Course 2 · Week 2 — ANOVA and non-linear extensions

Cheatsheet — biostats_courses

R. Heller

One-way ANOVA

```
aov(y ~ group, data = df) |> summary()
```

ANOVA is a linear model with a categorical predictor. The F-test compares between-group to within-group variance.

Contrasts with `emmeans`

```
library(emmeans)
emm <- emmeans(fit, ~ group)
pairs(emm, adjust = "tukey")   # all pairwise
contrast(emm, list(TrtVsCtrl = c(-1, 1, 1, 1) / 3))
```

Pre-specify contrasts before looking at the data; correct for multiplicity.

Two-way / factorial ANOVA

```
aov(y ~ A * B, data = df) |> summary()
emmip(fit, A ~ B)   # interaction plot
```

Interaction means "effect of A differs by level of B". Report the interaction first; main effects are conditional.

Repeated measures / blocking

- RCBD: `aov(y ~ treatment + Error(block))`.
- Repeated measures: move to a mixed model.

```
library(lme4); library(lmerTest)
lmer(y ~ treatment + time + (1 | subject), data = df)
```

GAMs — smooth non-linear terms

```
library(mgcv); library(gratia)
fit <- gam(y ~ s(x, k = 10) + z, data = df)
summary(fit)      # edf tells you how "wiggly"
draw(fit)         # smooth + CI
```

edf ≈ 1 ⟶ nearly linear; > 4 ⟶ clearly non-linear.

Non-linear regression (`nls`)

```
# Michaelis-Menten: y = Vmax * x / (K + x)
fit <- nls(y ~ Vmax * x / (K + x),
           data = df, start = list(Vmax = 1, K = 1))
```

Start values matter. If it fails, plot first to guess reasonable starts.

Decision rule for Week 2

- Categorical predictor, > 2 levels ⟶ ANOVA + contrasts.
- Factorial design ⟶ include interaction, report it first.
- Effect obviously curved ⟶ GAM with spline; else try `nls`.
- Repeated measures ⟶ mixed model, not repeated-measures ANOVA.

Common pitfalls

- Tukey HSD without pre-specified contrasts of interest.
- ANOVA $p < 0.05$ reported alone — without naming *which* groups differ.
- Forcing a GAM onto monotonic data that `nls` fits cleanly.
- Ignoring the random effect in clustered designs (pseudoreplication).

Further reading

- Wood, *Generalized Additive Models*, 2e.
- `emmeans` vignette.