NMSA407 Linear Regression

Presentation of plots used in the lectures

Michal Kulich

KPMS MFF UK

matfyz

July 24, 2023

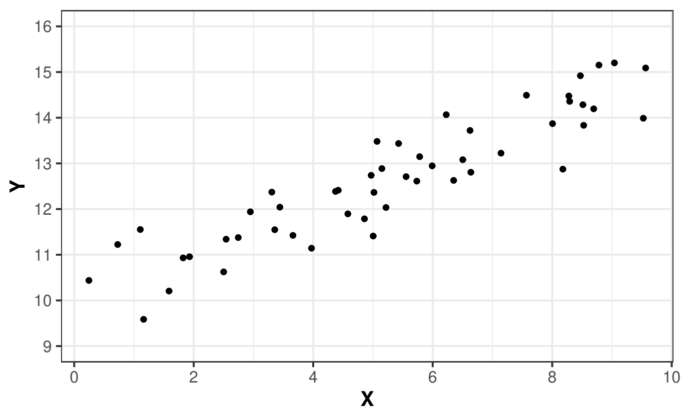


Figure: Scatterplot of two continuous variables in $\mathbb{R}^2$.

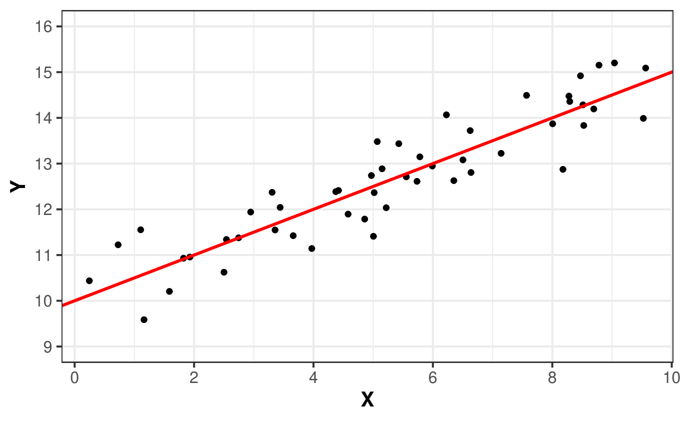


Figure: Scatterplot of two continuous variables in $\mathbb{R}^2$ with fitted line.

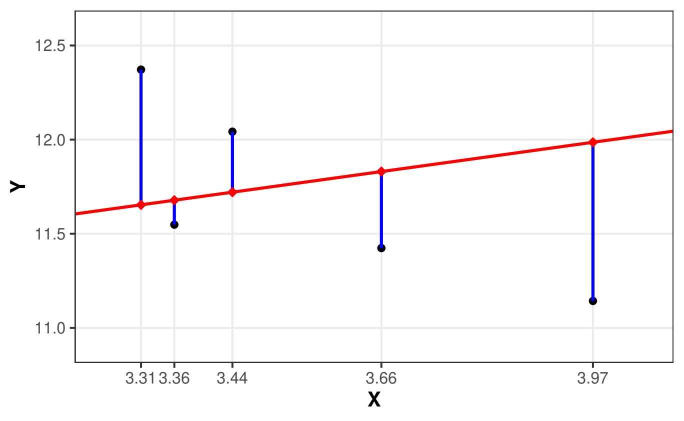


Figure: Zoom into the previous plot with visualized vertical distances (blue). Illustration of the least squares method.

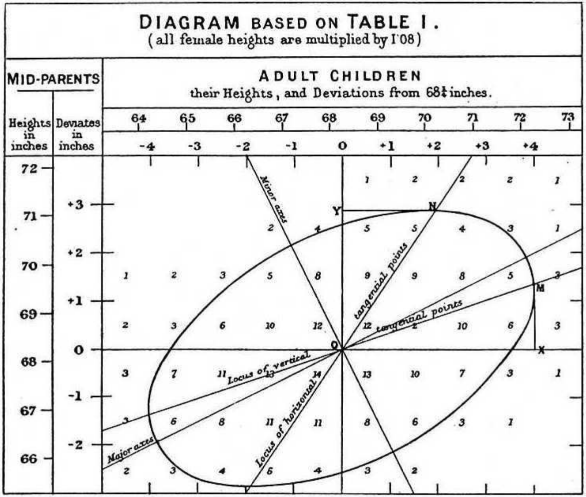
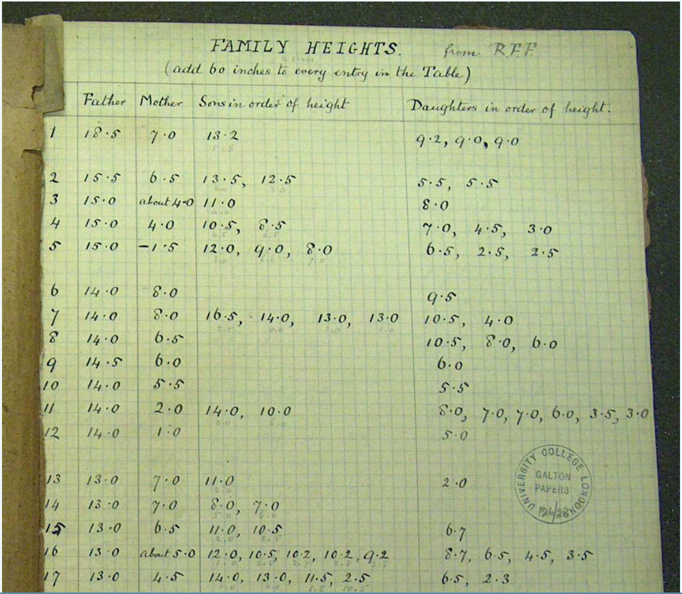


Figure: Galton height data: original visualization by the author.



FAMILY HEIGHTS. from RFF

(add 60 inches to every entry in the Table)

	Father	Mother	Sons in order of height	Daughters in order of height.
1	18·5	7·0	13·2	9·2, 9·0, 9·0
2	15·5	6·5	13·5, 12·5	5·5, 5·5
3	15·0	about 4·0	11·0	8·0
4	15·0	4·0	10·5, 8·5	7·0, 4·5, 3·0
5	15·0	−1·5	12·0, 9·0, 8·0	6·5, 2·5, 2·5
6	14·0	8·0		9·5
7	14·0	8·0	16·5, 14·0, 13·0, 13·0	10·5, 4·0
8	14·0	6·5		10·5, 8·0, 6·0
9	14·5	6·0		6·0
10	14·0	5·5		5·5
11	14·0	2·0	14·0, 10·0	8·0, 7·0, 7·0, 6·0, 3·5, 3·0
12	14·0	1·0		5·0
13	13·0	7·0	11·0	2·0
14	13·0	7·0	8·0, 7·0	
15	13·0	6·5	11·0, 10·5	6·7
16	13·0	about 5·0	12·0, 10·5, 10·2, 10·2, 9·2	8·7, 6·5, 4·5, 3·5
17	13·0	4·5	14·0, 13·0, 11·5, 2·5	6·5, 2·3

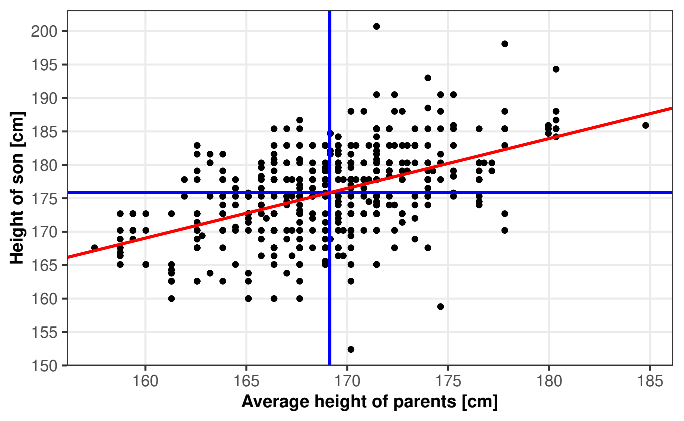


Figure: Modified Galton data with fitted least squares line (red). The slope of the line is ≈ 0.74. The means of the two variables are plotted as blue lines.

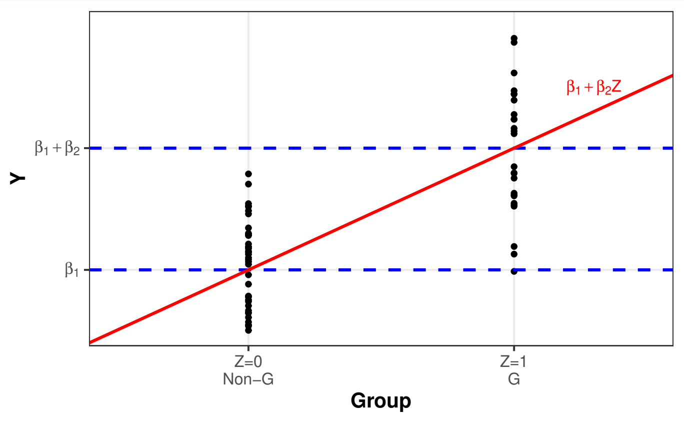


Figure: Two sample problem expressed as a linear regression model $\mathrm{E}\,Y = \beta_1 + \beta_2 Z$, where $Z = \mathbb{1}(G)$. The regression line has no interpretation except at $Z = 0$ or $Z = 1$.

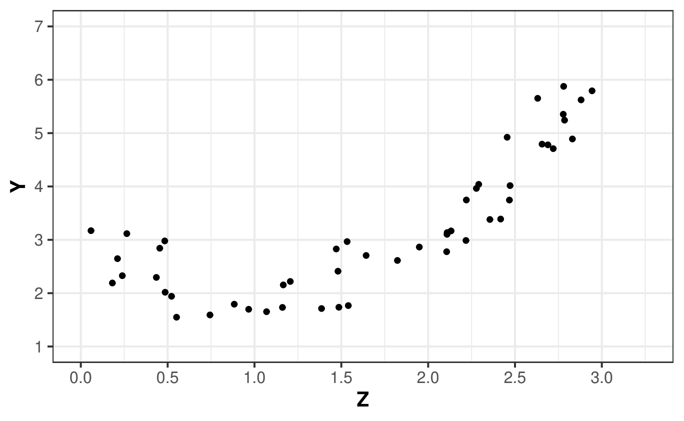


Figure: Data following a quadratic association.

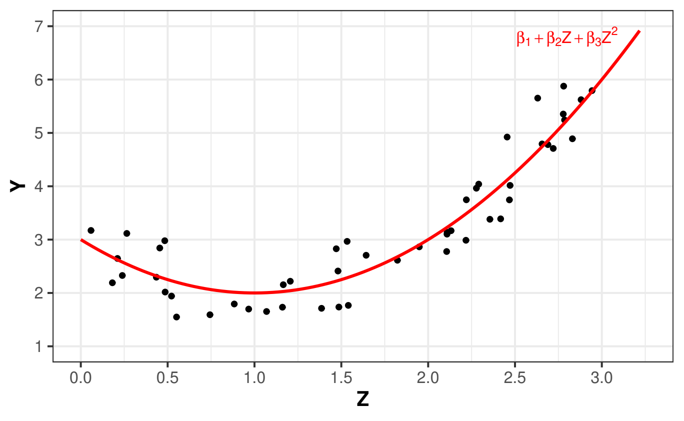


Figure: Data following a quadratic association with a fitted quadratic function.

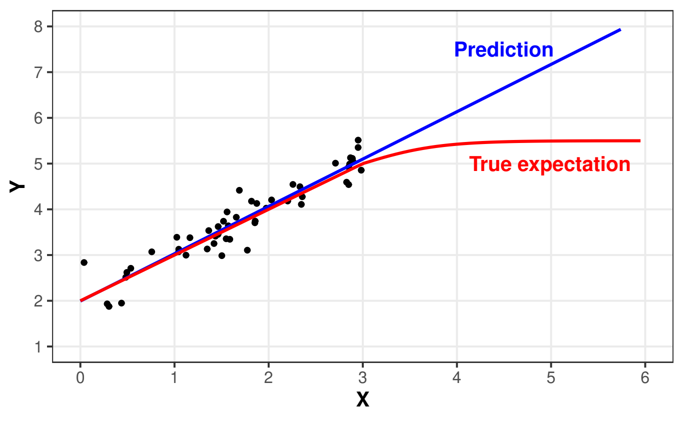


Figure: Extrapolation beyond the range of data: case 1.

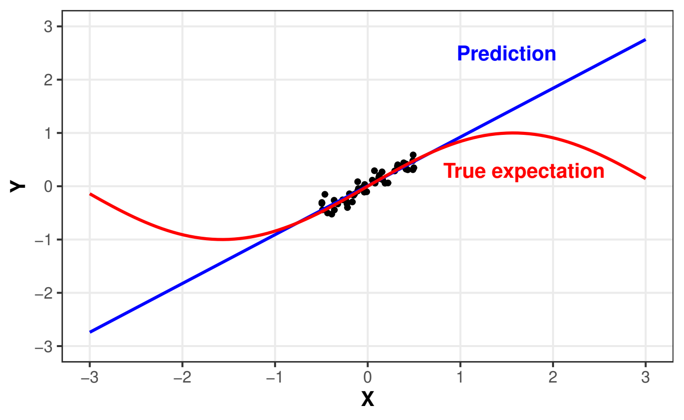


Figure: Extrapolation beyond the range of data: case 2.

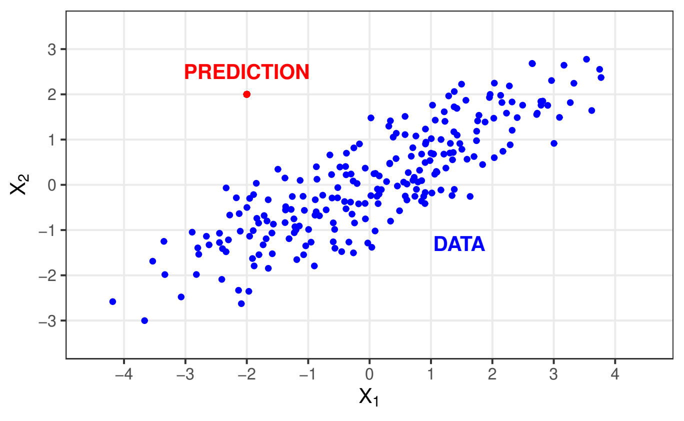


Figure: Extrapolation beyond the range of data: two covariates, the prediction is made within the range of both.

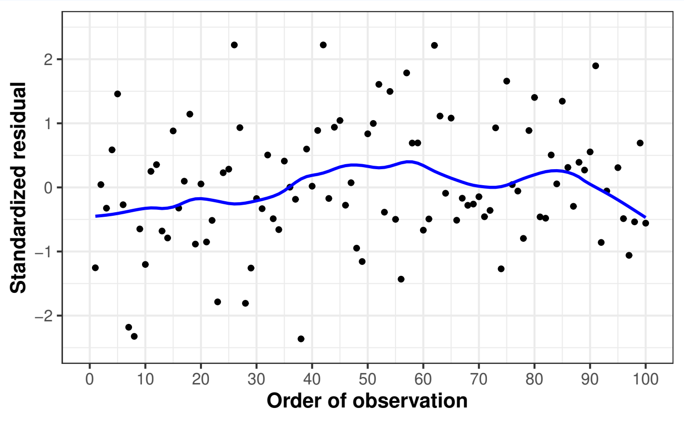


Figure: Standardized residuals against observation order: assumptions satisfied. Smoothed by lowess smoother with window over 1/4 of the data range (blue).

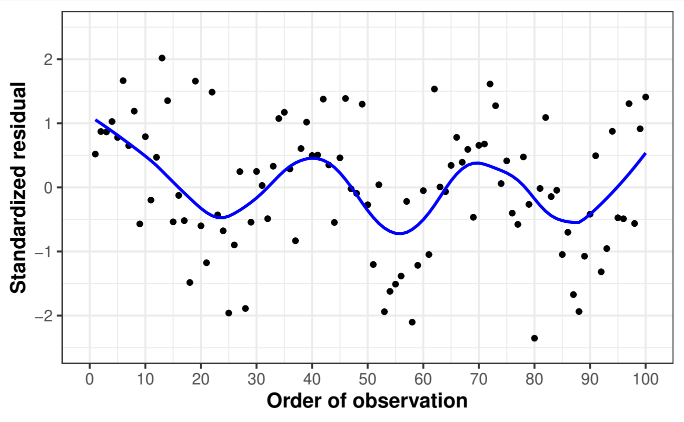


Figure: Standardized residuals against observation order: an uncaptured periodic effect. Smoothed by lowess smoother with window over 1/4 of the data range (blue).

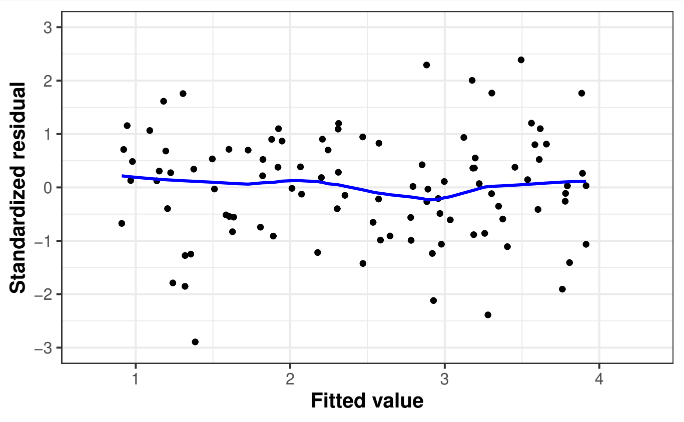


Figure: Standardized residuals against fitted values: assumptions satisfied. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

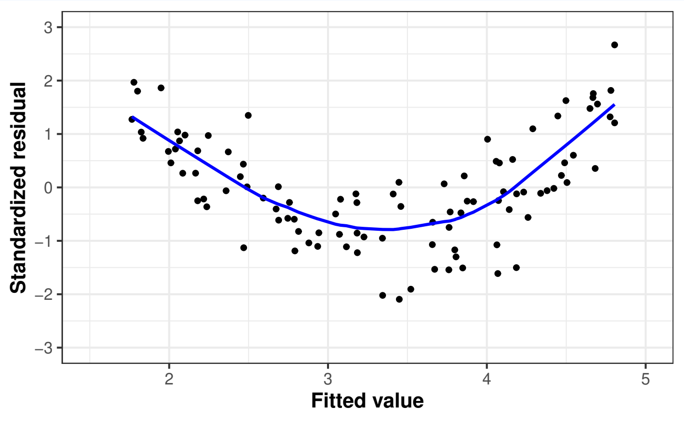


Figure: Standardized residuals against fitted values: omitted quadratic effect. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

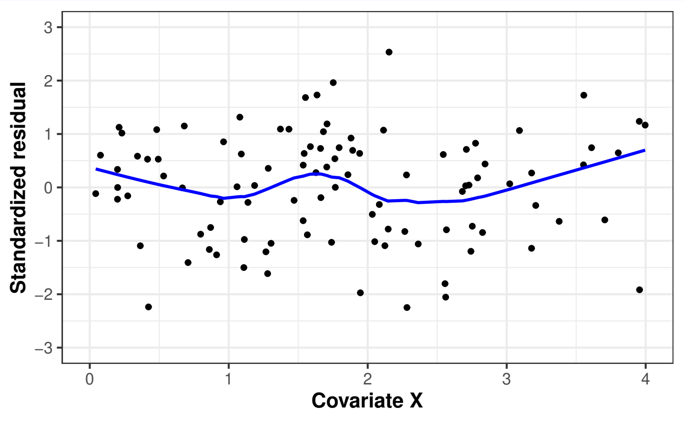


Figure: Standardized residuals against a covariate: assumptions satisfied. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

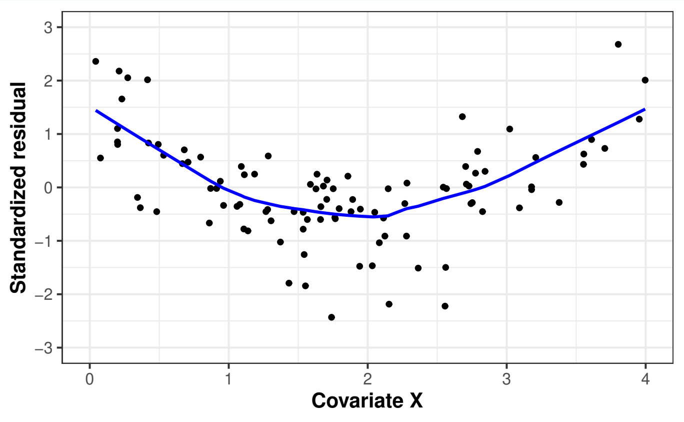


Figure: Standardized residuals against a covariate: omitted quadratic effect. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

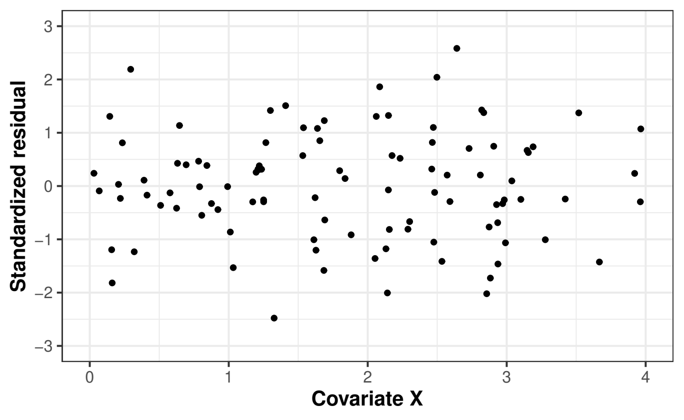


Figure: Standardized residuals against a covariate: assumptions satisfied.

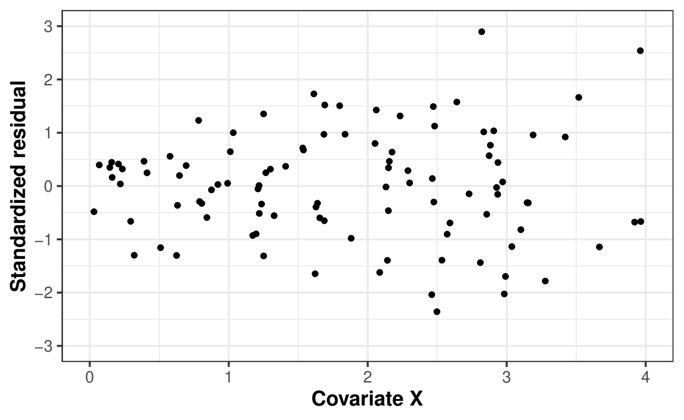


Figure: Standardized residuals against a covariate: mild increase of residual variance with covariate X.

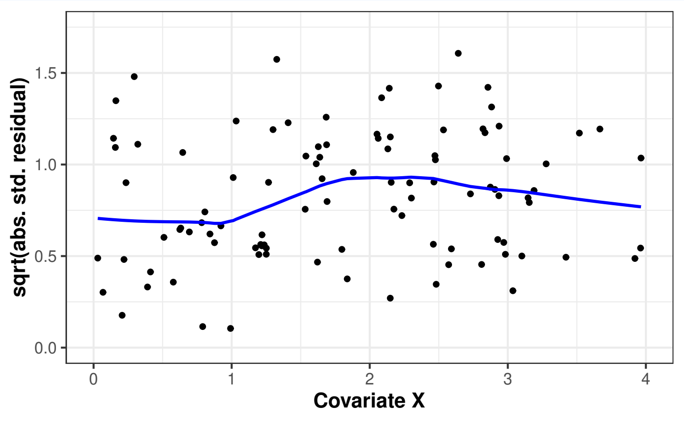


Figure: Square root of absolute standardized residuals against a covariate: assumptions satisfied. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

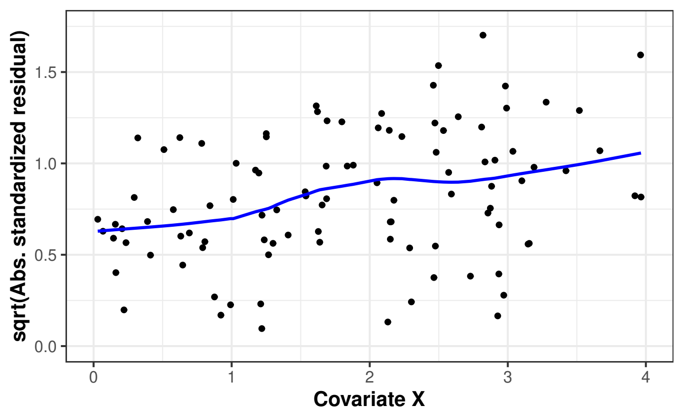


Figure: Square root of absolute standardized residuals against a covariate: mild increase of residual variance with covariate X. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

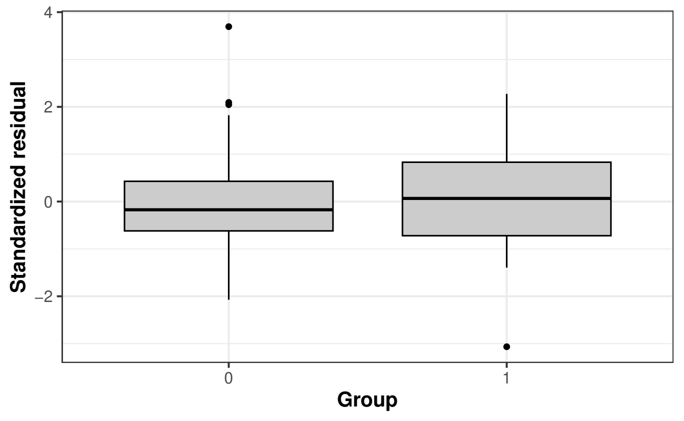


Figure: Boxplots of standardized residuals by a factor covariate: assumptions satisfied.

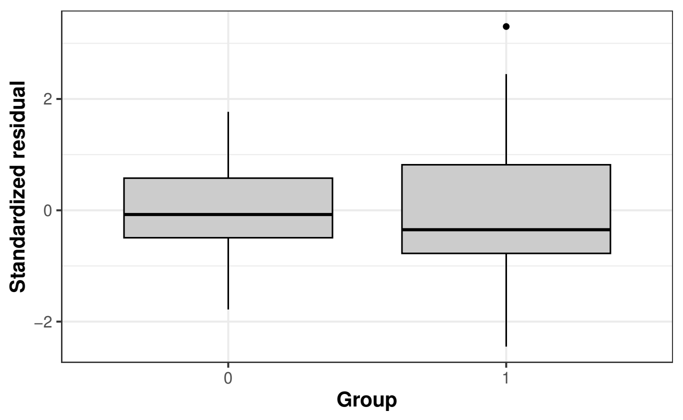


Figure: Boxplots of standardized residuals by a factor covariate: unequal variances of error terms and an omitted effect of another covariate.

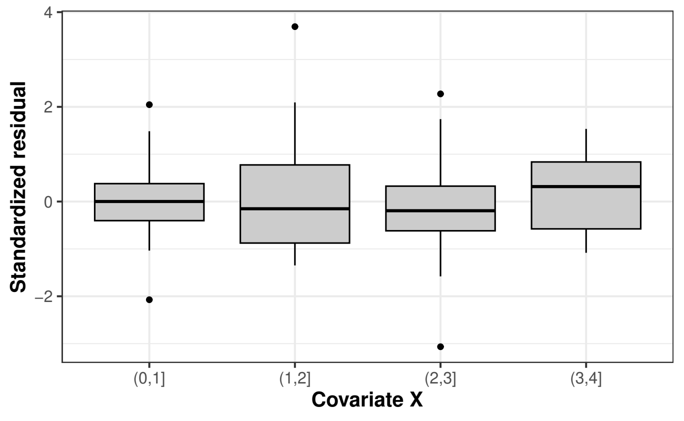


Figure: Boxplots of standardized residuals by a factorized continuous covariate: assumptions satisfied.

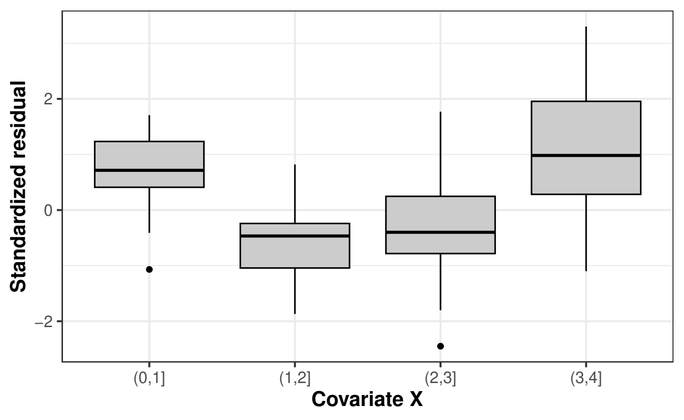


Figure: Boxplots of standardized residuals by a factorized continuous covariate: omitted quadratic effect and mildly increasing variance with X.

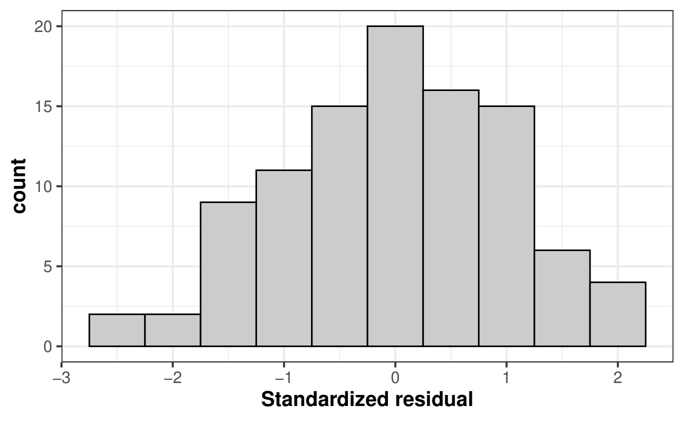


Figure: Histogram of standardized residuals: normal distribution of errors.

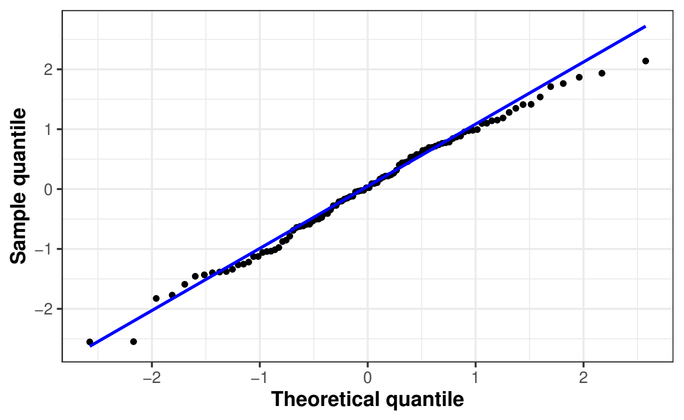


Figure: Q-Q plot of standardized residuals: normal distribution of errors.

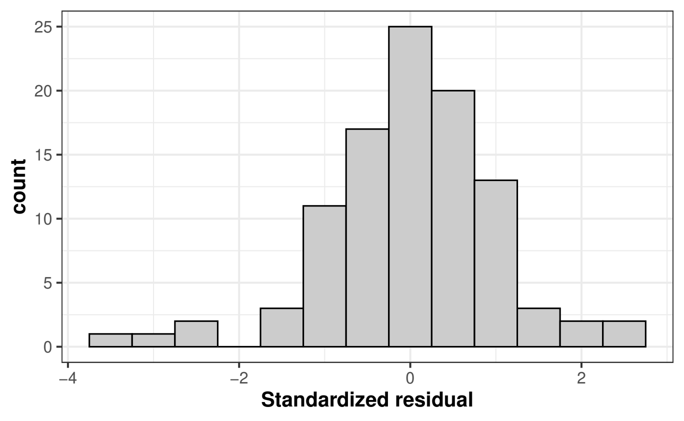


Figure: Histogram of standardized residuals: heavy-tailed distribution of errors (t_4).

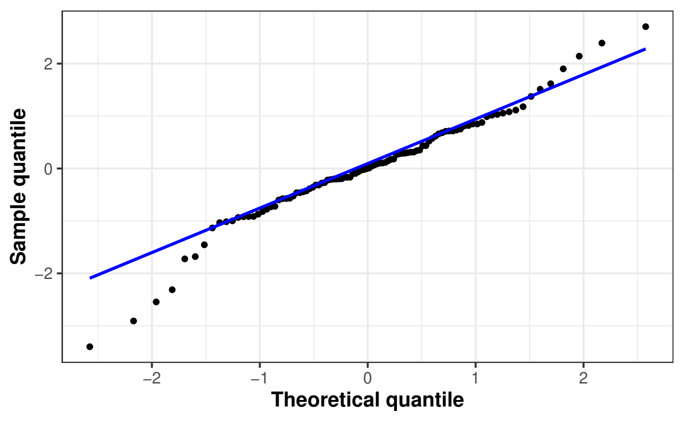


Figure: Q-Q plot of standardized residuals: heavy-tailed distribution of errors (t_4).

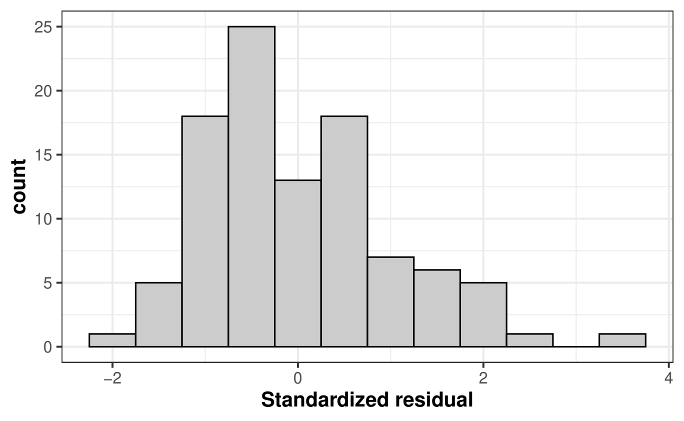


Figure: Histogram of standardized residuals: right-skewed distribution of errors (negative Gumbel).

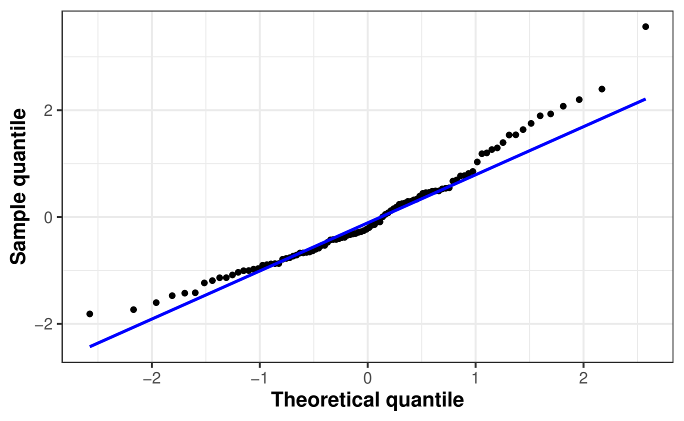


Figure: Q-Q plot of standardized residuals: right-skewed distribution of errors (negative Gumbel).

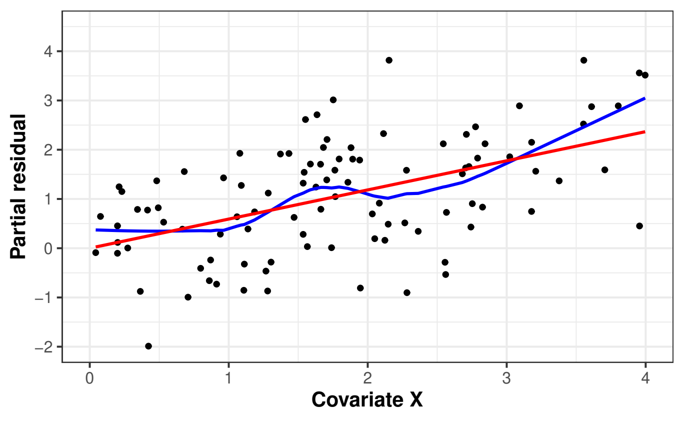


Figure: Partial residuals against a covariate: assumptions satisfied. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

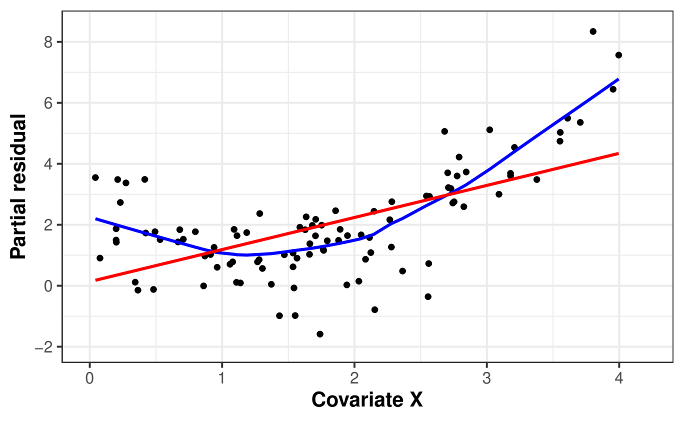


Figure: Partial residuals against a covariate: omitted quadratic effect. Smoothed by lowess smoother with window over 1/2 of the data range (blue).

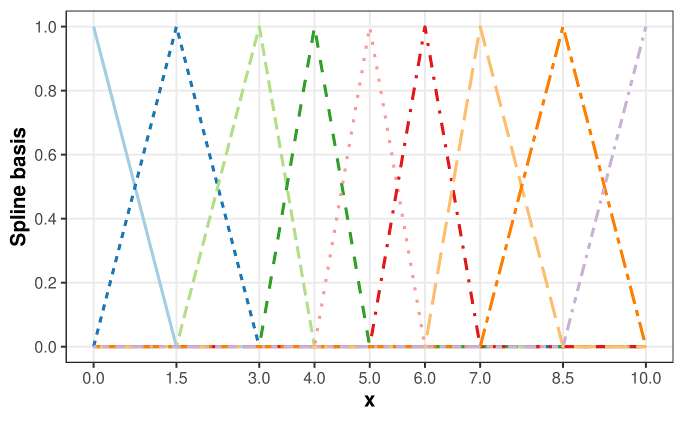


Figure: B-spline bases of degree 1 over the interval $\langle 0, 10 \rangle$.

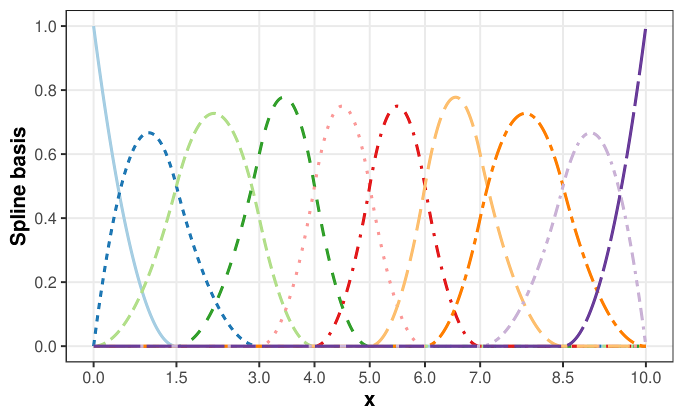


Figure: B-spline bases of degree 2 over the interval $\langle 0, 10 \rangle$.

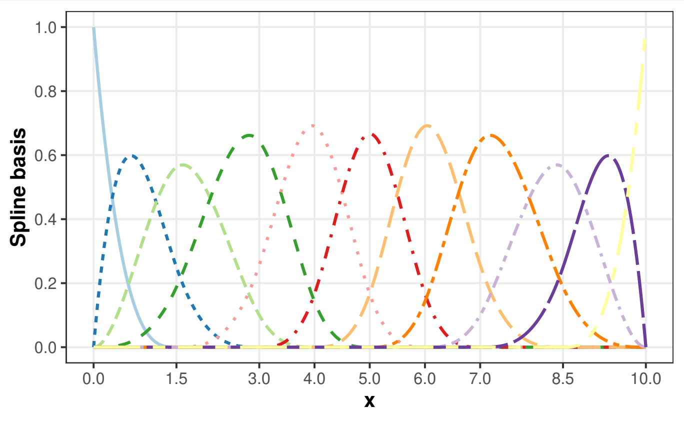


Figure: B-spline bases of degree 3 over the interval $\langle 0, 10 \rangle$.

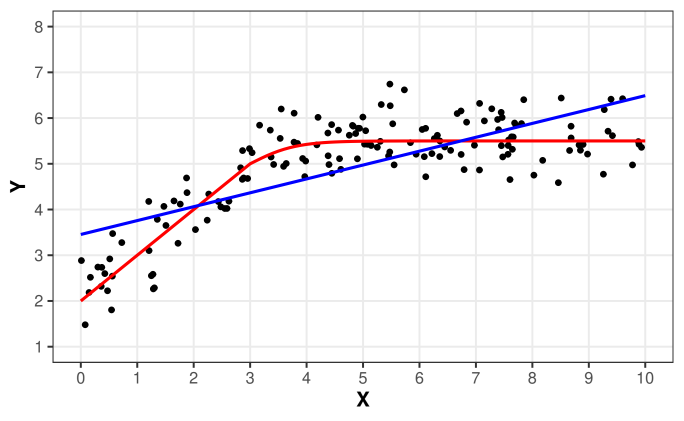


Figure: Logistic pipe function (red=true mean) fitted by regression line (blue).

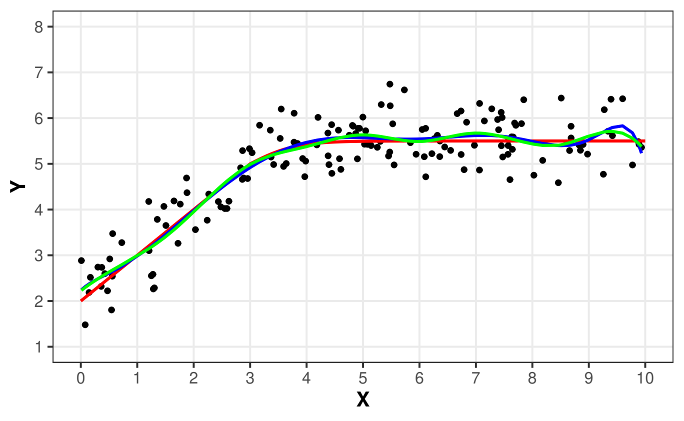


Figure: Logistic pipe function (red=true mean) fitted by 10th degree polynomial (blue) and 3rd degree spline with 7 inner knots (green).

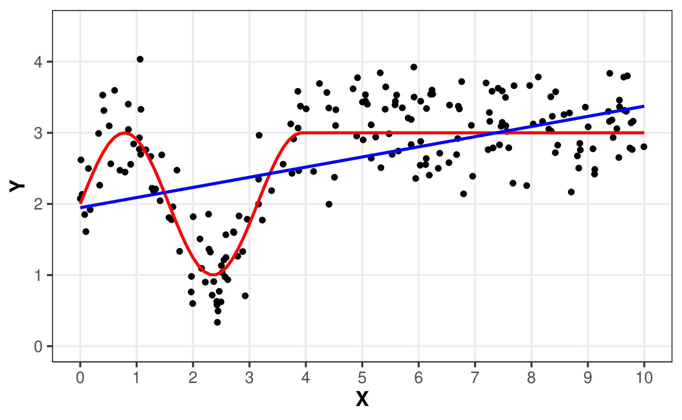


Figure: Sine-constant function (red=true mean) fitted by regression line (blue).

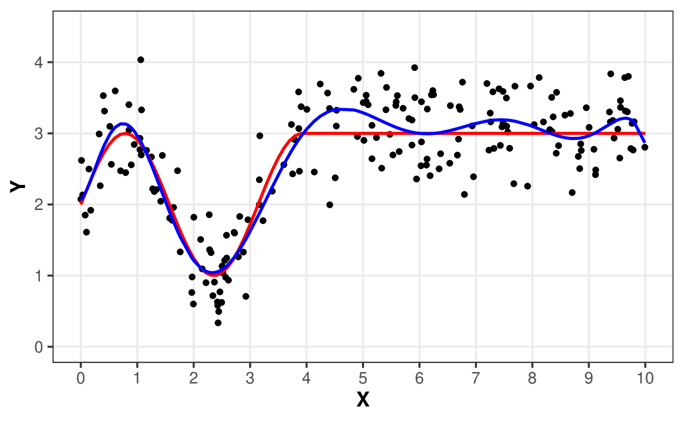


Figure: Sine-constant function (red=true mean) fitted by 10th degree polynomial (blue).

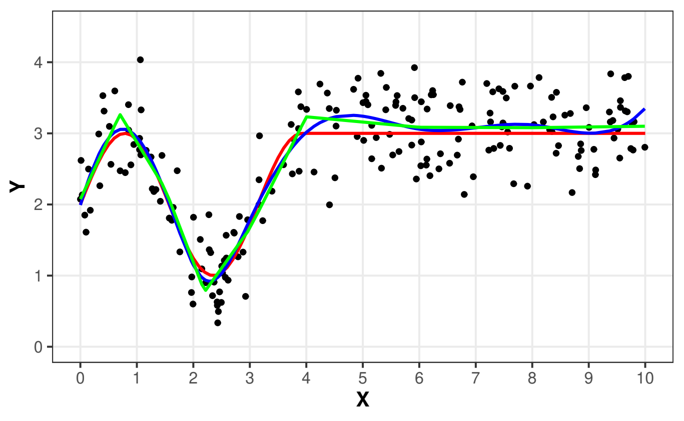


Figure: Sine-constant function (red=true mean) fitted by 1st degree (green) and 3rd degree (blue) spline with 7 inner knots.

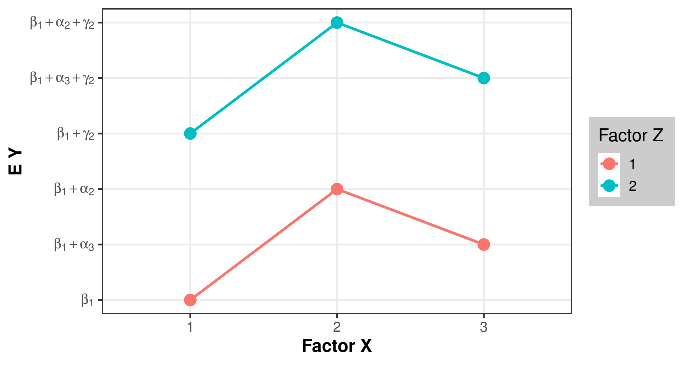


Figure: Means of two factors with no interaction - version 1.

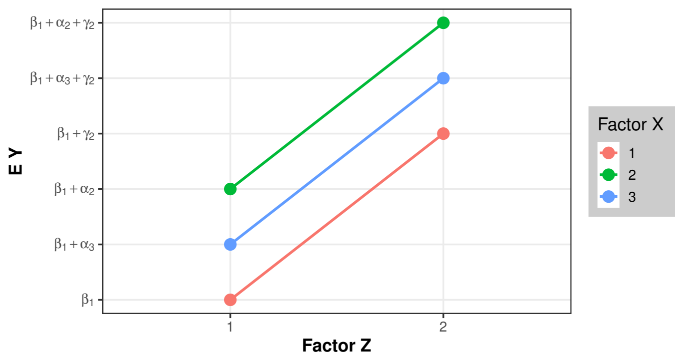


Figure: Means of two factors with no interaction - version 2.

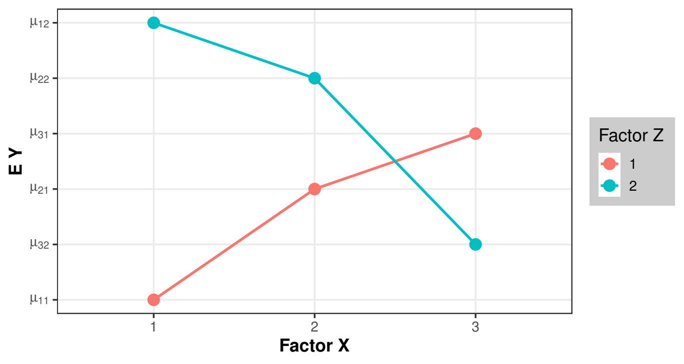


Figure: Means of two factors with interaction - version 1.

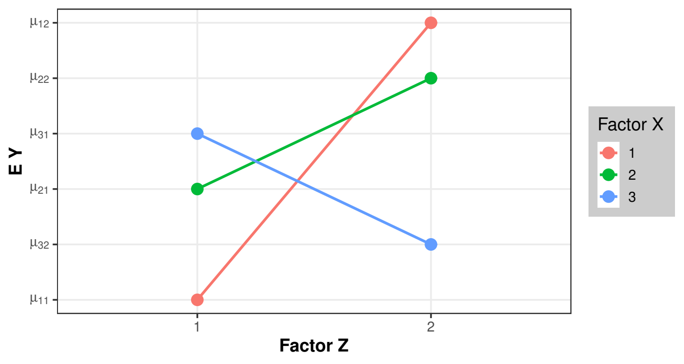


Figure: Means of two factors with interaction - version 2.