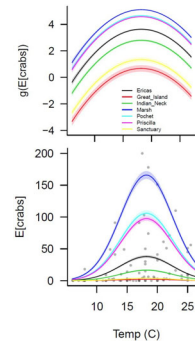


ECO 636 Applied Ecological Statistics

Week 4 – Multiple linear regression – recorded lecture



Meg Graham MacLean, PhD
Department of Environmental
Conservation

mgmaclean@umass.edu

2021 - Spring

Welcome to week 4s recorded lecture.

The Week

Tuesday

- Recap exercise from recorded lecture
- Regression!

Wednesday (Lab)

- No lab – Wellness Wednesday!

Thursday

- Standardizing
- Continue with regression/multiple linear regression

ECO 636 week 4 - Multiple linear regression - recorded lecture

2

Today we will build on what we started on Tuesday and continue with the salamanders just a little bit more to talk about standardization and what it can be useful for, then we will talk through multiple linear regression!

Simple linear regression example

- 331 salamanders
- Measured total length (TL) and snout-to-vent length (SVL)
- Tail length (Tail) = TL - SVL

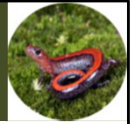


ECO 636 week 4 - Multiple linear regression - recorded lecture

3

So, back to those salamanders!

Example



4. Fit the model

- Algebraically
 - $y_i = \beta_0 + \beta_1 X_{1i} + e_i$
- In R:

```
> mAllo <- lm(Tail ~ SVL, data = mander)
> coef(mAllo)
(Intercept)      SVL 
 3.1490945    0.8942684
```

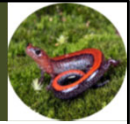
$$y = 3.15 + 0.89x$$

ECO 636 week 4 - Multiple linear regression - recorded lecture

4

On Tuesday we build a linear model to look at the relationship between snout-to-vent length and tail length and these were our coefficients.

Example



6. Interpret the results

```
> summary(mAllo)

Call:
lm(formula = Tail ~ SVL, data = mander)

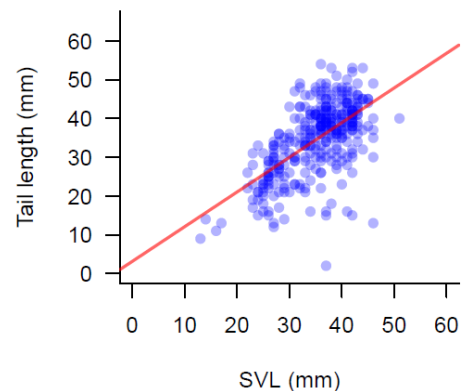
Residuals:
    Min       1Q   Median       3Q      Max
-34.237  -3.867   0.657   4.657  18.657

Coefficients:
(Intercept)   3.14909   2.32421   1.355   0.
SVL           0.89427   0.06502  13.754  <2e-

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.

Residual standard error: 7.477 on 329 degrees
Multiple R-squared:  0.3651, Adjusted R-squa
F-statistic: 189.2 on 1 and 329 DF, p-value:
```

Does the intercept make biological sense?



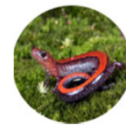
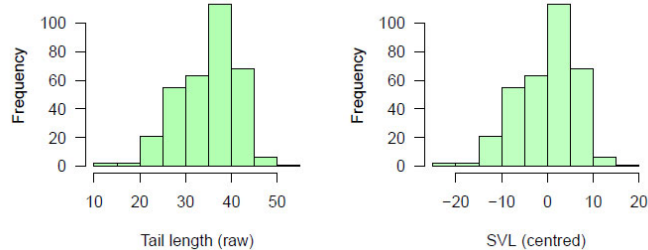
If we think about what this means in real life...

Does the intercept make sense to you?

Remember that in this case, the intercept is when SVL is zero... so... can we have a 3.15mm length tail when we have no snout-to-vent length? That would be a bodyless tail! Yikes. So, to me this intercept doesn't really have a biological meaning, so one of the things we can do is standardize our explanatory variable to give our intercept meaning.

Standardizing

- Centering
 - Centering data around 0 by subtracting the mean
 - $x'_i = x_i - \bar{x}$
 - Units remain the same!

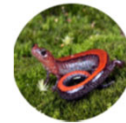
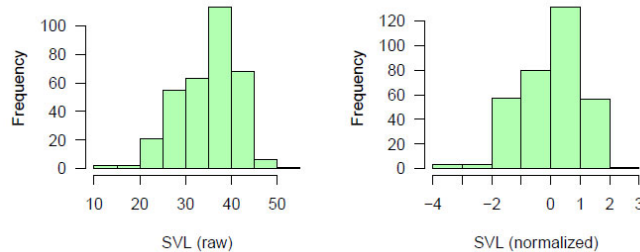


6

One version of a standardization is called centering – with centering we just subtract the mean from each observation to get our new centered value – then all of the values are around the mean. The nice thing about centering is that your values all retain their units, which make them more easily interpretable.

Standardizing

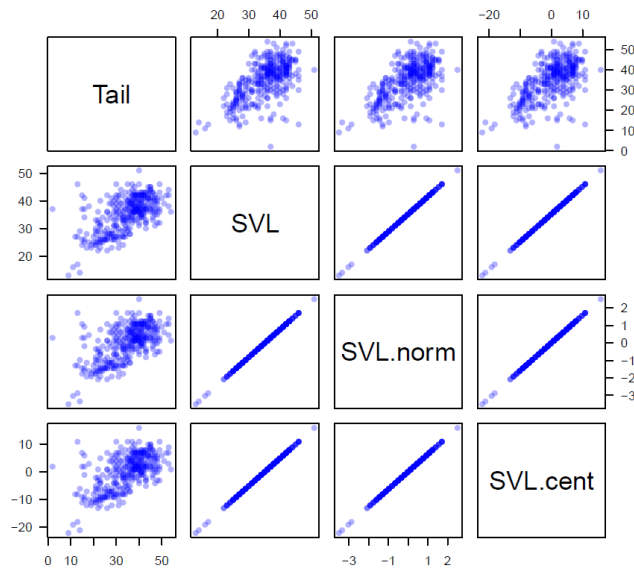
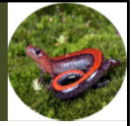
- Centering
- Normalization
 - Subtracting the mean and dividing by the sample sd
 - $x'_i = (x_i - \bar{x})/S_x$
 - One unit is now a standard deviation (a bit harder to interpret!)



7

Similarly, we can also normalize the data, where we subtract the mean and then divide by the standard deviation. This standardization is often used when you have multiple explanatory variables and they have very different ranges. By normalizing each of the variables, they will then have the same ranges and contribute equally to the model. However, one pretty big downside of normalization is that your values become unitless – so they become a bit harder to interpret later on.

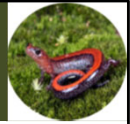
Standardizing



8

If we take a look at the correlation plot, you can see that both the normalized SVL values and the centered SVL values are perfectly correlated with the original SVL values, so we haven't changed those relationships at all.

Standardizing

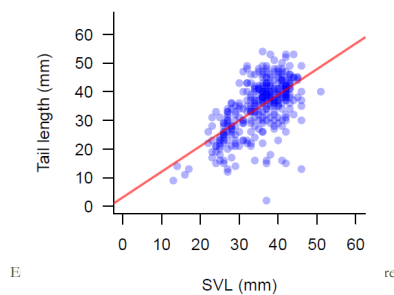


Original sample data:

```
> mAllo <- lm(Tail ~ SVL, data=mander)
>
> summary(mAllo)$coefficients
```

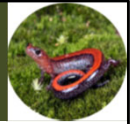
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.1490945	2.3242142	1.354907	1.763767e-01
SVL	0.8942684	0.0650175	13.754273	2.548099e-34

Intercept is the tail length when SVL = 0;
Slope is the change in tail length for every 1mm change in SVL



So here is our original SVL mode where our intercept is the same as our tail length when SVL is zero. Again, it looks like we could have a bodyless tail!

Standardizing

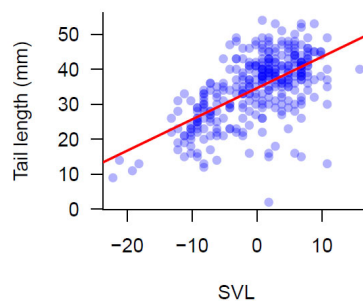


Centered data:

```
> mAllo.cent <- lm(Tail ~ SVL.cent, data=mander)
>
> summary(mAllo.cent)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	34.6132931	0.4109545	84.22659	1.047262e-224
SVL.cent	0.8942684	0.0650175	13.75427	2.548099e-34

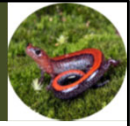
Intercept is the tail length for the *average* SVL - interesting!
Slope is the change in tail length for every 1mm change in SVL



10

But instead, what if we first center SVL and then run our linear model again with the centered values? Well, now our intercept is actually tail length for the average SVL! This seems more interesting and more biologically relevant! Also note that our estimate of the slope is exactly the same as in our previous model, so the relationship between SVL and tail length is unchanged with the new values.

Standardizing

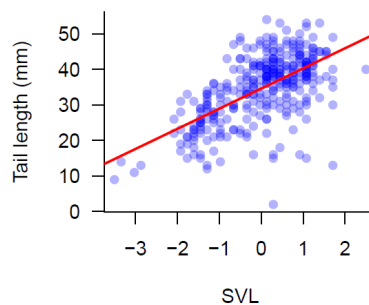


Normalized data:

```
> mAllo.norm <- lm(Tail ~ SVL.norm, data=mander)
>
> summary(mAllo.norm)$coefficients
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	34.613293	0.4109545	84.22659	1.047262e-224
SVL.norm	5.660938	0.4115766	13.75427	2.548099e-34

Intercept is the tail length for the *average* SVL
Slope is the change in tail length for every 1 *SD* change in SVL



11

Finally, if we look at our normalized data, notice the estimate for the intercept is the same as with our centered data, or represents the tail length for the average SVL, but now our slope has changed! That is because the slope now represents the change in tail length for every 1 standard deviation change in SVL – a slightly more difficult thing to interpret, but not impossible.

Review!

So far...

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	$\text{lm}(y \sim 1)$
Continuous	Single two-level factor	<i>t</i> -test	$\text{lm}(y \sim x)$
Continuous	Single multi-level factor	One-way ANOVA	$\text{lm}(y \sim x)$
Continuous	>1 multi-level factor (*)	Two-way ANOVA	$\text{lm}(y \sim x_1 * x_2)$
Continuous	Single continuous	Simple linear regression	$\text{lm}(y \sim x)$

Well, great! We've talked through some standardization methods, so let's move on to multiple linear regression! To review.. On Tuesday we talked through simple linear regression...

Review!

Next!

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	$\text{lm}(y \sim 1)$
Continuous	Single two-level factor	<i>t</i> -test	$\text{lm}(y \sim x)$
Continuous	Single multi-level factor	One-way ANOVA	$\text{lm}(y \sim x)$
Continuous	>1 multi-level factor (*)	Two-way ANOVA	$\text{lm}(y \sim x_1 * x_2)$
Continuous	Single continuous	Simple linear regression	$\text{lm}(y \sim x)$
Continuous	Multiple continuous	Multiple linear regression	$\text{lm}(y \sim x_1 * x_2)$

Estimating the relationship with multiple explanatory variables!

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

And now we will go on to multiple linear regression when we have multiple continuous explanatory variables!

Multiple linear regression

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

- β_0 is the intercept
- β_1 is the slope of the X_1 relationship
 - The change in \hat{y}_i with one unit change in X_1 at *any* value of X_2 (*additive model!*)
- β_2 is the slope of the X_2 relationship
 - The change in \hat{y}_i with one unit change in X_2 at *any* value of X_1 (*additive model!*)

The changes in \hat{y}_i with the change in one explanatory variable while the other(s) held constant are often called: **marginal effects**

ECO 636 week 4 - Multiple linear regression - recorded lecture

14

So, like we have in simple linear regression, beta-naught is our intercept, beta-one is the slope of the X_1 relationship for any values of X_2 – with the note here that we are only looking at the additive model 😊. Next, beta-two is the slope of the X_2 relationship for any value of X_1 .

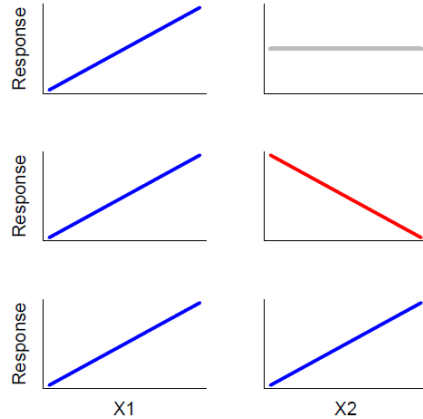
A note on terminology here – the changes in y , or our response variable, with the change in one explanatory variable while the others are held constant are often called the marginal effects of that variable.

Multiple linear regression

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

What are the possible general outcomes for our parameters?

$$\beta_1 > 0$$



$$\begin{aligned} \beta_2 &= 0 \\ \beta_2 &< 0 \\ \beta_2 &> 0 \end{aligned}$$

15

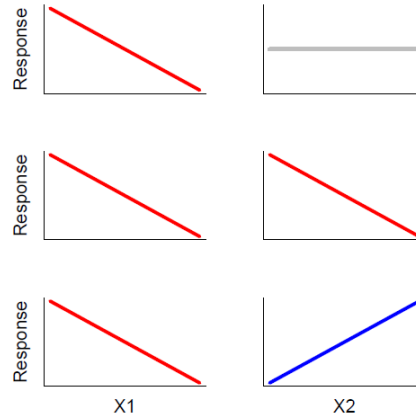
So, what are the possible outcomes of for our parameters? Well, there are nine of them – first, the marginal effects of X1 could be positive, so beta-one could be positive, with beta-two could be zero, negative or positive.

Multiple linear regression

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

What are the possible general outcomes for our parameters?

$$\beta_1 < 0$$



$$\begin{aligned} \beta_2 &= 0 \\ \beta_2 &< 0 \\ \beta_2 &> 0 \end{aligned}$$

16

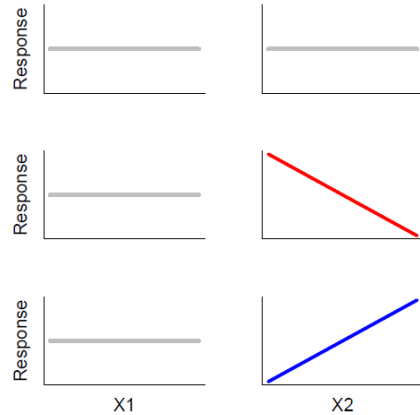
Similarly, beta-one could be negative, with the other three choices for beta-two.

Multiple linear regression

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

What are the possible general outcomes for our parameters?

$$\beta_1 = 0$$



$$\begin{aligned}\beta_2 &= 0 \\ \beta_2 &< 0 \\ \beta_2 &> 0\end{aligned}$$

17

And beta-one could also equal zero with all the same options for beta-two.

Example

Indigo snakes

- 92 indigo snakes
- We are interested in the variation in home range sizes of the snakes (hr.size)
 - We log transformed the home range data ($\log(\text{HR})$)
- Our proposed covariates are surrounding habitat structure (proportion)
 - Urban1.50
 - Upland1.50
 - Wetland1.50



ECO 636 week 4 - Multiple linear regression - recorded lecture

18

So, to explore how multiple linear regression works, let's try it with a new set of data. I'll have you follow along right now and then work with these data at the end of the recorded lecture. Remember that you can always download the pdf of the slides to use as a reference especially for the code.

So, in this example we are interested in the variation in home range size of some indigo snakes based on what their home ranges are made up of.

So, these data are very skewed, with a lot of small home range sizes and a few much larger home range sizes, so I have log transformed the home range data to make them more normally distributed. You will have to do this for your dataset as well.

The covariates that were measured from a land cover map, are the proportion of a 1.5 km radius circle around the nesting site is urban, upland, or wetland.

Example



Modeling process:

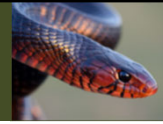
1. State the question/hypothesis
 - What is the question?
 - What are the variables (response and explanatory)?
2. Data exploration
3. Describe the model
 - In word form (should come from your question)
 - In mathematical form
 - Identify the assumptions of the model
4. Fit the model! (In R, of course 😊)
5. Evaluate the output
 - Model validation
 - Model selection
6. Interpret the results

ECO 636 week 4 - Multiple linear regression - recorded lecture

19

So, let's get started on the modeling process before I hand it off to you.

Example



1. State the question/hypothesis

Question:

Does habitat composition at the home range scale explain variation in home range size?

H_0 : There is no relationship between home range size and habitat composition.

Variables:

- Response:
 - Home range
- Explanatory
 - Urban percent
 - Upland percent
 - Wetland percent

ECO 636 week 4 - Multiple linear regression - recorded lecture

20

First, like we just talked about – the question is: does habitat composition at the home range scale explain variation in the home range size? And the null hypothesis is that there is no relationship between home range size and habitat composition.

We just chatted about our variables as well.

Example



Modeling process:

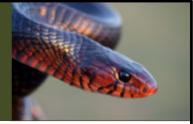
1. *State the question/hypothesis*
 - *What is the question?*
 - *What are the variables (response and explanatory)?*
2. Data exploration
3. Describe the model
 - In word form (should come from your question)
 - In mathematical form
 - Identify the assumptions of the model
4. Fit the model! (In R, of course 😊)
5. Evaluate the output
 - Model validation
 - Model selection
6. Interpret the results

ECO 636 week 4 - Multiple linear regression - recorded lecture

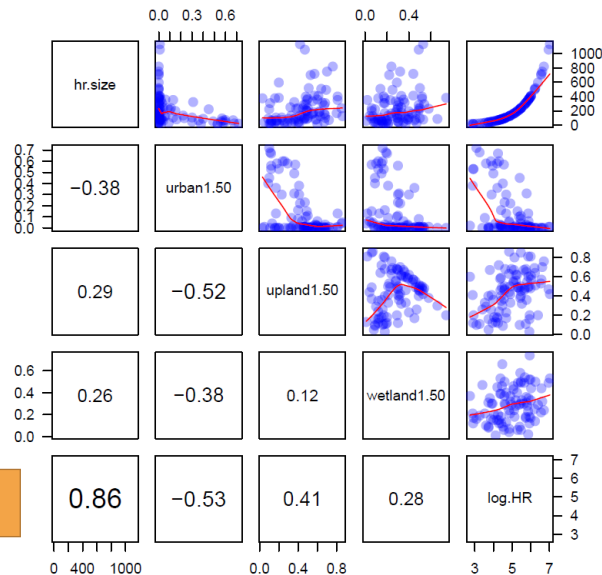
21

So lets do a little data exploration

Example



2. Data exploration



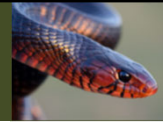
Some evidence of *collinearity* - we'll deal with this a bit later!

22

So I have done a little data exploration already, but wanted to point out one thing before moving on. Here is a co plot to start looking at covariation in the data...

When we look closely at the relationships between urban and upland, there is a little evidence of collinearity, but not terrible, so we will keep going for now, but we will learn how to check for and deal with collinearity next week.

Example



Modeling process:

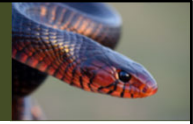
1. *State the question/hypothesis*
 - *What is the question?*
 - *What are the variables (response and explanatory)?*
2. *Data exploration*
3. Describe the model
 - In word form (should come from your question)
 - In mathematical form
 - Identify the assumptions of the model
4. Fit the model! (In R, of course 😊)
5. Evaluate the output
 - Model validation
 - Model selection
6. Interpret the results

ECO 636 week 4 - Multiple linear regression - recorded lecture

23

Ok, on to describing the model.

Example



3. Describe the model

- In words:
 - Does habitat composition at the home range scale explain variation in home range size?
- As a mathematical model:
 - $y_i = \beta_0 + \beta_1 X_{urbi} + \beta_2 X_{upi} + \beta_3 X_{weti} + e_i$
 - $H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$
- Assumptions?
 - Residuals are normally distributed
 - Constant variance (homogeneity)
 - Observations are independent
 - Predictors measured without error (fixed X)

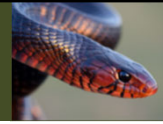
ECO 636 week 4 - Multiple linear regression - recorded lecture

24

So first, here is our mathematical model with three explanatory variables and our null hypothesis is that all of our betas are zero. What are our assumptions?

Yes, that's right, the same as always 😊.

Example

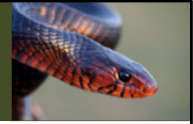


Modeling process:

1. *State the question/hypothesis*
 - *What is the question?*
 - *What are the variables (response and explanatory)?*
2. *Data exploration*
3. *Describe the model*
 - *In word form (should come from your question)*
 - *In mathematical form*
 - *Identify the assumptions of the model*
4. Fit the model! (In R, of course 😊)
5. Evaluate the output
 - Model validation
 - Model selection
6. Interpret the results

So on to fit the model!

Example



4. Fit the model

- Algebra: $y_i = \beta_0 + \beta_1 X_{urbi} + \beta_2 X_{upi} + \beta_3 X_{weti} + e_i$

- R:

```
> mG <- lm(log.HR ~ urban1.50 + upland1.50 + wetland1.50, data = indigos)
> model.matrix(~ urban1.50 + upland1.50 + wetland1.50, data = indigos)
```

	(Intercept)	urban1.50	upland1.50	wetland1.50
1	1	0.36	0.17	0.45
2	1	0.00	0.67	0.35
3	1	0.01	0.67	0.24
4	1	0.06	0.60	0.35
5	1	0.00	0.55	0.46
6	1	0.30	0.23	0.40

ECO 636 week 4 - Multiple linear regression - recorded lecture

26

Great – here again is our algebraic model, and how we create that model in R. Before we move on, let's look at one thing...

Do you remember looking at the model matrices for our two-way ANOVAs last week? Here is the model matrix for this multiple-linear regression. Notice that under each of our variables in the model matrix, we now have continuous values rather than zeros and ones! When we are looking at categorical variables, R assigns either zeros or ones to the Xs based on the group – however, when we look at the model matrix for continuous data, the Xs are now just their values! So all the numbers you see in this model matrix are the observations!

Example



4. Fit the model

$$y_i = \beta_0 + \beta_1 X_{urbi} + \beta_2 X_{upi} + \beta_3 X_{weti} + e_i$$

```
> summary(mG)

Call:
lm(formula = log.HR ~ urban1.50 + upland1.50 + wetland1.50, data = indigos)

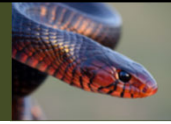
Residuals:
    Min       1Q   Median       3Q      Max
-2.63249 -0.52193  0.06665  0.59191  1.69868

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.6091     0.3502  13.160 < 2e-16 ***
urban1.50     -1.8465     0.5392  -3.424 0.000938 ***
upland1.50     0.9171     0.4999   1.835 0.069945 .
wetland1.50    0.6711     0.6086   1.103 0.273153
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8476 on 88 degrees of freedom
Multiple R-squared:  0.3091, Adjusted R-squared:  0.2855
F-statistic: 13.12 on 3 and 88 DF, p-value: 3.694e-07
```

So, now that we have our first model, or our full or global model, with all of the potential explanatory variables. Interesting! In this model it looks like our intercept is significantly different from zero – which I am not sure has a totally relevant biological explanation, but also that the slope of the relationship between the percent urban in the home range and home range size is significantly negative. Hm... but what about the other two explanatory variables, are they helping in this model at all?

Example



Modeling process:

1. *State the question/hypothesis*
 - *What is the question?*
 - *What are the variables (response and explanatory)?*
2. *Data exploration*
3. *Describe the model*
 - *In word form (should come from your question)*
 - *In mathematical form*
 - *Identify the assumptions of the model*
4. *Fit the model! (In R, of course 😊)*
5. **Evaluate the output**
 - **Model validation**
 - **Model selection**
6. **Interpret the results**

ECO 636 week 4 - Multiple linear regression - recorded lecture

28

So, let's evaluate our output!

Example



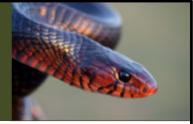
5. Evaluate the output

- Model validation - check assumptions!
 - Residuals are normally distributed
 - Constant variance (homogeneity)
 - Observations are independent
 - Predictors measured without error (fixed X)

```
> mG <- lm(log.HR ~ urban1.50 + upland1.50 + wetland1.50, data = indigos)
> summary(mG)
> plot(mG)
```

First, let's check our model assumptions.

Example

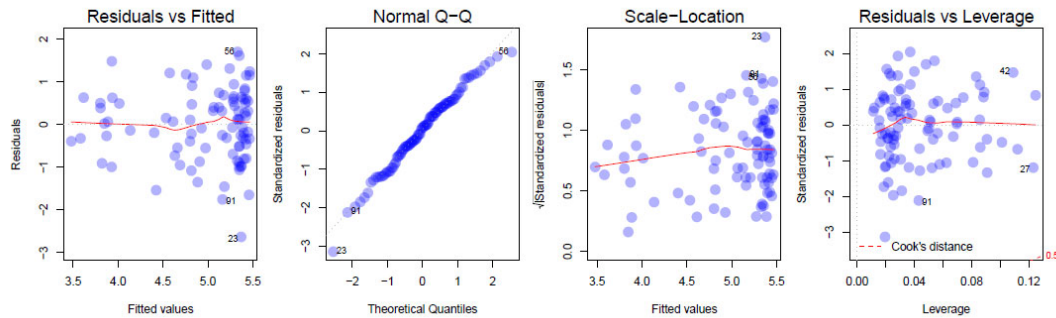


5. Evaluate the output

- Model validation - check assumptions!

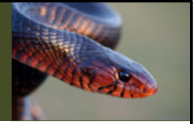
```
> mG <- lm(log.HR ~ urban1.50 + upland1.50 + wetland1.50, data = indigos)
> summary(mG)
> plot(mG)
```

Do we meet our assumptions?



What do you think? Do you think we meet our model assumptions? We look pretty good to me, so let's keep going on to model selection.

Example



5. Evaluate the output

- Model selection

- What is the best model? What other candidate models are there?

- Full (global) model: $y_i = \beta_0 + \beta_1 X_{urbi} + \beta_2 X_{upi} + \beta_3 X_{weti} + e_i$

- Single predictor models:

- $y_i = \beta_0 + \beta_1 X_{urbi} + e_i$

- $y_i = \beta_0 + \beta_2 X_{upi} + e_i$

- $y_i = \beta_0 + \beta_3 X_{weti} + e_i$

- Null model: $y_i = \beta_0 + e_i$

- Combination models!

- $y_i = \beta_0 + \beta_1 X_{urbi} + \beta_2 X_{upi} + e_i$

- $y_i = \beta_0 + \beta_1 X_{urbi} + \beta_3 X_{weti} + e_i$

- $y_i = \beta_0 + \beta_2 X_{upi} + \beta_3 X_{weti} + e_i$

ECO 636 week 4 - Multiple linear regression - recorded lecture

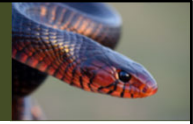
31

So... what is our best model? What other models should we try?

So, we already tried our full or global model... What other combinations can we try?

Well, we can try each explanatory variable on their own, we can try the null model (as we always should), and we can try any combination of two of the variables like shown here.

Example



5. Evaluate the output

- Model selection

- What is the best model? What other candidate models are there?

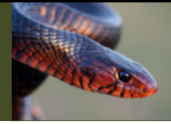
```
> UrUpWe<- lm(log.HR ~ urban1.50 + upland1.50 + wetland1.50, data = indigos)
> UrUp  <- lm(log.HR ~ urban1.50 + upland1.50, data = indigos)
> UrWe  <- lm(log.HR ~ urban1.50 + wetland1.50, data = indigos)
> UpWe  <- lm(log.HR ~ upland1.50 + wetland1.50, data = indigos)
> Ur    <- lm(log.HR ~ urban1.50, data=indigos)
> Up    <- lm(log.HR ~ upland1.50, data=indigos)
> We    <- lm(log.HR ~ wetland1.50, data=indigos)
> m0    <- lm(log.HR ~ 1, data=indigos)
```

How do we pick?

So, lets set up each of those version of the model.

Now, how do we pick?

Example



5. Evaluate the output

- Model selection
 - What is the best model?
 - AIC

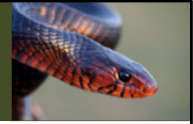
```
> fitList <- list(  
+ UrUpWe = lm(log.HR ~ urban1.50 + upland1.50 + wetland1.50, data = indigos),  
+ UrUp   = lm(log.HR ~ urban1.50 + upland1.50, data = indigos),  
+ UrWe   = lm(log.HR ~ urban1.50 + wetland1.50, data = indigos),  
+ UpWe   = lm(log.HR ~ upland1.50 + wetland1.50, data = indigos),  
+ Ur     = lm(log.HR ~ urban1.50, data=indigos),  
+ Up     = lm(log.HR ~ upland1.50, data=indigos),  
+ We     = lm(log.HR ~ wetland1.50, data=indigos),  
+ m0     = lm(log.HR ~ 1, data=indigos)  
+ )
```

ECO 636 week 4 - Multiple linear regression - recorded lecture

33

Well, let's add them all to a fit list and use AIC!

Example



5. Evaluate the output

- Model selection

- What is the best model? What other candidate models are there?

```
> (modtab <- aictab(fitList))
```

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
UrUp	4	236.29	0.00	0.38	0.38	-113.91
Ur	3	237.15	0.87	0.25	0.63	-115.44
UrUpWe	5	237.26	0.98	0.24	0.87	-113.28
UrWe	4	238.48	2.19	0.13	1.00	-115.01
UpWe	4	246.53	10.25	0.00	1.00	-119.04
Up	3	250.30	14.01	0.00	1.00	-122.01
We	3	259.50	23.21	0.00	1.00	-126.61
m0	2	264.72	28.43	0.00	1.00	-130.29

Huh, looks like maybe we should try the Urban and Upland model

ECO 636 week 4 - Multiple linear regression - recorded lecture

34

Excellent! Here are our results! It looks like our top model is just has Urban and Upland as explanatory variables, but I will admit that the full model is still within one delta AIC unit from this top model, so it's not too shabby either. A case could be made for any of these top three models.

But, it looks like it is worth at least checking out this top model!

Example



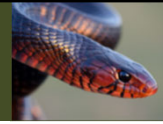
Modeling process:

1. *State the question/hypothesis*
 - *What is the question?*
 - *What are the variables (response and explanatory)?*
2. *Data exploration*
3. *Describe the model*
 - *In word form (should come from your question)*
 - *In mathematical form*
 - *Identify the assumptions of the model*
4. Fit the model! (In R, of course 😊)
5. Evaluate the output
 - Model validation
 - Model selection
6. Interpret the results

Now you try! Give the rest of the model fitting, evaluation and interpretation a try!

So, now that we know we may have a better model – let's try it all over again! This time I am going to have you go through the process on your own with the new model.

Example



Modeling process:

1. *State the question/hypothesis*
 - *What is the question?*
 - *What are the variables (response and explanatory)?*
2. *Data exploration*
3. *Describe the model*
 - *In word form (should come from your question)*
 - *In mathematical form*
 - *Identify the assumptions of the model*
4. *Fit the model! (In R, of course 😊)*
5. *Evaluate the output*
 - *Model validation*
 - *Model selection*
6. *Interpret the results*

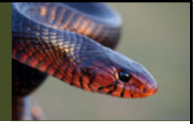
HINT!

ECO 636 week 4 - Multiple linear regression - recorded lecture

36

But - just to give you a little hint for the interpret the results section...

Example



6. Interpret the results

- How do we visualize our model?

$$y_i = \beta_0 + \beta_1 X_{urbi} + \beta_2 X_{upi} + e_i$$

```
> summary(mTop)$coefficients
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.854762   0.2705938 17.941140 2.600465e-31
urban1.50    -2.070404   0.5001409 -4.139642 7.885350e-05
upland1.50    0.863374   0.4981017  1.733329 8.650048e-02
```

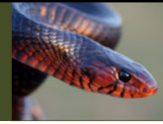
$$y_i = 4.85 - 2.07X_{urbi} + 0.86X_{upi}$$

- We can visualize our model by holding all but one X constant
 - Typically we use the mean of Xs we are holding constant

When you get to this part, you should end up with a model somewhat like this...

We can then put our estimated coefficients into the model. However, with two xs it becomes a little tricky to plot our model! We could plot a 3-d model OR we could visualize our model by plotting the effects of one of our explanatory variables at a time – this is usually how we do it. In this case, we usually set one of our xs to a constant, like zero or the mean of those values. So let's see what that might look like.

Example



6. Interpret the results

- Let's visualize the Urban relationship first

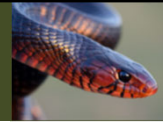
- $\bar{X}_{up} = 0.44$

$$y_i = 4.85 - 2.07X_{urbi} + 0.86 \times 0.44$$

```
> urb.seq <- seq(0,1,0.05) #these are proportions
> up.mean <- mean(indigos$upland1.50)
>
> #vary Urban keep Upland constant
> urb.df <- data.frame(urban1.50 = urb.seq, upland1.50 = up.mean)
```

So, in this example, we can visualize the urban relationship first by setting our upland habitat proportion to the mean of all of the values.

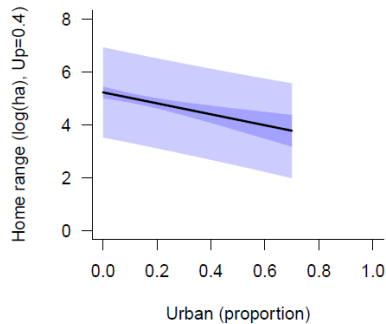
Example



$$y_i = 4.85 - 2.07X_{urbi} + 0.86 \times 0.44$$

- Show relationship and uncertainty in relationship

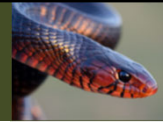
```
> #Predict Urban relationship  
> CI.urb <- predict(mTop, newdata=urb.df, interval="confidence")  
> PI.urb <- predict(mTop, newdata=urb.df, interval="prediction")
```



39

And then predicting on these new values, so with upland held constant at the mean. This is what the relationship should look like!

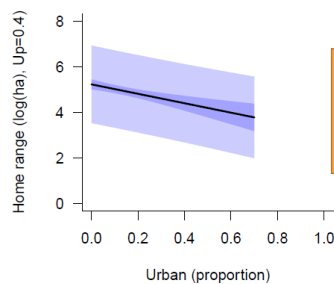
Example



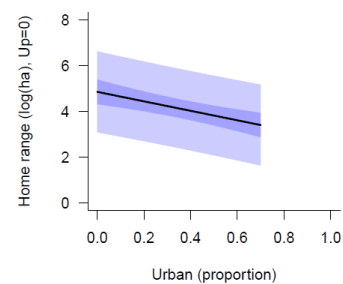
$$y_i = 4.85 - 2.07X_{urbi} + 0.86 \times 0.44$$

- Show relationship and uncertainty in relationship
- What about when there are no uplands?
 - $\bar{X}_{up} = 0$

```
> urb.df <- data.frame(urban1.50 = urb.seq, upland1.50 = 0)
```



WAIT! Our home range is still in log units - lets back transform the estimated home ranges!



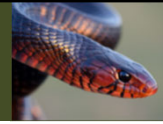
week 4 - Multiple linear regression - recorded lecture

We can also do the same thing, but with the upland category set to zero – so if there were no uplands, and that should look like...

This.

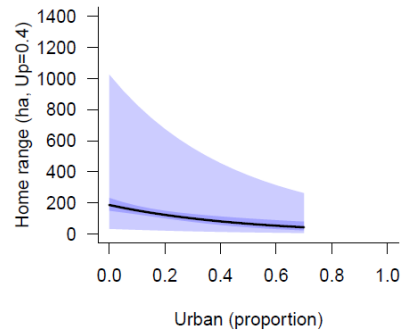
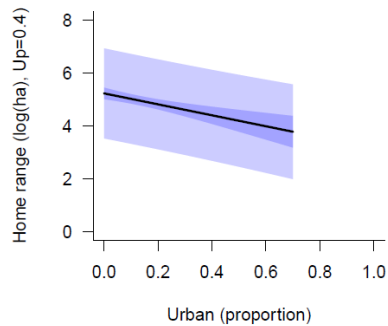
BUT WAIT! Our home ranges are still in log units! So lets be sure to back transform the estimated home ranges to actually see what these would look like in real units!

Example



$$y_i = 4.85 - 2.07X_{urbi} + 0.86 \times 0.44$$

- Show relationship and uncertainty in relationship



ECO 636 week 4 - Multiple linear regression - recorded lecture

41

There we go! This gives us a more realistic idea of how home ranges vary by urban proportion! Cool! You can use a similar technique to the upland predictor variable.

For next week:



For Tuesday:

- 1) Watch the recorded lecture and do the exercise
- 2) Complete the individual assessment on Moodle by 11:55pm Monday night.

Thanks and see you on Tuesday!