Course 2 · Week 3 — GLMs, ANCOVA, evaluation

Cheatsheet — biostats_courses

R. Heller

GLM link functions

Outcome	Family	Link	R
Continuous	gaussian	identity	`lm()`
Binary	binomial	logit	`glm(y ~ x, family = binomial)`
Count	poisson	log	`glm(y ~ x, family = poisson, offset = log(t))`
Overdispersed count	neg. binomial	log	`MASS::glm.nb`
Ordinal	cumulative	logit	`MASS::polr`
Nominal	multinomial	logit	`nnet::multinom`

Logistic regression

```
fit <- glm(y ~ x1 + x2, data = df, family = binomial)
exp(coef(fit))                      # odds ratios
exp(confint(fit))                   # 95% CI on OR
predict(fit, newdata = nd, type = "response")
```

- Check for perfect separation (huge SEs).
- Interpret ORs cautiously; RR is more intuitive for the audience.

ANCOVA in an RCT

Adjust for baseline; **do not** analyse the change score.

```
lm(y_followup ~ arm + y_baseline, data = trial)
```

More efficient than simple t-test on change scores when baseline and follow-up are correlated.

Poisson / negative binomial

```
glm(cases ~ x + offset(log(person_years)),
    family = poisson, data = df)
MASS::glm.nb(cases ~ x + offset(log(person_years)), data = df)
```

Check dispersion: `sum(residuals(fit, type = "pearson")^2) / df.residual`. If > 1.5, switch to NB.

Evaluation — calibration + discrimination

Metric	Means	R
Calibration plot	predicted vs observed	`rms::val.prob`, manual bin
ROC / AUC	rank ordering	`pROC::roc(y, phat)`
Brier score	overall accuracy	`mean((phat - y)^2)`

Metric	Means	R
Calibration slope / intercept	systematic bias	from logistic recalibration

```
library(pROC)
roc_obj <- roc(y, phat)
auc(roc_obj); ci.auc(roc_obj)
plot(roc_obj)
```

Decision rule for Week 3

- Binary outcome ⟶ logistic; report OR + 95% CI.
- Count outcome ⟶ Poisson; check overdispersion; NB if needed.
- Trial analysis ⟶ ANCOVA, not change score.
- Prediction model ⟶ calibration curve first, ROC second, decision curves third.

Common pitfalls

- Quoting AUC without calibration (a discriminating but miscalibrated model is dangerous).
- Ignoring offsets in count data.
- Using ordinal logit when the proportional-odds assumption fails.
- Presenting logistic regression coefficients on the log-odds scale without OR.

Further reading

- Harrell, *RMS*, ch. 10–12.
- Steyerberg, *Clinical Prediction Models*.