

# Course 1 · Week 4 – Two-group comparisons, associations, reporting

## Cheatsheet – biostats\_courses

R. Heller

### Choosing a comparison

Design	Continuous outcome	Binary outcome
Two independent groups	<code>t.test(y ~ g)</code> (Welch); Wilcoxon rank-sum	<code>prop.test / fisher.test</code>
Paired / pre-post	<code>t.test(pre, post, paired = TRUE)</code> ; Wilcoxon signed-rank	McNemar
> 2 groups, continuous	<code>aov(y ~ g)</code> ; Kruskal-Wallis	chi-square

### Effect sizes

Statistic	What it reports
Mean difference + 95% CI	primary for continuous
Cohen's <i>d</i>	standardised mean difference
Hedges' <i>g</i>	small-sample correction of <i>d</i>
Risk ratio (RR) / odds ratio (OR)	two proportions
Risk difference	absolute, clinically intuitive

```
effectsize::cohens_d(y ~ g)
```

### Two proportions

```
prop.test(c(tA, tB), c(nA, nB)) # asymptotic
fisher.test(matrix(c(tA, nA - tA,
                    tB, nB - tB), 2)) # exact, small cells
```

Report RR (or OR) with 95% CI, not just the *p*-value.

### Correlation

Method	Captures	Assumptions
Pearson	linear association, continuous	bivariate normal, no outliers
Spearman	monotonic association	rank-based, robust
Kendall	concordant/discordant pairs	robust, slow on large data

```
cor.test(x, y, method = "spearman")
```

## Non-parametric tests

Test	Replaces
Wilcoxon rank-sum (Mann-Whitney)	two-sample $t$
Wilcoxon signed-rank	paired $t$
Kruskal-Wallis	one-way ANOVA
Sign test	paired $t$ , when even ranks fail

## Power and sample size

```
pwr::pwr.t.test(d = 0.5, power = 0.80, sig.level = 0.05,  
               type = "two.sample")
```

Family	Function
two-sample $t$	<code>pwr.t.test</code>
two proportions	<code>pwr.2p.test</code> , <code>pwr.2p2n.test</code>
correlation	<code>pwr.r.test</code>
one-way ANOVA	<code>pwr.anova.test</code>

Simulation-based power for anything the textbooks skip: `simr::powerSim`.

## Reporting with `gtsummary`

```
library(gtsummary)  
trial |>  
  tbl_summary(by = arm, statistic = list(all_continuous() ~ "{mean} ({sd})")) |>  
  add_p() |>  
  add_overall()
```

## Decision rule for Week 4

- Ask: one-group, two-group, paired, or many-group?
- Check: normal-ish, or should I use a rank-based test?
- Report: point estimate + 95% CI + effect size;  $p$ -value last, not first.

## Common pitfalls

- Equal-variance  $t$ -test (default in some languages) when variances differ.
- Reporting OR when the audience expects RR (or vice versa).
- Forgetting that the  $p$ -value of a paired test depends on the pairing.
- Chaining correlations until one is “significant”.

## Further reading

- Altman, Machin et al., *Statistics with Confidence*.
- `gtsummary` docs.