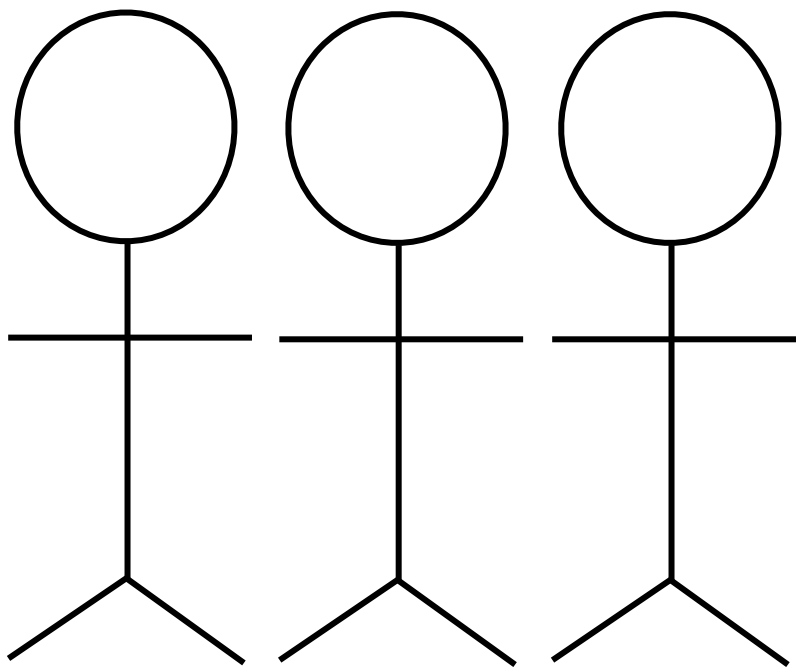# Computing information criteria for multilevel models using different sample sizes

Wendy Christensen

University of California, Los Angeles

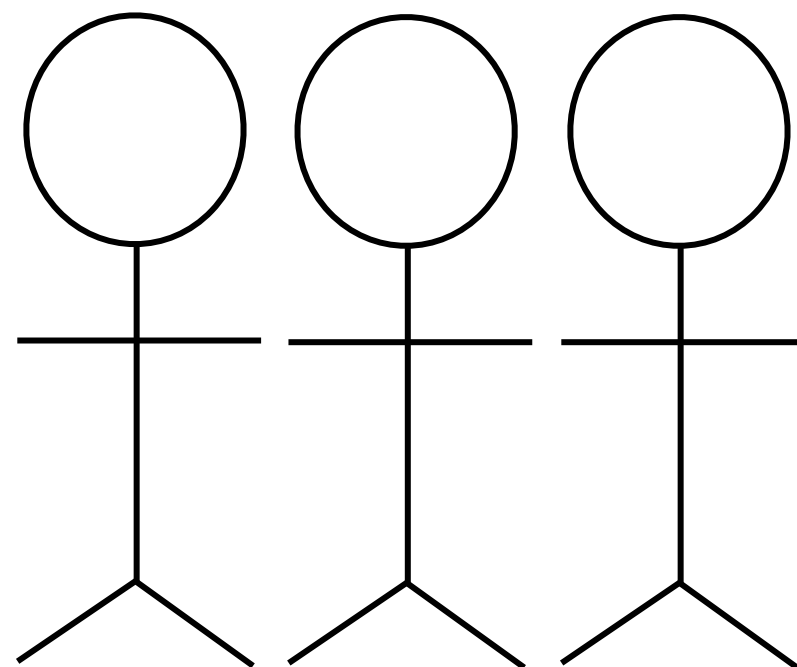School 1
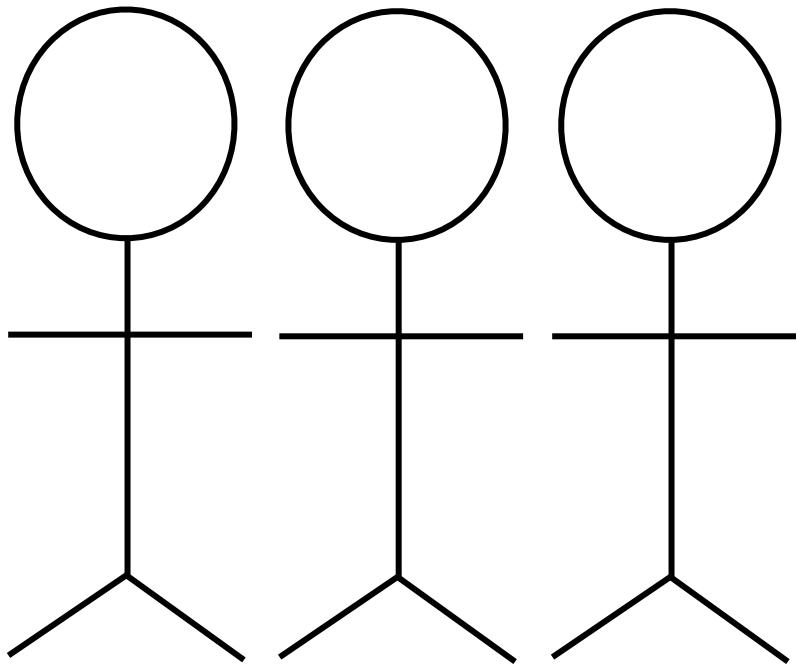
School 2

S1 S2 S3

S4 S5 S6

School 1
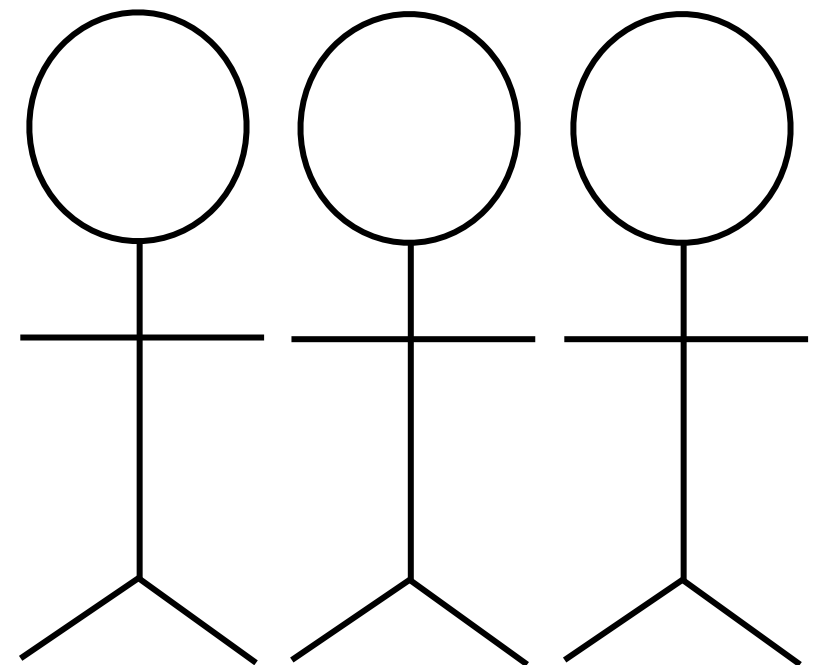
School 2

N = 6

S1 S2 S3

S4 S5 S6

School 1

m = 2

School 2

N = 6
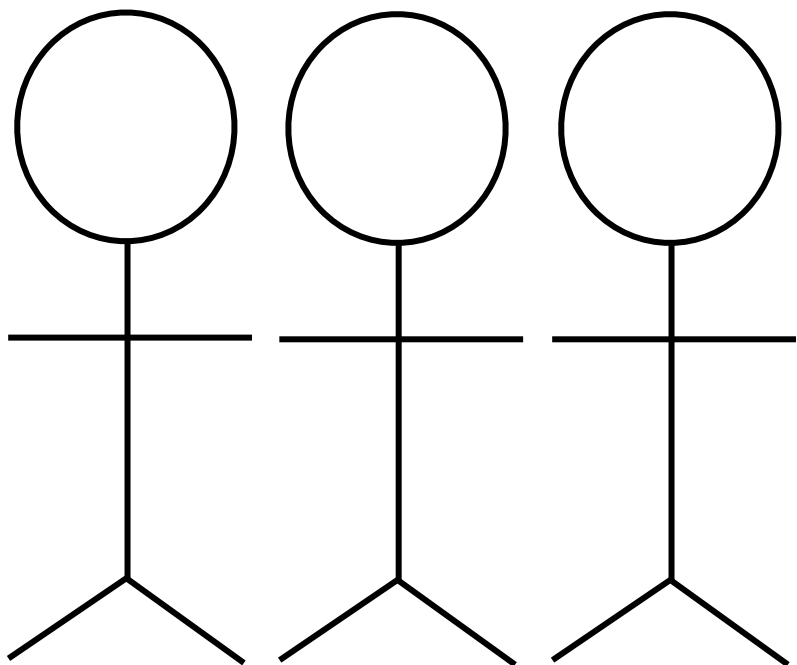
S1   S2   S3

S4   S5   S6
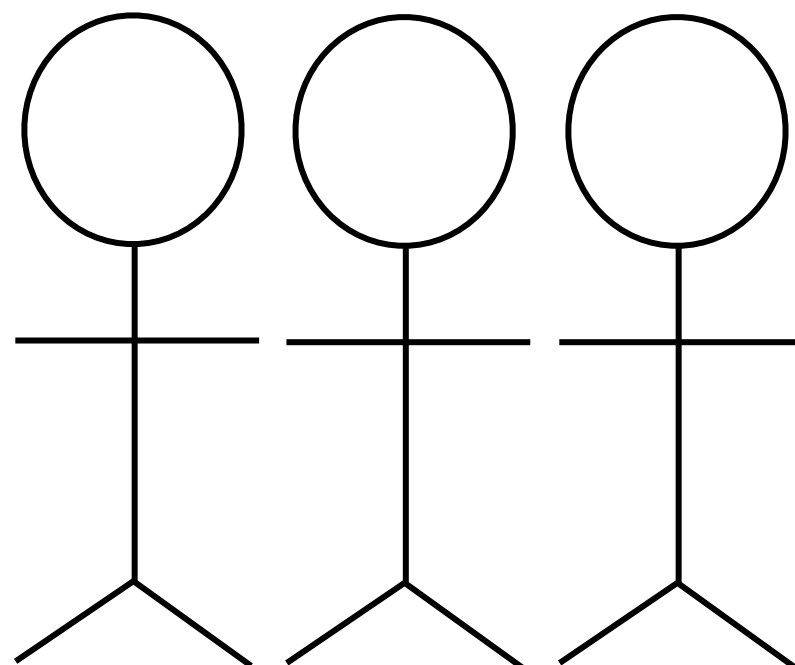
Model selection using information criteria

①   Model 1   Model 2   Model 3

Model selection using information criteria

①   Model 1   Model 2   Model 3

②   Deviance 1   Deviance 2   Deviance 3

Model selection using information criteria

① Model 1   Model 2   Model 3

② Deviance 1   Deviance 2   Deviance 3

③ I.C. 1   I.C. 2   I.C. 3

Model selection using information criteria

①   Model 1   Model 2   Model 3

②   Deviance 1   Deviance 2   Deviance 3

③   I.C. 1   I.C. 2   I.C. 3
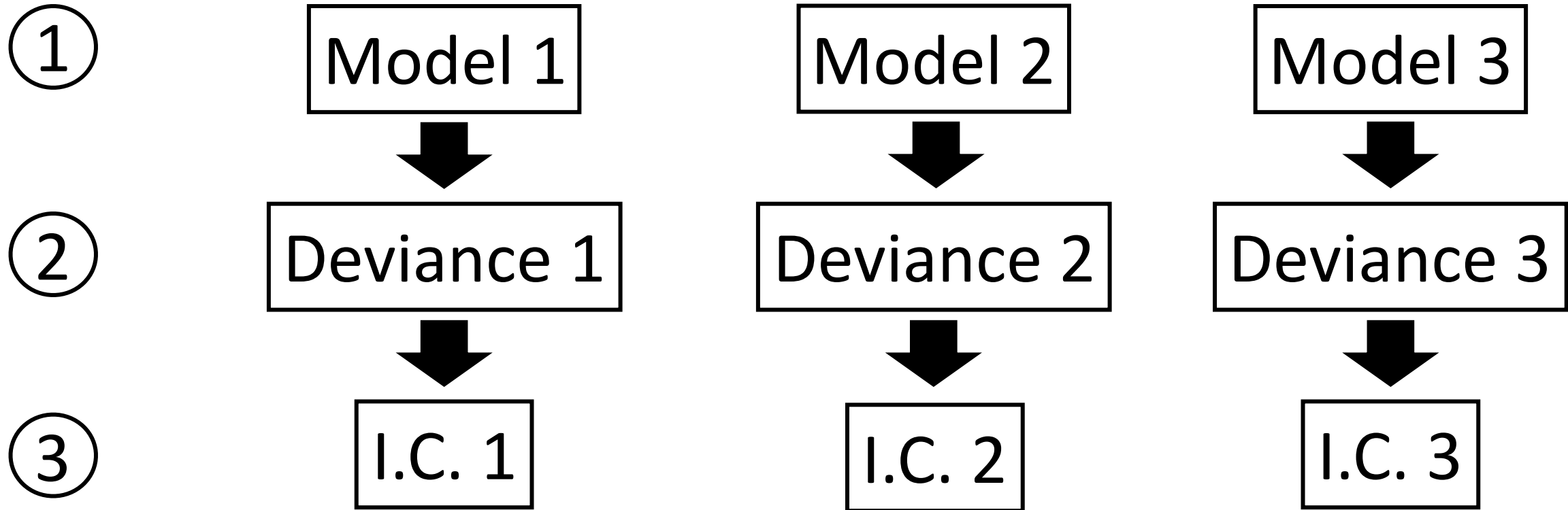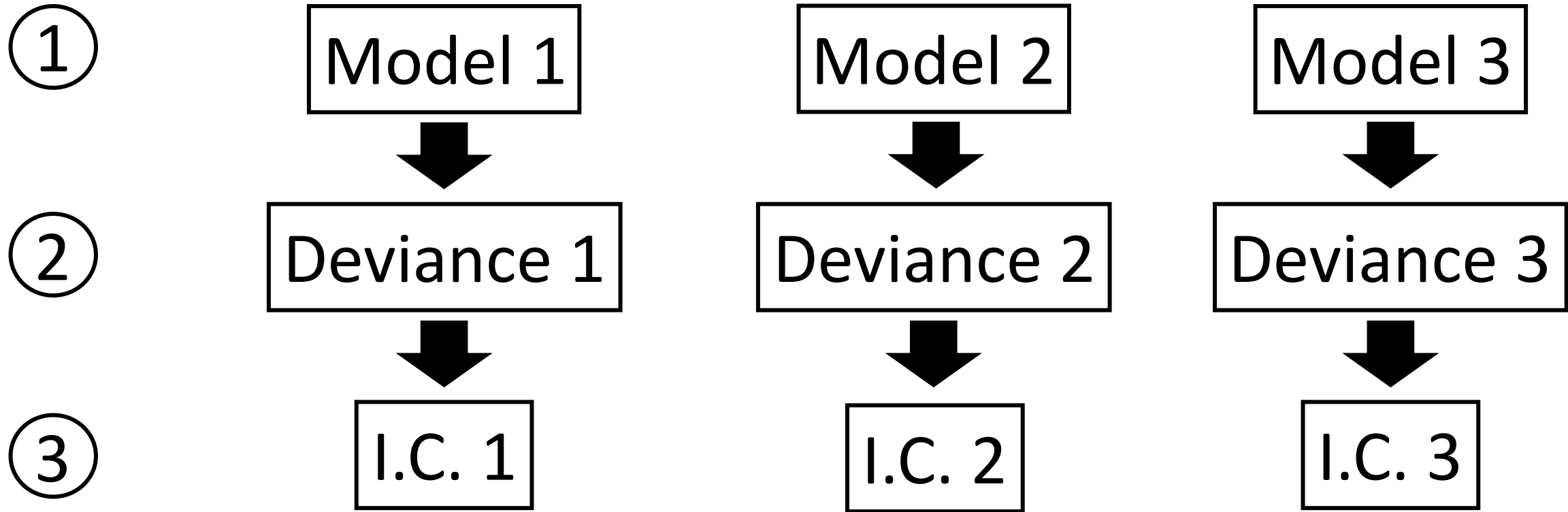
④   Smallest I.C. value = "best" model

Choice of sample size matters!

AIC

Deviance + $2*k$

BIC

Deviance + $\log(N)*k$

Choice of sample size matters!

AIC
Deviance + 2*k

BIC
Deviance + log(N)*k

Model deviance

Choice of sample size matters!

AIC
Deviance + 2*k

BIC
Deviance + log(N)*k

Model deviance

Number of model parameters

Choice of sample size matters!
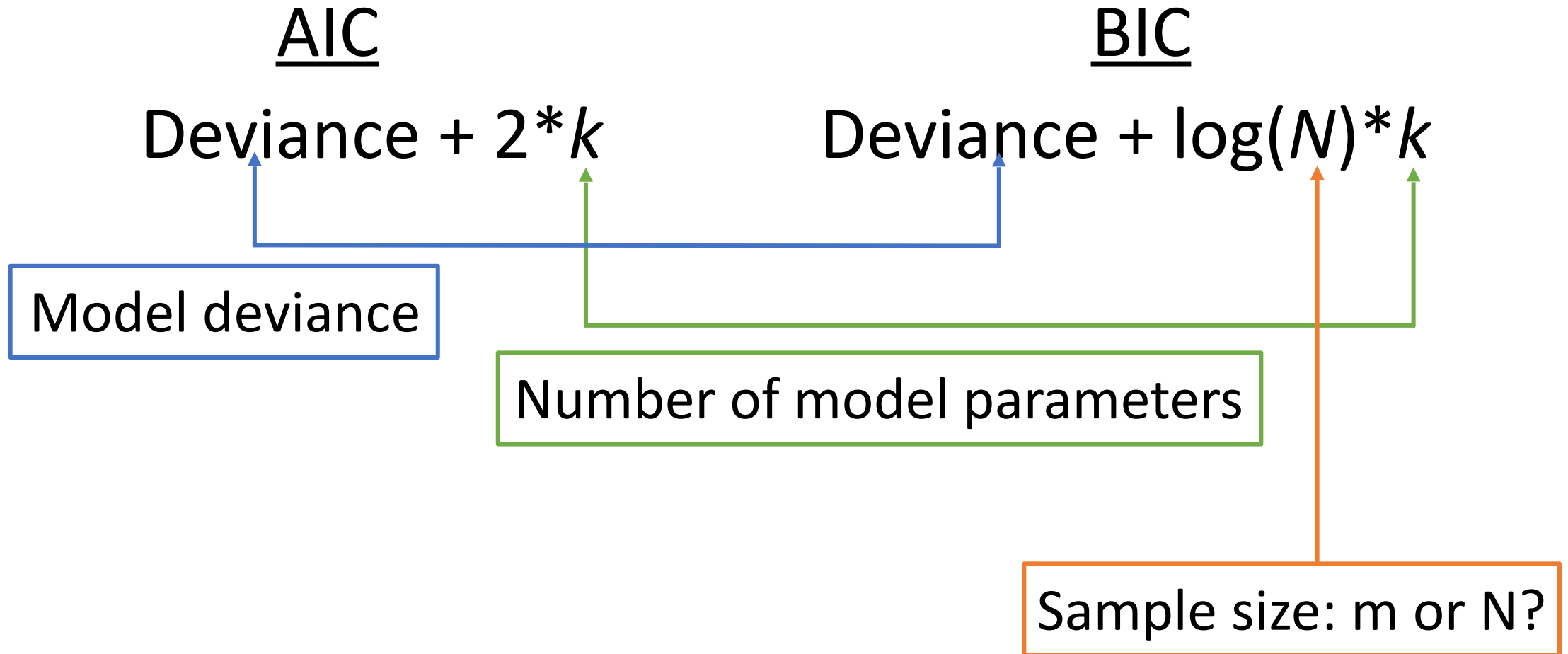
AIC
$$\text{Deviance} + 2*k$$

BIC
$$\text{Deviance} + \log(N)*k$$

Model deviance

Number of model parameters

Sample size: m or N?

Example using the nlme package[1][2]

Information criteria

Parameters

N

m

```
Linear mixed-effects model fit by maximum likelihood
 Data: alcohol1
      AIC      BIC    logLik
  676.156 686.672 -335.078

Random effects:
 Formula: ~1 | id
          (Intercept)  Residual
StdDev:    0.7509085 0.7494974

Fixed effects: alcuse ~ 1
                 Value  Std.Error  DF  t-value p-value
(Intercept) 0.9219549 0.09590253 164 9.613458        0

Standardized Within-Group Residuals:
       Min         Q1        Med        Q3       Max
-1.8865485 -0.3075977 -0.3066575 0.6136653 2.8567495

Number of Observations: 246
Number of Groups: 82
```

[1] Singer & Willett (2003): *Applied Longitudinal Data Analysis*
[2] UCLA Statistical Consulting Group: https://stats.idre.ucla.edu/r/examples/alda/r-applied-longitudinal-data-analysis-ch-4/

Computing BIC using m (number of clusters)

```
str(model)
```

Examine to find necessary elements from output

```
logLikelihood <- model$logLik
deviance <- -2*logLikelihood

groups <- head(model$dims$ngrps,n=1)
m <- unname(groups, force = FALSE)

N <- model$dims$N

BIC_m <- deviance+log(m)*3

BIC_N <- deviance+log(N)*3
```

Computing BIC using m (number of clusters)

```
str(model)

logLikelihood <- model$logLik
deviance <- -2*logLikelihood

groups <- head(model$dims$ngrps,n=1)
m <- unname(groups, force = FALSE)

N <- model$dims$N

BIC_m <- deviance+log(m)*3

BIC_N <- deviance+log(N)*3
```

Isolate model log-likelihood

Compute model deviance

Computing BIC using m (number of clusters)

```
str(model)

logLikelihood <- model$logLik
deviance <- -2*logLikelihood

groups <- head(model$dims$ngrps,n=1)
m <- unname(groups, force = FALSE)

N <- model$dims$N

BIC_m <- deviance+log(m)*3

BIC_N <- deviance+log(N)*3
```

Obtain number of clusters

Computing BIC using m (number of clusters)

```
str(model)

logLikelihood <- model$logLik
deviance <- -2*logLikelihood

groups <- head(model$dims$ngrps,n=1)
m <- unname(groups, force = FALSE)

N <- model$dims$N

BIC_m <- deviance+log(m)*3

BIC_N <- deviance+log(N)*3
```

Compute BIC(m)

683.376

Computing BIC using N (total observations)

```
str(model)

logLikelihood <- model$logLik
deviance <- -2*logLikelihood

groups <- head(model$dims$ngrps,n=1)
m <- unname(groups, force = FALSE)

N <- model$dims$N

BIC_m <- deviance+log(m)*3

BIC_N <- deviance+log(N)*3
```

Matches output!

686.6719...

AIC	BIC	logLik
676.156	686.672	-335.078

Thanks!

Wendy Christensen

wchristensen@ucla.edu

Code available at: https://github.com/wendychristensen