sided")</pre>
1	≥ 2	Multinomial test	Yes, use with $N \leq 200$	<pre># df is a long-format data table w/columns for subject (S) and N-category outcome (Y) library(XNomial) # for xmulti df\$S = factor(df\$S) # Subject id is nominal (unused) df\$Y = factor(df\$Y) # Y is an outcome of ≥ 2 categories xt = xtabs(~ Y, data=df) # make counts xmulti(xt, rep(1/length(xt), length(xt)), statName="Prob") # the following gives the same result library(RVAideMemoire) # for multinomial.test multinomial.test(df\$Y)</pre>
		One-sample Pearson chi-squared test	No, use with $N > 200$	<pre># df is a long-format data table w/columns for subject (S) and N-category outcome (Y) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$Y = factor(df\$Y) # Y is an outcome of ≥ 2 categories xt = xtabs(~ Y, data=df) # make counts chisq.test(xt)</pre>

Proportion & Association

One sample

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

Proportion & Association

Two 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/subject (S), categorical factor (X) and outcome (Y) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$X = factor(df\$X) # X is a factor of m ≥ 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/subject (S), categorical factor (X) and outcome (Y) library(RVAideMemoire) # for G.test df\$S = factor(df\$S) # Subject id is nominal (unused) df\$X = factor(df\$X) # X is a factor of m ≥ 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/subject (S), categorical factor (X) and outcome (Y) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$X = factor(df\$X) # X is a factor of m ≥ 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 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 the 'x', 'y', and 'z' outcomes for each of 'a' and 'b'. Fisher's exact test indicated a statistically significant association between X and Y ($p < .0001$)."
		G-test	No, use with $N > 200$	"Table 1 shows the counts of the 'x', 'y', and 'z' outcomes for each of 'a' and 'b'. A G-test indicated a statistically significant association between X and Y ($G(2) = 21.40, p < .0001$)."
		Two-sample Pearson chi-squared test	No, use with $N > 200$	"Table 1 shows the counts of the 'x', 'y', and 'z' outcomes for each of 'a' and 'b'. A two-sample Pearson Chi-Squared test indicated a statistically significant association between X and Y ($\chi^2(2, N=60) = 19.88, p < .0001$)."

Table 1

		Y		
		x	y	z
X	a	3	26	1
	b	14	9	7

1F2LBs_multinomial.csv

Proportion & Association

Post hoc comparisons

Proportion & Association

Post hoc tests – One sample

Samples	Response Categories	Omnibus Test	Contrast Test	R Code
1	≥ 2	Multinomial test	Pairwise binomial tests	<pre># For Y's response categories (x,y,z), test each in pairwise fashion against chance. xy = binom.test(c(sum(df\$Y == "x"), sum(df\$Y == "y")), p=1/2) # proportion of "x" vs. "y" rows xz = binom.test(c(sum(df\$Y == "x"), sum(df\$Y == "z")), p=1/2) # proportion of "x" vs. "z" rows yz = binom.test(c(sum(df\$Y == "y"), sum(df\$Y == "z")), p=1/2) # proportion of "y" vs. "z" rows p.adjust(c(xy\$p.value, xz\$p.value, yz\$p.value), method="holm") # or, equivalently, if xt is a table of counts for each category of response 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.</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 (x,y,z), # test each proportion against chance. x = binom.test(sum(df\$Y == "x"), nrow(df), p=1/3) # proportion of "x" rows y = binom.test(sum(df\$Y == "y"), nrow(df), p=1/3) # proportion of "y" rows z = binom.test(sum(df\$Y == "z"), nrow(df), p=1/3) # proportion of "z" rows p.adjust(c(x\$p.value, y\$p.value, z\$p.value), method="holm")</pre>

Proportion & Association

Post hoc tests – One sample

Samples	Response Categories	Omnibus Test	Contrast Test	Report
1	≥2	Multinomial test	Pairwise binomial tests	“Pairwise comparisons using exact binomial tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the proportions of ‘x’ and ‘z’ and ‘y’ and ‘z’ were statistically significantly different ($p < .05$), but that the proportions of ‘x’ and ‘y’ were not.”
1	≥2	One-sample Pearson chi-squared test	Pairwise chi-squared tests	“Pairwise comparisons using Pearson Chi-Squared tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the proportions of ‘x’ and ‘z’ were statistically significantly different ($\chi^2(1, N=52) = 5.02, p < .05$), as were the proportions of ‘y’ and ‘z’ ($\chi^2(1, N=43) = 14.96, p < .001$), but that the proportions of ‘x’ and ‘y’ were only marginal ($\chi^2(1, N=25) = 3.24, p = .072$).”
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, indicated that the proportions of ‘y’ and ‘z’ were statistically significantly different from chance ($p < .05$), but not the proportion of ‘x’. Specifically, the proportion of ‘y’ at 8 was significantly lower than 20, and the proportion of ‘z’ at 35 was significantly higher than 20.”

Proportion & Association

Post hoc tests – Two 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.</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 (x,y,z), test each proportion of Y within each level of X against # chance. # df is a long-format data table w/subject (S), categorical factor (X) and outcome (Y). ax = binom.test(sum(df[df\$X == "a",]\$Y == "x"), nrow(df[df\$X == "a",]), p=1/3) ay = binom.test(sum(df[df\$X == "a",]\$Y == "y"), nrow(df[df\$X == "a",]), p=1/3) az = binom.test(sum(df[df\$X == "a",]\$Y == "z"), nrow(df[df\$X == "a",]), p=1/3) p.adjust(c(ax\$p.value, ay\$p.value, az\$p.value), method="holm") bx = binom.test(sum(df[df\$X == "b",]\$Y == "x"), nrow(df[df\$X == "b",]), p=1/3) by = binom.test(sum(df[df\$X == "b",]\$Y == "y"), nrow(df[df\$X == "b",]), p=1/3) bz = binom.test(sum(df[df\$X == "b",]\$Y == "z"), nrow(df[df\$X == "b",]), p=1/3) p.adjust(c(bx\$p.value, by\$p.value, bz\$p.value), method="holm")</pre>

Proportion & Association

Post hoc tests – Two 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, one for each 2×2 crosstabulation of X and Y, corrected with Holm's sequential Bonferroni procedure, indicated that there were statistically significant associations for (a,b) × (x,y) ($p < .001$) and (a,b) × (y,z) ($p < .01$), but not for (a,b) × (x,z)."
2	≥2	G-test	Pairwise G-tests	"Pairwise comparisons using G-tests, corrected with Holm's sequential Bonferroni procedure, indicated that (a,x) vs. (a,y), (a,z) vs. (b,x), (a,z) vs. (a,y), (b,y) vs. (a,y), and (b,z) vs. (a,y) were all statistically significantly different ($p < .05$). Furthermore, (a,x) vs. (b,x) ($p = .055$) was marginal, as was (a,z) vs. (b,y) ($p = .060$). The eight other possible pairwise comparisons were not statistically significant."
2	≥2	Two-sample Pearson chi-squared test	Pairwise chi-squared tests	"Pairwise comparisons using Pearson Chi-Square tests, corrected with Holm's sequential Bonferroni procedure, indicated that (a,x) vs. (a,y), (a,z) vs. (b,x), (a,z) vs. (a,y), (b,y) vs. (a,y), (b,z) vs. (a,y) were all statistically significantly different ($p < .05$). Furthermore, (a,x) vs. (b,x) ($p = .076$) was marginal. The nine other possible 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, indicated that, within 'a', the proportions of 'x', 'y', and 'z' were all significantly different from chance (i.e., 10 each) ($p < .01$). Three similar tests within 'b' indicated that none of the proportions of 'x', 'y', or 'z' significantly differed from chance (i.e., 10 each)."

Assumptions

./data/Assumptions/

ANOVA Assumptions

Normality test of response in each condition

Assumption	Test Name	Context of Use	R Code
Normality	Shapiro-Wilk test (on the response in each condition)	t-test, ANOVA	<pre># assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y shapiro.test(df[df\$X1 == "a" & df\$X2 == "a",]\$Y) # condition a,a shapiro.test(df[df\$X1 == "a" & df\$X2 == "b",]\$Y) # condition a,b shapiro.test(df[df\$X1 == "b" & df\$X2 == "a",]\$Y) # condition b,a shapiro.test(df[df\$X1 == "b" & df\$X2 == "b",]\$Y) # condition b,b</pre>
Normality	Anderson-Darling test (on the response in each condition)	t-test, ANOVA	<pre># assume df has two factors (X1,X2) each w/two levels (a,b) and continuous response Y library(nortest) # for ad.test ad.test(df[df\$X1 == "a" & df\$X2 == "a",]\$Y) # condition a,a ad.test(df[df\$X1 == "a" & df\$X2 == "b",]\$Y) # condition a,b ad.test(df[df\$X1 == "b" & df\$X2 == "a",]\$Y) # condition b,a ad.test(df[df\$X1 == "b" & df\$X2 == "b",]\$Y) # condition b,b</pre>

ANOVA Assumptions

Normality test of response in each condition

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 assumption of conditional normality, a Shapiro-Wilk test was run on the response Y for each combination of levels of factors X1 and X2. All combinations were found to be statistically non-significant except condition (b,b), which showed a statistically significant deviation from normality ($W = .794, p < .01$).”
Normality	Anderson-Darling test (on the response in each condition)	<i>t</i> -test, ANOVA	“To test the assumption of conditional normality, an Anderson-Darling test was run on the response Y for each combination of levels of factors X1 and X2. All combinations were found to be statistically non-significant except condition (b,b), which showed a statistically significant deviation from normality ($A = 1.417, p < .001$).”

ANOVA Assumptions

Normality test of model residuals

Assumption	Test Names	Context of Use	R Code
<i>Extract residuals</i>		Between-subjects models (independent-samples t-test, ANOVA, LM)	<pre># df has two between-Ss factors (X1,X2) each w/two levels (a,b) and continuous response (Y) m = aov(Y ~ X1*X2, data=df) # make anova model r = residuals(m) ## ... continue to common code, below ##</pre>
<i>Extract residuals</i>		Within-subjects models (paired-samples t-test, repeated measures ANOVA)	<pre># df has two within-Ss factors (X1,X2) each w/two levels (a,b) and continuous response (Y) library(afex) m = aov_ez(dv="Y", within=c("X1","X2"), id="S", type=3, data=df) # make rm-anova model r = residuals(m\$lm) ## ... continue to common code, below ##</pre>
<i>Extract residuals</i>		Linear mixed models (LMM) with between- or within- subjects factors	<pre># df has two factors (X1,X2) each w/two levels (a,b) and continuous response (Y) library(lme4) library(lmerTest) m = lmer(Y ~ X1*X2 + (1 S), data=df) # make linear mixed model r = residuals(m) ## ... continue to common code, below ##</pre>
Normality	Shairo-Wilk test, Anderson-Darling test (on residuals)	<i>Common code</i>	<pre>## common code ## mean(r); sum(r) # both should be ~0 plot(r[1:length(r)]); 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 (we want p>.05) library(nortest) ad.test(r) # Anderson-Darling test (we want p>.05)</pre>

ANOVA Assumptions

Normality test of model residuals

Assumption	Test Name	Context of Use	Report
Normality	Shapiro-Wilk test (on residuals)	Independent-samples <i>t</i> -test, ANOVA, LM	“To test the normality assumption, a Shapiro-Wilk test was run on the residuals of a between-subjects full-factorial ANOVA model. The test was statistically non-significant ($W = .988, p = .798$), indicating compliance with the normality assumption. A plot of residuals, histogram of residuals, and Q-Q plot all visually confirm the same (Figure 1a).”
Normality	Anderson-Darling test (on residuals)	Independent-samples <i>t</i> -test, ANOVA, LM	“To test the normality assumption, an Anderson-Darling test was run on the residuals of a between-subjects full-factorial ANOVA model. The test was statistically non-significant ($A = 0.329, p = .510$), indicating compliance with the normality assumption. A plot of residuals, histogram of residuals, and Q-Q plot all visually confirm the same (Figure 1a).”
Normality	Shapiro-Wilk test (on residuals)	Paired-samples <i>t</i> -test, repeated measures ANOVA, LMM	“To test the normality assumption, a Shapiro-Wilk test was run on the residuals of a within-subjects full-factorial ANOVA model. The test was statistically non-significant ($W = .988, p = .798$), indicating compliance with the normality assumption. A plot of residuals, histogram of residuals, and Q-Q plot all visually confirm the same (Figure 1b).”
Normality	Anderson-Darling test (on residuals)	Paired-samples <i>t</i> -test, repeated measures ANOVA, LMM	“To test the normality assumption, an Anderson-Darling test was run on the residuals of a within-subjects full-factorial ANOVA model. The test was statistically non-significant ($A = 0.329, p = .510$), indicating compliance with the normality assumption. A plot of residuals, histogram of residuals, and Q-Q plot all visually confirm the same (Figure 1b).”

Figure 1a

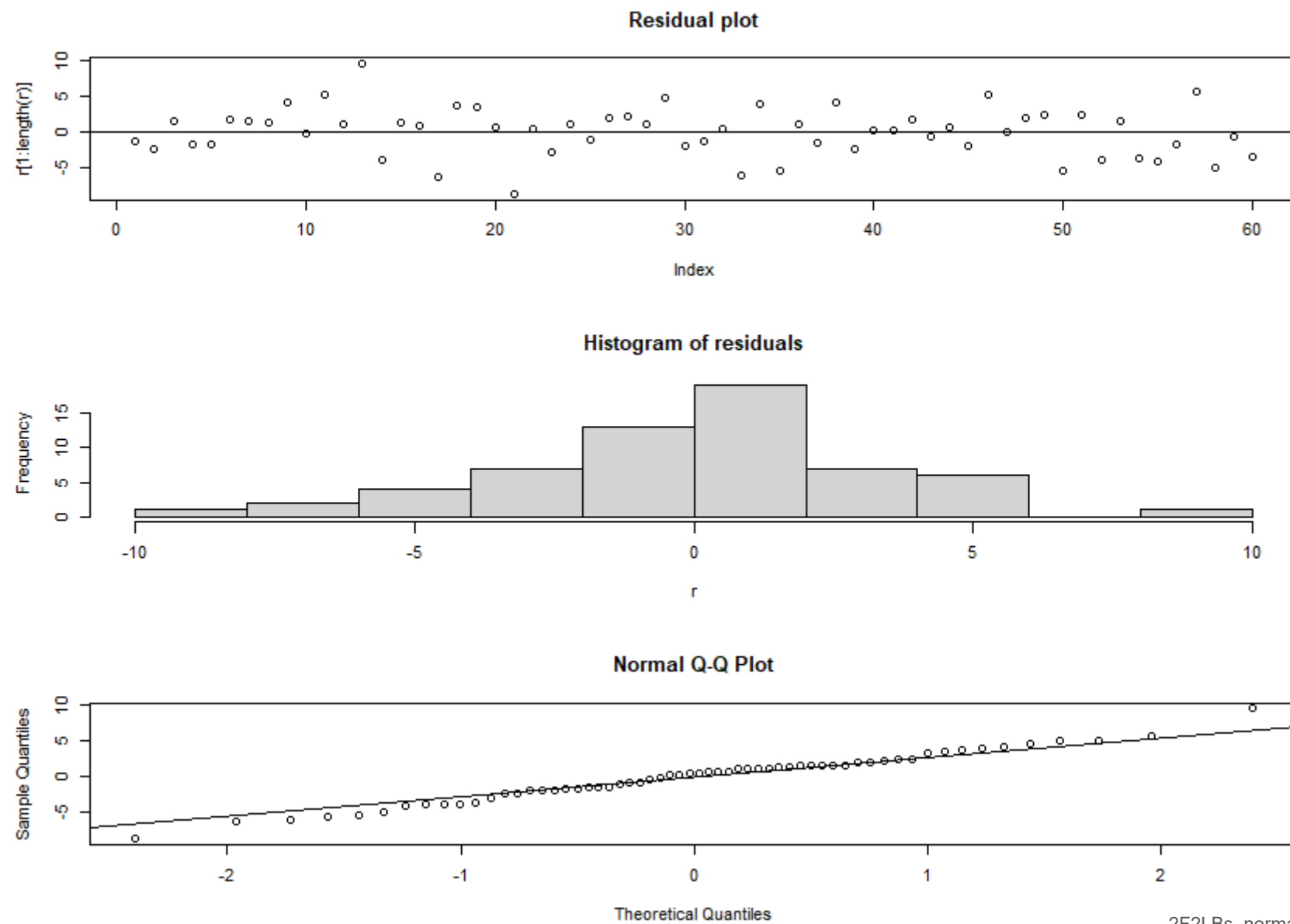
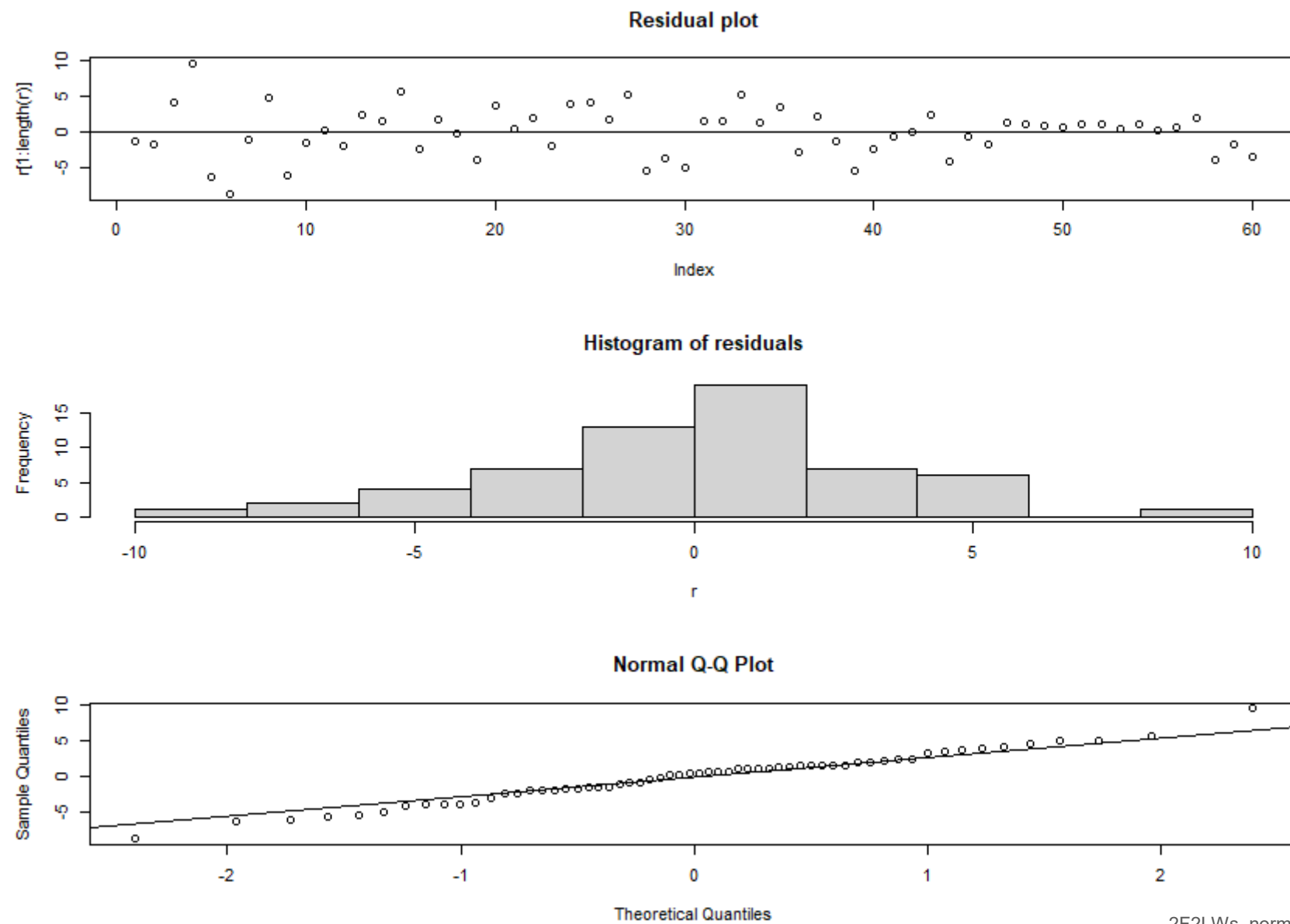


Figure 1b



ANOVA Assumptions

Homoscedasticity and sphericity

Assumption	Test Name	Context of Use	R Code
Homoscedasticity, a.k.a. homogeneity of variance	Levene's test	Any ANOVA model with at least one between-subjects factor	<pre> # df has subjects (S), one between-Ss factor (X1), one within-Ss factor (X2), and # continuous response (Y) library(afex) # for aov_ez library(performance) # for check_homogeneity m = aov_ez(dv="Y", between="X1", within="X2", id="S", 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... t.test(Y ~ X1, data=df, var.equal=FALSE) # ...or a White-adjusted ANOVA for >1 factor library(car) # for Anova Anova(m\$lm, type=3, white.adjust=TRUE) # shows between-Ss factors only </pre>
Sphericity	Mauchly's test of sphericity	Repeated measures ANOVA	<pre> # df has subjects (S), one between-Ss factor (X1), one within-Ss factor (X2), and # continuous response (Y) library(afex) # for aov_ez m = aov_ez(dv="Y", between="X1", within="X2", id="S", type=3, data=df) s = summary(m) s\$sphericity.tests # Mauchly's sphericity tests s\$univariate.tests # if p>.05, use these unadjusted DFs and p-values anova(m) # if p<.05, use these G-G adjusted DFs and p-values </pre>

ANOVA Assumptions

Homoscedasticity and sphericity

Assumption	Test Name	Context of Use	Report
Homoscedasticity, a.k.a. homogeneity of variance	Levene's test	Any ANOVA model with at least one between-subjects factor	“To test the homoscedasticity assumption, Levene's test was run on a mixed factorial ANOVA model with a between-subjects factor X1 and a within-subjects factor X2. The test was statistically non-significant ($p = .802$), indicating no homoscedasticity violation.”
Sphericity	Mauchly's test of sphericity	Repeated measures ANOVA	“To test the sphericity assumption, Mauchly's test of sphericity was run on a mixed factorial ANOVA model with a between-subjects factor X1 and a within-subjects factor X2. The test was statistically significant for both X2 ($W = .637$, $p < .01$) and X1×X2 ($W = .637$, $p < .01$), indicating sphericity violations. Accordingly, the Greenhouse-Geisser correction was used when reporting ANOVA results.”

Distributions

./data/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># assume df has one factor (X) w/two levels (a,b) and continuous response Y library(MASS) # for fitdistr fa = fitdistr(df[df\$X == "a"],\$Y, "normal")\$estimate # create fit for X.a ks.test(df[df\$X == "a"],\$Y, "pnorm", mean=fa[1], sd=fa[2]) fb = fitdistr(df[df\$X == "b"],\$Y, "normal")\$estimate # create fit for X.b ks.test(df[df\$X == "b"],\$Y, "pnorm", mean=fb[1], sd=fb[2])</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># assume df has one factor (X) w/two levels (a,b) and continuous response Y library(MASS) # for fitdistr fa = fitdistr(df[df\$X == "a"],\$Y, "lognormal")\$estimate # create fit for X.a ks.test(df[df\$X == "a"],\$Y, "plnorm", meanlog=fa[1], sdlog=fa[2]) fb = fitdistr(df[df\$X == "b"],\$Y, "lognormal")\$estimate # create fit for X.b ks.test(df[df\$X == "b"],\$Y, "plnorm", meanlog=fb[1], sdlog=fb[2])</pre>
Poisson	lambda (λ): lambda	<code>_pois:</code> <code>dpois</code> <code>ppois</code> <code>qpois</code> <code>rpois</code>	<pre># assume df has one factor (X) w/two levels (a,b) and nonnegative integer response Y library(MASS) # for fitdistr fa = fitdistr(df[df\$X == "a"],\$Y, "Poisson")\$estimate # create fit for X.a ks.test(df[df\$X == "a"],\$Y, "ppois", lambda=fa[1]) fb = fitdistr(df[df\$X == "b"],\$Y, "Poisson")\$estimate # create fit for X.b ks.test(df[df\$X == "b"],\$Y, "ppois", lambda=fb[1]) # if var/mean > 1.15, we have overdispersion; if so, use quasipoisson or a nbinom GLM var(df[df\$X == "a"],\$Y) / abs(ean(df[df\$X == "a"],\$Y)) > 1.15 var(df[df\$X == "b"],\$Y) / abs(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># assume df has one factor (X) w/two levels (a,b) and nonnegative integer response Y library(MASS) # for fitdistr fa = fitdistr(df[df\$X == "a"],\$Y, "negative binomial", lower=1e-6)\$estimate # fit for X.a ks.test(df[df\$X == "a"],\$Y, "pnbinom", size=fa[1], mu=fa[2]) fb = fitdistr(df[df\$X == "b"],\$Y, "negative binomial", lower=1e-6)\$estimate # fit for X.b ks.test(df[df\$X == "b"],\$Y, "pnbinom", size=fb[1], mu=fb[2])</pre>

Distributions

Distribution	Parameterization	R Distribution Funcs	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 2 shows the distributions of response Y for both levels of factor X. To test whether these distributions were normally distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .158, p = .404$), as was the test for level ‘b’ ($D = .104, p = .867$), indicating non-detectable deviations from a normal distribution for both levels.”
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 3 shows the distributions of response Y for both levels of factor X. To test whether these distributions were lognormally distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .096, p = .918$), as was the test for level ‘b’ ($D = .161, p = .375$), indicating non-detectable deviations from a lognormal distribution for both levels.”
Poisson	lambda (λ): <code>lambda</code>	<code>_pois:</code> <code>dpois</code> <code>ppois</code> <code>qpois</code> <code>rpois</code>	“Figure 4 shows the distributions of response Y for both levels of factor X. To test whether these distributions were Poisson distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .247, p = .052$), as was the test for level ‘b’ ($D = .235, p = .074$), indicating non-detectable deviations from a Poisson distribution.”
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 5 shows the distributions of response Y for both levels of factor X. To test whether these distributions were negative binomially distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .110, p = .864$), as was the test for level ‘b’ ($D = .107, p = .882$), indicating non-detectable deviations from a negative binomial distribution for both levels.”

Figure 2

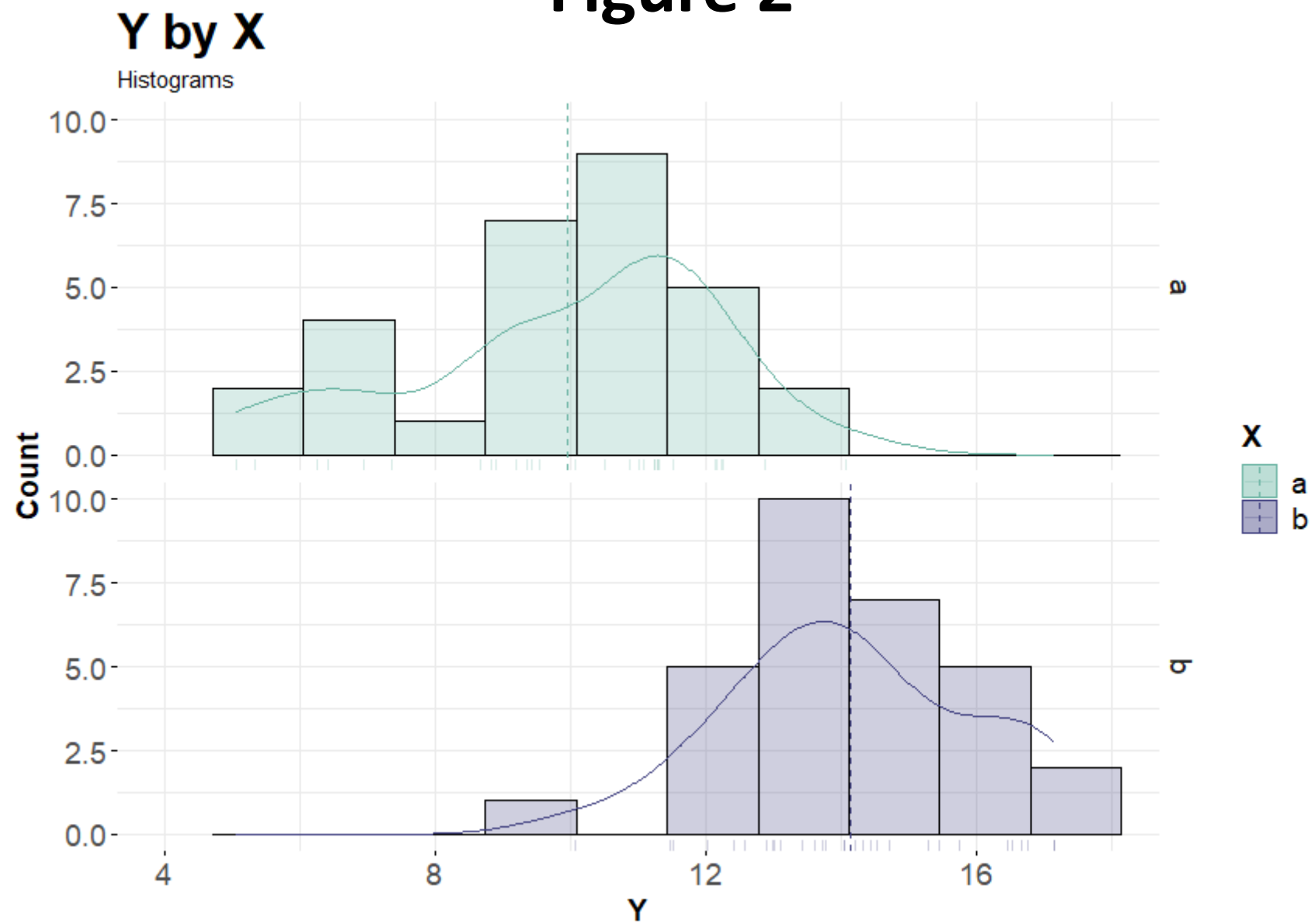


Figure 3

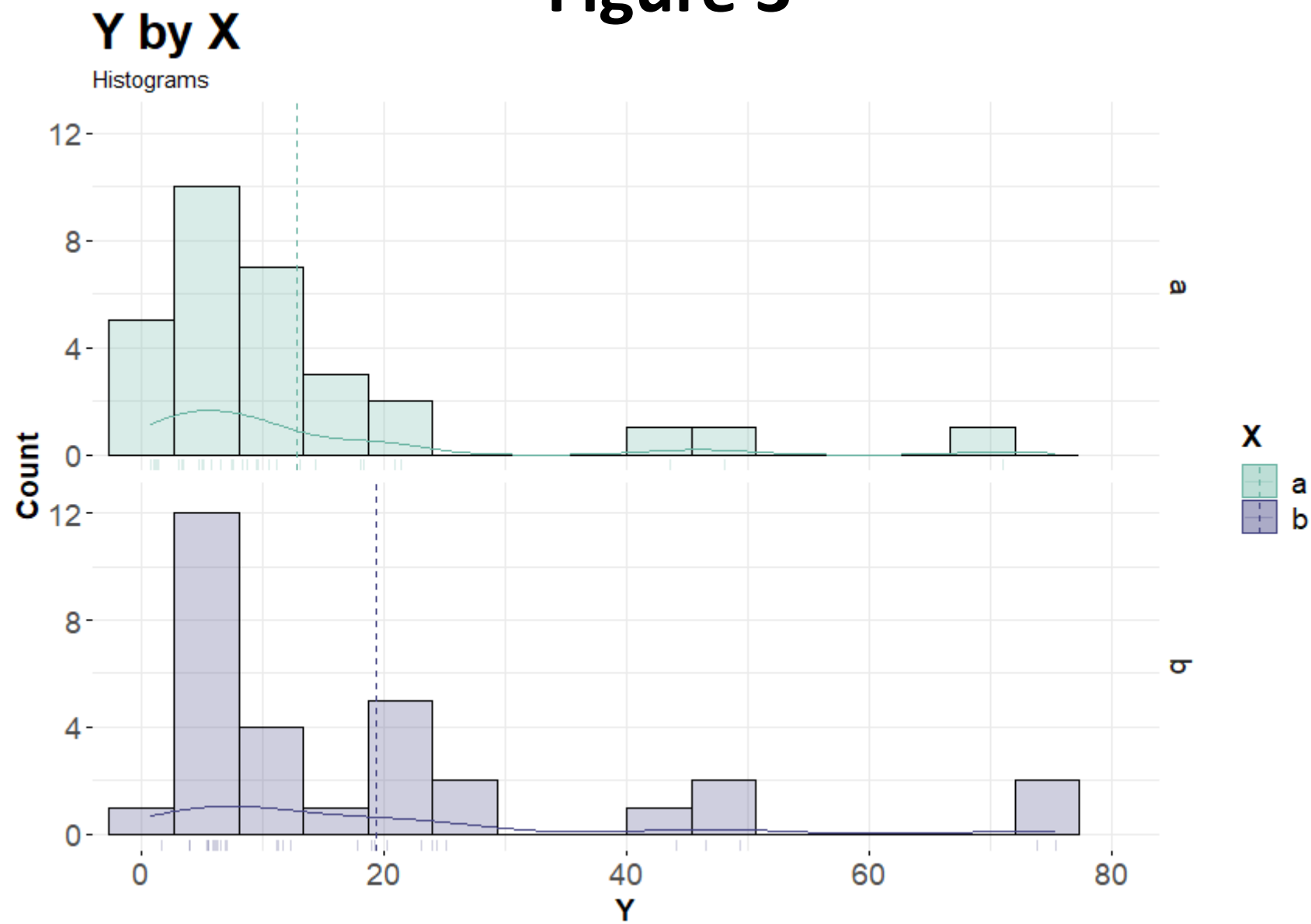


Figure 4

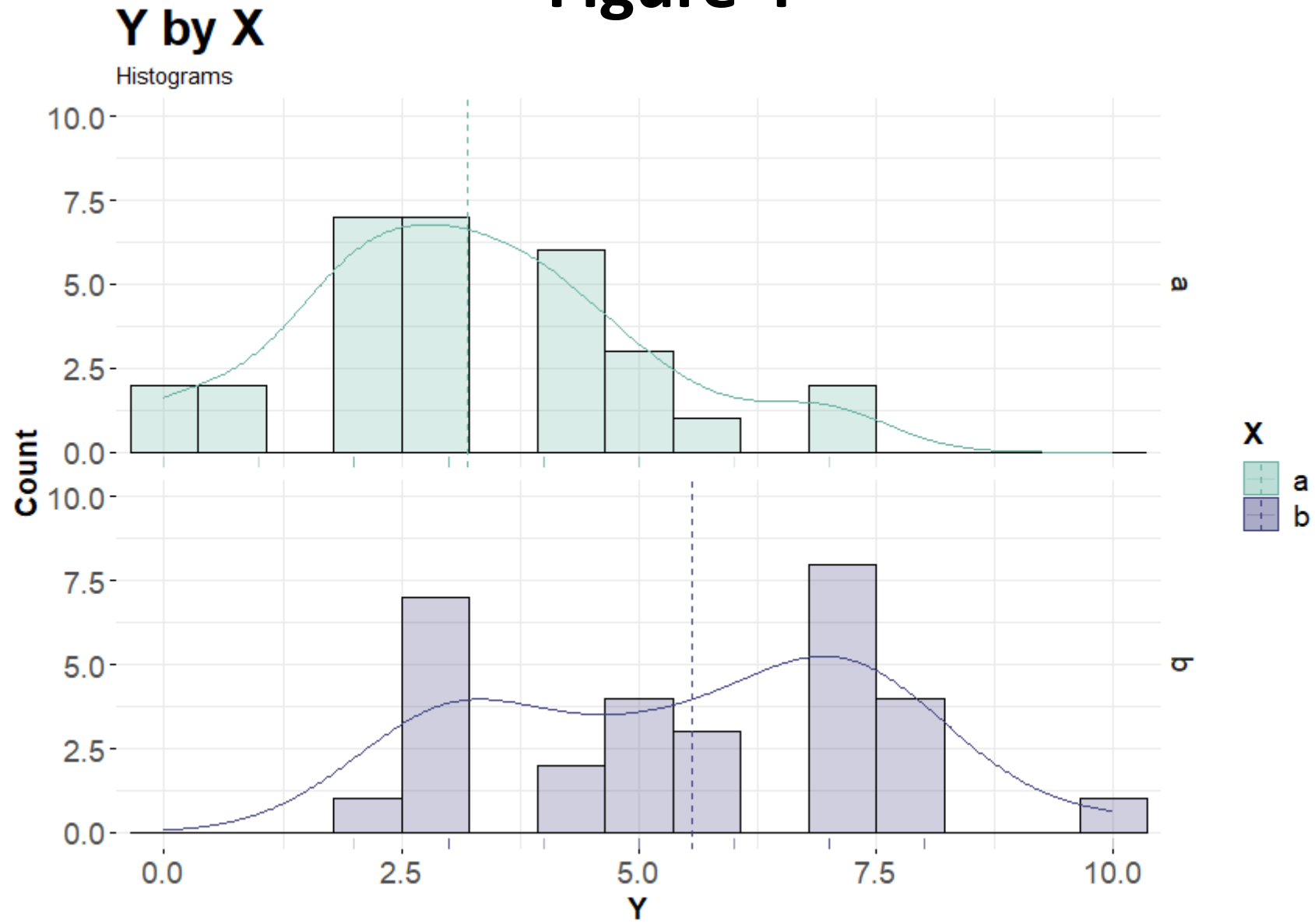
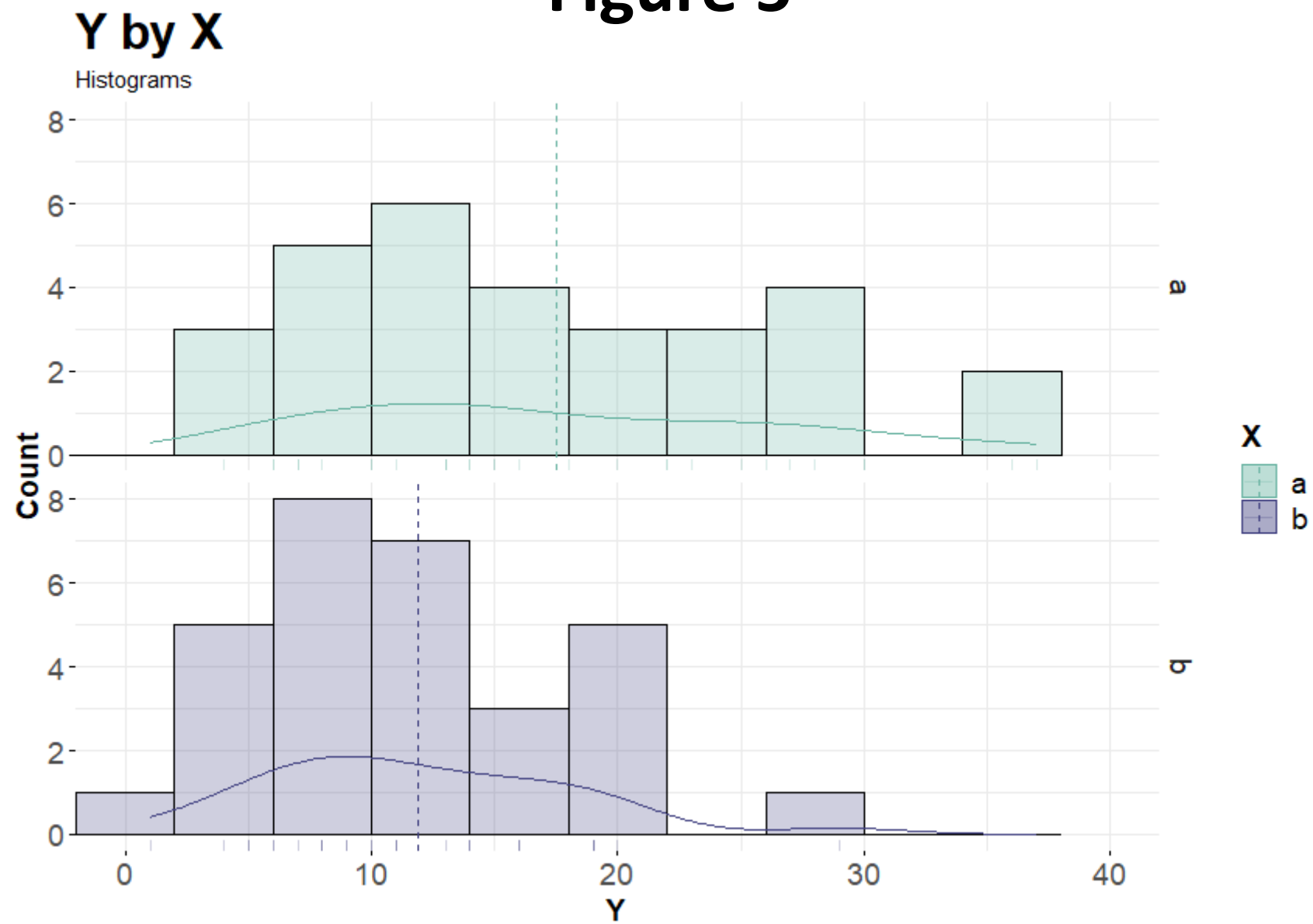


Figure 5



Distribution Tests

Distribution	Parameterization	R Distribution Fns	R Code
Exponential	rate (λ): rate	<code>_exp:</code> <code>dexp</code> <code>pexp</code> <code>qexp</code> <code>rexp</code>	<pre># assume df has one factor (X) w/two levels (a,b) and continuous response Y library(MASS) # for fitdistr fa = fitdistr(df[df\$X == "a",]\$Y, "exponential")\$estimate # create fit for X.a ks.test(df[df\$X == "a",]\$Y, "pexp", rate=fa[1]) fb = fitdistr(df[df\$X == "b",]\$Y, "exponential")\$estimate # create fit for X.b ks.test(df[df\$X == "b",]\$Y, "pexp", rate=fb[1])</pre>
Gamma	shape (α): shape rate (β): rate	<code>_gamma:</code> <code>dgamma</code> <code>pgamma</code> <code>qgamma</code> <code>rgamma</code>	<pre># assume df has one factor (X) w/two levels (a,b) and continuous response Y library(MASS) # for fitdistr fa = fitdistr(df[df\$X == "a",]\$Y, "gamma")\$estimate # create fit for X.a ks.test(df[df\$X == "a",]\$Y, "pgamma", shape=fa[1], rate=fa[2]) fb = fitdistr(df[df\$X == "b",]\$Y, "gamma")\$estimate # create fit for X.b ks.test(df[df\$X == "b",]\$Y, "pgamma", shape=fb[1], rate=fb[2])</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 6 shows the distributions of response Y for both levels of factor X. To test whether these distributions were exponentially distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .107$, $p = .849$), as was the test for level ‘b’ ($D = .119$, $p = .742$), indicating non-detectable deviations from an exponential distribution for both levels.”
Gamma	shape (α): <code>shape</code> rate (β): <code>rate</code>	<code>_gamma:</code> <code>dgamma</code> <code>pgamma</code> <code>qgamma</code> <code>rgamma</code>	“Figure 7 shows the distributions of response Y for both levels of factor X. To test whether these distributions were Gamma distributed, a Kolmogorov-Smirnov test was run on Y for both levels of X. The test for level ‘a’ was statistically non-significant ($D = .116$, $p = .773$), as was the test for level ‘b’ ($D = .143$, $p = .526$), indicating non-detectable deviations from a Gamma distribution for both levels.”

Figure 6

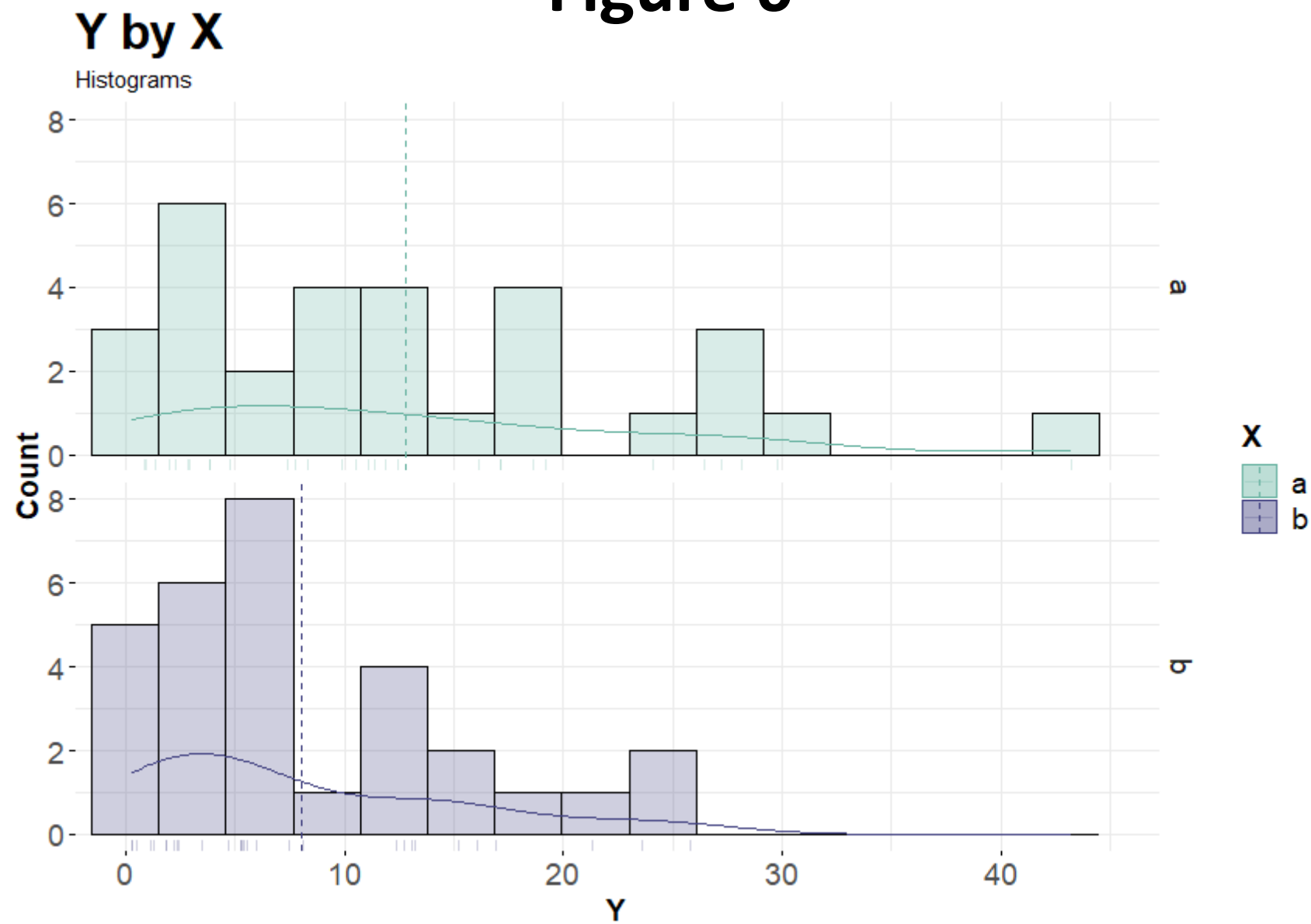
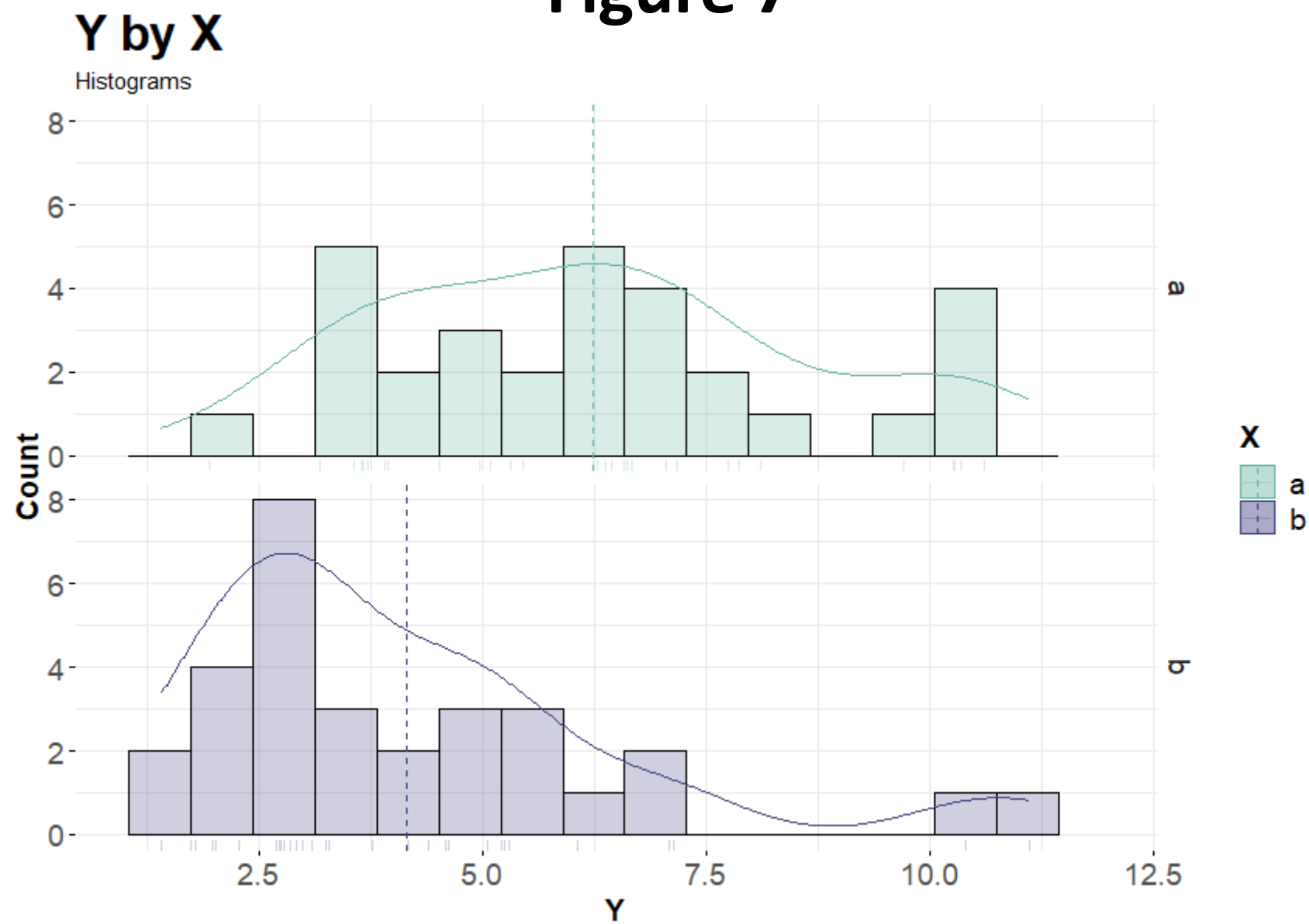


Figure 7



Parametric Tests

./data/Parametric/

Parametric Tests

One factor

Factors	Levels	Between or Within Subjects	Test Name	R Code
1	2	Between	Independent-samples <i>t</i> -test	<pre># df has subjects (S), one between-Ss factor (X) w/levels (a,b), and continuous response (Y) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$X = factor(df\$X) # X is a 2-level factor t.test(Y ~ X, data=df, var.equal=TRUE) # use var.equal=FALSE if heteroscedastic</pre>
1	2	Within	Paired-samples <i>t</i> -test	<pre># df has subjects (S), one within-Ss factor (X) w/levels (a,b), and continuous response (Y) library(reshape2) # for dcast df\$S = factor(df\$S) # Subject id is nominal df\$X = factor(df\$X) # X is a 2-level factor df2 <- dcast(df, S ~ X, value.var="Y") # make wide-format table t.test(df2\$a, df2\$b, paired=TRUE) # homoscedasticity is irrelevant for a paired-samples t-test</pre>
1	≥2	Between	One-way ANOVA	<pre># df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and continuous response (Y) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$X = factor(df\$X) # X is a 3-level factor m = aov(Y ~ X, data=df) # fit model anova(m)</pre>
1	≥2	Within	One-way repeated measures ANOVA	<pre># df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y) library(afex) # for aov_ez df\$S = factor(df\$S) # Subject id is nominal df\$X = factor(df\$X) # X is a 3-level factor m = aov_ez(dv="Y", within="X", id="S", type=3, data=df) s = summary(m) s\$sphericity.tests # Mauchly's sphericity tests s\$univariate.tests # if p>.05, use these unadjusted DFs and p-values anova(m) # if p<.05, use these G-G adjusted DFs and p-values # the following also performs the equivalent repeated measures ANOVA m0 = aov(Y ~ X + Error(S/X), data=df) summary(m0)</pre>

Parametric Tests

One factor

Factors	Levels	Between or Within Subjects	Test Name	Report
1	2	Between	Independent-samples <i>t</i> -test	“The mean of ‘a’ was 14.63 ($SD = 2.13$) and of ‘b’ was 11.01 ($SD = 1.75$). This difference was statistically significant according to an independent-samples <i>t</i> -test ($t(58) = 7.18, p < .0001$).”
1	2	Within	Paired-samples <i>t</i> -test	“The mean of ‘a’ was 13.15 ($SD = 2.53$) and of ‘b’ was 14.37 ($SD = 2.16$). This difference was statistically significant according to a paired-samples <i>t</i> -test ($t(29) = -2.14, p < .05$).”
1	≥ 2	Between	One-way ANOVA	“The mean of ‘a’ was 13.74 ($SD = 2.84$), of ‘b’ was 14.15 ($SD = 2.65$), and of ‘c’ was 9.08 ($SD = 4.29$). These differences were statistically significant according to a one-way ANOVA ($F(2, 57) = 14.18, p < .0001$).”
1	≥ 2	Within	One-way repeated measures ANOVA	“The mean of ‘a’ was 14.04 ($SD = 2.98$), of ‘b’ was 11.95 ($SD = 1.98$), and of ‘c’ was 11.40 ($SD = 2.75$). Mauchly’s test of sphericity indicated no sphericity violation ($W = .926, p = .499$), allowing for an uncorrected repeated measures ANOVA, which showed statistically significant differences ($F(2, 38) = 6.57, p < .01$).”

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

Parametric Tests

Multiple between-Ss factors

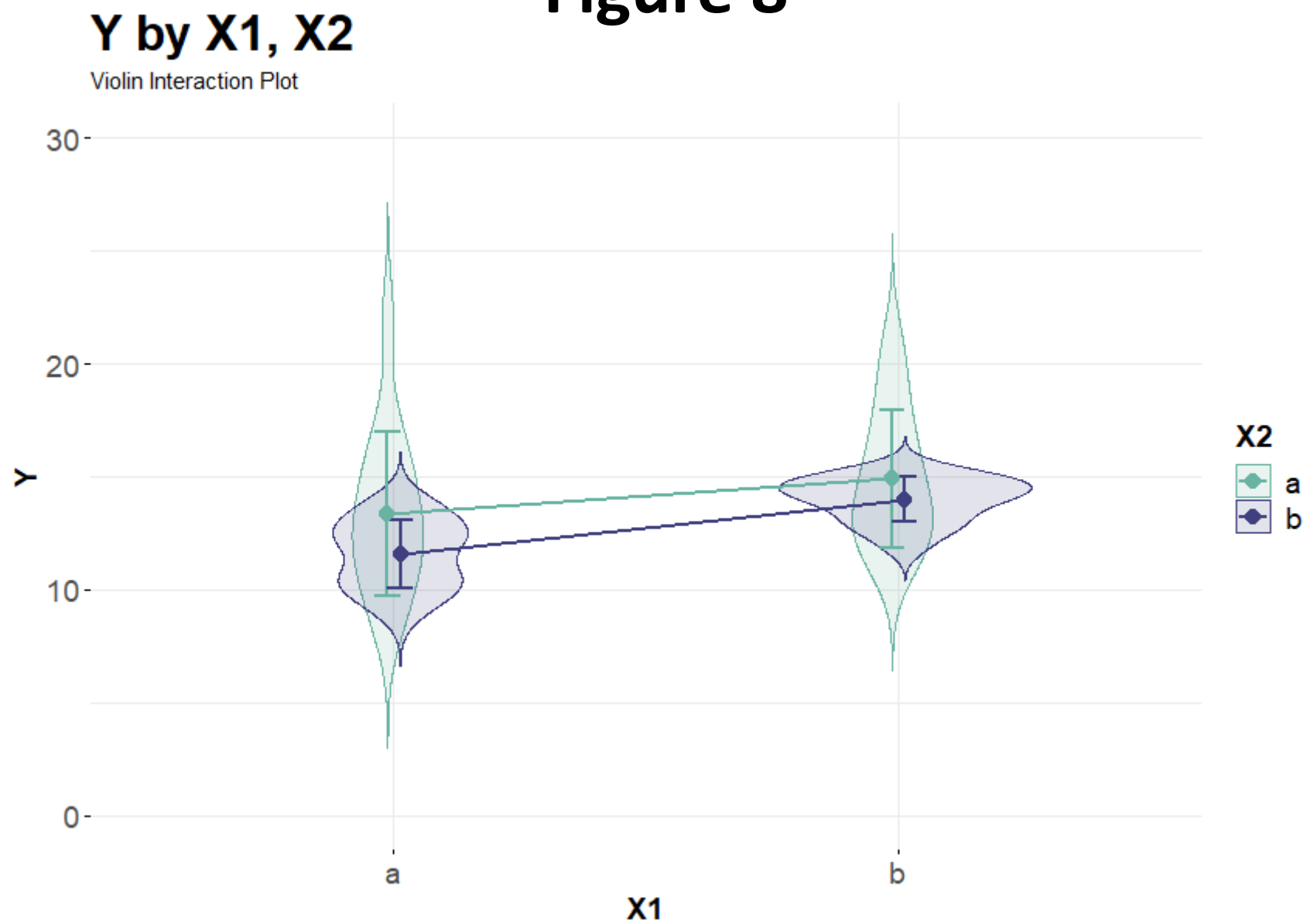
Factors	Levels	Between or Within Subjects	Test Name	R Code
≥2	≥2	Between	Factorial ANOVA	<pre># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(afex) # for aov_ez library(performance) # for check_homogeneity library(car) # for Anova df\$S = factor(df\$S) # Subject id is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor m = aov_ez(dv="Y", between=c("X1","X2"), id="S", type=3, data=df) print(check_homogeneity(m)) # Levene's test anova(m) # if p>.05 from Levene's test Anova(m\$lm, type=3, white.adjust=TRUE) # if p<.05 from Levene's test # the following also performs the equivalent factorial ANOVA m0 = aov(Y ~ X1*X2, data=df) anova(m0)</pre>
			Linear Model (LM)	<pre># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor m = lm(Y ~ X1*X2, data=df) anova(m)</pre>

Parametric Tests

Multiple between-Ss factors

Factors	Levels	Between or Within Subjects	Test Name	Report
≥2	≥2	Between	Factorial ANOVA	“Figure 8 shows a violin interaction plot with ± 1 standard deviation error bars for X1 and X2. A factorial ANOVA indicated a statistically significant effect on Y of X1 ($F(1, 56) = 9.35, p < .01$) and X2 ($F(1, 56) = 4.30, p < .05$), but not of the X1×X2 interaction ($F(1, 56) = 0.42, n.s.$).”
			Linear Model (LM)	“Figure 8 shows a violin interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance indicated a statistically significant effect on Y of X1 ($F(1, 56) = 9.35, p < .01$) and X2 ($F(1, 56) = 4.30, p < .05$), but not of the X1×X2 interaction ($F(1, 56) = 0.42, n.s.$).”

Figure 8



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 subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(afex) # for aov_ez df\$S = factor(df\$S) # Subject id is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor m = aov_ez(dv="Y", within=c("X1","X2"), id="S", type=3, data=df) s = summary(m) s\$sphericity.tests # Mauchly's sphericity tests s\$univariate.tests # if p>.05, use these unadjusted DFs and p-values anova(m) # if p<.05, use these G-G adjusted DFs and p-values # the following also performs the equivalent repeated measures ANOVA, but does not address sphericity m0 = aov(Y ~ X + Error(S/X), data=df) summary(m0) </pre>

Parametric Tests

Multiple within-Ss factors

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

Parametric Tests

Multiple within-Ss factors

Factors	Levels	Between or Within Subjects	Test Name	R Code
≥2	≥2	Within	Linear Mixed Model (LMM)*	<pre> # df has subjects (S), 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\$S = factor(df\$S) # Subject id is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = lmer(Y ~ X1*X2 + (1 S), data=df) Anova(m, type=3, test.statistic="F") </pre>

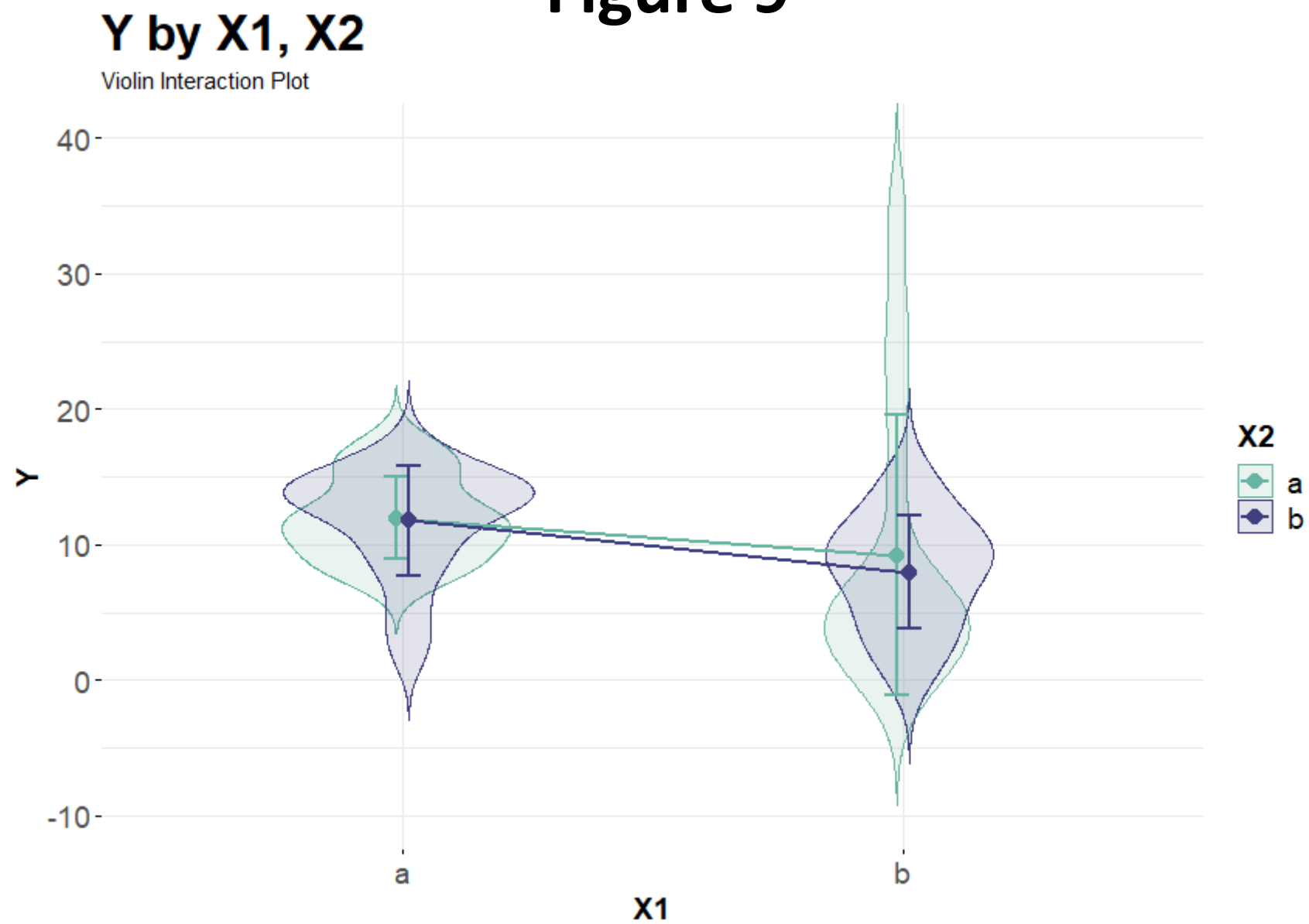
*The LMM sample code uses a random intercept for *subject* (S). There are also random slope models, which are used when the response changes at different rates for each subject over a repeated factor. A 2-minute random slope example of county population growth over time can be seen here (<https://www.youtube.com/watch?v=YDe6F7CXjWw>). 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	Linear Mixed Model (LMM)	“Figure 9 shows a violin interaction plot with ± 1 standard deviation error bars for X1 and X2. A linear mixed model analysis of variance indicated a statistically significant effect on Y of X1 ($F(1, 42) = 4.42, p < .05$), but not of X2 ($F(1, 42) = 0.22, n.s.$), or of the X1×X2 interaction ($F(1, 42) = 0.12, n.s.$).”

Figure 9



Parametric Tests

Post hoc 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 subjects (S), one between-Ss factor (X) w/levels (a,b,c), and continuous response (Y) library(multcomp) # for glht, mcp summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm")) # m is from aov # the following also performs the same contrast tests library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X, adjust="tukey")</pre>
1	≥ 2	One-way repeated measures ANOVA	Paired samples <i>t</i> -test	Within	<pre># df has subjects (S), 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 # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X, adjust="tukey")</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	<p>“Pairwise comparisons using independent-samples <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ and ‘c’ ($t(57) = 4.41, p < .0001$) and ‘b’ and ‘c’ ($t(57) = 4.80, p < .0001$) were statistically significantly different, but that ‘a’ and ‘b’ were not ($t(57) = -0.39, n.s.$).”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that ‘a’ and ‘c’ ($t(57) = 4.41, p < .0001$) and ‘b’ and ‘c’ ($t(57) = 4.80, p < .0001$) were statistically significantly different, but that ‘a’ and ‘b’ were not ($t(57) = -0.39, n.s.$).”</p>
1	≥ 2	One-way repeated measures ANOVA	Paired samples <i>t</i> -test	Within	<p>“Three <i>post hoc</i> paired-samples <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ and ‘b’ ($t(19) = 3.14, p < .05$) and ‘a’ and ‘c’ ($t(19) = 3.40, p < .01$) were statistically significantly different, but that ‘b’ and ‘c’ were not ($t(19) = 0.65, n.s.$).”</p>

Note: “Tukey’s HSD” test stands for “Tukey’s Honestly Significant Difference” test.

Parametric Tests

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Test	Test Name	B/W	R Code
≥2	≥2	Factorial ANOVA	Independent samples <i>t</i> -test	Btwn	<pre># df has subjects (S), 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 afex::aov_ez # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
≥2	≥2	Linear Model (LM)	Independent samples <i>t</i> -test	Btwn	<pre># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # m is from aov or lm # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
≥2	≥2	Factorial repeated measures ANOVA	Paired samples <i>t</i> -test	Within	<pre># df has subjects (S), 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 # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
≥2	≥2	Linear Mixed Model (LMM)	Paired samples <i>t</i> -test	Within	<pre># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # m is from lmer # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</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	<p>“Pairwise comparisons using independent-samples <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (b,a) <i>vs.</i> (a,b) showed a statistically significant difference ($t(56) = 3.63, p < .01$), and that (a,b) <i>vs.</i> (b,b) was marginal ($t(56) = -2.62, p = .057$). The other four comparisons showed no statistically significant differences.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that that (b,a) <i>vs.</i> (a,b) showed a statistically significant difference ($t(56) = 3.63, p < .01$), and that (a,b) <i>vs.</i> (b,b) was marginal ($t(56) = -2.62, p = .054$). The other four comparisons showed no statistically significant differences.”</p>
≥2	≥2	Factorial repeated measures ANOVA	Paired samples <i>t</i> -test	Within	<p>“Six <i>post hoc</i> paired-samples <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,a) <i>vs.</i> (b,b) showed a statistically significant difference ($t(14) = 3.10, p < .05$), and that (a,b) <i>vs.</i> (b,b) was marginal ($t(14) = 2.65, p = .096$). The other four comparisons showed no statistically significant differences.”</p>
≥2	≥2	Linear Mixed Model (LMM)	Paired samples <i>t</i> -test	Within	<p>“Pairwise comparisons using paired-samples <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no pairs were statistically significantly different.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no pairs were statistically significantly different.”</p>

Note: “Tukey’s HSD” test stands for “Tukey’s Honestly Significant Difference” test.

Variance-Covariance Structures

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

Variance-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. Default for nlme::lme.	<pre># df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and continuous response (Y) library(nlme) # for lme library(car) # for Anova library(emmeans) # for emmeans df\$S = factor(df\$S) # Subject id is nominal df\$X = factor(df\$X) # X is a 3-level factor contrasts(df\$X) <- "contr.sum" m = lme(Y ~ X, random=~1 S, data=df, weights=varIdent(form=~X)) # 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") # post hoc tests</pre>
DIAG	Diagonal	All variances can differ, and all covariances are zero. Default in SPSS.	<pre># See R Code for ID. Only the blue model building line changes to: m = lme(Y ~ X, random=~1 S, data=df, weights=varIdent(form=~1 X)) # DIAG</pre>
CS	Compound symmetry	All variances are equal, and all covariances are equal. Default in JMP.	<pre># See R Code for ID. Only the blue model building line changes to: m = lme(Y ~ X, random=~1 S, data=df, correlation=corCompSymm(form=~1 S)) # CS</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 S, data=df, correlation=corCompSymm(form=~1 S), 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). Above, when `weights=varIdent(form=~X)`, the variances are required to be equal, which is the default when `weights` is unspecified. When `weights=varIdent(form=~1|X)`, the variances are allowed to differ.

Variance-Covariance Structures

Optional when fitting linear mixed models (LMMs)

Abbreviation	Name	Description	R Code
AR1	First-order autoregressive	All variances are equal, and all covariances decrease the further apart they are.	<pre># See R Code for ID. Only the blue model building line changes to: m = lme(Y ~ X, random=~1 S, data=df, correlation=corAR1(form=~1 S)) # AR1</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 S, data=df, correlation=corAR1(form=~1 S), weights=varIdent(form=~1 X)) # ARH1</pre>
ARMA11	Autoregressive moving average	All variances are equal, and all covariances decrease the further apart they are, 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 S, data=df, correlation=corARMA(form=~1 S, p=1, q=1)) # ARMA11 # Note that (p,q)=(1,0) would be AR1. The 'q' parameter determines the moving average.</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 S, data=df, correlation=corARMA(form=~1 S, 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 S, data=df, correlation=corARMA(form=~1 S, p=2, q=0), weights=varIdent(form=~1 X)) # TPH</pre>
UN	Unstructured	All variances and covariances can differ. A completely unpatterned matrix. Default for <code>lme4::lmer</code> .	<pre># See R Code for ID. Only the blue model building line changes to: m = lme(Y ~ X, random=~1 S, data=df, correlation=corSymm(form=~1 S), weights=varIdent(form=~1 X))</pre>

Note: See `?nlme::corClasses` and `?nlme::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>.

Nonparametric Tests

./data/Nonparametric/

Nonparametric Tests

One factor

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

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

Nonparametric Tests

One factor

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

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 subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(ARTool) df\$S = factor(df\$S) # Subject id is nominal (unused) df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor m = art(Y ~ X1*X2, data=df) anova(m)</pre>
			Generalized Linear Model (GLM)*	See “Generalized Linear (Mixed) Models: Distributions and canonical links”
≥2	≥2	Within	Aligned Rank Transform (ART)**	<pre># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(ARTool) df\$S = factor(df\$S) # Subject id is nominal df\$X1 = factor(df\$X1) # X1 is a 2-level factor df\$X2 = factor(df\$X2) # X2 is a 2-level factor m = art(Y ~ X1*X2 + (1 S), data=df) # S is a random factor anova(m)</pre>
			Generalized Linear Mixed Model (GLMM)*	See “Generalized Linear (Mixed) Models: Distributions and canonical links”

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

**The Aligned Rank Transform within-subjects sample code uses a random intercept for *subject* (S). There are also random slope models, which are used when the response changes at different rates for each subject over a repeated factor. A 2-minute random slope example of county population growth over time can be seen here (<https://www.youtube.com/watch?v=YDe6F7CXjWw>). 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 10 shows boxplots for all levels of $X1 \times X2$. A nonparametric analysis of variance based on the Aligned Rank Transform indicated no statistically significant effect on Y of X1 ($F(1, 56) = 1.98, n.s.$), or of X2 ($F(1, 56) = 0.81, n.s.$), but there was a statistically significant $X1 \times X2$ interaction ($F(1, 56) = 13.65, p < .001$).”
			Generalized Linear Model (GLM)	See “Generalized Linear (Mixed) Models: Distributions and canonical links”
≥2	≥2	Within	Aligned Rank Transform (ART)	“Figure 11 shows boxplots for all levels of $X1 \times X2$. A nonparametric analysis of variance based on the Aligned Rank Transform indicated no statistically significant effect on Y of X1 ($F(1, 42) = 1.85, n.s.$) or of the $X1 \times X2$ interaction ($F(1, 42) = 1.93, n.s.$), but there was a statistically significant effect of X2 ($F(1, 42) = 24.97, p < .0001$).”
			Generalized Linear Mixed Model (GLMM)	See “Generalized Linear (Mixed) Models: Distributions and canonical links”

Figure 10

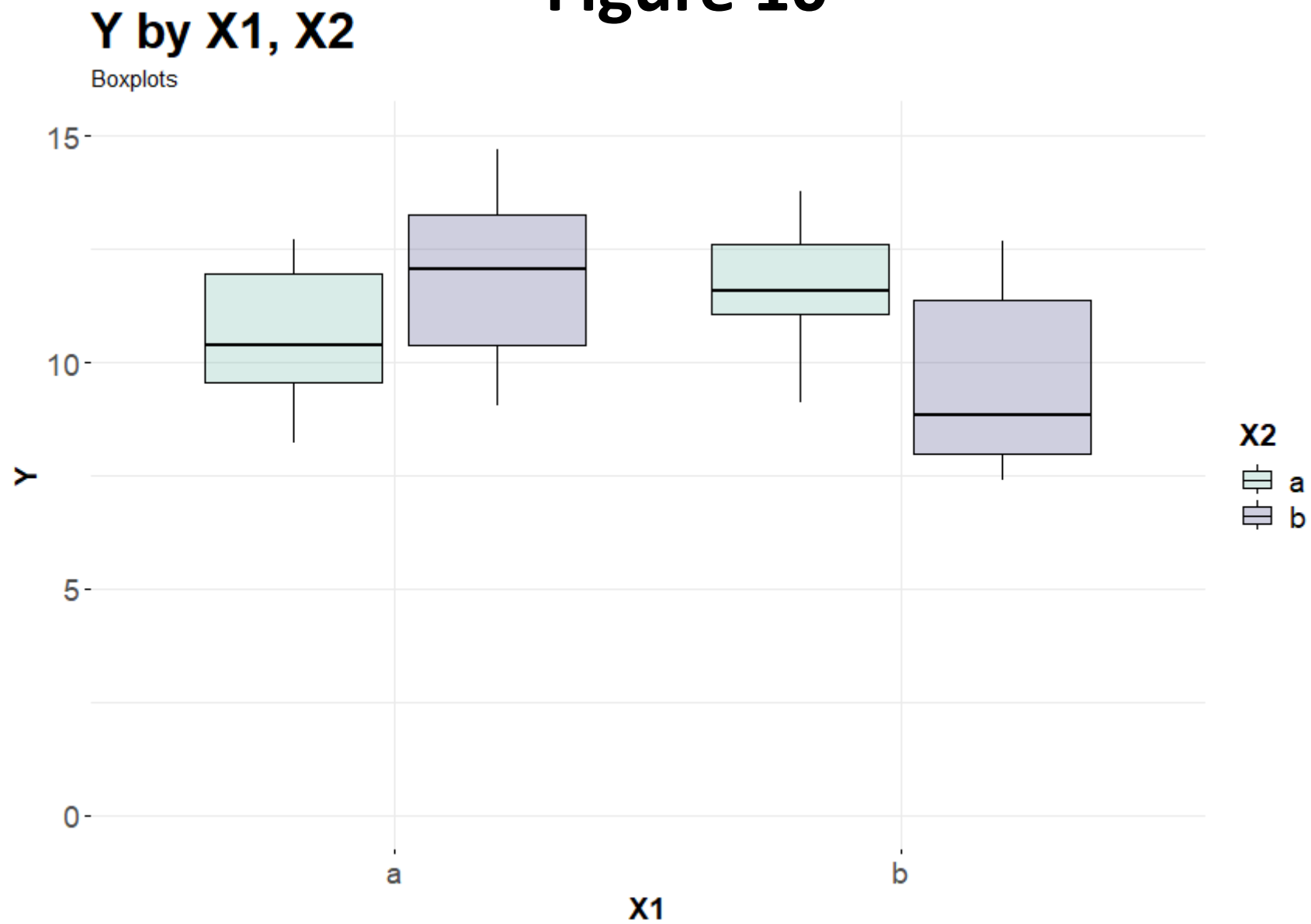
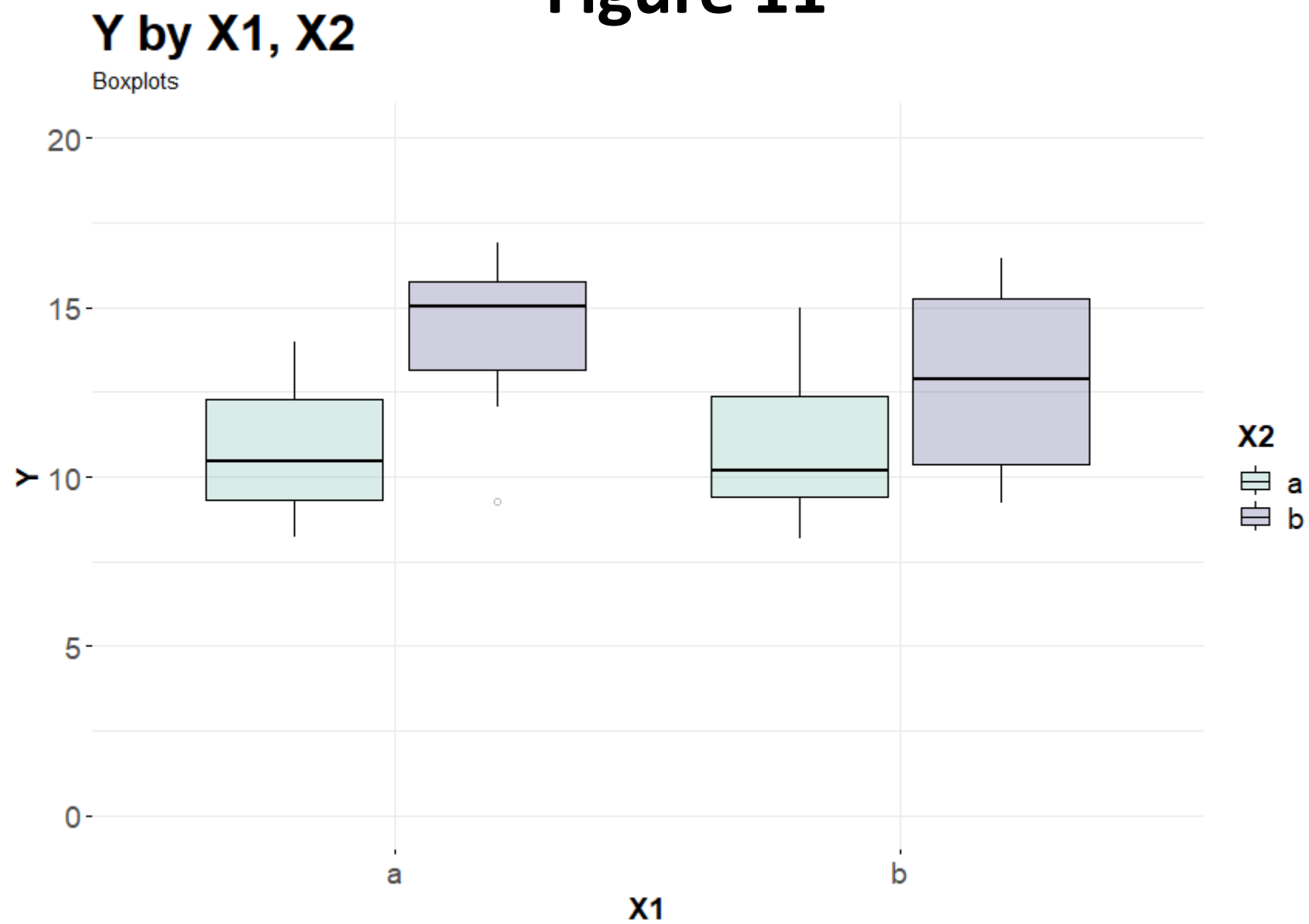


Figure 11



Nonparametric Tests

Post hoc 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 subjects (S), 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) # a vs. b ac = wilcox.test(df[df\$X == "a",]\$Y, df[df\$X == "c",]\$Y) # a vs. c bc = wilcox.test(df[df\$X == "b",]\$Y, df[df\$X == "c",]\$Y) # b vs. c p.adjust(c(ab\$p.value, ac\$p.value, bc\$p.value), method="holm") # 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 subjects (S), 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, S ~ X, value.var="Y") # make wide-format table ab = wilcox.test(df2\$a, df2\$b, paired=TRUE) # a vs. b ac = wilcox.test(df2\$a, df2\$c, paired=TRUE) # a vs. c bc = wilcox.test(df2\$b, df2\$c, paired=TRUE) # b vs. c p.adjust(c(ab\$p.value, ac\$p.value, bc\$p.value), method="holm") # 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, indicated that ‘a’ and ‘b’ ($Z = 2.95, p < .01$) and ‘a’ and ‘c’ ($Z = 3.87, p < .001$) were statistically significantly different, but that ‘b’ and ‘c’ were marginal ($Z = 1.73, p = .086$).”
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, indicated that ‘a’ and ‘c’ ($Z = -2.73, p < .05$) and ‘b’ and ‘c’ ($Z = -2.31, p < .05$) were statistically significantly different, but that ‘a’ and ‘b’ were not ($Z = -0.82, n.s.$).”

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 subjects (S), two factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(dplyr) # for %>% pipe art.con(m, ~ X1*X2, adjust="holm") %>% # run ART-C for X1 × X2 summary() %>% # add significance stars to the output mutate(sig. = symnum(p.value, corr=FALSE, na=FALSE, cutpoints = c(0, .001, .01, .05, .10, 1), symbols = c("****", "***", "**", ".", " ")))</pre>
			Mann-Whitney U test	Btwn	<pre># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) aa_ab = wilcox.test(df[df\$X1 == "a" & df\$X2 == "a",]\$Y, df[df\$X1 == "a" & df\$X2 == "b",]\$Y) aa_ba = wilcox.test(df[df\$X1 == "a" & df\$X2 == "a",]\$Y, df[df\$X1 == "b" & df\$X2 == "a",]\$Y) aa_bb = wilcox.test(df[df\$X1 == "a" & df\$X2 == "a",]\$Y, df[df\$X1 == "b" & df\$X2 == "b",]\$Y) ab_ba = wilcox.test(df[df\$X1 == "a" & df\$X2 == "b",]\$Y, df[df\$X1 == "b" & df\$X2 == "a",]\$Y) ab_bb = wilcox.test(df[df\$X1 == "a" & df\$X2 == "b",]\$Y, df[df\$X1 == "b" & df\$X2 == "b",]\$Y) ba_bb = wilcox.test(df[df\$X1 == "b" & df\$X2 == "a",]\$Y, df[df\$X1 == "b" & df\$X2 == "b",]\$Y) p.adjust(c(aa_ab\$p.value, aa_ba\$p.value, aa_bb\$p.value, ab_ba\$p.value, ab_bb\$p.value, ba_bb\$p.value), method="holm") # p-values wilcoxonZ(df[df\$X1 == "a" & df\$X2 == "a",]\$Y, df[df\$X1 == "a" & df\$X2 == "b",]\$Y) # Z-scores wilcoxonZ(df[df\$X1 == "a" & df\$X2 == "a",]\$Y, df[df\$X1 == "b" & df\$X2 == "a",]\$Y) wilcoxonZ(df[df\$X1 == "a" & df\$X2 == "a",]\$Y, df[df\$X1 == "b" & df\$X2 == "b",]\$Y) wilcoxonZ(df[df\$X1 == "a" & df\$X2 == "b",]\$Y, df[df\$X1 == "b" & df\$X2 == "a",]\$Y) wilcoxonZ(df[df\$X1 == "a" & df\$X2 == "b",]\$Y, df[df\$X1 == "b" & df\$X2 == "b",]\$Y) wilcoxonZ(df[df\$X1 == "b" & df\$X2 == "a",]\$Y, df[df\$X1 == "b" & df\$X2 == "b",]\$Y)</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	“ <i>Post hoc</i> pairwise comparisons conducted with the ART-C procedure, and corrected with Holm’s sequential Bonferroni procedure, indicated that (a,b) vs. (b,b) showed a statistically significant difference ($t(56) = 3.42, p < .01$), as did (b,a) vs. (b,b) ($t(56) = 3.21, p < .05$). No other pairwise comparisons were significantly different.”
				Within	“ <i>Post hoc</i> pairwise comparisons conducted with the ART-C procedure, and corrected with Holm’s sequential Bonferroni procedure, indicated that (a,a) vs. (a,b) showed a statistically significant difference ($t(42) = -4.43, p < .001$), as did (a,b) vs. (b,a) ($t(42) = 4.38, p < .001$). The test of (a,a) vs. (b,b) was marginal ($t(42) = -2.49, p = .067$), as was the test of (b,a) vs. (b,b) ($t(42) = -2.44, p = .067$). No other pairwise comparisons were statistically significant or marginal.”
			Mann-Whitney <i>U</i> test	Btwn	“Six <i>post hoc</i> Mann-Whitney <i>U</i> tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,b) vs. (b,b) showed a statistically significant difference ($Z = 2.76, p < .05$), and that (b,a) vs. (b,b) did as well ($Z = 2.97, p < .05$). The other four comparisons showed no statistically significant differences.”

Nonparametric Tests

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Test	Test Name	B/W	R Code
≥2	≥2	Aligned Rank Transform (ART)	Wilcoxon signed-rank test	Within	<pre> # df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(reshape2) # for dcast library(rcompanion) # for wilcoxonZ df2 <- dcast(df, S ~ X1 + X2, value.var="Y") # make wide-format table aa_ab = wilcox.test(df2\$a_a, df2\$a_b, paired=TRUE) aa_ba = wilcox.test(df2\$a_a, df2\$b_a, paired=TRUE) aa_bb = wilcox.test(df2\$a_a, df2\$b_b, paired=TRUE) ab_ba = wilcox.test(df2\$a_b, df2\$b_a, paired=TRUE) ab_bb = wilcox.test(df2\$a_b, df2\$b_b, paired=TRUE) ba_bb = wilcox.test(df2\$b_a, df2\$b_b, paired=TRUE) p.adjust(c(aa_ab\$p.value, aa_ba\$p.value, aa_bb\$p.value, ab_ba\$p.value, ab_bb\$p.value, ba_bb\$p.value), method="holm") # p-values wilcoxonZ(df2\$a_a, df2\$a_b, paired=TRUE) # Z-scores wilcoxonZ(df2\$a_a, df2\$b_a, paired=TRUE) wilcoxonZ(df2\$a_a, df2\$b_b, paired=TRUE) wilcoxonZ(df2\$a_b, df2\$b_a, paired=TRUE) wilcoxonZ(df2\$a_b, df2\$b_b, paired=TRUE) wilcoxonZ(df2\$b_a, df2\$b_b, paired=TRUE) </pre>

Nonparametric Tests

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Test	Test Name	B/W	Report
≥2	≥2	Aligned Rank Transform (ART)	Wilcoxon signed-rank test	Within	<p>“Six <i>post hoc</i> Wilcoxon signed-rank tests, corrected with Holm’s sequential Bonferroni procedure, indicated that (a,a) vs. (a,b) showed a statistically significant difference ($Z = -3.18, p < .01$), as did (a,b) vs. (b,a) ($Z = 3.01, p < .01$). The test of (a,a) vs. (b,b) was marginal ($Z = -2.33, p = .072$). The other three comparisons showed no statistically significant differences.”</p>

Generalized Linear (Mixed) Models

./data/GLM/

./data/GLMM/

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
Gaussian	<code>lm()</code> ^[1] <code>glm()</code> <code>family=gaussian</code>	<code>lme4::lmer()</code> ^[2] <code>lme4::glmer()</code> <code>family=gaussian</code>
Binomial	<code>glm()</code> <code>family=binomial</code>	<code>lme4::glmer()</code> <code>family=binomial</code>
Multinomial	<code>nnet::multinom()</code>	<code>lme4::glmer()</code> <code>family=poisson</code> <code>#after multinomial-Poisson transform</code> ^[3]
Ordinal	<code>MASS::polr()</code>	<code>ordinal::clmm()</code> <code>RVAideMemoire::Anova.clmm()</code>
Poisson	<code>glm()</code> <code>family=poisson</code> <code>#or family=quasipoisson</code> ^[4]	<code>lme4::glmer()</code> <code>family=poisson</code>
Zero-inflated Poisson ^[5]	<code>pscl::zeroinfl()</code> <code>dist="poisson"</code>	<code>glmmTMB::glmmTMB()</code> <code>family=poisson</code> <code>ziformula=~...</code> ^[6]
Negative binomial	<code>MASS::glm.nb()</code>	<code>lme4::glmer.nb()</code>
Zero-inflated negative binomial ^[5]	<code>pscl::zeroinfl()</code> <code>dist="negbin"</code>	<code>glmmTMB::glmmTMB()</code> <code>family=nbinom2</code> <code>ziformula=~...</code> ^[6]
Exponential	<code>glm()</code> <code>family=Gamma</code> <code>#or family=Gamma(link="log")</code> ^[7]	<code>lme4::glmer()</code> <code>family=Gamma</code> <code>#or family=Gamma(link="log")</code>
Gamma	<code>glm()</code> <code>family=Gamma</code> <code>#or family=Gamma(link="log")</code>	<code>lme4::glmer()</code> <code>family=Gamma</code> <code>#or family=Gamma(link="log")</code>

Footnotes [1] – [7] appear on next slide.

Footnotes

- [1] `lm` and `glm` are from the base statistics package built into R.
- [2] `lmer` and `glmer` are both from the `lme4` package. A `glmer` call with `family=gaussian` will fail to run with a message that one should just use `lmer` for an LMM, which is equivalent to `glmer` with `family=gaussian`.
- [3] See below for more on the multinomial-Poisson transform (Baker 1994).
- [4] Use `family=quasipoisson` when mild overdispersion is present. If overdispersion is large, use negative binomial regression. Note that `family=quasipoisson` cannot be used with `glmer`.
- [5] Use zero-inflated variants of Poisson or negative binomial regression when distributions show large numbers of zeros relative to other counts.
- [6] Set the `ziformula` parameter to the fixed effects in the model, e.g., `ziformula=~X1*X2`.
- [7] The canonical link function for the Gamma family is the inverse function. If this fails, e.g., due to the discontinuity at zero, then using `family=Gamma(link="log")` is a viable alternative.

There is no `family=multinomial` option in `glmer`. There is no GLMM equivalent to the `nnet::multinom` function for GLM. There are a handful of options for polytomous responses with repeated measures, e.g., based on Markov Chain Monte-Carlo (MCMC) simulations in package `MCMCglmm`, but these approaches deviate considerably from the consistent GLMM approach. Fortunately, Baker (1994) showed the equivalence of multinomial logistic regression and Poisson regression via the multinomial-Poisson transformation. After transformation, `family=poisson` can be used with `glmer`.

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> Equivalent to the linear model (LM) or linear mixed model (LMM)	<pre>library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glm(Y ~ X1*X2, data=df, family=gaussian) Anova(m, type=3, test.statistic="F")</pre>	<pre>library(lme4) # for glmer, lmer library(lmerTest) library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" #m = glmer(Y ~ X1*X2 + (1 S), data=df, family=gaussian) #deprecated m = lmer(Y ~ X1*X2 + (1 S), data=df) # equivalent, use instead Anova(m, type=3, test.statistic="F")</pre>
Binomial	logit	<i>Logistic regression:</i> Dichotomous responses (i.e., nominal responses with two categories)	<pre>library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = factor(df\$Y) # nominal response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glm(Y ~ X1*X2, data=df, family=binomial) Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = factor(df\$Y) # nominal response contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmer(Y ~ X1*X2 + (1 S), data=df, family=binomial) Anova(m, type=3)</pre>

Notes: A normal distribution is also known as a Gaussian distribution. Logistic regression is also known as binomial regression.

GLM / GLMM

Distributions and canonical links

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

Figure 12a

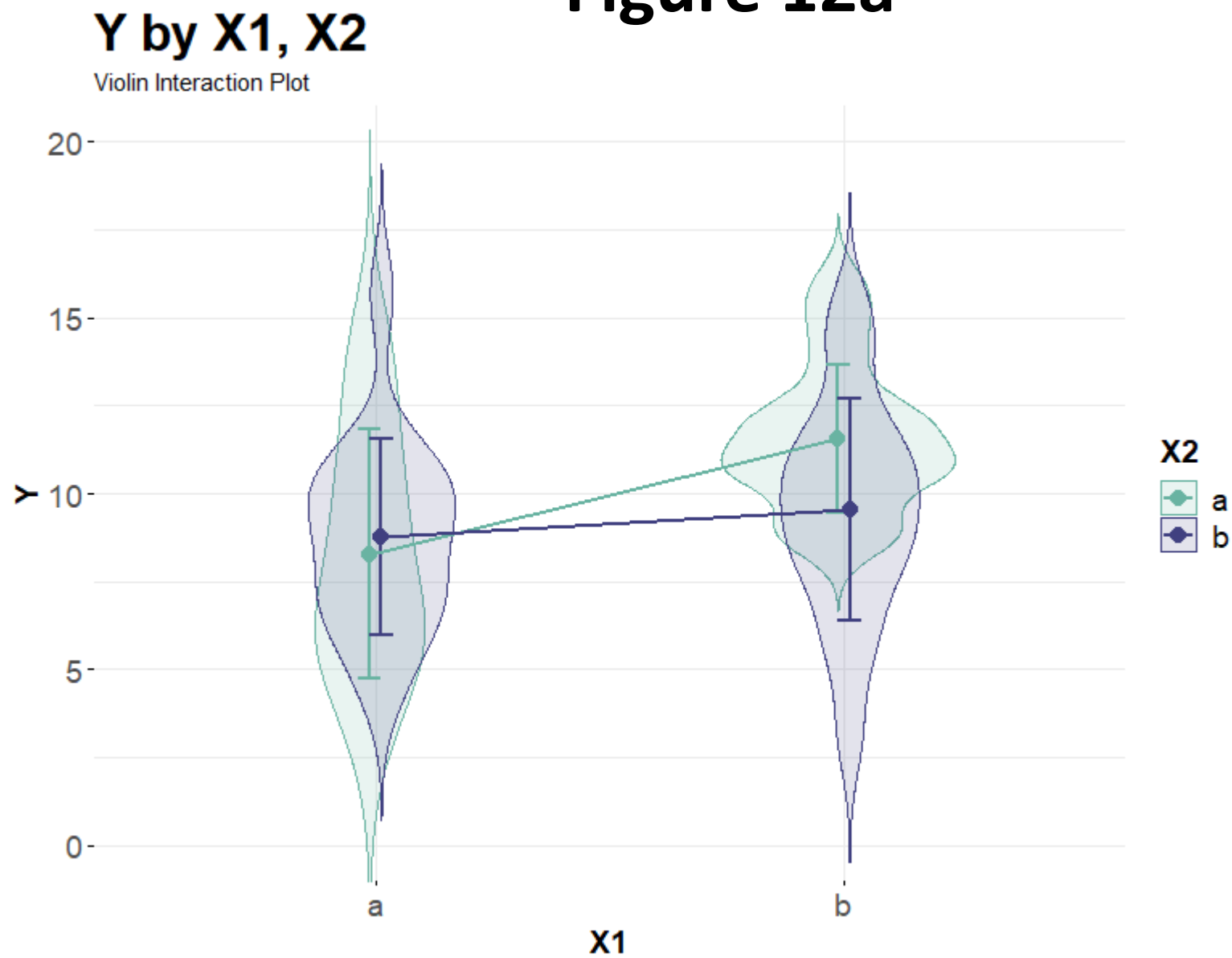


Figure 12b

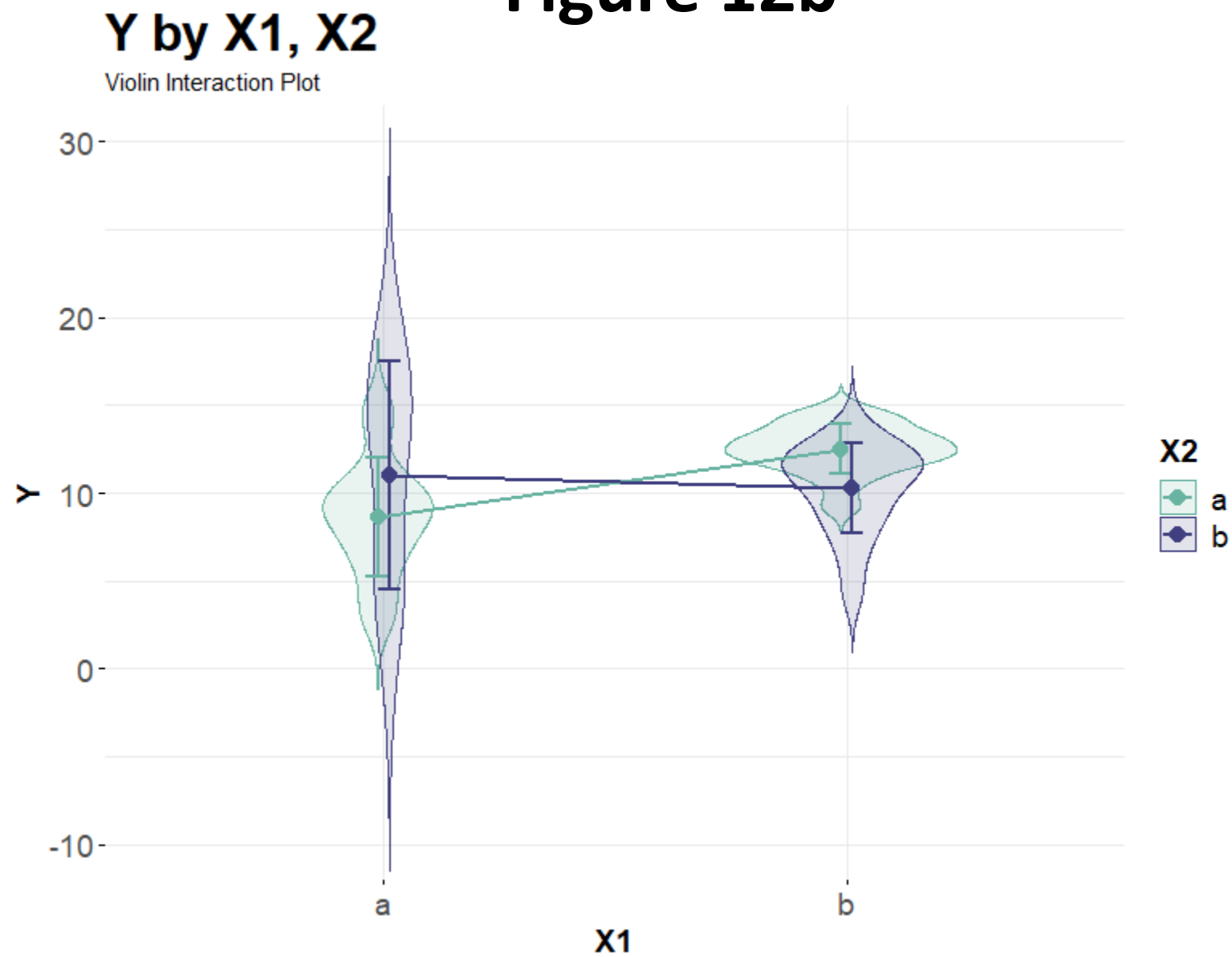


Figure 13a

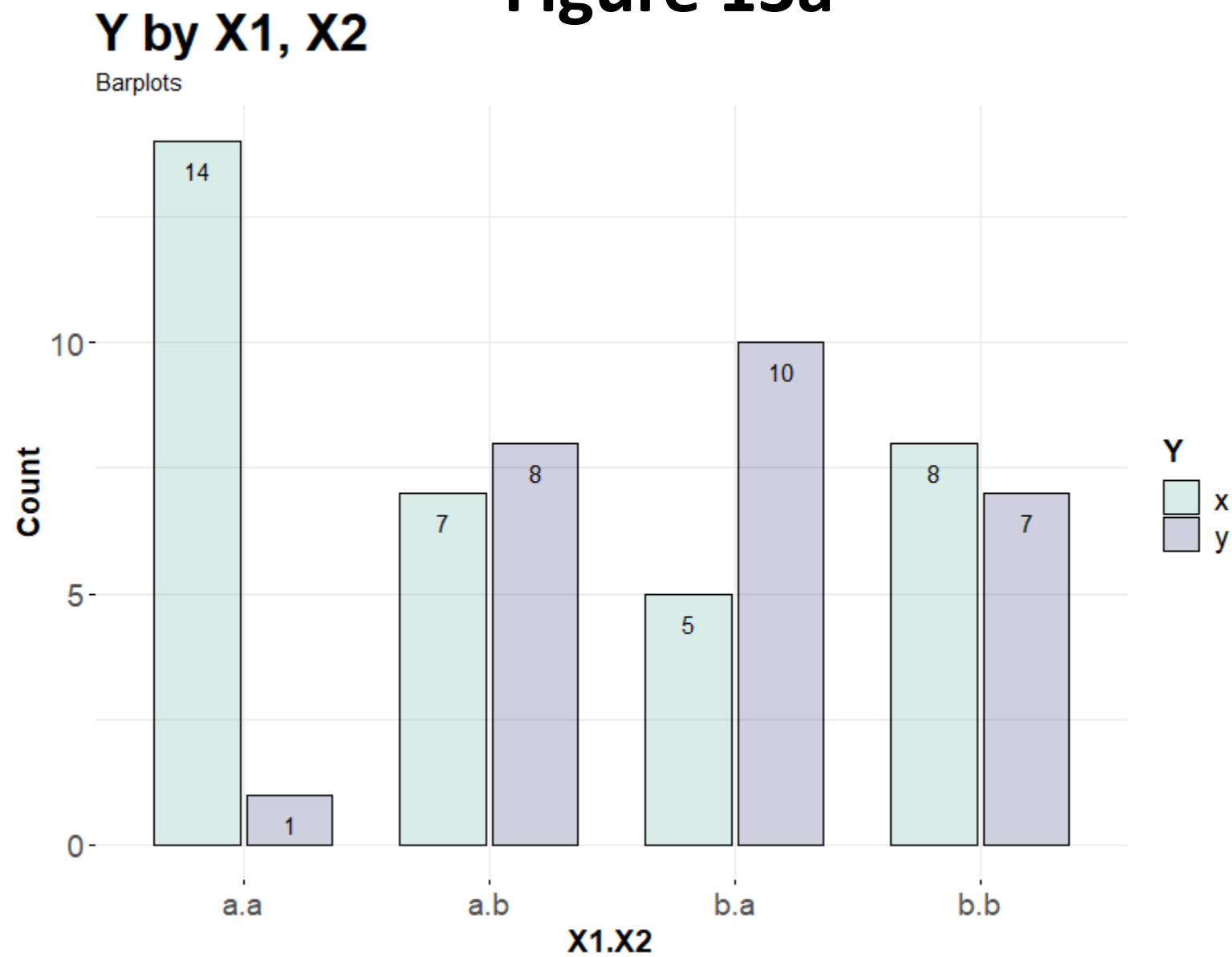
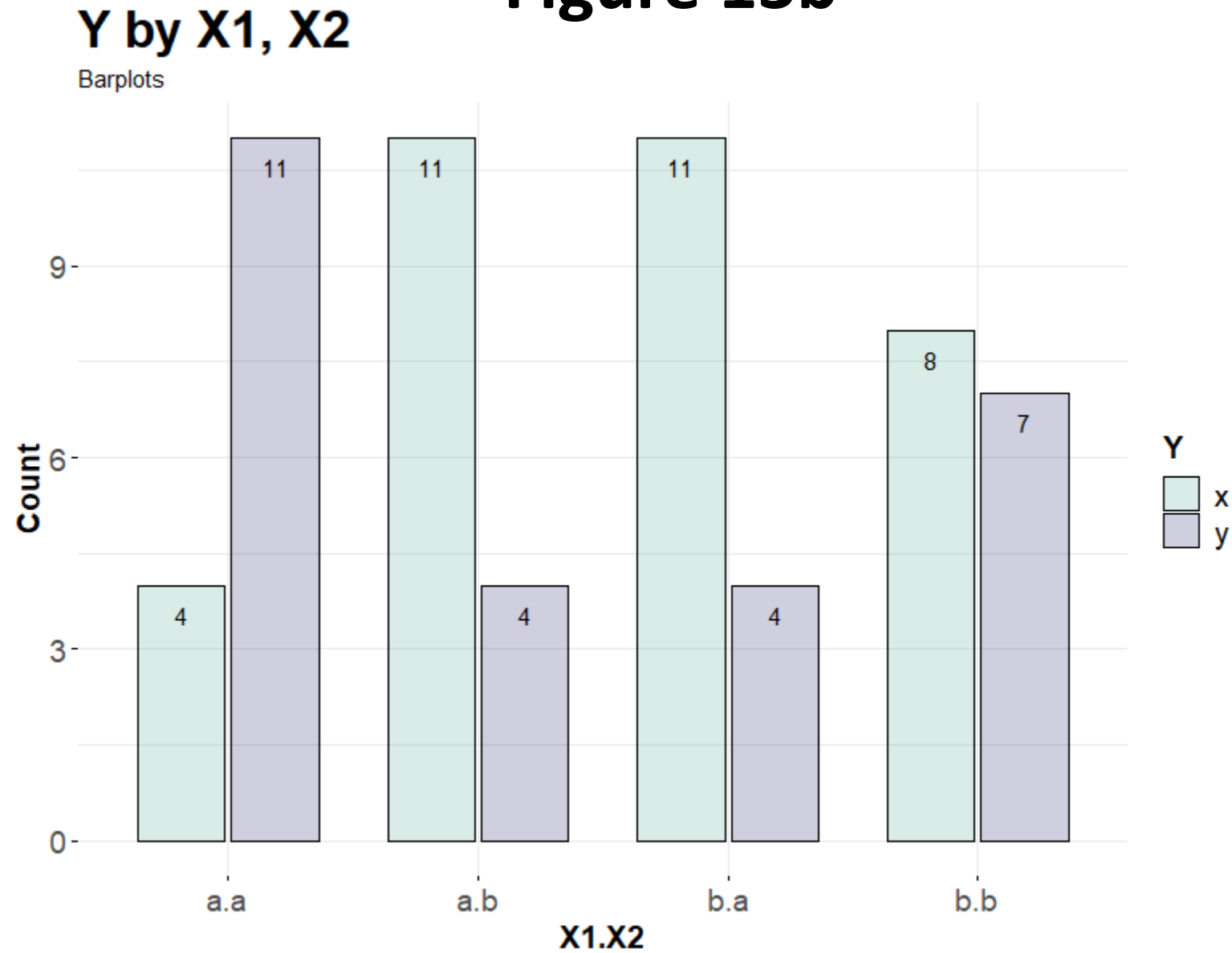


Figure 13b



GLM / GLMM

Distributions and canonical links

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

Notes: Multinomial logistic regression is also known as nominal logistic regression. The multinomial-Poisson transformation (Baker 1994) is tested in `m-P.R` in the accompanying code files.

GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	Report for GLM (between-Ss)	Report for GLMM (within-Ss)
Multinomial	logit	<i>Multinomial logistic regression:</i> polytomous responses (i.e., nominal responses with more than two categories)	“Figure 14a shows the number of ‘x’, ‘y’, and ‘z’ outcomes for each level of X1 and X2. An analysis of variance based on multinomial logistic regression indicated a statistically significant effect on Y of X1 ($\chi^2(2, N=60) = 10.46, p < .01$), of X2 ($\chi^2(2, N=60) = 15.21, p < .001$), and a marginal X1×X2 interaction ($\chi^2(2, N=60) = 5.09, p = .078$).”	“Figure 14b shows the number of ‘x’, ‘y’, and ‘z’ outcomes for each level of X1 and X2. An analysis of variance based on mixed Poisson regression using the multinomial-Poisson transformation (Baker 1994) indicated no statistically significant effect on Y of X1 ($\chi^2(2, N=60) = 2.32, n.s.$) or X2 ($\chi^2(2, N=60) = 0.30, n.s.$), but there was a statistically significant X1×X2 interaction ($\chi^2(2, N=60) = 11.39, p < .01$).”

Figure 14a

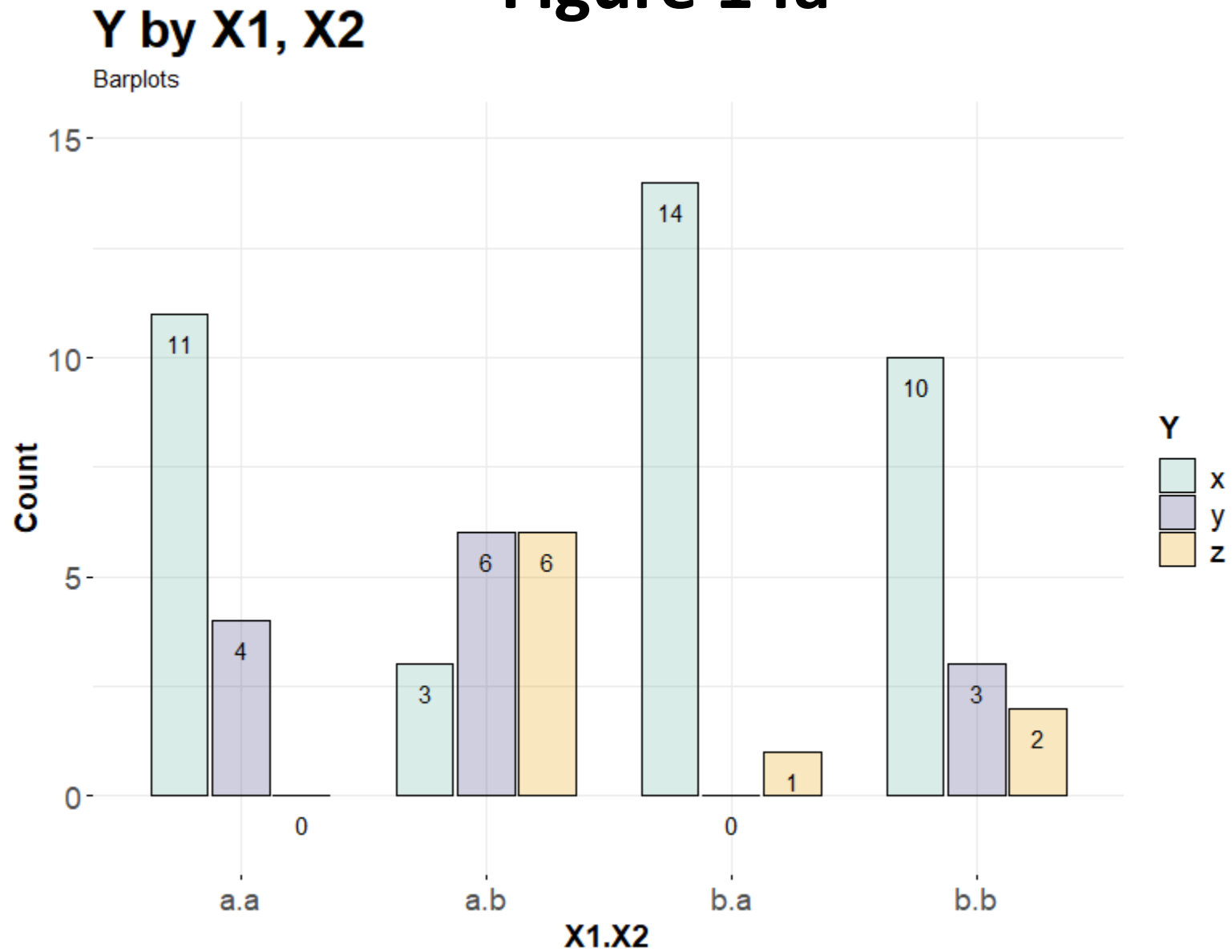
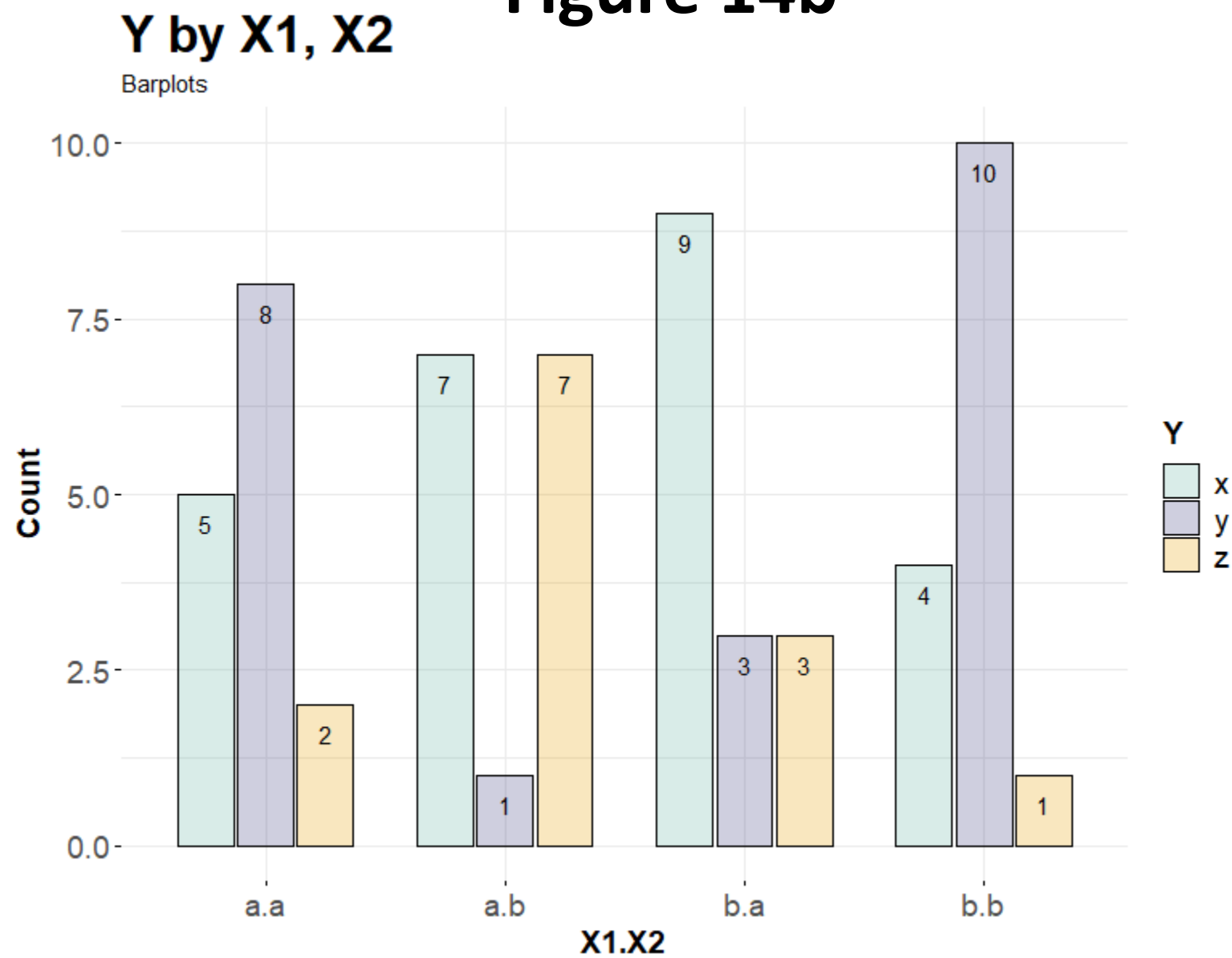


Figure 14b



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 scales)	<pre>library(MASS) # for polr library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) df\$Y = ordered(df\$Y) # ordinal response contrasts(df\$X1) <- "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\$S = factor(df\$S) 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 S), data=df2, link="logit") # or "probit" Anova.clmm(m)</pre>
Poisson	log	<i>Poisson regression:</i> Counts, rare events (e.g., gesture recognition errors, 3-pointers per quarter, number of “F” grades)	<pre>library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glm(Y ~ X1*X2, data=df, family=poisson) # use family=quasipoisson if overdispersed Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmer(Y ~ X1*X2 + (1 S), data=df, family=poisson) Anova(m, type=3)</pre>

Notes: Ordinal logistic regression is also known as cumulative logistic regression. 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 `AER::dispersiontest(m)`, where “m” is a fitted model with `family=poisson`. One can also use `(var(Y)/abs(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 `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)
Ordinal	cumulative logit	<i>Ordinal logistic regression:</i> Ordinal responses (e.g., Likert scales)	“Figure 15a shows the distribution of Likert responses (1-7) for each combination of X1 and X2. An analysis of variance based on ordinal logistic regression indicated a statistically significant effect on Y of X2 ($\chi^2(1, N=60) = 6.14, p < .05$), but not of X1 ($\chi^2(1, N=60) = 1.65, n.s.$) or of the X1×X2 interaction ($\chi^2(1, N=60) = 0.05, n.s.$).”	“Figure 15b shows the distribution of Likert responses (1-7) for each combination of X1 and X2. An analysis of variance based on mixed ordinal logistic regression indicated no statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 0.02, n.s.$) or of X2 ($\chi^2(1, N=60) = 1.00, n.s.$), but there was a statistically significant X1×X2 interaction ($\chi^2(1, N=60) = 4.22, p < .05$).”
Poisson	log	<i>Poisson regression:</i> Counts, rare events (e.g., gesture recognition errors, 3-pointers per quarter, number of “F” grades)	“Figure 16a shows a violin interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance based on Poisson regression indicated a statistically significant effect on Y of the X1×X2 interaction ($\chi^2(1, N=60) = 3.84, p < .05$), but not of either X1 ($\chi^2(1, N=60) = 0.17, n.s.$) or X2 ($\chi^2(1, N=60) = 1.19, n.s.$).”	“Figure 16b shows a violin interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance based on mixed Poisson regression indicated a statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 18.10, p < .0001$), but not of X2 ($\chi^2(1, N=60) = 1.47, n.s.$) or of the X1×X2 interaction ($\chi^2(1, N=60) = 0.26, n.s.$).”

Figure 15a

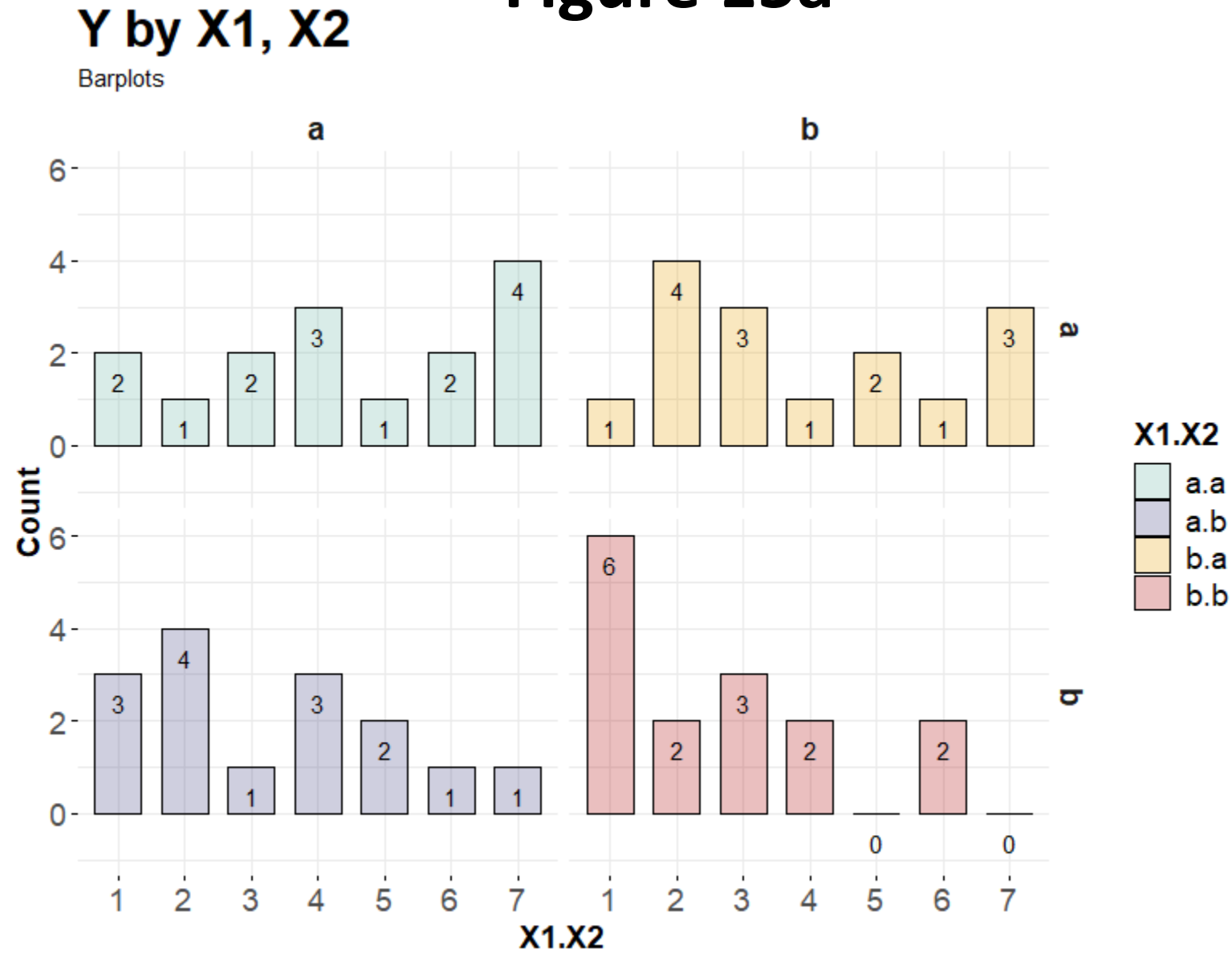


Figure 15b

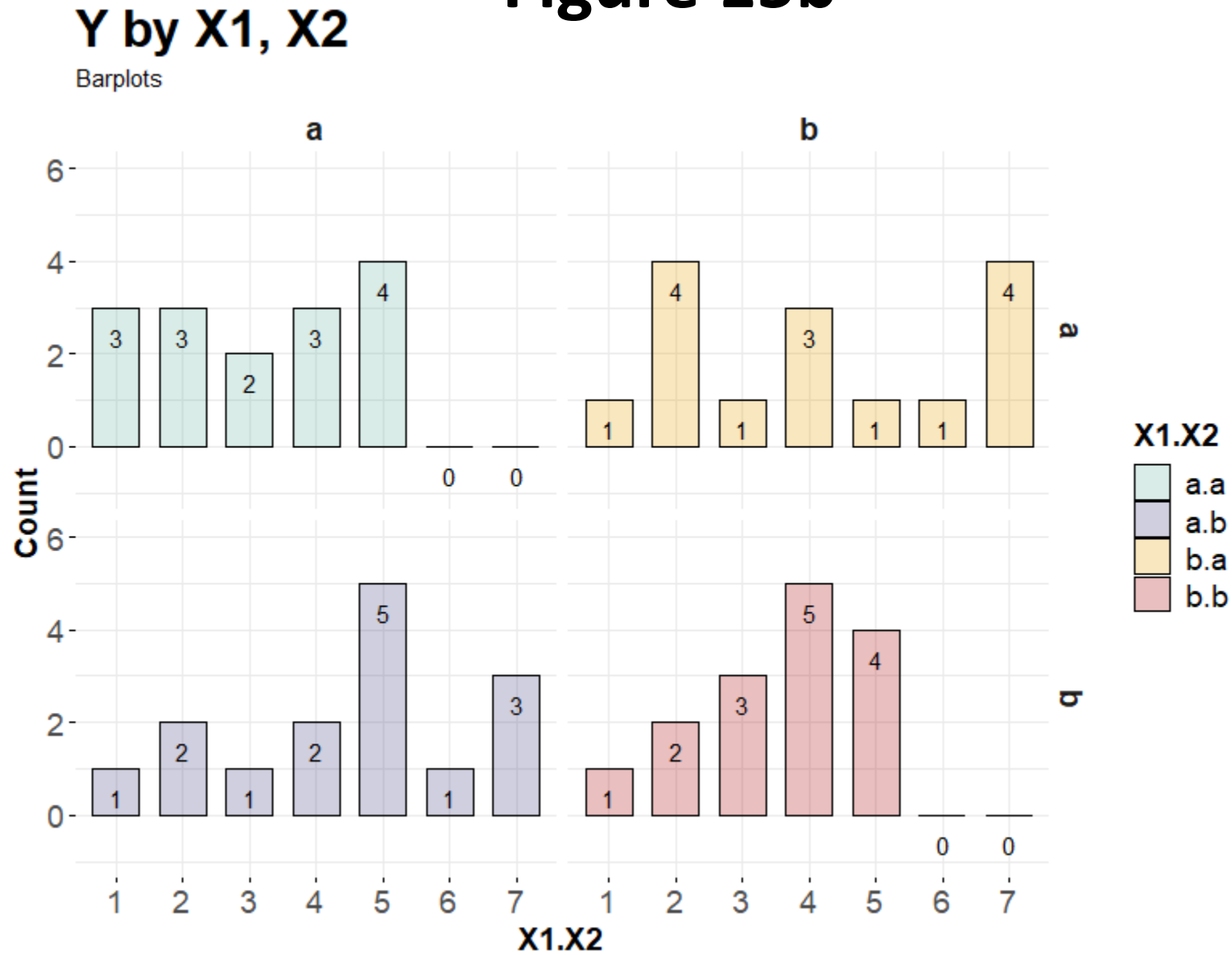


Figure 16a

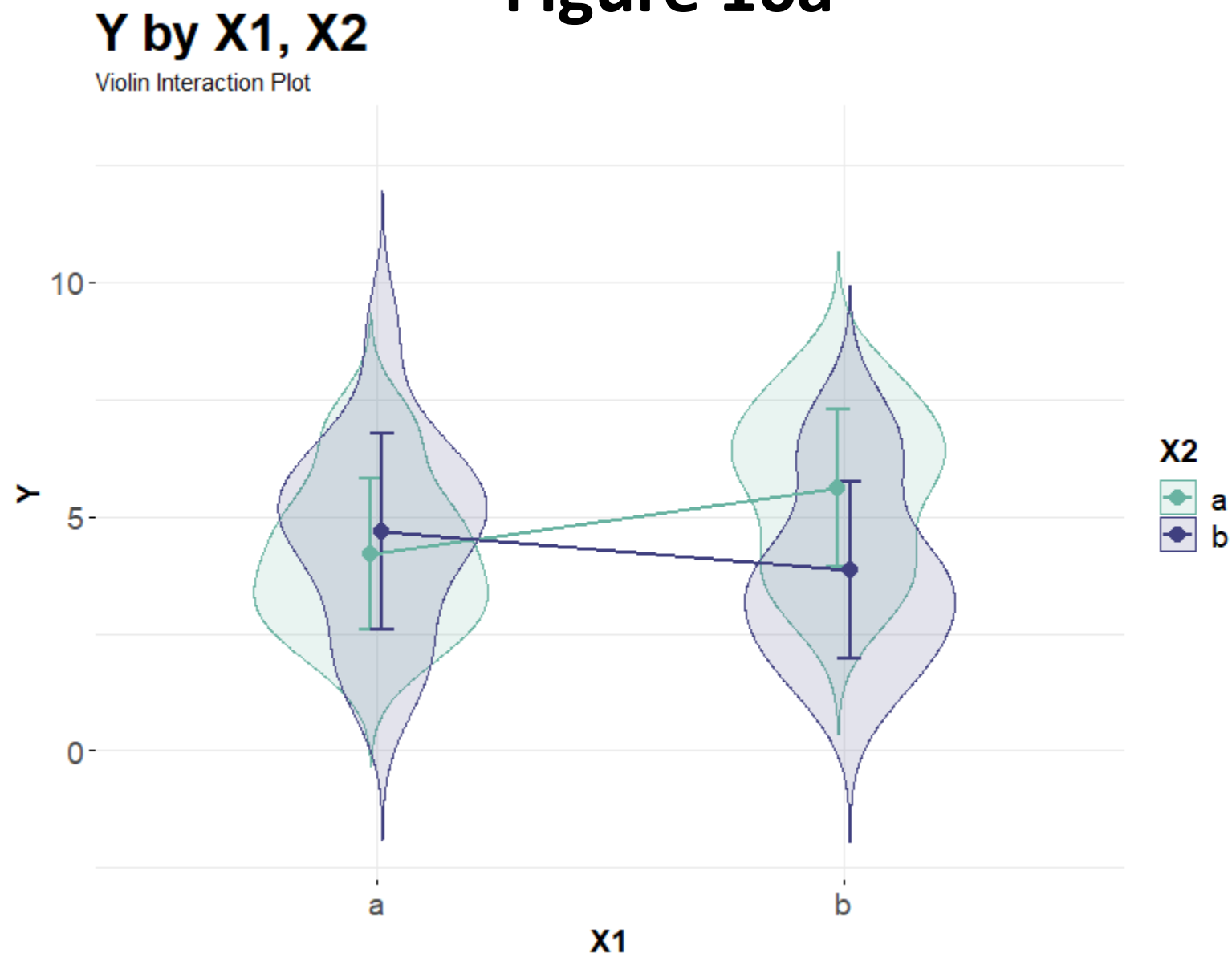
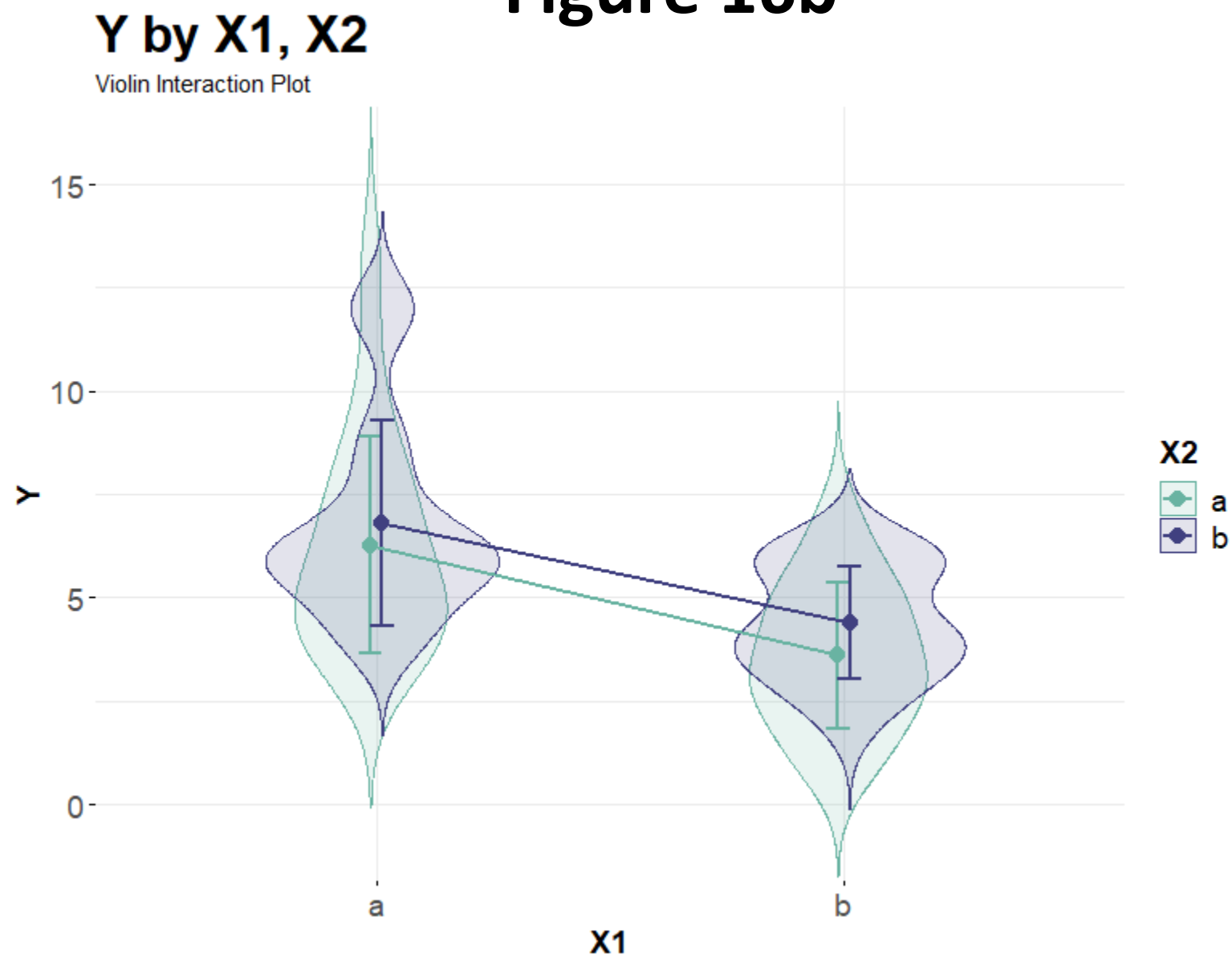


Figure 16b



GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss)	R code for GLMM (within-Ss)
Zero-Inflated Poisson	log	<i>Zero-inflated Poisson regression:</i> Counts and rare events with a large proportion of zeroes	<pre>library(pscl) # for zeroinfl library(car) # for Anova library(performance) # for check_zeroinflation df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = zeroinfl(Y ~ X1*X2, data=df, dist="poisson") check_zeroinflation(m) Anova(m, type=3)</pre>	<pre>library(glmmTMB) # for glmmTMB library(car) # for Anova library(performance) # for check_zeroinflation df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmmTMB(Y ~ X1*X2 + (1 S), data=df, family=poisson, ziformula=~X1*X2) check_zeroinflation(m) Anova(m, type=3)</pre>
Negative Binomial	log	<i>Negative binomial regression:</i> Same as Poisson but for use in the presence of overdispersion	<pre>library(MASS) # for glm.nb library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "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\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmer.nb(Y ~ X1*X2 + (1 S), data=df) Anova(m, type=3)</pre>

Notes: Negative binomial regression should be used when count data is overdispersed, which is when the variance of the distribution is greater than its mean. To test for overdispersion, one can use `AER::dispersiontest(m)`, where “m” is a fitted model with `family=poisson`. One can also use `(var(Y)/abs(mean(Y))) > 1.15`, where “Y” is the response in each condition.

GLM / GLMM

Distributions and canonical links

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

Figure 17a

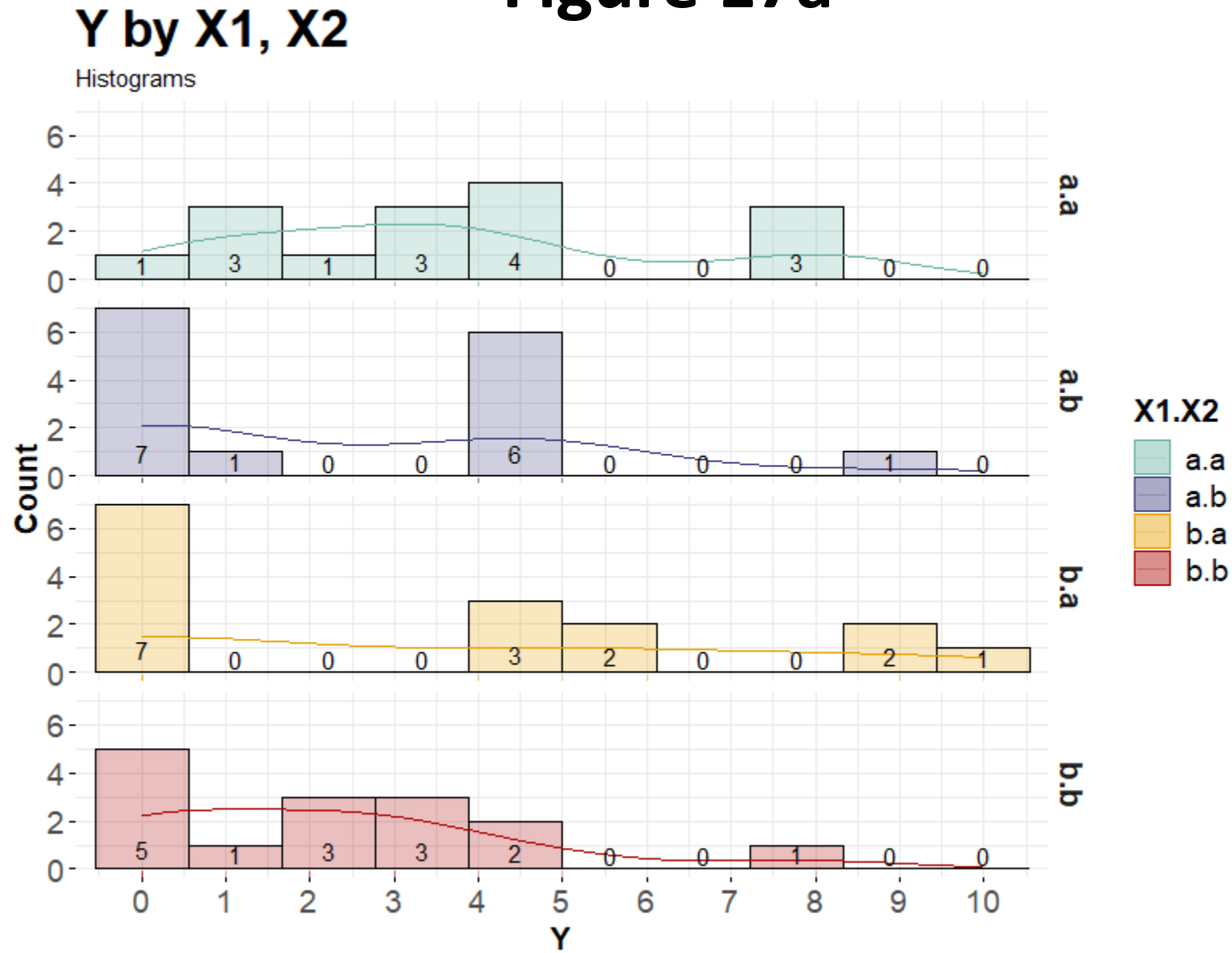


Figure 17b

Y by X1, X2

Histograms

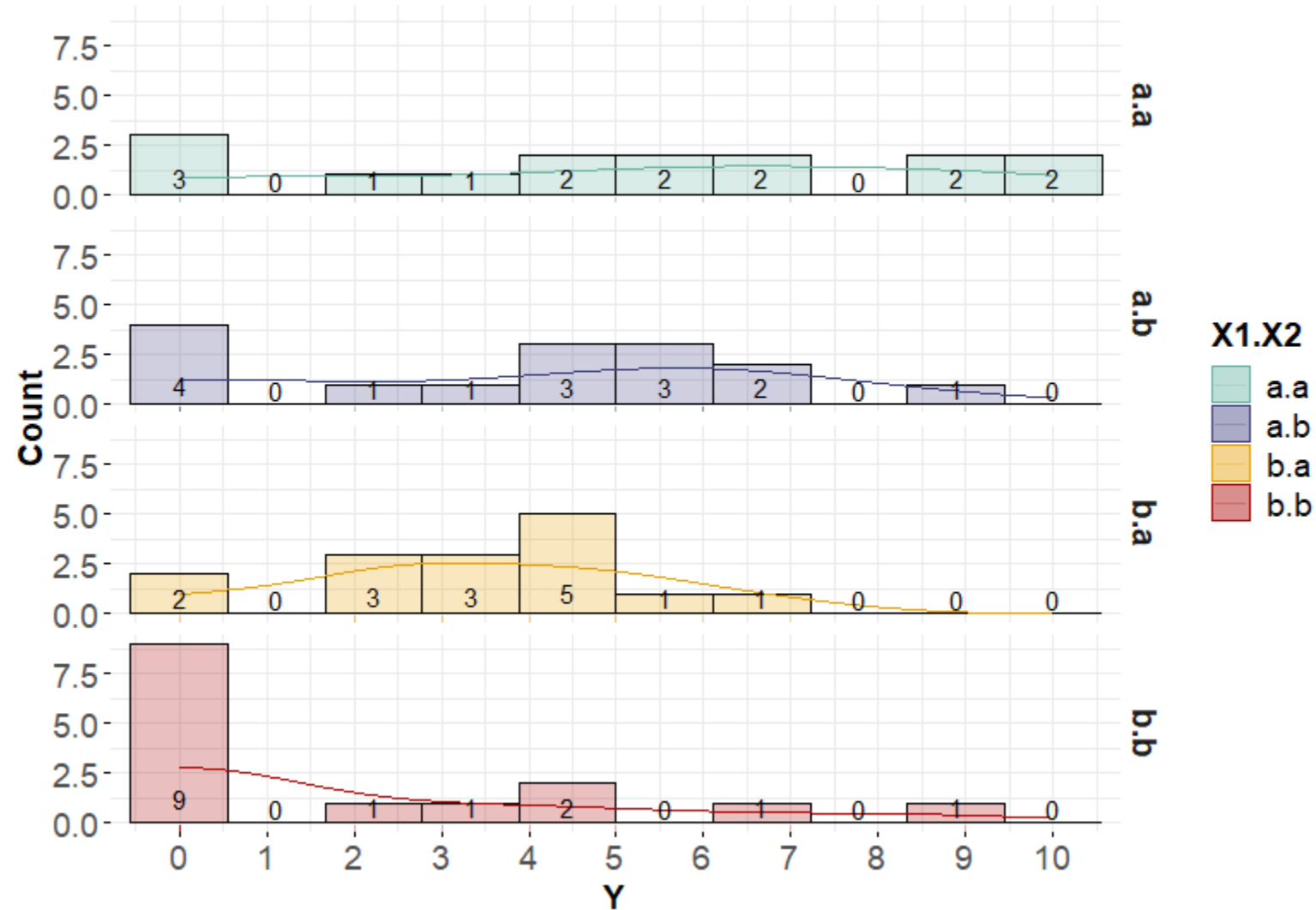


Figure 18a

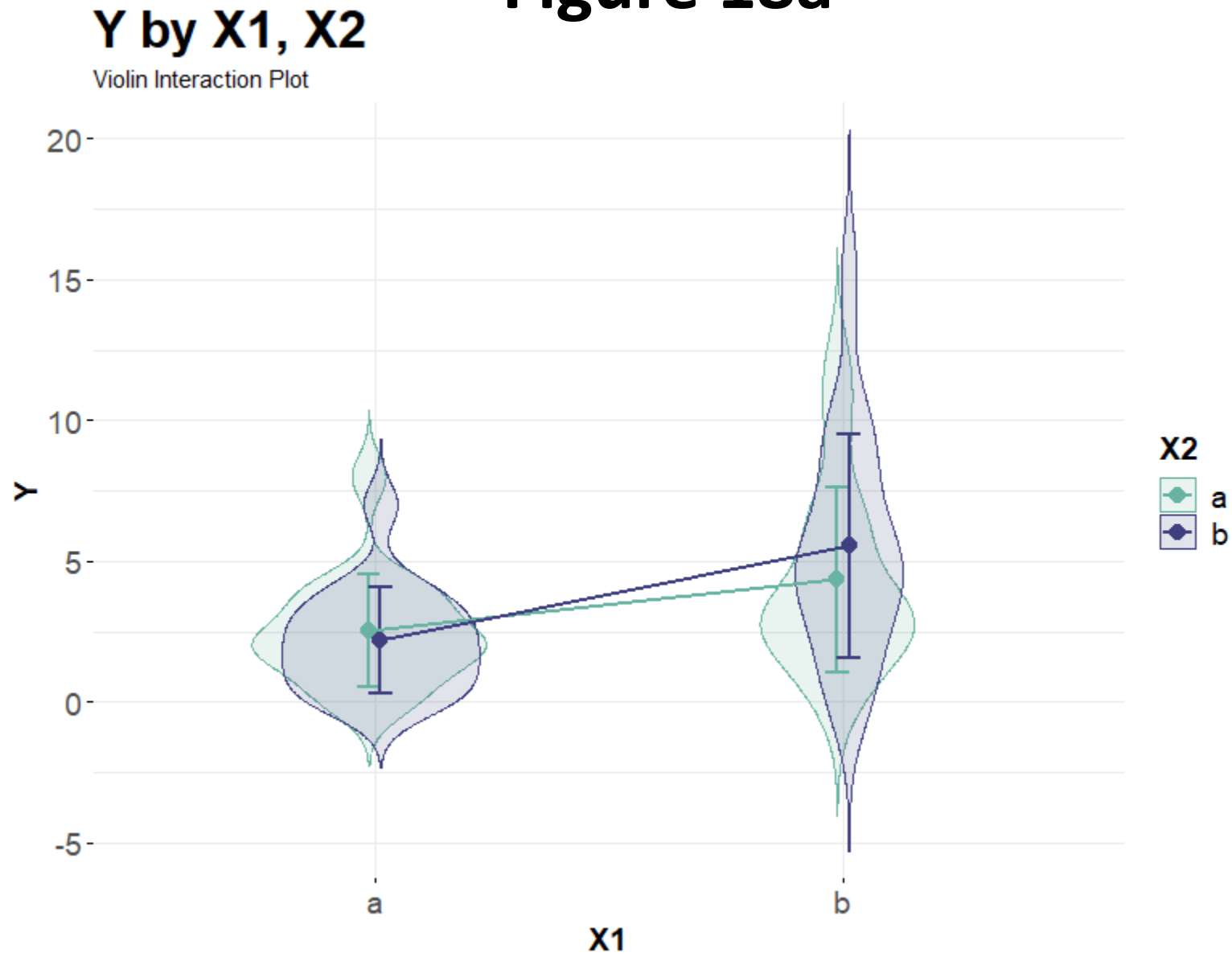
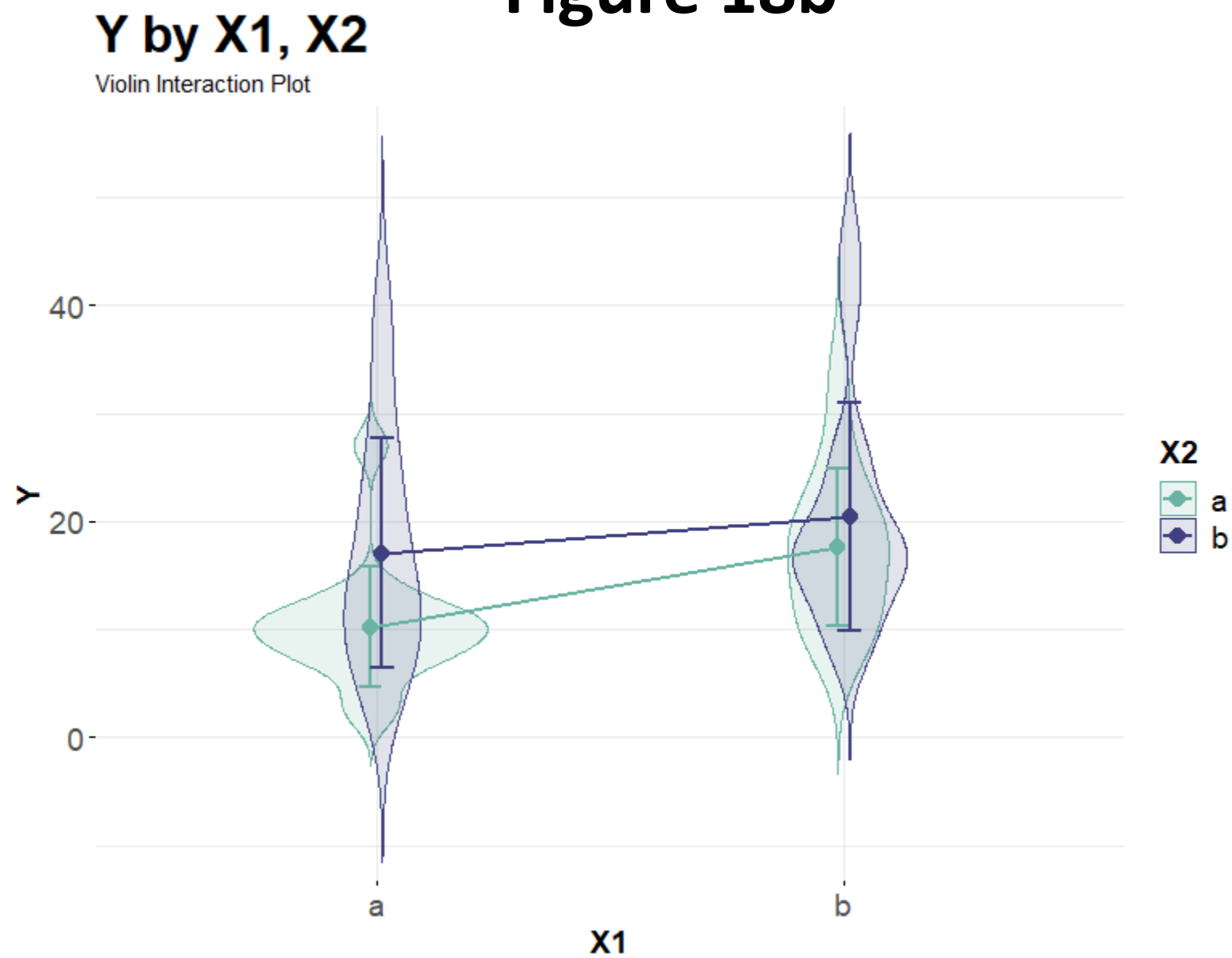


Figure 18b



GLM / GLMM

Distributions and canonical links

Distribution	Link	Typical Uses	R code for GLM (between-Ss)	R code for GLMM (within-Ss)
Zero-Inflated Negative Binomial	log	<i>Zero-inflated negative binomial regression:</i> Same as Zero-Inflated Poisson but for use in the presence of overdispersion	<pre>library(pscl) # for zeroinfl library(car) # for Anova library(performance) # for check_zeroinflation df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = zeroinfl(Y ~ X1*X2, data=df, dist="negbin") check_zeroinflation(m) Anova(m, type=3)</pre>	<pre>library(glmmTMB) # for glmmTMB library(car) # for Anova library(performance) # for check_zeroinflation df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmmTMB(Y ~ X1*X2 + (1 S), data=df, family=nbinom2, ziformula=~X1*X2) check_zeroinflation(m) Anova(m, type=3)</pre>
Gamma, including Exponential	inverse, but log is often used because 1/0 is undefined	<i>Gamma regression:</i> Exponentially distributed responses (e.g., income, wait times)	<pre>library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glm(Y ~ X1*X2, data=df, family=Gamma) # family=Gamma(link="log") is an alternative Anova(m, type=3)</pre>	<pre>library(lme4) # for glmer library(lmerTest) library(car) # for Anova df\$S = factor(df\$S) df\$X1 = factor(df\$X1) df\$X2 = factor(df\$X2) contrasts(df\$X1) <- "contr.sum" contrasts(df\$X2) <- "contr.sum" m = glmer(Y ~ X1*X2 + (1 S), data=df, family=Gamma) # family=Gamma(link="log") is an alternative 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 GLMM sample code uses a random intercept for *subject* (*S*). There are also random slope models, which are used when the response changes at different rates for each subject over the repeated factor(s). A random slope example of county population growth over time can be seen here (<https://www.youtube.com/watch?v=YDe6F7CXjWw>). 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)
Zero-Inflated Negative Binomial	log	<i>Zero-inflated negative binomial regression:</i> Same as Zero-Inflated Poisson but for use in the presence of overdispersion	“Figure 19a shows histograms of Y by X1 and X2. An analysis of variance based on zero-inflated negative binomial regression indicated no statistically significant effects on Y of X1 ($\chi^2(1, N=60) = 0.43, n.s.$), X2 ($\chi^2(1, N=60) = 1.28, n.s.$), or the X1×X2 interaction ($\chi^2(1, N=60) = 0.10, n.s.$).”	“Figure 19b shows histograms of Y by X1 and X2. An analysis of variance based on zero-inflated mixed negative binomial regression indicated a statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 31.09, p < .0001$) and a marginal X1×X2 interaction ($\chi^2(1, N=60) = 3.17, p = .075$), but no detectable effect of X2 ($\chi^2(1, N=60) = 0.26, n.s.$).”
Gamma, including Exponential	inverse, but log is often used because 1/0 is undefined	<i>Gamma regression:</i> Exponentially distributed responses (e.g., income, wait times)	“Figure 20a shows a violin interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance based on Gamma regression indicated no statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 0.40, n.s.$) or X2 ($\chi^2(1, N=60) = 0.58, n.s.$), but the X1×X2 interaction was marginal ($\chi^2(1, N=60) = 3.26, p = .071$).”	“Figure 20b shows a violin interaction plot with ± 1 standard deviation error bars for X1 and X2. An analysis of variance based on mixed Gamma regression indicated no statistically significant effect on Y of X1 ($\chi^2(1, N=60) = 1.25, n.s.$) or of X2 ($\chi^2(1, N=60) = 0.29, n.s.$), but there was a statistically significant X1×X2 interaction ($\chi^2(1, N=60) = 10.84, p < .001$).”

Figure 19a

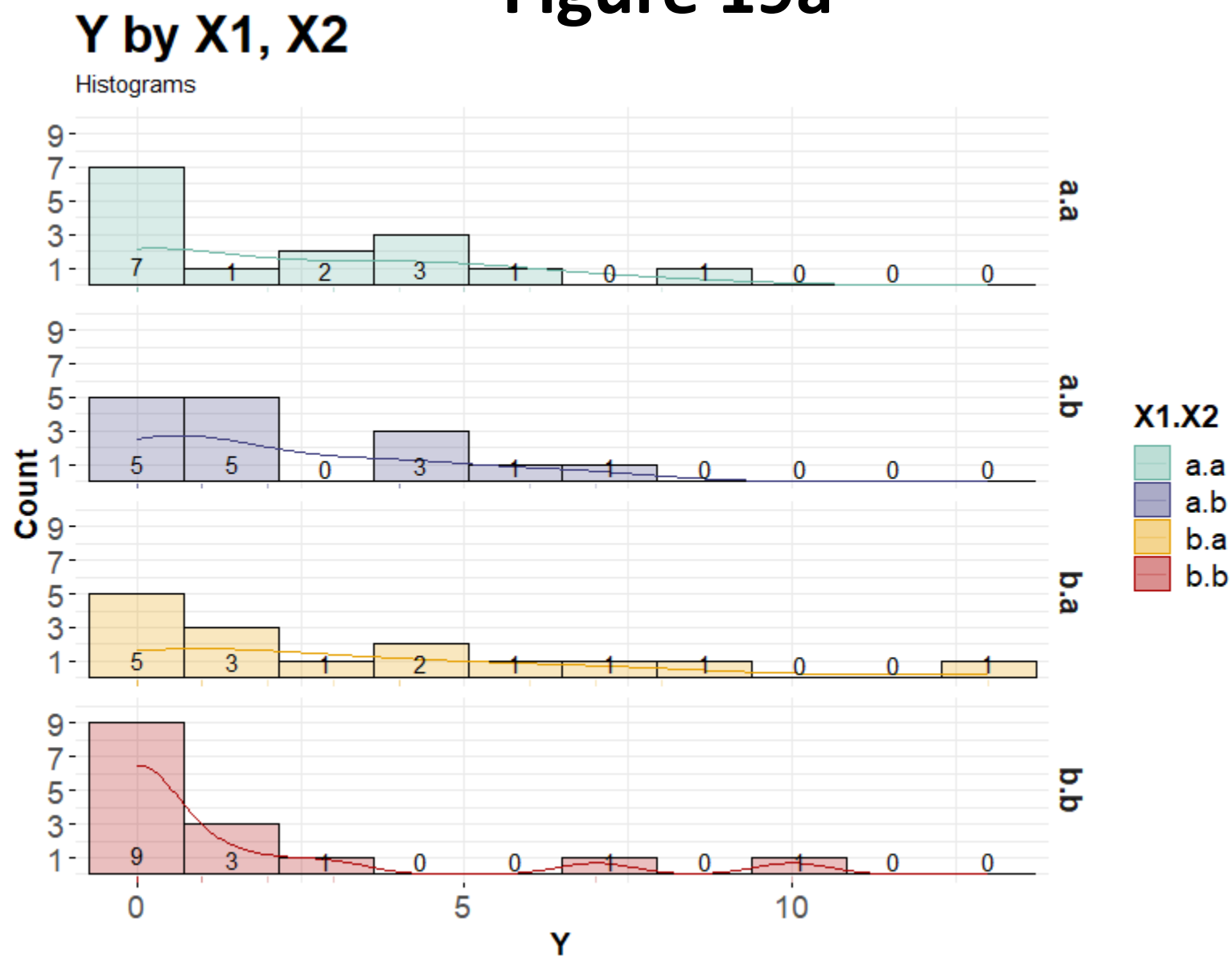


Figure 19b

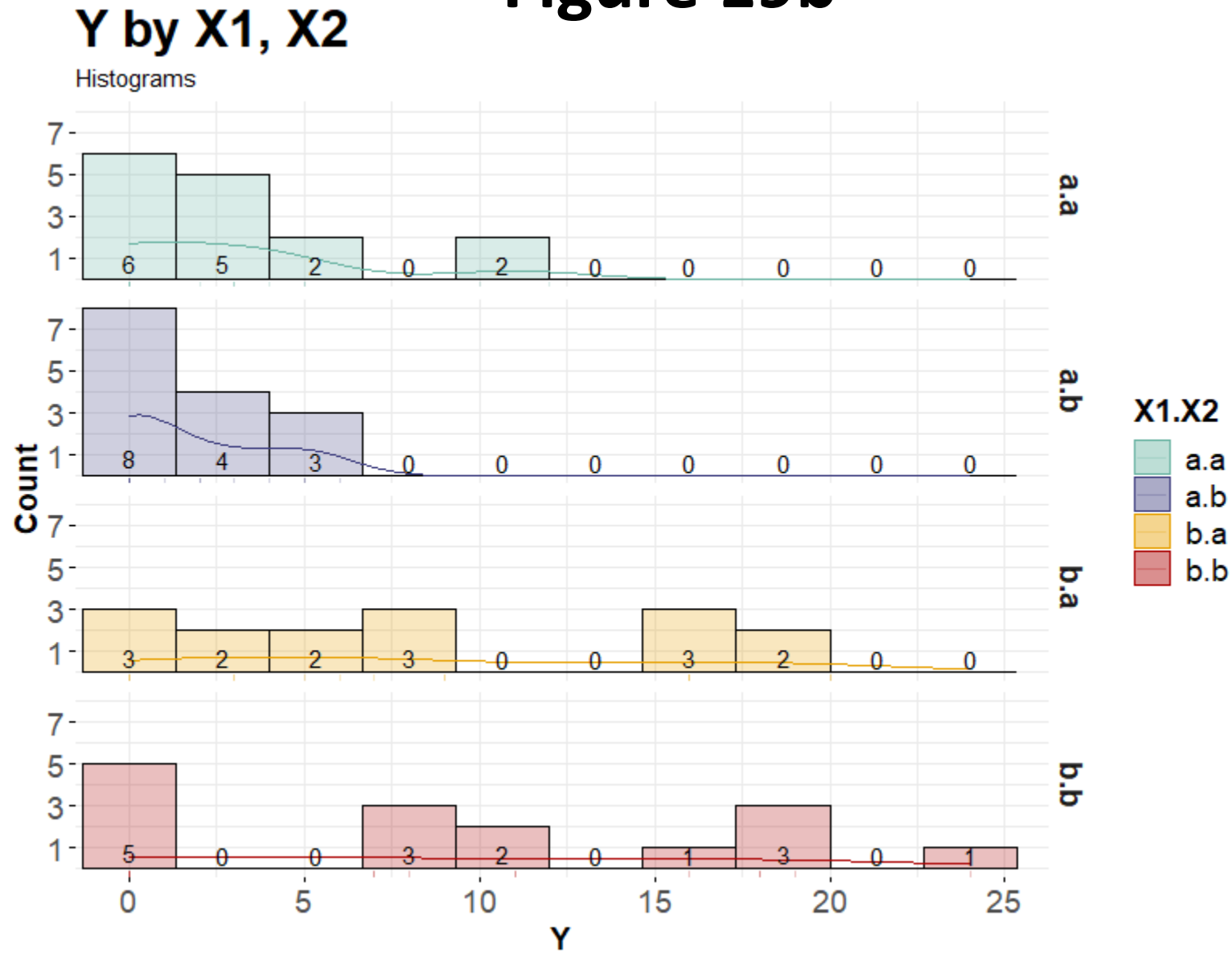


Figure 20a

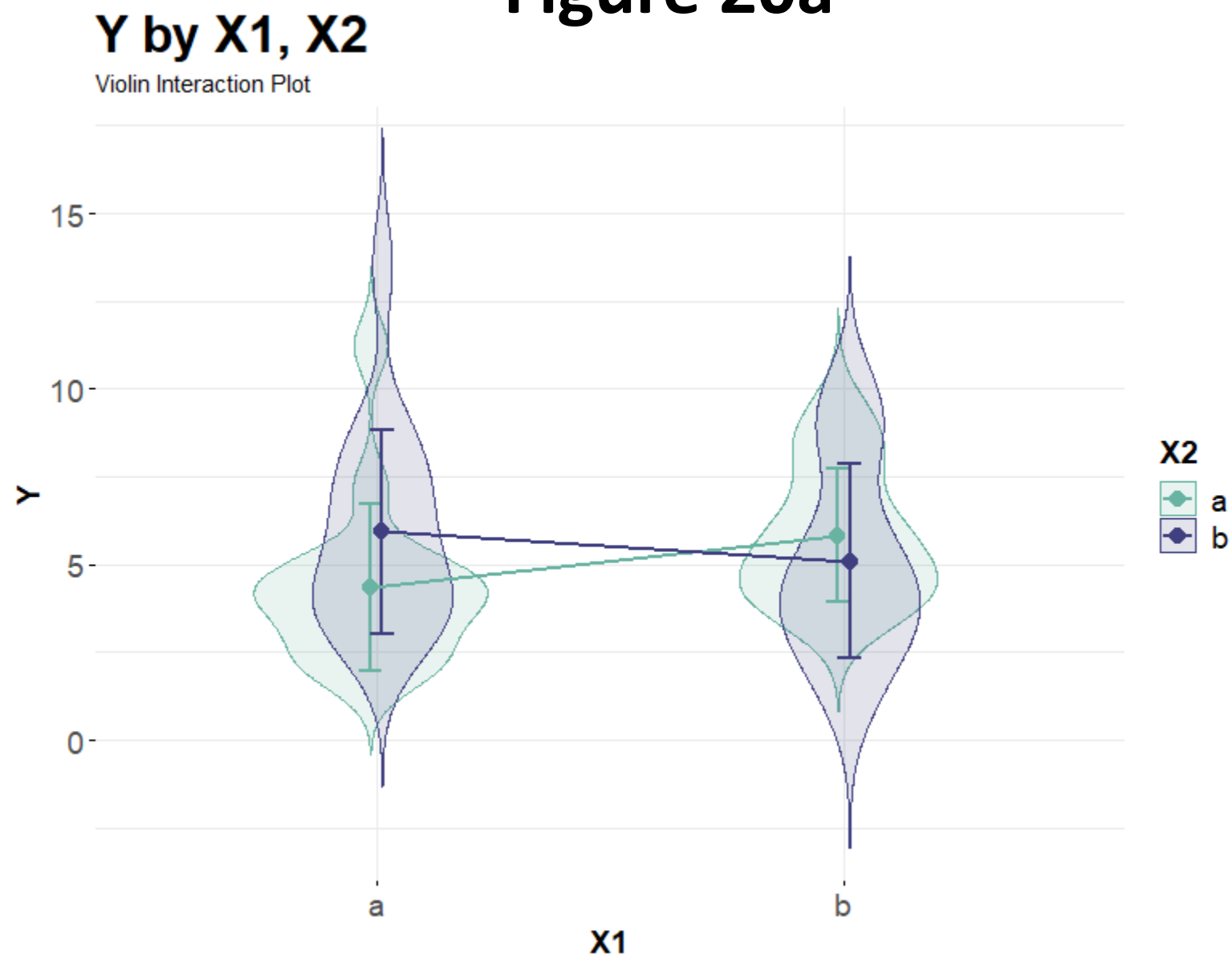
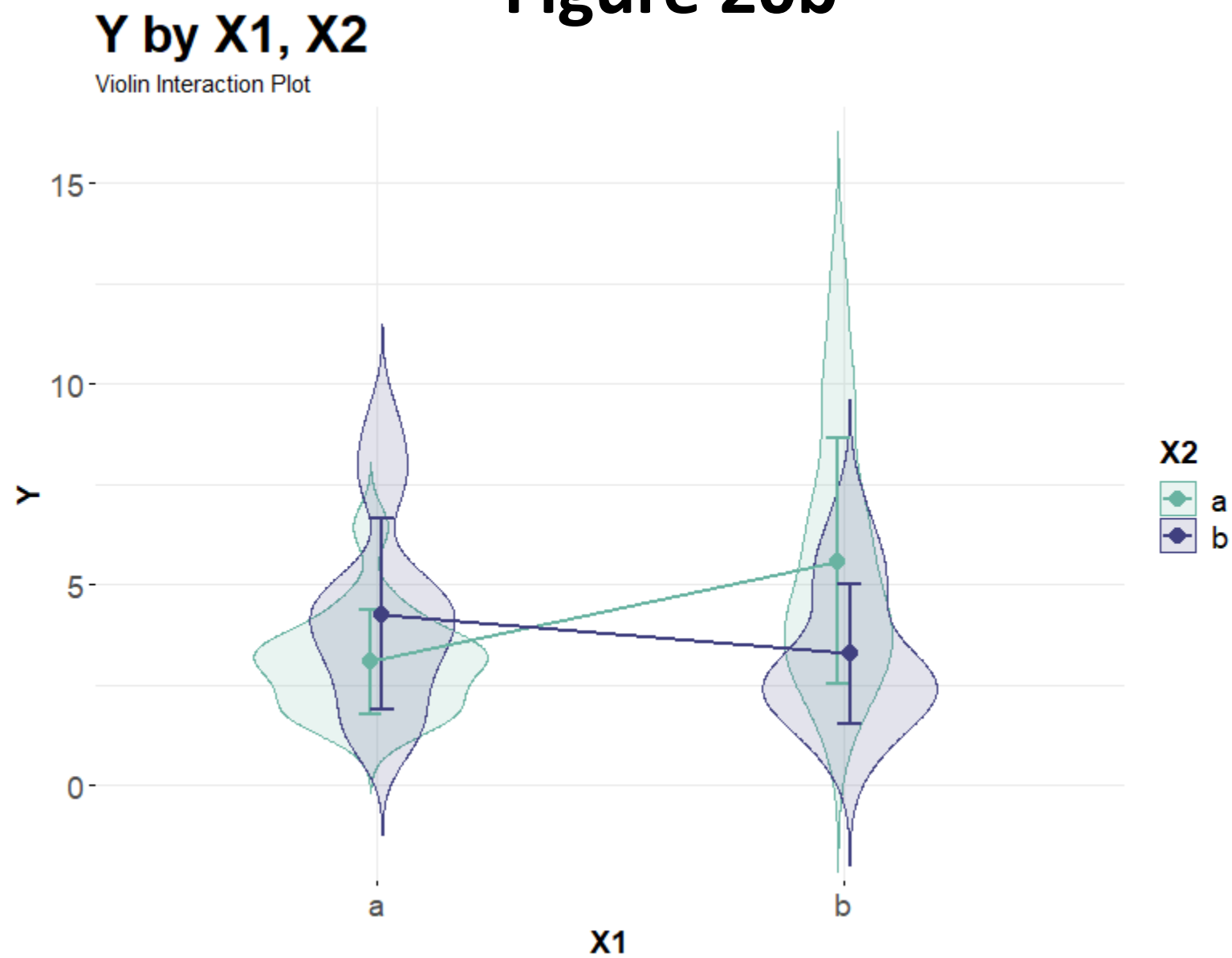


Figure 20b



Generalized Linear (Mixed) Models

Post hoc comparisons

GLM / GLMM

Post hoc pairwise comparisons – One factor

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

Note: 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	Linear regression	Z-test	Btwn	“Pairwise comparisons using <i>Z</i> -tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘b’ difference was statistically significant ($Z = 2.52, p < .05$), but those of ‘a’ vs. ‘c’ ($Z = 1.16, n.s.$) and ‘b’ vs. ‘c’ ($Z = -1.36, n.s.$) were not.”
			<i>t</i> -test		<p>“Pairwise comparisons using <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘b’ difference was statistically significant ($t(57) = 2.52, p < .05$), but those of ‘a’ vs. ‘c’ ($t(57) = 1.16, n.s.$) and ‘b’ vs. ‘c’ ($t(57) = -1.36, n.s.$) were not.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘b’ difference was statistically significant ($t(57) = 2.52, p < .05$), but those of ‘a’ vs. ‘c’ ($t(57) = 1.16, n.s.$) and ‘b’ vs. ‘c’ ($t(57) = -1.36, n.s.$) were not.”</p>
			Z-test	Within	“Pairwise comparisons using <i>Z</i> -tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ vs. ‘b’ was not statistically significantly different ($Z = 0.27, n.s.$), but ‘a’ vs. ‘c’ ($Z = 2.38, p = .052$) and ‘b’ vs. ‘c’ ($Z = 2.11, p = .070$) were marginal.”
			<i>t</i> -test		<p>“Pairwise comparisons using <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that ‘a’ vs. ‘b’ was not statistically significantly different ($t(38) = -0.27, n.s.$), but ‘a’ vs. ‘c’ ($t(38) = -2.38, p = .068$) and ‘b’ vs. ‘c’ ($t(38) = -2.11, p = .083$) were marginal.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that neither ‘a’ vs. ‘b’ ($t(38) = -0.27, n.s.$) nor ‘b’ vs. ‘c’ ($t(38) = -2.11, n.s.$) were detectably different, but ‘a’ vs. ‘c’ was marginal ($t(38) = -2.38, p = .057$).”</p>

GLM / GLMM

Post hoc pairwise comparisons – One factor

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

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	R Code
1	≥2	Multinomial logistic regression	Pairwise multinomial logistic regression as chi-squared tests	Btwn	<pre> # df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and polytomous response (Y) library(nnet) # for multinom library(car) # for Anova df2 <- df[df\$X == "a" df\$X == "b",] # a vs. b df2\$X = factor(df2\$X) contrasts(df2\$X) <- "contr.sum" m = multinom(Y ~ X, data=df2) ab = Anova(m, type=3) df2 <- df[df\$X == "a" df\$X == "c",] # a vs. c df2\$X = factor(df2\$X) contrasts(df2\$X) <- "contr.sum" m = multinom(Y ~ X, data=df2) ac = Anova(m, type=3) df2 <- df[df\$X == "b" df\$X == "c",] # b vs. c df2\$X = factor(df2\$X) contrasts(df2\$X) <- "contr.sum" m = multinom(Y ~ X, data=df2) bc = Anova(m, type=3) p.adjust(c(ab\$`Pr(>Chisq)` , ac\$`Pr(>Chisq)` , bc\$`Pr(>Chisq)`), method="holm") </pre>

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	R Code
1	≥2	Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)	Pairwise mixed Poisson regression as chi-squared tests	Within	<pre> # df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and polytomous response (Y) # df2 was produced by mlogit.data and has a logical response (Y) and alt factor library(lme4) # for glmer library(lmerTest) library(car) # for Anova df3 <- df2[df2\$X == "a" df2\$X == "b",] # a vs. b df3\$X = factor(df3\$X) contrasts(df3\$X) <- "contr.sum" m = glmer(Y ~ X*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) ab = a[grep(":alt", rownames(a)),] df3 <- df2[df2\$X == "a" df2\$X == "c",] # a vs. c df3\$X = factor(df3\$X) contrasts(df3\$X) <- "contr.sum" m = glmer(Y ~ X*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) ac = a[grep(":alt", rownames(a)),] df3 <- df2[df2\$X == "b" df2\$X == "c",] # b vs. c df3\$X = factor(df3\$X) contrasts(df3\$X) <- "contr.sum" m = glmer(Y ~ X*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) bc = a[grep(":alt", rownames(a)),] p.adjust(c(ab\$`Pr(>Chisq)`, ac\$`Pr(>Chisq)`, bc\$`Pr(>Chisq)`), method="holm") </pre>

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	Report
1	≥2	Multinomial logistic regression	Pairwise multinomial logistic regression as chi-squared tests	Btwn	“Three <i>post hoc</i> multinomial logistic regressions, one for each pair of X’s levels, corrected with Holm’s sequential Bonferroni procedure, indicated that the number of ‘x’, ‘y’, and ‘z’ responses was statistically significantly different for ‘a’ vs. ‘b’ ($\chi^2(2, N=40) = 11.81, p < .01$), ‘a’ vs. ‘c’ ($\chi^2(2, N=40) = 6.62, p < .05$), and ‘b’ vs. ‘c’ ($\chi^2(2, N=40) = 17.33, p < .001$).”
		Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)	Pairwise mixed Poisson regression as chi-squared tests	Within	“Three <i>post hoc</i> mixed Poisson regressions, one for each pair of X’s levels, corrected with Holm’s sequential Bonferroni procedure, indicated that the number of ‘x’, ‘y’, and ‘z’ responses was marginally different for ‘a’ vs. ‘c’ ($\chi^2(2, N=40) = 7.82, p = .060$), but ‘a’ vs. ‘b’ ($\chi^2(2, N=40) = 4.06, n.s.$) and ‘b’ vs. ‘c’ ($\chi^2(2, N=40) = 3.48, n.s.$) were not detectably different.”

GLM / GLMM

Post hoc pairwise comparisons – One factor

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

GLM / GLMM

Post hoc pairwise comparisons – One factor

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

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	Report
1	≥2	Poisson regression	Z-test	Btwn	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘b’ vs. ‘c’ ($Z = -2.81, p < .05$) and ‘a’ vs. ‘c’ ($Z = -2.33, p < .05$) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = 0.50, n.s.$).”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘b’ vs. ‘c’ difference was statistically significant ($Z = -2.81, p < .05$), but the ‘a’ vs. ‘c’ difference was only marginal ($Z = -2.33, p = .053$), and the ‘a’ vs. ‘b’ difference was not statistically significant ($Z = 0.50, n.s.$).”</p>
				Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference was statistically significant ($Z = 3.21, p < .01$), but differences for ‘a’ vs. ‘b’ and ‘b’ vs. ‘c’ were not.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ difference was statistically significant ($Z = 3.21, p < .01$), but differences for ‘a’ vs. ‘b’ and ‘b’ vs. ‘c’ were not.”</p>

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	R Code
1	≥2	Zero-inflated Poisson regression	Z-test	Btwn	<pre># df has subjects (S), one between-Ss factor (X) w/levels (a,b,c), and count response (Y) library(multcomp) # for adjusted library(emmeans) # for as.glht, pairs, emmeans summary(as.glht(pairs(emmeans(m, ~ X))), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X, adjust="tukey")</pre>
				Within	<pre># df has subjects (S), one within-Ss factor (X) w/levels (a,b,c), and count response (Y) library(multcomp) # for glht, mcp summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm")) # the following also performs the equivalent contrast tests library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X, adjust="tukey")</pre>

GLM / GLMM

Post hoc pairwise comparisons – One factor

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

GLM / GLMM

Post hoc pairwise comparisons – One factor

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

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	Report
1	≥2	Negative binomial regression	Z-test	Btwn	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ ($Z = -2.81, p < .01$) and ‘b’ vs. ‘c’ ($Z = -3.21, p < .01$) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = 0.40, n.s.$).”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ ($Z = -2.81, p < .05$) and ‘b’ vs. ‘c’ ($Z = -3.21, p < .01$) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = 0.40, n.s.$).”</p>
				Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ ($Z = 4.14, p < .001$) and ‘b’ vs. ‘c’ ($Z = 3.98, p < .001$) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = 0.20, n.s.$).”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ ($Z = 4.14, p < .001$) and ‘b’ vs. ‘c’ ($Z = 3.98, p < .001$) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = 0.20, n.s.$).”</p>
1	≥2	Zero-inflated negative binomial regression	Z-test	Btwn	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ ($Z = -2.35, p < .05$) and ‘b’ vs. ‘c’ ($Z = -2.90, p < .05$) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = 0.78, n.s.$).”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ ($Z = -2.35, p < .05$) and ‘b’ vs. ‘c’ ($Z = -2.90, p < .05$) differences were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = 0.78, n.s.$).”</p>
				Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference was statistically significant ($Z = 3.16, p < .01$), but not the ‘a’ vs. ‘b’ ($Z = 1.53, n.s.$) or the ‘b’ vs. ‘c’ ($Z = 1.96, n.s.$) differences.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ difference was statistically significant ($Z = 3.16, p < .01$), but not the ‘a’ vs. ‘b’ ($Z = 1.53, n.s.$) or the ‘b’ vs. ‘c’ ($Z = 1.96, n.s.$) differences.”</p>

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	R Code
1	≥2	Gamma regression, including Exponential regression	Z-test	Btwn, Within	<pre># df has subjects (S), one factor (X) w/levels (a,b,c), and continuous response (Y) library(multcomp) # for glht, mcp summary(glht(m, mcp(X="Tukey")), test=adjusted(type="holm"))</pre>
			t-test	Btwn	<pre># df has subjects (S), one factor (X) w/levels (a,b,c), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X, adjust="tukey")</pre>
			Z-test	Within	<pre># df has subjects (S), one factor (X) w/levels (a,b,c), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X, adjust="tukey")</pre>

GLM / GLMM

Post hoc pairwise comparisons – One factor

Factors	Levels	Omnibus Model	Test	B/W	Report
1	≥2	Gamma regression, including Exponential regression	Z-test	Btwn, Within	“Pairwise comparisons using <i>Z</i> -tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘b’ difference was statistically significant ($Z = 3.44, p < .01$), but differences for the other two pairwise comparisons were not.”
			<i>t</i> -test	Btwn	<p>“Pairwise comparisons using <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘b’ difference was statistically significant ($t(57) = 3.44, p < .01$), but differences for the other two pairwise comparisons were not.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘b’ difference was statistically significant ($t(57) = 3.44, p < .01$), but differences for the other two pairwise comparisons were not.”</p>
			Z-test	Within	<p>“Pairwise comparisons using <i>Z</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the ‘a’ vs. ‘c’ difference ($Z = -3.87, p < .001$) and the ‘b’ vs. ‘c’ difference ($Z = -3.63, p < .001$) were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = -0.30, n.s.$).”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the ‘a’ vs. ‘c’ difference ($Z = -3.87, p < .001$) and the ‘b’ vs. ‘c’ difference ($Z = -3.63, p < .001$) were statistically significant, but not the ‘a’ vs. ‘b’ difference ($Z = -0.30, n.s.$).”</p>

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	<pre># df has subjects (S), two factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
≥2	≥2	Logistic regression	Z-test	Btwn, Within	<pre># df has subjects (S), two factors (X1,X2) each w/levels (a,b), and dichotomous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Linear regression	<i>t</i> -test	Btwn	<p>“Pairwise comparisons using <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant ($t(56) = -3.05, p < .05$). The (b,a) vs. (a,b) difference was marginal ($t(56) = 2.59, p = .061$). The other pairwise comparisons were not statistically significantly different or marginal.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant ($t(56) = -3.05, p < .05$), and the the (b,a) vs. (a,b) difference was marginal ($t(56) = 2.59, p = .057$). The other pairwise comparisons were not statistically significantly different or marginal.”</p>
				Within	<p>“Pairwise comparisons using <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was marginal ($t(42) = -2.71, p = .058$). No other pairwise comparisons were statistically significant or marginal.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant ($t(42) = -2.71, p < .05$). The other pairwise comparisons were not statistically significant.”</p>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Logistic regression	Z-test	Btwn	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = -2.85, p < .05$). Furthermore, the (a,a) vs. (a,b) difference was marginal ($Z = -2.40, p = .083$). No other pairwise comparisons were detectably different or marginal.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = -2.85, p < .05$). Furthermore, the (a,a) vs. (a,b) difference was marginal ($Z = -2.40, p = .078$). No other pairwise comparisons were detectably different or marginal.”</p>
				Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) and (a,a) vs. (a,b) differences were both marginal (both: $Z = 2.45, p = .086$). No other pairwise comparisons were detectably different or marginal.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) and (a,a) vs. (a,b) differences were both marginal (both: $Z = 2.45, p = .068$). No other pairwise comparisons were detectably different or marginal.”</p>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Multinomial logistic regression	Pairwise multinomial logistic regression as chi-squared tests	Btwn	<pre> # df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and polytomous response (Y) library(nnet) # for multinom library(car) # for Anova df\$X12 = with(df, interaction(X1,X2)) # make combined factor for contrasts df2 <- df[df\$X12 == "a.a" df\$X12 == "a.b",] # aa vs. ab df2\$X12 = factor(df2\$X12) contrasts(df2\$X12) <- "contr.sum" m = multinom(Y ~ X12, data=df2) aa_ab = Anova(m, type=3) df2 <- df[df\$X12 == "a.a" df\$X12 == "b.a",] # aa vs. ba df2\$X12 = factor(df2\$X12) contrasts(df2\$X12) <- "contr.sum" m = multinom(Y ~ X12, data=df2) aa_ba = Anova(m, type=3) df2 <- df[df\$X12 == "a.a" df\$X12 == "b.b",] # aa vs. bb df2\$X12 = factor(df2\$X12) contrasts(df2\$X12) <- "contr.sum" m = multinom(Y ~ X12, data=df2) aa_bb = Anova(m, type=3) df2 <- df[df\$X12 == "a.b" df\$X12 == "b.a",] # ab vs. ba df2\$X12 = factor(df2\$X12) contrasts(df2\$X12) <- "contr.sum" m = multinom(Y ~ X12, data=df2) ab_ba = Anova(m, type=3) # ... Continued on next slide ... # </pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Multinomial logistic regression	Pairwise multinomial logistic regression as chi-squared tests	Btwn	<pre> # ... Continued from previous slide ... # df2 <- df[df\$X12 == "a.b" df\$X12 == "b.b",] # ab vs. bb df2\$X12 = factor(df2\$X12) contrasts(df2\$X12) <- "contr.sum" m = multinom(Y ~ X12, data=df2) ab_bb = Anova(m, type=3) df2 <- df[df\$X12 == "b.a" df\$X12 == "b.b",] # ba vs. bb df2\$X12 = factor(df2\$X12) contrasts(df2\$X12) <- "contr.sum" m = multinom(Y ~ X12, data=df2) ba_bb = Anova(m, type=3) p.adjust(c(aa_ab\$`Pr(>Chisq)` , aa_ba\$`Pr(>Chisq)` , aa_bb\$`Pr(>Chisq)` , ab_ba\$`Pr(>Chisq)` , ab_bb\$`Pr(>Chisq)` , ba_bb\$`Pr(>Chisq)`), method="holm") </pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)	Pairwise mixed Poisson regression as chi-squared tests	Within	<pre> # df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and polytomous response (Y) # df2 was produced by mlogit.data and has a logical response (Y) and alt factor library(lme4) # for glmer library(lmerTest) library(car) # for Anova df2\$X12 = with(df2, interaction(X1,X2)) # make combined factor for contrasts df3 <- df2[df2\$X12 == "a.a" df2\$X12 == "a.b",] # aa vs. ab df3\$X12 = factor(df3\$X12) contrasts(df3\$X12) <- "contr.sum" m = glmer(Y ~ X12*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) aa_ab = a[grep(":alt", rownames(a)),] df3 <- df2[df2\$X12 == "a.a" df2\$X12 == "b.a",] # aa vs. ba df3\$X12 = factor(df3\$X12) contrasts(df3\$X12) <- "contr.sum" m = glmer(Y ~ X12*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) aa_ba = a[grep(":alt", rownames(a)),] df3 <- df2[df2\$X12 == "a.a" df2\$X12 == "b.b",] # aa vs. bb df3\$X12 = factor(df3\$X12) contrasts(df3\$X12) <- "contr.sum" m = glmer(Y ~ X12*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) aa_bb = a[grep(":alt", rownames(a)),] # ... Continued on next slide ... # </pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)	Pairwise mixed Poisson regression as chi-squared tests	Within	<pre> # ... Continued from previous slide ... # df3 <- df2[df2\$X12 == "a.b" df2\$X12 == "b.a",] # ab vs. ba df3\$X12 = factor(df3\$X12) contrasts(df3\$X12) <- "contr.sum" m = glmer(Y ~ X12*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) ab_ba = a[grep("alt", rownames(a)),] df3 <- df2[df2\$X12 == "a.b" df2\$X12 == "b.b",] # ab vs. bb df3\$X12 = factor(df3\$X12) contrasts(df3\$X12) <- "contr.sum" m = glmer(Y ~ X12*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) ab_bb = a[grep("alt", rownames(a)),] df3 <- df2[df2\$X12 == "b.a" df2\$X12 == "b.b",] # ba vs. bb df3\$X12 = factor(df3\$X12) contrasts(df3\$X12) <- "contr.sum" m = glmer(Y ~ X12*alt + (1 S), data=df3, family=poisson) a = Anova(m, type=3) ba_bb = a[grep("alt", rownames(a)),] p.adjust(c(aa_ab\$`Pr(>Chisq)` , aa_ba\$`Pr(>Chisq)` , aa_bb\$`Pr(>Chisq)` , ab_ba\$`Pr(>Chisq)` , ab_bb\$`Pr(>Chisq)` , ba_bb\$`Pr(>Chisq)`), method="holm") </pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Multinomial logistic regression	Pairwise multinomial logistic regression as chi-squared tests	Btwn	“Six <i>post hoc</i> multinomial logistic regressions, one for each combination of the levels of X1 and X2, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (a,b) difference was statistically significant ($\chi^2(2, N=30) = 13.58, p < .01$), as was the (a,b) vs. (b,a) difference ($\chi^2(2, N=30) = 20.00, p < .001$). No other pairwise comparisons were statistically significant.”
		Multinomial logistic regression via the multinomial-Poisson transformation (Baker 1994)	Pairwise mixed Poisson regression as chi-squared tests	Within	“Six <i>post hoc</i> mixed Poisson regressions, one for each combination of the levels of X1 and X2, corrected with Holm’s sequential Bonferroni procedure, indicated that the difference in the number of ‘x’, ‘y’, and ‘z’ responses was marginal for (a,b) vs. (b,b) ($\chi^2(2, N=30) = 8.68, p = .078$). No other pairwise comparisons were detectably different or marginal.”

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Ordinal logistic regression	Z-test	Btwn, Within	<pre># df has subjects (S), two factors (X1,X2) each w/levels (a,b), and ordinal response (Y) library(multcomp) # for adjusted library(emmeans) # for as.glht, pairs, emmeans summary(as.glht(pairs(emmeans(m, ~ X1*X2))), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
≥2	≥2	Poisson regression	Z-test	Btwn, Within	<pre># df has subjects (S), two factors (X1,X2) each w/levels (a,b), and count response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
≥2	≥2	Zero-inflated Poisson regression	Z-test	Btwn	<pre># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and count response (Y) library(multcomp) # for adjusted library(emmeans) # for as.glht, pairs, emmeans summary(as.glht(pairs(emmeans(m, ~ X1*X2))), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
				Within	<pre># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and count response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

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

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Zero-inflated Poisson regression	Z-test	Btwn	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant or marginal.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no comparisons were statistically significant or marginal.”</p>
				Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = 2.89, p < .05$). No other comparisons were detectably different.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = 2.89, p < .05$). No other comparisons were detectably different.”</p>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Negative binomial regression	Z-test	Btwn, Within	<pre># df has subjects (S), two factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
≥2	≥2	Zero-inflated negative binomial regression	Z-test	Btwn	<pre># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and count response (Y) library(multcomp) # for adjusted library(emmeans) # for as.glht, pairs, emmeans summary(as.glht(pairs(emmeans(m, ~ X1*X2))), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
				Within	<pre># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and count response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Negative binomial regression	Z-test	Btwn	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the differences between (a,a) vs. (b,b) ($Z = -2.82, p < .05$) and (a,b) vs. (b,b) ($Z = -3.25, p < .01$) were statistically significant. In addition, the difference between (b,a) vs. (a,b) was marginal ($Z = 2.34, p = .078$). No other differences were statistically significant or marginal.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the differences between (a,a) vs. (b,b) ($Z = -2.82, p < .05$) and (a,b) vs. (b,b) ($Z = -3.25, p < .01$) were statistically significant. In addition, the difference between (b,a) vs. (a,b) was marginal ($Z = 2.34, p = .090$). No other differences were statistically significant or marginal.”</p>
				Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the differences between (a,a) vs. (b,a) ($Z = -3.06, p < .05$), (a,a) vs. (a,b) ($Z = -2.85, p < .05$), and (a,a) vs. (b,b) ($Z = -3.89, p < .001$) were all statistically significant. No other comparisons were statistically significant.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the differences between (a,a) vs. (b,a) ($Z = -3.06, p < .05$), (a,a) vs. (a,b) ($Z = -2.85, p < .05$), and (a,a) vs. (b,b) ($Z = -3.89, p < .001$) were all statistically significant. No other comparisons were statistically significant.”</p>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Zero-inflated negative binomial regression	Z-test	Btwn	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no comparisons were statistically significant.”</p>
				Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that these differences were statistically significant: (a,a) vs. (b,a) ($Z = -2.95, p < .01$); (a,a) vs. (b,b) ($Z = -3.87, p < .001$); (b,a) vs. (a,b) ($Z = 4.04, p < .001$); and (a,b) vs. (b,b) ($Z = -4.81, p < .0001$). No other comparisons were significantly different.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that these differences were statistically significant: (a,a) vs. (b,a) ($Z = -2.95, p < .05$); (a,a) vs. (b,b) ($Z = -3.87, p < .001$); (b,a) vs. (a,b) ($Z = 4.04, p < .001$); and (a,b) vs. (b,b) ($Z = -4.81, p < .0001$). No other comparisons were significantly different.”</p>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	R Code
≥2	≥2	Gamma regression, including exponential regression	t-test	Btwn	<pre># df has subjects (S), two between-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>
			Z-test	Within	<pre># df has subjects (S), two within-Ss factors (X1,X2) each w/levels (a,b), and continuous response (Y) library(multcomp) # for glht library(emmeans) # for emm, emmeans summary(glht(m, emm(pairwise ~ X1*X2)), test=adjusted(type="holm")) # or, using the Tukey HSD correction instead of Holm's emmeans(m, pairwise ~ X1*X2, adjust="tukey")</pre>

GLM / GLMM

Post hoc pairwise comparisons – Multiple factors

Factors	Levels	Omnibus Model	Test	B/W	Report
≥2	≥2	Gamma regression, including exponential regression	t-test	Btwn	<p>“Pairwise comparisons using <i>t</i>-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that no comparisons were statistically significant.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that no comparisons were statistically significant.”</p>
			Z-test	Within	<p>“Pairwise comparisons using Z-tests, corrected with Holm’s sequential Bonferroni procedure, indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = 3.18, p < .01$). Also, the (b,a) vs. (b,b) difference was statistically significant ($Z = -2.89, p < .05$). No other pairwise comparisons were statistically significant.”</p> <p>“Simultaneous pairwise comparisons using Tukey’s HSD test indicated that the (a,a) vs. (b,a) difference was statistically significant ($Z = 3.18, p < .01$). Also, the (b,a) vs. (b,b) difference was significant ($Z = -2.89, p < .05$). No other pairwise comparisons were statistically significant.”</p>

Bibliography

of Canonical Sources

Bibliography

Proportion & Association:

- 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, no. 302, pp. 157-175. <https://doi.org/10.1080/14786440009463897>
- Sokal, R.R. and Rohlf, F.J. (1981). *Biometry: The Principles and Practice of Statistics in Biological Research*. Oxford, England: W. H. Freeman. <https://books.google.com/books?id=C-OTQgAACAAJ>

Distributions:

- Anderson, T.W. and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* 23 (2), pp. 193-212. <https://www.jstor.org/stable/2236446>
- Anderson, T.W. and Darling, D.A. (1954). A test of goodness of fit. *Journal of the American Statistical Association* 49 (268), pp. 765-769. <https://www.jstor.org/stable/2281537>
- Conover, W.J. (1972). A Kolmogorov goodness-of-fit test for discontinuous distributions. *Journal of the American Statistical Association* 67 (339), pp. 591-596. <https://doi.org/10.2307/2284444>
- Kolmogorov, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* 4, pp. 83-91.
- Massey, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* 46 (253), pp. 68-78. <https://doi.org/10.2307/2280095>
- 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>
- Smirnov, H. (1939). Sur les écarts de la courbe de distribution empirique. *Recueil Mathématique (Matematicheskii Sbornik)* 6, pp. 3-26.

Bibliography

ANOVA Assumptions:

- 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>
- 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>

Linear Models (LM):

- Fisher, R.A. (1921). On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron* 1 (4), pp. 3-32. <https://hdl.handle.net/2440/15169>
- Fisher, R.A. (1925). *Statistical Methods for Research Workers*. Edinburgh, Scotland: Oliver and Boyd.
- 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>

Bibliography

Linear Mixed Models (LMM):

- 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>
- 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>
- Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: 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>

Classic Nonparametric Tests:

- 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>
- 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>
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin* 1 (6), pp. 80-83. <https://doi.org/10.2307/3001968>

Bibliography

Aligned Rank Transform (ART):

- 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: ACM Press, pp. 754-768. <https://doi.org/10.1145/3472749.3474784>
- 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>
- 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>
- 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: ACM Press, pp. 143-146. <https://doi.org/10.1145/1978942.1978963>

Generalized Linear Models (GLM):

- 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>

Generalized Linear Mixed Models (GLMM):

- 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>

Bibliography

Binomial Regression:

- 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>

Multinomial Regression:

- 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>
- 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>

Multinomial-Poisson Transformation:

- Baker, S.G. (1994). The multinomial-Poisson transformation. *The Statistician* 43 (4), pp. 495-504. <https://doi.org/10.2307/2348134>
- 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>

Bibliography

Ordinal Regression:

- Agresti, Alan (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>

Poisson Regression:

- Vermunt, J.K. (1997). *Log-linear Models for Event Histories*. Thousand Oaks, CA: Sage Publications. <https://us.sagepub.com/en-us/nam/log-linear-models-for-event-histories/book6350>
- 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

Negative Binomial Regression:

- Hilbe, J.M. (2011). *Negative Binomial Regression, 2nd ed*. Cambridge, UK: Cambridge University Press.
<https://www.amazon.com/Negative-Binomial-Regression-Joseph-Hilbe/dp/0521198151>

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>
- 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: ACM Press, pp. 754-768. <https://doi.org/10.1145/3472749.3474784>
- 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.