Applied Regression in R: Cheatsheet

https://sociology-fa-cu.github.io/applied-regression-r/
Also check out https://avehtari.github.io/ROS-Examples/ and https://marginaleffects.com/

Creating Regression Models

Simple linear regression	lm(eval ~ beauty, data = teachers)
Linear regression with categorical predictor	lm(eval ~ gender, data = teachers)
Multiple linear regression	lm(eval ~ beauty + gender, data = teachers)
Multiple linear regression with interaction	lm(eval ~ beauty * gender, data = teachers)

You can save your model by assigning it a name, e.g. model <- lm(eval ~ beauty + gender, data = teachers)
Categorical variables automatically converted to dummy variables

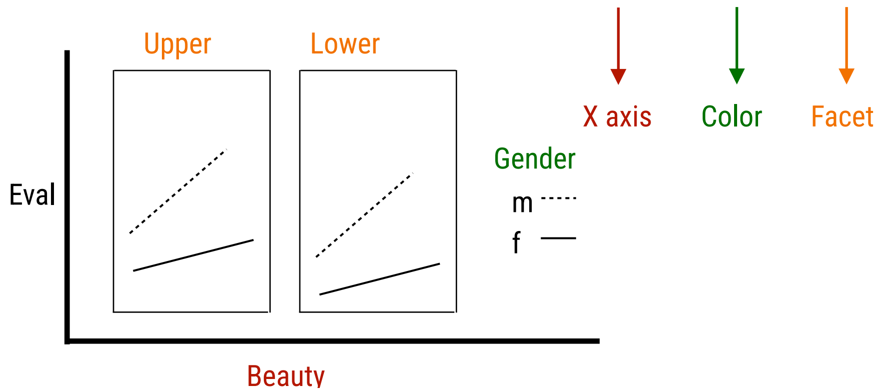
Summarizing Regression Models

Text summary of your model	summary(model)
Regression coefficients as tidy data frame	broom::tidy(model, conf.int = TRUE)
Fit indices (e.g. R^2) as tidy data frame	broom::glance(model)
Original data plus predicted values, residuals, etc.	broom::augment(model)

Visualizing Regression Models

Plot expected values of outcome against predictor(s)	marginaleffects::plot_predictions(model, condition = “beauty”)
Plot slopes of regression line/curve against predictor(s)	marginaleffects::plot_slopes(model, variables = “beauty”, by = “beauty”)
Plot group comparisons	marginaleffects::plot_comparisions(model, variables = “gender”, by = “gender”)

model = lm(eval ~ beauty * gender + division, data = teachers)
plot_predictions(model, condition = c(“beauty”, “gender”, “division”))



Interpreting Regression Models

Marginal effect	Effect of predictor on the outcome for individual observations	Effect of attractiveness on evaluation scores for each lecturer	marginaleffects::slopes(model)
Average marginal effect	Average effect of predictor on the outcome across all observations	Average effect of attractiveness on evaluation scores for all lecturer	marginaleffects::avg_slopes(model)
Conditional average marginal effect	Average effect of predictor on the outcome across all observations conditional on another outcome	Average effect of attractiveness on evaluation scores for male and female lecturers separately	marginaleffects::avg_slopes(model, by = “gender”)
Marginal effect at the mean	Effect of predictor on the outcome for an average observation	Effect of attractiveness on evaluation scores for a lecturer with an average evaluation score	marginaleffects::avg_slopes(model, newdata = “mean”)

Also see comparisons()/avg_comparisons() for comparing mean differences between groups, ratios of means and custom contrasts, as well as predictions()/avg_predictions() for computing predicted values.

Linear Regression Models Assumptions (in Order of Importance)

Validity & Reliability	All variables are measured with high validity and reliability.	If not, regression coefficients will be biased
Representativeness	Sample is representative of the population.	If not, regression coefficients will be biased
Linearity & additivity	Relationship between predictors & outcome can be expressed as linear combination of regression coefficients (and you chose the right combination, i.e. you have accounted for possible non-linear relationships	If not, regression coefficients will be biased
Independence of errors	Errors are independent	If not, standard errors of regression coefficients will be biased
Homoskedasticity	Errors have constant variance for all levels of the outcome	If not, standard errors of regression coefficients will be biased
Normality	Errors are normally distributed	If not, standard errors of regression coefficients will be biased (in small samples); prediction intervals will be biased

Errors = Difference between true and predicted values. Residuals = Difference between observed and predicted values. True values are unknown, so residuals are used as substitute for errors in model checking.

Checking Model Assumptions

Validity & Reliability	Psychometric methods, e.g. factor analysis	
Representativeness	Check against know population data (e.g. census). Or use sampling methods that guarantees representativeness (probability sampling).	
Linearity & additivity	Residuals vs fitted plot	performance::check_model(model) or plot(model, which = 1)
Independence of errors	No easy way, think about possible sources of dependence (e.g. students going to the same school)	
Homoskedasticity	(Square Root of) Residuals vs fitted plot	performance::check_model(model) or plot(model, which = 3)
Normality	Histogram of residuals or Q-Q plot	performance::check_model(model) or plot(model, which = 2)

Relaxing Model Assumptions

Validity & Reliability	Gather higher quality data or at least use models to account for low reliability/validity	Packages like lavaan, psych or mirt
Representativeness	Use survey weights to diminish differences between sample and population	Package survey
Linearity & additivity	Categorization, Simple Polynomials, Splines	lm(eval ~ poly(beauty, 2), data = teachers); lm(eval ~ splines::bs(beauty, df = 3), data = teachers)
Independence of errors	Clustered (Robust) Standard Errors	avg_slopes(model, vcov = ~prof)
Homoskedasticity	Robust Standard Errors	avg_slopes(model, vcov = “HC3”)
Normality	Bootstrapping	marginaleffects::avg_slopes(model, by = "gender") |> marginaleffects::inferences(method = "boot")