

A blue-toned background image of a financial candlestick chart. The chart features a grid with horizontal and vertical lines. A thick, curved white line, likely a moving average or trend line, arches across the chart. A straight white line with a downward slope is labeled '61.6%: 99.19'. Two specific price points are highlighted with white boxes: '104.19' and '86.72'. The chart shows several candlesticks, some with long wicks, indicating price volatility.

# Model and Variable Selection

DR JOSH HODGE

[JHODGE@IC.AC.UK](mailto:JHODGE@IC.AC.UK)



# Intended Learning Outcomes

---

- Define collinearity and select methods for investigating it
- Interpret the variance-inflation factor and use it as a tool to choose variables
- Recall the procedures of the model selection approaches
- Execute and interpret the hypothesis testing and Information Theoretic (IT) approaches to model selection

GEORGE E P BOX  
20<sup>TH</sup> CENTURY  
STATISTICIAN

*All models are wrong,  
but some are useful*

# The Dilemma

---

- Measured many variables and want to include them all in a single modelling approach
- Interested in which covariates (and/or their interactions) are driving the response variable
  - E.g., you're interested in changes of the number of saltmarsh sparrows and measured (12 explanatory variables):
    - % coverage of *Juncus gerardii*, Shrub, *Spartina patens*, *Distichlis spicata*, Bare ground, Other vegetation, *Phragmites australis*, Tall sedge, Water, *Spartina alterniflora* (short), *Spartina alterniflora* (tall)
    - Height of thatch, Max vegetation
    - Vegetation stem density
- What would you do?

# The Steps

---

- Find the optimal set of covariates
- Steps:
  1. Check for redundancy (collinearity) amongst the explanatory variables and remove those variables
  2. Fit your reduced model and decide on model selection procedures (if terms are non-significant)
- Model selection is political and highly debated: some disciplines and journals advocate for certain methods
  - Debate among the staff at Imperial

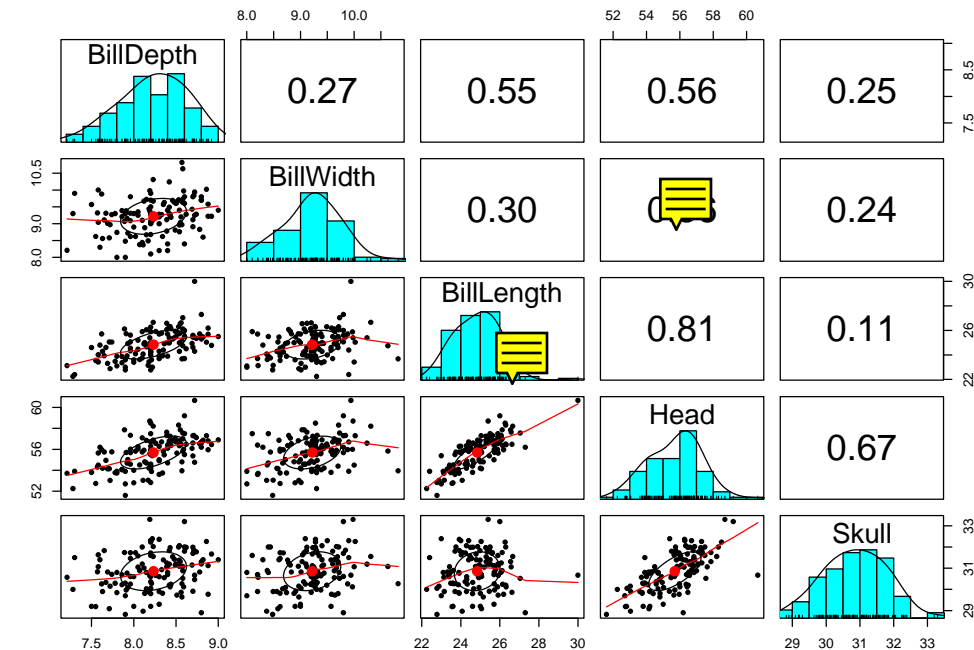
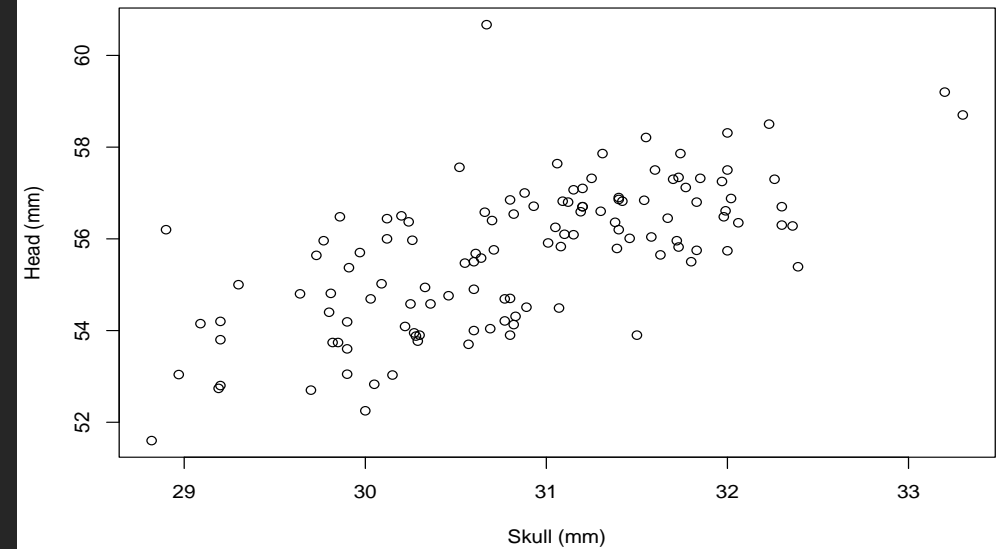


# Collinearity

---

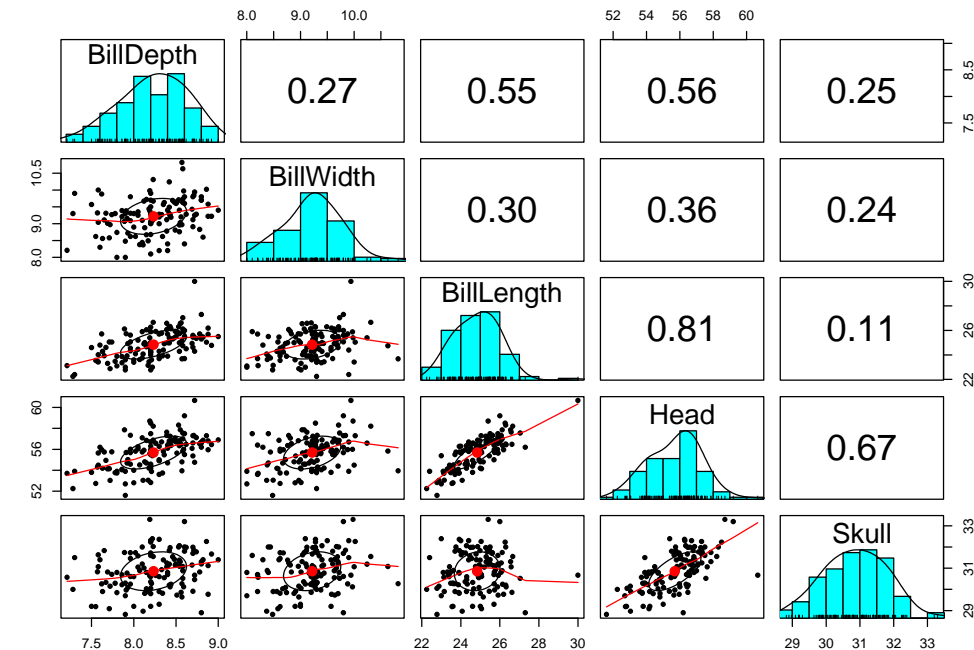
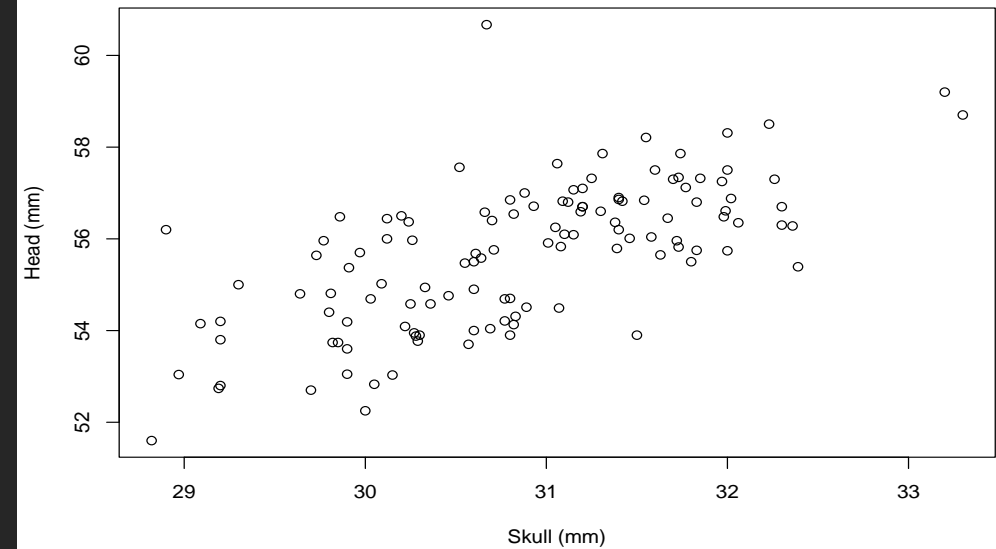
# Collinearity

- Correlation among our explanatory variables:
  - $\text{Body Mass} \sim \text{Head} + \text{Skull} + \text{Bill Depth} + \text{Bill Width} + \text{Bill Length}$
- If ignored increases Type II errors (false negatives)
  - Inflates standard errors via reducing in degrees of freedom (statistical power)
  - Confusing statistical analysis where nothing significant
- Methods to detect:
  - Plotting (pairwise plots amongst all variables)
  - Correlation tests
  - Variance-Inflation Factor (VIF)



# Collinearity

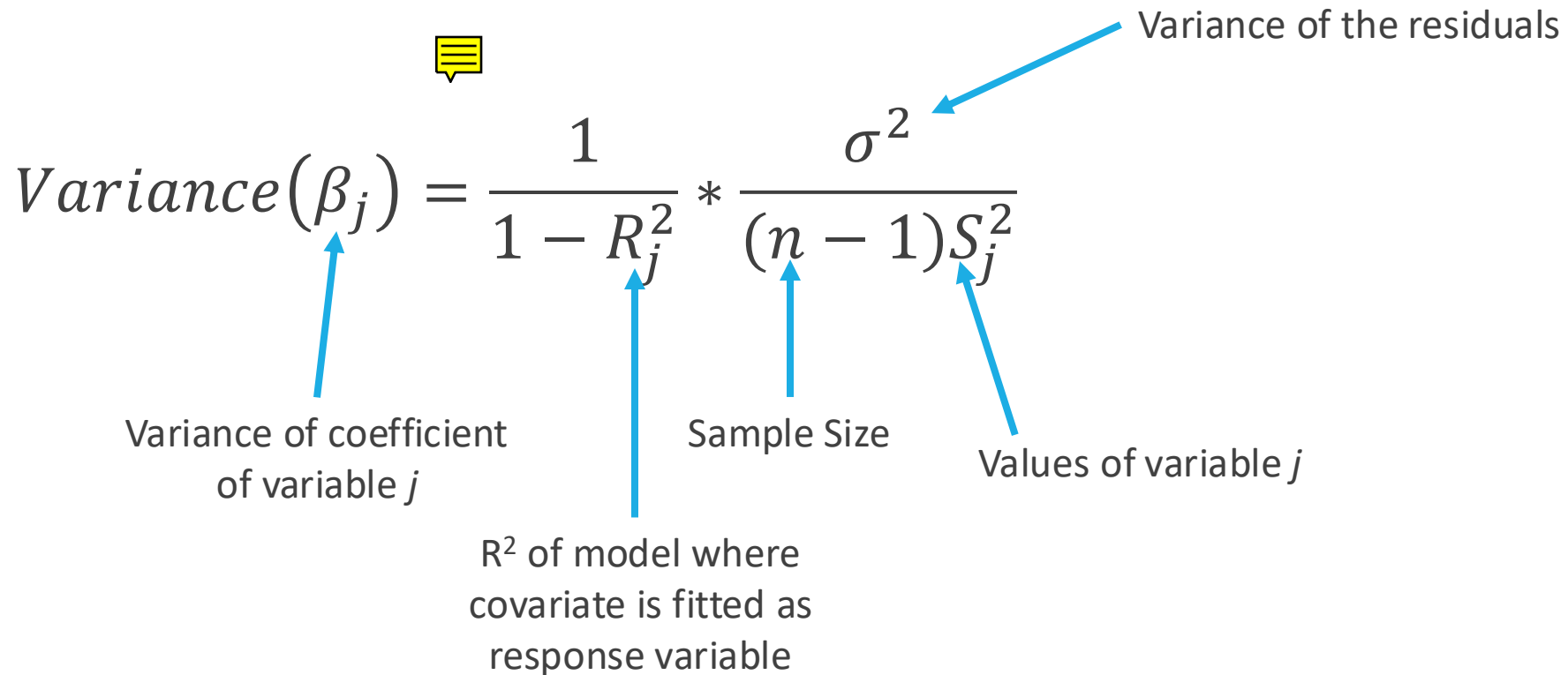
- Correlation among our explanatory variables:
  - $\text{Body Mass} \sim \text{Head} + \text{Skull} + \text{Bill Depth} + \text{Bill Width} + \text{Bill Length}$
- If ignored increases Type II errors (false negatives)
  - Inflates standard errors via reducing in degrees of freedom (statistical power)
  - Confusing statistical analysis where nothing significant
- Methods to detect:
  - Plotting (pairwise plots amongst all variables)
  - Correlation tests
  - **Variance-Inflation Factor (VIF)**





# Variance-Inflation Factor (VIF)

---



The diagram illustrates the formula for the Variance-Inflation Factor (VIF) for a coefficient  $\beta_j$ . The formula is presented as:

$$\text{Variance}(\beta_j) = \frac{1}{1 - R_j^2} * \frac{\sigma^2}{(n - 1)S_j^2}$$

Annotations with blue arrows point to the components of the formula:

- Variance of coefficient of variable  $j$** : Points to  $\text{Variance}(\beta_j)$ .
- $R^2$  of model where covariate is fitted as response variable**: Points to  $R_j^2$ .
- Sample Size**: Points to  $n$ .
- Values of variable  $j$** : Points to  $S_j^2$ .
- Variance of the residuals**: Points to  $\sigma^2$ .

A yellow speech bubble icon is located above the fraction  $\frac{1}{1 - R_j^2}$ .

# Variance-Inflation Factor (VIF)

---

$$\text{Variance}(\beta_j) = \frac{1}{1 - R_j^2} * \frac{\sigma^2}{(n - 1)S_j^2}$$

Diagram illustrating the components of the Variance-Inflation Factor (VIF) formula:

- $\text{Variance}(\beta_j)$ : Variance of coefficient of variable  $j$
- $R_j^2$ :  $R^2$  of model where covariate is fitted as response variable
- $n$ : Sample Size
- $S_j^2$ : Values of variable  $j$
- $\sigma^2$ : Variance of the residuals

# Example 1

---

- Essentially, VIF is based on the  $R^2$  of a series of models where each covariate becomes the response variable, e.g.
  - Body Mass ~ [Head + Skull + Bill Depth + Bill Width + Bill Length]
    - Head ~ Skull + Bill Depth + Bill Width + Bill Length
    - Skull ~ Head + Bill Depth + Bill Width + Bill Length
    - Bill Depth ~ Head + Skull + Bill Width + Bill Length
    - Bill Width ~ Head + Skull + Bill Depth + Bill Length
    - Bill Length ~ Head + Skull + Bill Depth + Bill Width
- Single VIF score for each variable

Variable	VIF
Head	2899993
Skull	98681.63
Bill Depth	1.54
Bill Width	1.17
Bill Length	162667.6

*Extracted from R*

# What do we do with these numbers?

---

- Sequentially drop variable with the highest VIF, recalculate VIF and repeat until all variables are below a given threshold:
  - 10, 5 or 3 (depending on how conservative) → I'd go with 3.

Variable	VIF
Head	2899993
Skull	98681.63
Bill Depth	1.54
Bill Width	1.17
Bill Length	162667.6



Variable	VIF
Skull	1.11
Bill Depth	1.53
Bill Width	1.16
Bill Length	1.50



# Example 2

■ Number of Sparrows ~

Covariate	P-value (full model)	VIF	P-value (collinearity removed)
% <i>Juncus gerardii</i>	0.0203	44.9953	0.0001
% Shrub	0.9600	2.7818	0.0568
Height of thatch	0.9989	1.6712	0.8263
% <i>Spartina patens</i>	0.0640	159.3506	0.3312
% <i>Distichlis spicata</i>	0.0527	53.7545	0.2538
% Bare ground	0.0666	12.0586	0.8908
% Other vegetation	0.0730	5.8170	0.9462
% <i>Phragmites australis</i>	0.0715	3.7490	0.2734
% Tall sedge	0.2160	4.4093	0.4313
% Water	0.0568	17.0677	0.6942
% <i>Spartina alterniflora</i> (short)	0.0549	121.4637	0.2949
% <i>Spartina alterniflora</i> (tall)	0.0960	159.3828	
Maximum vegetation height	0.2432	6.1200	
Vegetation stem density	0.7219	3.2064	

Zuur, Ieno & Elphick (2010). *A protocol for data exploration to avoid common statistical problems*

# What next? – Model Selection

---

- After removing collinear variables, we can move onto model selection procedures
  
- Three approaches in ecology and evolution:
  1. *A priori* specification of a model and present the model from the outset (do nothing)
  2. Classic model selection using criteria
    - What approach? What criteria? What terms?
  3. *A priori* specification of a multiple models and comparison
    - Information Theoretic (IT) approach

The background is a dark, blurred image featuring financial data visualizations. It includes a line graph with white circular markers and a bar chart with orange bars. Some data points are labeled with numbers like 183.102, 154.178, and 245.5. The overall aesthetic is professional and data-oriented.

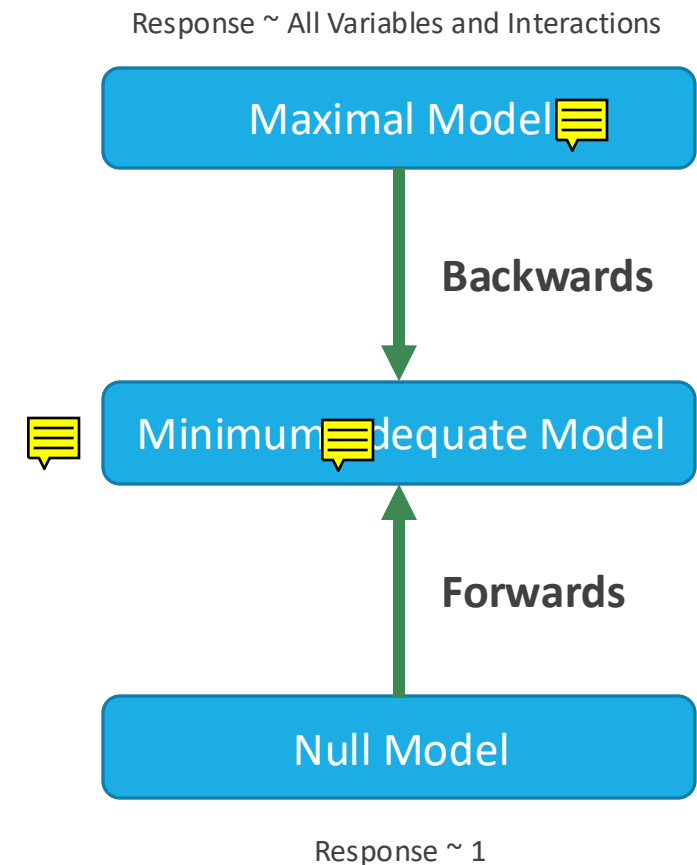
# Model Selection

---

# Classic Model Selection – The Approach

---

- Finding the optimal model = *the minimum adequate model*
- Stepwise (forward or backward selection)
- Backwards is preferable due to *a priori* specification of the maximal
- Forwards, and to a certain extent backwards, is criticized for data dredging
- Backwards builds a maximal model and sequentially removes terms (higher order interactions first and then main effects)
- Debated whether only interactions should be removed
- Removal based on hypothesis testing or information criteria





# Classic Model Selection – The Criteria

---

## 1. Hypothesis testing procedure

- Removing terms hierarchically based on highest p-values
- Starting with higher order interactions to main effects
- Removing terms will reduce the  $R^2$  (explanatory power) of your model and you evaluate whether this removal is negligible or not.

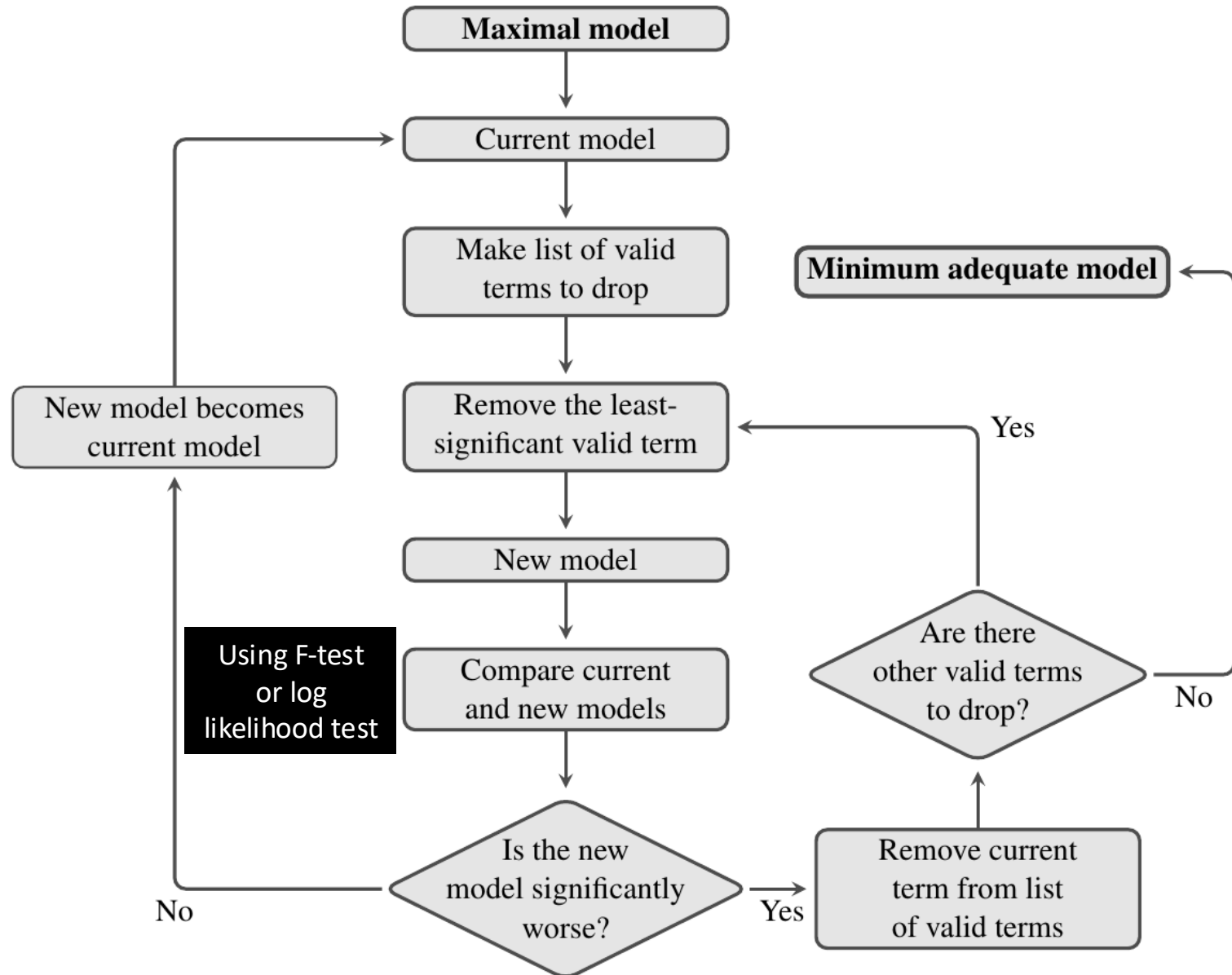
## 2. Information criteria

- Estimators of prediction error = quality of the model
  - Akaike Information Criterion (AIC)
  - Bayesian Information Criterion (BIC)

# Hypothesis Testing Procedure

## Rules:

1. Can't remove a significant term
2. Can't remove a main effect if part of an interaction
3. Always performed iteratively (removing one term at a time)



# Example

```
Call:
lm(formula = log(BMR..W.) ~ log(Mass..g.) * Food * Climate, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.86449	-0.24211	-0.00055	0.19882	0.95438

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.87740	0.13229	-29.310	< 2e-16 ***
log(Mass..g.)	0.72847	0.01960	37.165	< 2e-16 ***
FoodI	0.43199	0.17374	2.486	0.0141 *
ClimateTR	-1.11265	0.70226	-1.584	0.1154 .
log(Mass..g.):FoodI	-0.15781	0.03427	-4.605	9.2e-05 ***
log(Mass..g.):ClimateTR	0.12950	0.07351	1.762	0.0803 .
FoodI:ClimateTR	0.55724	0.71932	0.775	0.4399 .
log(Mass..g.):FoodI:ClimateTR	-0.01837	0.08316	-0.221	0.8255 .

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3669 on 138 degrees of freedom

(6 observations deleted due to missingness)

Multiple R-squared: 0.9692, Adjusted R-squared: 0.9676

F-statistic: 620 on 7 and 138 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: log(BMR..W.)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
log(Mass..g.)	1	576.75	576.75	4285.0054	< 2.2e-16 ***
Food	1	2.75	2.75	20.4488	1.307e-05 ***
Climate	1	0.28	0.28	2.0470	0.154770
log(Mass..g.):Food	1	2.60	2.60	19.3475	2.162e-05 ***
log(Mass..g.):Climate	1	1.44	1.44	10.7063	0.001349 **
Food:Climate	1	0.29	0.29	2.1493	0.144909
log(Mass..g.):Food:Climate	1	0.01	0.01	0.0488	0.825485
Residuals	138	18.57	0.13		

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



Analysis of Variance Table

Model 1: log(BMR..W.) ~ log(Mass..g.) \* Food \* Climate

Model 2: log(BMR..W.) ~ log(Mass..g.) + Food + Climate + log(Mass..g.):Food + log(Mass..g.):Climate + Food:Climate

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	138	18.574				
2	139	18.581	-1	-0.00688	0.0488	0.8255

# Example 2

	14		11		2
Covariate	<i>P</i> -value (full model)	VIF	<i>P</i> -value (collinearity removed)		<i>P</i> -value (reduced model)
% <i>Juncus gerardii</i>	0.0203	44.9953	0.0001		0.00004
% Shrub	0.9600	2.7818	0.0568		0.0727
Height of thatch	0.9989	1.0712	0.8205		
% <i>Spartina patens</i>	0.0640	159.3506	0.3312		
% <i>Distichlis spicata</i>	0.0527	53.7545	0.2538		
% Bare ground	0.0666	12.0586	0.8908		
% Other vegetation	0.0730	5.8170	0.9462		
% <i>Phragmites australis</i>	0.0715	3.7490	0.2734		
% Tall sedge	0.2160	4.4093	0.4313		
% Water	0.0568	17.0677	0.6942		
% <i>Spartina alterniflora</i> (short)	0.0549	121.4637	0.2949		
% <i>Spartina alterniflora</i> (tall)	0.0960	159.3828			
Maximum vegetation height	0.2432	6.1200			
Vegetation stem density	0.7219	3.2064			




# Mixed Model Selection

---




# Mixed Effects Model Selection

---

1. Specify the fixed effects for maximal model and fit using the `glS()` function.
2. Specify the random effect structure and fit using `lmer()` adding the argument `REML=TRUE`.
3. Compare `glS()` and `lmer()` model using log-likelihood test 
4. Iteratively remove the random effects and test sequentially using the log-likelihood test
5. Iteratively remove the fixed effects and test sequentially using the log-likelihood test
6. Refit the minimum adequate model with `REML=FALSE`, validate the model and interpret the coefficients.
7. Present the results in a table and make a supporting figure

# Example – Owl Data

---

- Sampled different 599 nestlings from 27 nests
- Measured:
  - Number of calls to a visit from a parent per nestling 
  - The sex of the visiting parent
  - How long it took the parent to arrive
  - Whether the owls were in a deprived or satiated treatment
  - The nest id  
- Research question:
  - What factors influence the number of calls of nestlings?

## Steps 1-3

```
M1<- gls(NegPerChick~FoodTreatment*SexParent*ArrivalTime, data=Owls)
M2<- lmer(NegPerChick~FoodTreatment*SexParent*ArrivalTime+(1|Nest), data=Owls, REML = TRUE)
lrtest(M1, M2)
```

Likelihood ratio test

Model 1: NegPerChick ~ FoodTreatment \* SexParent \* ArrivalTime

Model 2: NegPerChick ~ FoodTreatment \* SexParent \* ArrivalTime + (1 |  
Nest)

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	9	-1108.4			
2	10	-1100.0	1	16.902	3.935e-05 ***
---					

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-likelihood has increased by 8.4 = model accuracy/quality has significantly improved



## Step 4

Linear mixed model fit by REML ['lmerMod']

Formula: NegPerChick ~ FoodTreatment \* SexParent \* ArrivalTime + (1 | Nest)

Data: Owls

REML criterion at convergence: 2200

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.9010	-0.6884	-0.2621	0.5146	4.9993

Random effects:

Groups	Name	Variance	Std.Dev.
Nest	(Intercept)	0.2301	0.4797
	Residual	2.1344	1.4610

Number of obs: 599, groups: Nest, 27

Nest variance > 0 and accounts for non-independence so should keep

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	7.51039	1.69859	4.422
FoodTreatmentSatiated	-2.46045	2.42787	-1.013
SexParentMale	1.23164	2.16643	0.569
ArrivalTime	-0.22294	0.06827	-3.265
FoodTreatmentSatiated:SexParentMale	-2.68301	3.22559	-0.832
FoodTreatmentSatiated:ArrivalTime	0.06219	0.09792	0.635
SexParentMale:ArrivalTime	-0.04835	0.08699	-0.556
FoodTreatmentSatiated:SexParentMale:ArrivalTime	0.11279	0.13006	0.867

Non-significant highest order interaction

## Step 5

## Step 4

Linear mixed model fit by REML ['lmerMod']

Formula: NegPerChick ~ FoodTreatment \* SexParent \* ArrivalTime + (1 | Nest)

Data: Owls

REML criterion at convergence: 2200

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.9010	-0.6884	-0.2621	0.5146	4.9993

Random effects:

Groups	Name	Variance	Std.Dev.
Nest	(Intercept)	0.2301	0.4797
	Residual	2.1344	1.4610

Number of obs: 599, groups: Nest, 27

Fixed effects:

(Intercept)	7.51039
FoodTreatmentSatiated	-2.46045
SexParentMale	1.23164
ArrivalTime	-0.22294
FoodTreatmentSatiated:SexParentMale	-2.68301
FoodTreatmentSatiated:ArrivalTime	0.06219
SexParentMale:ArrivalTime	-0.04835
FoodTreatmentSatiated:SexParentMale:ArrivalTime	0.11279

Nest variance > 0 and accounts for non-independence so should keep

Analysis of Variance Table

		npar	Sum Sq	Mean Sq	F value
FoodTreatment		1	93.993	93.993	44.0375
SexParent		1	0.241	0.241	0.1130
ArrivalTime		1	80.782	80.782	37.8480
FoodTreatment:SexParent		1	0.447	0.447	0.2096
FoodTreatment:ArrivalTime		1	8.161	8.161	3.8237
SexParent:ArrivalTime		1	0.003	0.003	0.0014
FoodTreatment:SexParent:ArrivalTime		1	1.605	1.605	0.7521
	Estimate	St			
	0.09792		0.635		
	0.08699		-0.556		
	0.13006		0.867		

Non-significant highest order interaction

## Step 5

## Step 5

```
M3<- update(M2,.~.-FoodTreatment:SexParent:ArrivalTime)
lrtest(M2, M3)
```

Likelihood ratio test

Model 1: NegPerChick ~ FoodTreatment \* SexParent \* ArrivalTime + (1 | Nest)

Model 2: NegPerChick ~ FoodTreatment + SexParent + ArrivalTime + (1 | Nest) + FoodTreatment:SexParent + FoodTreatment:ArrivalTime + SexParent:ArrivalTime

	#Df	LogLik	Df	Chisq	Pr(>Chisq)
1	10	-1100.0			
2	9	-1099.2	1	1.4928	0.2218

Log-likelihood has increased by 0.8, no differences between the models but M3 is simpler than M2, repeat step 5

## Step 6

```
M7 <- update(M6, REML=FALSE)
summary(M7)
```

```
Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: NegPerChick ~ FoodTreatment + SexParent + ArrivalTime + (1 | Nest)
Data: Owls
```

AIC	BIC	logLik	deviance	df.resid
2194.2	2220.6	-1091.1	2182.2	593


Scaled residuals:

Min	1Q	Median	3Q	Max
-1.8537	-0.6805	-0.2237	0.5265	4.9483

Random effects:

Groups	Name	Variance	Std.Dev.
Nest	(Intercept)	0.2392	0.4891
Residual		2.1200	1.4560

Number of obs: 599, groups: Nest, 27

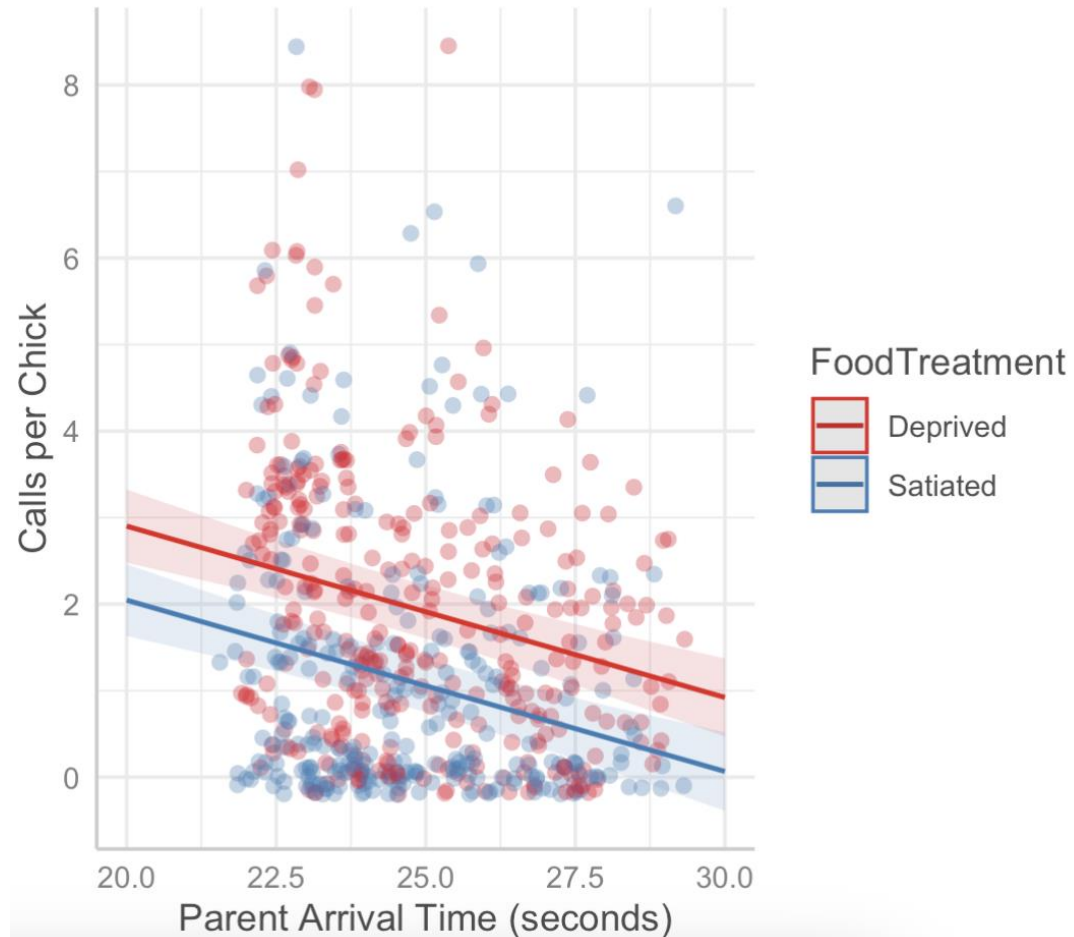
Nest explains 10.14% of variation  
in calls per chick 

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6.86017	0.80656	8.505
FoodTreatmentSatiated	-0.85527	0.12513	-6.835
SexParentMale	0.08250	0.13259	0.622
ArrivalTime	-0.19795	0.03206	-6.173

Chicks had fewer calls when satiated and when their parents took longer to arrive. A one second increase in arrival time decreased the number of calls per chick by  $-0.20 \pm 0.03$  ( $t = -6.17$ ). Sex of the parent had no effect on chick calls ( $0.08 \pm 0.13$ ,  $t = 0.62$ ).

## Step 7



```
require(ggeffects)
M7_plot <- ggpredict(M7, terms=c("ArrivalTime", "FoodTreatment"))
plot(M7_plot, add.data = T, jitter = TRUE) +
  labs(y="Calls per Chick", x="Parent Arrival Time (seconds)", title="")
```

<i>Predictors</i>	<b>Calls per Chick</b>		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	6.86	5.28 – 8.44	<b>&lt;0.001</b>
FoodTreatment [Satiated]	-0.86	-1.10 – -0.61	<b>&lt;0.001</b>
SexParent [Male]	0.08	-0.18 – 0.34	0.534
ArrivalTime	-0.20	-0.26 – -0.13	<b>&lt;0.001</b>
<b>Random Effects</b>			
$\sigma^2$	2.12		
$\tau_{00}$ Nest	0.24		
ICC	0.10		
$N_{\text{Nest}}$	27		
Observations	599		
Marginal $R^2$ / Conditional $R^2$	0.118 / 0.207		

```
require(sjPlot)
tab_model(M7, dv.labels = "Calls per Chick")
```

# sjPlot – tab\_model()

Calls per Chick			
Predictors	Estimates	CI	p
(Intercept)	6.86	5.28 – 8.44	<0.001
FoodTreatment [Satiated]	-0.86	-1.10 – -0.61	<0.001
SexParent [Male]	0.08	-0.18 – 0.34	0.534
ArrivalTime	-0.20	-0.26 – -0.13	<0.001
Random Effects			
$\sigma^2$	2.12	Residual Variance	
$\tau_{00}$ Nest	0.24	Nest Variance	
ICC	0.10	Proportion explained	
N <sub>Nest</sub>	27		
Observations	599		
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.118 / 0.207		

Predictors	Maximal Model			Minimum Adequate Model		
	Estimates	CI	p	Estimates	CI	p
(Intercept)	7.51	4.20 – 10.83	<0.001	6.86	5.28 – 8.44	<0.001
FoodTreatment [Satiated]	-2.47	-7.21 – 2.27	0.307	-0.86	-1.10 – -0.61	<0.001
SexParent [Male]	1.23	-3.00 – 5.46	0.567	0.08	-0.18 – 0.34	0.534
ArrivalTime	-0.22	-0.36 – -0.09	0.001	-0.20	-0.26 – -0.13	<0.001
FoodTreatment [Satiated] × SexParent [Male]	-2.71	-9.01 – 3.58	0.397			
FoodTreatment [Satiated] × ArrivalTime	0.06	-0.13 – 0.25	0.521			
SexParent [Male] × ArrivalTime	-0.05	-0.22 – 0.12	0.577			
(FoodTreatment [Satiated] × SexParent [Male]) × ArrivalTime	0.11	-0.14 – 0.37	0.378			
<b>Random Effects</b>						
$\sigma^2$	2.11			2.12		
$\tau_{00}$	0.21 Nest			0.24 Nest		
ICC	0.09			0.10		
N	27 Nest			27 Nest		
Observations	599			599		
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.125 / 0.205			0.118 / 0.207		

tab\_model(M2, M7, dv.labels=c("Maximal Model", "Minimum Adequate Model"))

# Information Criteria

---

MODEL SELECTION

# Information Criteria – AIC

---



$$AIC = -2 * \log Lik + 2 * p$$



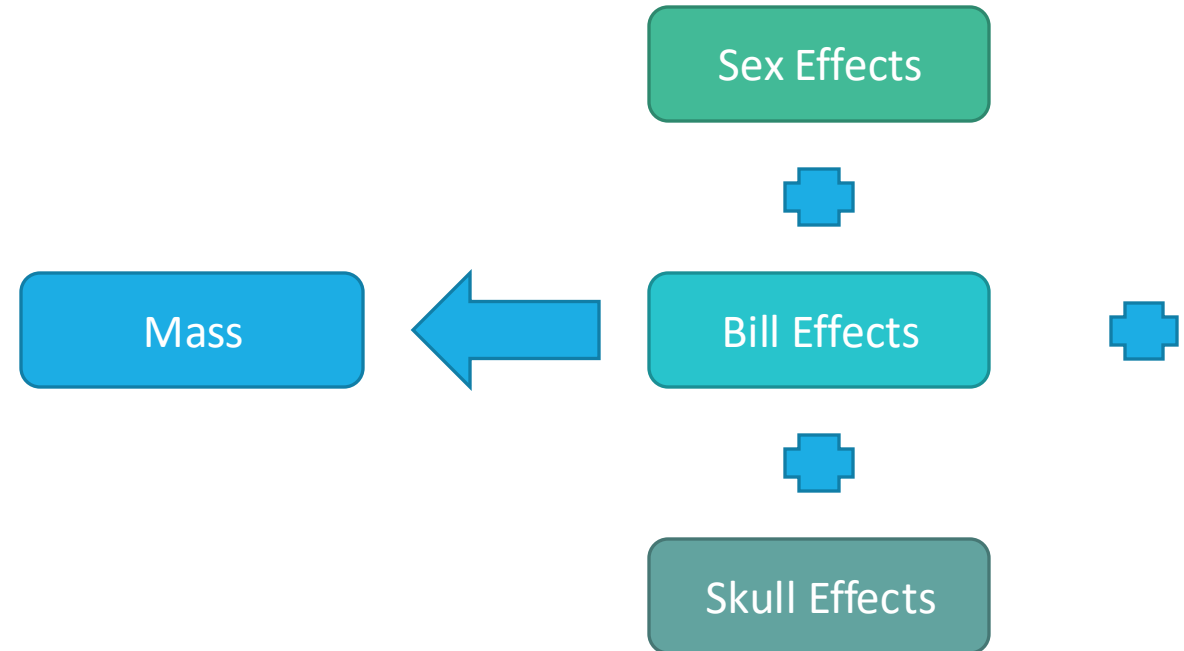
- Combines model fit via log-likelihood and model complexity (p= number of parameters estimated).
- AIC's are used to compared models built on the same response variable and dataset
- Preferable to avoid using AIC on models with small sample sizes (<50)
  - Can account for this by using AICc instead
- Choose the model with the lowest AIC value
- step function automates the selection procedure (see practical later)
- AIC is integral in the Information Theoretic (IT) Approach



# The Information Theoretic (IT) Approach

---

- Gaining popularity in ecological fields and dictates that we thoroughly assess the data with a range of plausible ecological hypotheses
- Using *a priori* knowledge construct 10-15 models
- Models are compared with AIC and the most interesting/important range of models are presented
- Different to hypothesis testing as it produces a range of models rather than just one



# The Information Theoretic (IT) Approach

---

- Gaining popularity in ecological fields and dictates that we thoroughly assess the data with a range of plausible ecological hypotheses
- Using *a priori* knowledge construct 10-15 models
- Models are compared with AIC and the most interesting/important range of models are presented
- Different to hypothesis testing as it produces a range of models rather than just one

Model	Description
Skull + Bill Length + Bill Depth + Bill Width + Sex	All variables (additive)
Skull + Bill Length + Bill Depth + Bill Width + Sex + Sex:Skull + Sex:Bill Length + Sex:Bill Depth + Sex:Bill Width	All variables (interactions with sex)
Skull + Bill Length + Bill Depth + Bill Width	All variables (no sex)
Skull + Sex	Skull and sex
Skull + Sex + Sex:Skull	Skull interacting with sex
Skull	Just Skull
Bill Length + Bill Depth + Bill Width + Sex	Bill traits and sex
Bill Length + Bill Depth + Bill Width + Sex + Sex:Bill Length + Sex:Bill Depth + Sex:Bill Width	Bill traits interaction with sex
Bill Length + Bill Depth + Bill Width	Just Bill traits

# The Information Theoretic (IT) Approach

Model	Description	AIC	AIC Differences	Akaike Weights
Skull + Bill Length + Bill Depth + Bill Width + Sex	All variables (additive)	665.70	0	0.65
Skull + Bill Length + Bill Depth + Bill Width + Sex + Sex:Skull + Sex:Bill Length + Sex:Bill Depth + Sex:Bill Width	All variables (interactions with sex)	671.41	5.71	0.04
Skull + Bill Length + Bill Depth + Bill Width	All variables (no sex)	667.17	1.47	0.31
Skull + Sex	Skull and sex	691.82	26.12	0
Skull + Sex + Sex:Skull	Skull interacting with sex	693.79	28.09	0
Skull	Just Skull	693.35	27.65	0
Bill Length + Bill Depth + Bill Width + Sex	Bill traits and sex	704.11	38.41	0
Bill Length + Bill Depth + Bill Width + Sex + Sex:Bill Length + Sex:Bill Depth + Sex:Bill Width	Bill traits interaction with sex	707.86	42.16	0
Bill Length + Bill Depth + Bill Width	Just Bill traits	702.76	37.06	0
Sex	Just sex	721.15	55.44	0

# The Information Theoretic (IT) Approach

---

- Models  $< 2$  AICs from best model are indistinguishable
- From the AICs, we can calculate Akaike Weights
- The Akaike Weights represent the relative likelihood of that model
- If we were to repeat the experiment a large number of times, in 65% of cases the model with all variable (additive) would be the optimal model, and in 31% of cases the all variable but no sex would be the optimal model.
- For the IT approach, we'd present and interpret both models **OR**

Model	Description	AIC	AIC Diff	Akaike Weights
Skull + Bill Length + Bill Depth + Bill Width + Sex	All variables (additive)	665.70	0	0.65
Skull + Bill Length + Bill Depth + Bill Width + Sex + Sex:Skull + Sex:Bill Length + Sex:Bill Depth + Sex:Bill Width	All variables (interactions with sex)	671.41	5.71	0.04
Skull + Bill Length + Bill Depth + Bill Width	All variables (no sex)	667.17	1.47	0.31

# Model Averaging

---

- Averaging of the coefficients of models  $< 2$  AICs from the best model (lowest AIC).
- Two approaches:
  - Conditional averaging – average coefficients in which the variable exists

	Skull	Bill Length	Bill Width	Bill Depth	Sex
Model1	0.25	1.34	1.44	1.56	0.5
Model2	0.19	1.34	1.47	1.86	--
Average	0.22	1.34	1.46	1.71	0.5

# Model Averaging

---

- Averaging of the coefficients of models  $< 2$  AICs from the best model (lowest AIC).
- Two approaches:
  - Conditional averaging – average coefficients in which the variable exists
  - Full averaging – average coefficients whereby missed selection receive 0

	Skull	Bill Length	Bill Width	Bill Depth	Sex
Model1	0.25	1.34	1.44	1.56	0.5
Model2	0.19	1.34	1.47	1.86	0
Average	0.22	1.34	1.46	1.71	0.25

- This discourse reiterates the message of:

GEORGE E P BOX  
20<sup>TH</sup> CENTURY  
STATISTICIAN

*All models are wrong,  
but some are useful*

# Summary

---

- Investigating and adjusting statistical models according to collinearity is vital to reduce Type II errors
- The variance-inflation factor can be used to whittle down the number of variables in a model but the threshold for this is debatable (3, 5 or 10)
- Model selection is a controversial topic in statistics and ecological modelling and various approaches exist; from do nothing to data dredging approaches like stepwise forward selection
- Stepwise backwards selection involved removing terms sequentially from a model evaluating the reduced model to the previous model in finding the minimum adequate model
- The Information Theoretic (IT) approach produces a range of candidate models based on prior knowledge, of which all should be interpreted and presented