

operators

+	addition
-	subtraction
*	multiplication
/	division
^	taking powers
<-	assignment

& AND
| OR
! NOT

== equality
!= inequality
> greater than
>= greater than or equal to
< less than
<= less than or equal to

Functions

sqrt()	- Square root
round()	- Round a number
log()	- Logarithm → $\log(x, \text{base})$
exp()	- Exponentiation
abs()	- Absolute value

A lot of arguments have "defaults"

Variables

variable classes

variable	example	picture
numeric	8	
logical	TRUE	
character	"a bunny"	

variable assignment

`age <- 34`

pivot_longer()

Converts from wide to long format by decreasing the number of columns and increasing the number of rows

`data %>% pivot_longer(cols, names_to="key", values_to="val")`

Says which columns to combine into the new one
Defines the name of the new column
Defines what to call the new column of values

pivot_wider()

Converts from long to wide format by increasing the number of columns and decreasing the number of rows

`data %>% pivot_wider(names_from="key", values_from="val")`

Defines where the names for the new columns come from
Defines where the values in the new columns come from

vectors are lists of variables of the same class:

`name <- c("bunny", "gladly", "flopsy")`

Data manipulation

```
> data
   name colour height bunnyk rearrank doggyrank
1  bunny  grey    20      1     3      2
2  glady  purple  18      1     2      1
3  flopsy red     28      1     1      3
4  shadow  yellow  20      1     3      2
5  doggy  blue    24      1     2      1
6  cuddly  blue    17      2     NA     NA
7  doggle  blue    19      1     3      1
```

`data$height` selects the variable `height` from the tibble called `data`
`data[1,]` selects the first entire row from the tibble
`data$height[1]` selects the first case (row) of `height` from the tibble
`data[c(1,4,7), c("age","gender")]` selects rows 1, 4, and 7 and age/gender columns
`data[1:3, 2:4]` selects rows 1, 2, and 3 and columns 2, 3, and 4
`data$tall <- data$height > 19` creates a new variable called `tall` which is true if `height` is over 19
`data$tall <- NULL` removes the variable `tall`
`data[data$height>19,]` selects the rows from data for which the height is over 19

`ggplot(mapping = aes(x=xvar, y=yvar, colour=colour)) + geom_jitter(size=3) + facet_wrap(~var) + theme_bw() + labs(title = "Title", subtitle = "Subtitle", x = "X", y = "Y")`

Geoms that specify what the plot should look like (i.e., what kind of plot it is). Geoms take arguments that affect how the parts of them look.

Facets allow you to break your data into different panels, where each panel corresponds to one of the values of the given variable.

Themes make the plot look nice in terms of background colors and things like that

You can also s

graphics

Plots are made of aesthetic mappings that specify how the data in dataset map onto the axes and aesthetic properties like colour

geom_solid(size=3) + Geoms that specify what the plot should look like (i.e., what kind of plot it is). Geoms take arguments that affect how the parts of them look.

Facets allow you to break your data into different panels, where each panel corresponds to one of the values of the given variable.

Themes make the plot look nice in terms of background colors and things like that

One variable, continuous: title, etc

Two variables, continuous: mean, median, mode

geom_histogram()

geom_density()

geom_dotplot()

Two variables, one discrete: `fill: geom_jitter()`

kines: `scale_fill_brewer(palette="palettename")`

geom_boxplot()

geom_violin()

Dot plot of good food ratings

jitter

geom_jitter()

geom_boxplot()

Dot plot of good food ratings

density

geom_dotplot()

Density of good food ratings

Dot plot of good food ratings

boxplot

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

geom_violin()

Dot plot of good food ratings

density

geom_boxplot()

Dot plot of good food ratings

density

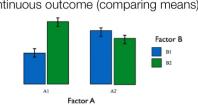
geom_violin()

problem of multiple comparisons
ANOVA only tells you that there is a difference, not which groups are different
running many t-tests inflates Type I error
solve with Bonferroni or Holm corrections

Bonferroni: simple, but very conservative.
 $p' = \frac{p}{n}$
adjusted p-value
number of tests
original p-value
Holm: better, more complicated
library(lsr)
posthocPairwiseT(x=modelName)
default: Holm
Bonferroni: $p.adjust.method = "bonferroni"$

posthoc test: ran after ANOVA, don't have particular hypothesis, need a correction.
planned comparison one of only a few tests you planned in advance, no correction needed but be honest and careful!

Two-way ANOVA
Has two factors instead of one, with a continuous outcome (comparing means)



Data must be in long form!

Between groups SS_{df}: how much variation in the outcome is there after taking into account the variation associated with both factors?
Within groups SS_{df}: how much variation in the outcome is there after taking into account the variation associated from one another?

$$MS_{\text{B}} = SS_{\text{B}}/(G-1)$$

$$MS_{\text{R}} = SS_{\text{R}}/(N-R-C+1)$$

$$F = MS_{\text{B}}/MS_{\text{R}}$$

Running interactions in R
Indicate with : in between the two variables
av(outcome ~ pred1 + pred2 + pred1:pred2, data=d)

Or can use shorthand to tell R to do all three:

	Df	Sum Sq	Mean Sq	F Statistic	Pr(>F)
Food	2	6360.2	3180.1	1285.1	0.382 ***
type:Food	4	5551.1	1388.3	16.675	1.47e-11 ***
Residuals	171	34231.83			

Each factor and the interaction have an F-statistic and can be significant or not

Effect size for ANOVA
Same as one-way: eta-squared
But partial eta squared needs to be interpreted
etaSquared(model)

η^2 : proportion of total variance due to each factor

Partial η^2 : proportion of variance due to each factor assuming you pretend the effect size of all the others is zero

3 myModel <- lm(y ~ x1 + x2 + x1:x2, dataset)
this command returns a model object which also prints out slope of each factor (coefficients) as well as the intercept

4 Hypothesis test 1: Is the overall model significantly different to what you'd expect if H0 were true?
H0: No relationship between X and Y
H1: Relationship between X and Y as in the best-fitting regression equation
- look at overall F statistic and p-value

Hypothesis test 2: Is a specific predictor (or interaction term) significant?
- look at p-value and slope for each predictor
- R does t-tests for each predictor against outcome

Effect size: given by R^2 , which captures proportion of the variance in Y accounted for by the model
- the same as the correlation coefficient r^2 (squared for the model with only that predictor)

5 standardized coefficients β : allow you to compare weight of variables
- calculated doing regression on z-scores

3 Assumption: residuals are normal
- check with Shapiro-Wilk on the residuals
- if violated, run Kruskal-Wallis
get residuals: `model <- lm(y ~ x1 + x2 + x1:x2, dataset)`
Shapiro-Wilk: `shapiro.test(model$residuals)`
Kruskal-Wallis: `kruskal.test(outcome ~ predictor, data=d)`
library(rstatix)
kruskalTest(outcome ~ predictor, data=d)

4 Assumption: variance homogeneous
Use Levene Test to check
If violated, do Welch one-way ANOVA
Levene Test:
library(car)
LeveneTest(outcome ~ predictor, data=d)

Welch one-way ANOVA
oneway.test(outcome ~ predictor, data=d)

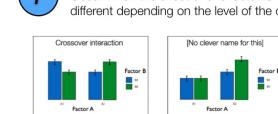
6 Running in R
add + to formula to indicate multiple factors
av(outcome ~ pred1 + pred2, data=d)

type:Food Residuals DF Sum Sq Mean Sq F value Pr(>F)
2 201 100.1285 50.064 0.454 ***
3 175 45852.3190 282.422 <2e-16 ***
4 171 113

Each group has an F-statistic and can be significant or not
If a group is significant, that's a **main effect**

Different from testing each factor separately because the residuals are different!

7 Interactions
Occur when the effect of one factor is different depending on the level of the other



Direction of effect of a factor depends on the level of the other factor
Magnitude of effect of a factor depends on the level of the other factor

Same idea, just get an interaction sum

1 Correlation: compares two numeric variables



2 Regression: compares two or more numeric variables (and evaluates which of them most influence an outcome)

Linear regression: finds line of best fit to the data for one predictor X and one outcome Y, each datapoint is given by:

$$Y_i = b_0 + b_1 X_i + \epsilon_i$$

b_0 : slope of the line
 b_1 : intercept
 ϵ_i : residual, i.e., deviation from line

this is the line that minimizes the deviations between the line and the data (SS_{res})

$$SS_{res} = \sum_i (Y_i - \hat{Y}_i)^2$$

SS_{res} is small
SS_{res} is large

multiple regression: more than two predictors

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \epsilon_i$$

for each predictor k , figures out its slope and how much it affects the outcome

Interactions included by multiplying factors: $X_1 \times X_2$ is an interaction between X_1 and X_2 .

3 Hypothesis test 1: Is the overall model significantly different to what you'd expect if H0 were true?
H0: No relationship between X and Y
H1: Relationship between X and Y as in the best-fitting regression equation
- look at overall F statistic and p-value

Hypothesis test 2: Is a specific predictor (or interaction term) significant?
- look at p-value and slope for each predictor
- R does t-tests for each predictor against outcome

Effect size: given by R^2 , which captures proportion of the variance in Y accounted for by the model
- the same as the correlation coefficient r^2 (squared for the model with only that predictor)

5 standardized coefficients β : allow you to compare weight of variables
- calculated doing regression on z-scores

1 Linearity test by looking at the residuals on each fixed value. If even across all, probably linear
> library(car)
> residualPlots(model)
or
> plot(model, which=1)



Measures outerness

$$D_i = \frac{\epsilon_i^2}{K-1} \times \frac{h_{ii}}{1-h_{ii}}$$

Measures leverage

Many rules of thumb:

Common rule is if Cook's distance is greater than about 0.1 or 1. In this subject we'll be more conservative: if $> 2/N$, where $K=3$ of coefficients, you might have an issue

2 Normality of the residuals: test by extracting the residuals and then using standard tests of normality

> resids <- rstandard(model)

Q-Q plot

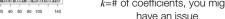
> qqnorm(resids)

Shapiro-Wilk

> shapiro.test(resids)

Much larger than 2 or 3... possible problem?
Don't use VIF on models with interactions!

3 Identify high-influence points: calculate Cook's Distance, which captures both outliers and leverage. High-influence points are high on both



Measures outerness

$$D_i = \frac{\epsilon_i^2}{K-1} \times \frac{h_{ii}}{1-h_{ii}}$$

Measures leverage

Many rules of thumb:

Common rule is if Cook's distance is greater than about 0.1 or 1. In this subject we'll be more conservative: if $> 2/N$, where $K=3$ of coefficients, you might have an issue

4 Identify collinearity: variables that are highly collinear contribute similar information, and can make the regression not robust. Quantify with Variance Inflation Factor which can tell you how the correlation is muddling up your estimate of the coefficients

$$VIF_i = \frac{1}{1-R_{i,i}^2}$$

where variable i is the outcome and the others are predictors

> vif(model)

VIF = 1 is great!

Using assumptions of linear regression

meta-analysis: statistical analysis of a large collection of analysis results from individual studies to integrate findings.

Narrative Review:
- Expert subjectively decides which studies should be included in the review, and how they should be weighted

- Strategy of search for studies generally not disclosed

- Focus on qualitative description of prior results, rather than on quantification.

- Expert subjectively reaches a conclusion about the overall trend of prior results

Systematic Review:
- collects empirical evidence based on pre-specified eligibility criteria to answer a specific research question.

- uses transparent and predefined study inclusion/exclusion and search strategies

- Most systematic reviews include a meta-analysis.

To do a meta-analysis, you pretty much need to do a systematic review first

3. Interpreting Forest Plots

• Visual representation of individual study effect size and their confidence intervals.

• Shows the meta-analysis combination (pooled effect) with its own confidence interval.

• Key Feature: Meta-analysis confidence interval is typically shorter (e.g., ~30%) than individual study confidence intervals, indicating increased precision. ↗ ↗

• Combines results from multiple studies to provide a more precise estimate of an effect.

• Can reveal statistically significant effects that individual studies might miss due to insufficient power (e.g., vote counting vs. meta-analysis example).

• Historically, could have led to earlier recognition of risks (e.g., SIDS and infant sleeping position)

• Addresses issues like "vote counting" where individual studies might be non-significant, but the combined evidence is significant. ↗ ↗

6. Estimating Reliability in Practice

a. Test-retest reliability: Correlate scores from original test and retest.

• Problems: Not good for transitory traits, carryover effects, true score changes, participant dropout.

b. Alternate-forms reliability: Correlate scores from two alternate forms of the test.

• Problems: Difficulty ensuring true parallelism, may not fully fix carryover effects.

c. Split-half reliability: Split test into two parallel subtests and correlate them.

• Problems: Parallel subtest assumption often questionable, tends to deflate reliability estimate due to halved length.

d. Cronbach's α (alpha):

• Losely, the mean of all possible split-half reliabilities, scaled to the full test.

• Historically popular and easy to implement.

• Provides a lower-bound (conservative) estimate for reliability.

• Has limitations; modern alternatives like McDonald's ω may be preferable. ↗ ↗ ↗

2. Psychological Assessment Defined

• Encompasses multiple data sources from one or more assessment methods (e.g., psychological tests, behavioral observation, interviews, questionnaires, rating scales, behavioral simulations, etc.).

• Aims to make generalizations about specific individuals ($n=1$), unlike research which generalizes about samples. ↗ ↗ ↗

4. Validity and Reliability

• Validity: A test measures what it purports to measure.

• Reliability: A test has consistency in measurement.

• Often judged by consistent scores upon re-examination (same test/repeated conditions, or different sets of equivalent items).

• A test must be reliable to be valid, but a reliable test is not necessarily valid. ↗

reliability

• Classical Test Theory (CTT) and reliability

* Fundamental Equation 1: Observed Score (x_i) = True Score (τ_i) + Error (ϵ_i).

* Error is unsystematic variance.

* Assumptions:

* On average, errors = zero.

* Errors do not correlate with each other.

* Errors do not correlate with true scores.

* Fundamental Equation 2 (Variance): $\sigma_x^2 = \sigma_\tau^2 + \sigma_\epsilon^2$ (Observed variance = True score variance + Error variance).

* Fundamental Equation 3 (Reliability Coefficient): Reliability (r_{xx}) = $\frac{\sigma_\tau^2}{\sigma_x^2} = \frac{\text{True Score Variance}}{\text{Observed Score Variance}}$ = Signal / Signal + Noise

* In practice, estimate r_{xx}^2 using the sample reliability coefficient (r_{xx}).

* Four Perspectives on Reliability:

* Proportions of Variance: Ratio of true score variance to observed score variance, or lack of error variance.

* Correlations: Squared correlation between observed and true scores, or lack of squared correlation between observed and error scores. ↗ ↗ ↗ ↗

Forest plot (blobbogram)

Study name and sample size

Lucky (Total N = 44)

Neluck (Total N = 36)

MA combination

Effect size from Dr. Lucy's study

Lines represent 95% Confidence Intervals

Spearman-Brown prophecy formula:
 $r'_{xx} = \frac{nr_{xx}}{1 + (n-1)r_{xx}}$

Where

• r'_{xx} is the reliability of the expanded test

• r_{xx} is the reliability of the original test

• n is the expansion factor (so if $n=2$ we double the number of items, $n=3$ we halve the number of items, $n=1$ we leave the number of items the same)

expansion factor: $n = \frac{r'_{xx}(1-r_{xx})}{r_{xx}(1-r'_{xx})}$

Where

• r'_{xx} is the reliability of the expanded test

• r_{xx} is the reliability of the original test

• n is the expansion factor (so if $n=2$ we double the number of items, $n=3$ we halve the number of items, $n=1$ we leave the number of items the same)

We should make our revised test 2.25 times as long as the original test.

↳ Content Validity:

Def.: content has validity to the extent its content adequately covers the full domain of the construct it is assessing.

• Related concept: Face validity

↳ Criterion Validity:

Def.: content has validity to the extent it seems to be validly correlated with some criterion variable.

↳ Construct Validity:

Def.: content has validity to the extent its content adequately covers the full domain of the construct it is assessing.

↳ Content Validity:

Def.: content has validity to the extent its content adequately covers the full domain of the construct it is assessing.

↳ Criterion Positive:

Def.: content has validity to the extent its content adequately covers the full domain of the construct it is assessing.

↳ Criterion Negative:

Def.: content has validity to the extent its content adequately covers the full domain of the construct it is assessing.

↳ Sensitivity:

Def.: content has sensitivity to correctly detect positive cases.

↳ Specificity:

Def.: content has specificity to correctly detect negative cases.

↳ Specificity:

Def.: content has specificity to correctly detect negative cases.

↳ Positive Predictive Power (PPP): Probability a positive test result indicates a positive case

PPP = $\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$

↳ Negative Predictive Power:

NPP = $\frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}}$

Prevalence: Probability a random case is criterion positive

Prevalence = $\frac{\text{True Positives} + \text{False Negatives}}{\text{True Positives} + \text{False Positives} + \text{False Negatives} + \text{True Negatives}}$

Forest plot (blobbogram)

Study name and sample size

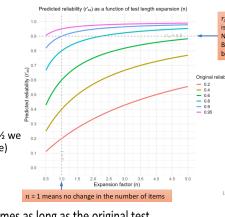
Lucky (Total N = 44)

Neluck (Total N = 36)

MA combination

Effect size from Dr. Lucy's study

Lines represent 95% Confidence Intervals



↳ $n > 3$ is needed for maximum benefit

↳ Numeracy: Bernoulli Sampling

Def.: content has validity to the extent its content adequately covers the full domain of the construct it is assessing.

• Validity: content interpretation and use of tests, not merely the test items themselves.

• Validity: content is a continuum and must be given relevant evidence and theory

• Validity: is a continuum and not a binary category

• Validity tends to require reference to some external criterion

↳ Criterion Validity:

Def.: content has validity to the extent it seems to be validly correlated with some criterion variable.

• Criterion Validity: Face validity

• Criterion Validity: Content validity

• Criterion Validity: Construct validity

• Criterion Validity: Content validity

• Criterion Validity: Construct validity

• Criterion Validity: Content validity

• Criterion Validity: Construct validity

• Criterion Validity: Content validity

• Criterion Validity: Construct validity

• Criterion Validity: Content validity