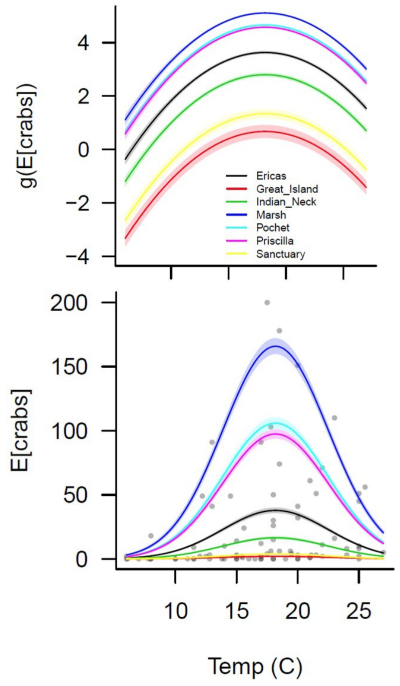
ECO 636
Applied Ecological Statistics

Week 4 – Simple linear regression



Meg Graham MacLean, PhD

Department of Environmental Conservation

mgmaclean@umass.edu

2021 - Spring

The Week

Tuesday

- Recap exercise from recorded lecture
- Regression!

Wednesday (Lab)

- No lab – Wellness Wednesday!

Thursday

- Standardizing
- Continue with regression/multiple linear regression

Example

4. Fit the model

- Algebra: $y_i = \beta_0 + \beta_{1(g)} Sex_{1i(g)} + \beta_{2(g)} Network_{2i(g)} + \beta_{3(g)} Sex_{1i(g)} Network_{2i(g)} + e_i$
- R:

```
> mSexPopI <- lm(Weight ~ Sex*Network, data = vole)
```

Example

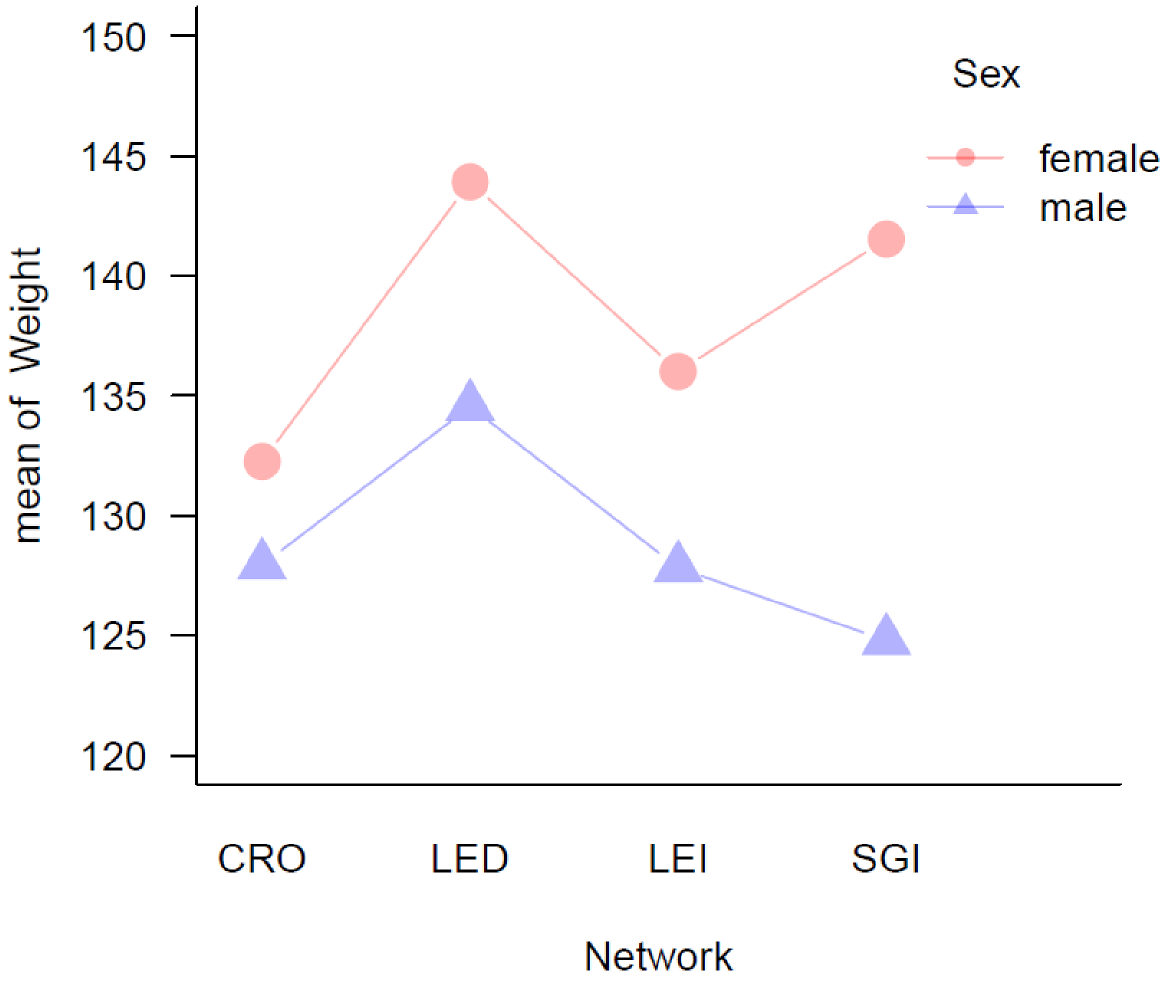
4. Fit the model

```
> mSexPopI <- lm(Weight ~ Sex*Network, data = vole)
> summary(mSexPopI)$coefficients
                     Estimate Std. Error    t value    Pr(>|t|)
(Intercept)        132.257874   2.420647 54.6374109 0.000000000
Sexmale             -4.320937   3.350052 -1.2898119 0.197225833
NetworkLED          11.662989   5.222473  2.2332313 0.025614880
NetworkLEI           3.756412   5.207807  0.7213039 0.470784673
NetworkSGI           9.274929   3.431808  2.7026362 0.006922025
Sexmale:NetworkLED  -5.047295   7.226390 -0.6984532 0.484953662
Sexmale:NetworkLEI  -3.874140   7.021521 -0.5517522 0.581163570
Sexmale:NetworkSGI -12.415873   4.750006 -2.6138646 0.009001901
```

```
> (with(vole, tapply(Weight,list(Sex,Network),mean)))
            CRO      LED      LEI      SGI
female 132.2579 143.9209 136.0143 141.5328
male   127.9369 134.5526 127.8192 124.7960
```

Example

4. Fit the model – does it make sense?



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

Example

5. Evaluate the output

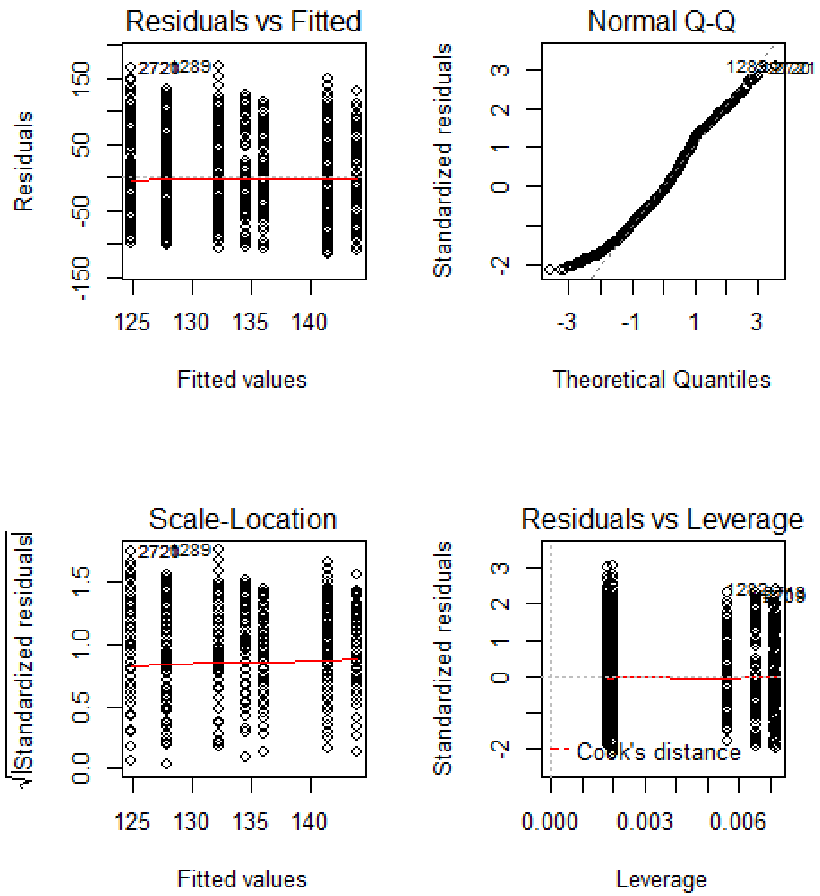
- Model validation - check assumptions!
  - Residuals are normally distributed
  - Constant variance (homogeneity)
  - Observations are independent
  - Predictors measured without error (fixed X)

```
> par(mfrow=c(2,2), oma=c(0,0,0,0))
> plot(mSexPop)
```

What do you see?

Example

```
> par(mfrow=c(2,2), oma=c(0,0,0,0))
> plot(mSexPop)
```



Do we meet our assumptions?

Example

5. Evaluate the output

- Model validation - check fit

```
> summary(modList[[1]])

  Call:
  lm(formula = Weight ~ Sex * Network, data = vole)

  Residuals:
       Min        1Q    Median        3Q       Max
  -116.533  -37.937   -7.258   38.467  167.742

  Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
  (Intercept)         132.258      2.421  54.637  < 2e-16 ***
  Sexmale              -4.321      3.350  -1.290  0.19723
  NetworkLED           11.663      5.222   2.233  0.02561 *
  NetworkLEI            3.756      5.208   0.721  0.47078
  NetworkSGI            9.275      3.432   2.703  0.00692 **
  Sexmale:NetworkLED   -5.047      7.226  -0.698  0.48495
  Sexmale:NetworkLEI   -3.874      7.022  -0.552  0.58116
  Sexmale:NetworkSGI  -12.416      4.750  -2.614  0.00900 **
  ---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  Residual standard error: 54.56 on 2715 degrees of freedom
  Multiple R-squared:  0.01335, Adjusted R-squared:  0.01081
  F-statistic: 5.249 on 7 and 2715 DF,  p-value: 5.76e-06
```

Example

5. Evaluate the output

- *Model validation*
- Model selection – what other models could explain water vole weight?
  - Interaction model: $y_i = \beta_0 + \beta_{1(g)} Sex_{1i(g)} + \beta_{2(g)} Network_{2i(g)} + \beta_{3(g)} Sex_{1i(g)} Network_{2i(g)} + e_i$
  - Additive model: $y_i = \beta_0 + \beta_{1(g)} Network_{1i(g)} + \beta_{2(g)} Sex_{2i(g)} + e_i$
  - Network only: $y_i = \beta_0 + \beta_{1(g)} Network_{1i(g)} + e_i$
  - Sex only: $y_i = \beta_0 + \beta_{2(g)} Sex_{2i(g)} + e_i$
  - Null model! $y_i = \beta_0 + e_i$

Example

5. Evaluate the output

- *Model validation*
- Model selection – what other models could there be to explain water vole weight?
  - Use AIC to compare models

```
> modList <- list()
> modList[["mSexPopI"]] <- lm(Weight ~ Sex * Network, data=vole)
> modList[["mSexPop"]]  <- lm(Weight ~ Sex + Network, data=vole)
> modList[["mSex"]]     <- lm(Weight ~ Sex,           data=vole)
> modList[["mPop"]]     <- lm(Weight ~ Network,       data=vole)
> modList[["m0"]]       <- lm(Weight ~ 1,             data=vole)
```

Example

5. Evaluate the output

- *Model validation*
- Model selection – what other models could there be to explain water vole weight?

```
> (aictab(modList))

  Model selection based on AICc:

           K     AICc Delta_AICc AICcWt Cum.Wt        LL
  mSexPopI 9 29517.65       0.00   0.47   0.47 -14749.79
  mSexPop  6 29518.59       0.94   0.29   0.76 -14753.28
  mSex     3 29518.99       1.34   0.24   1.00 -14756.49
  mPop     5 29539.80      22.15   0.00   1.00 -14764.89
  m0       2 29540.19      22.54   0.00   1.00 -14768.09
```

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

Example

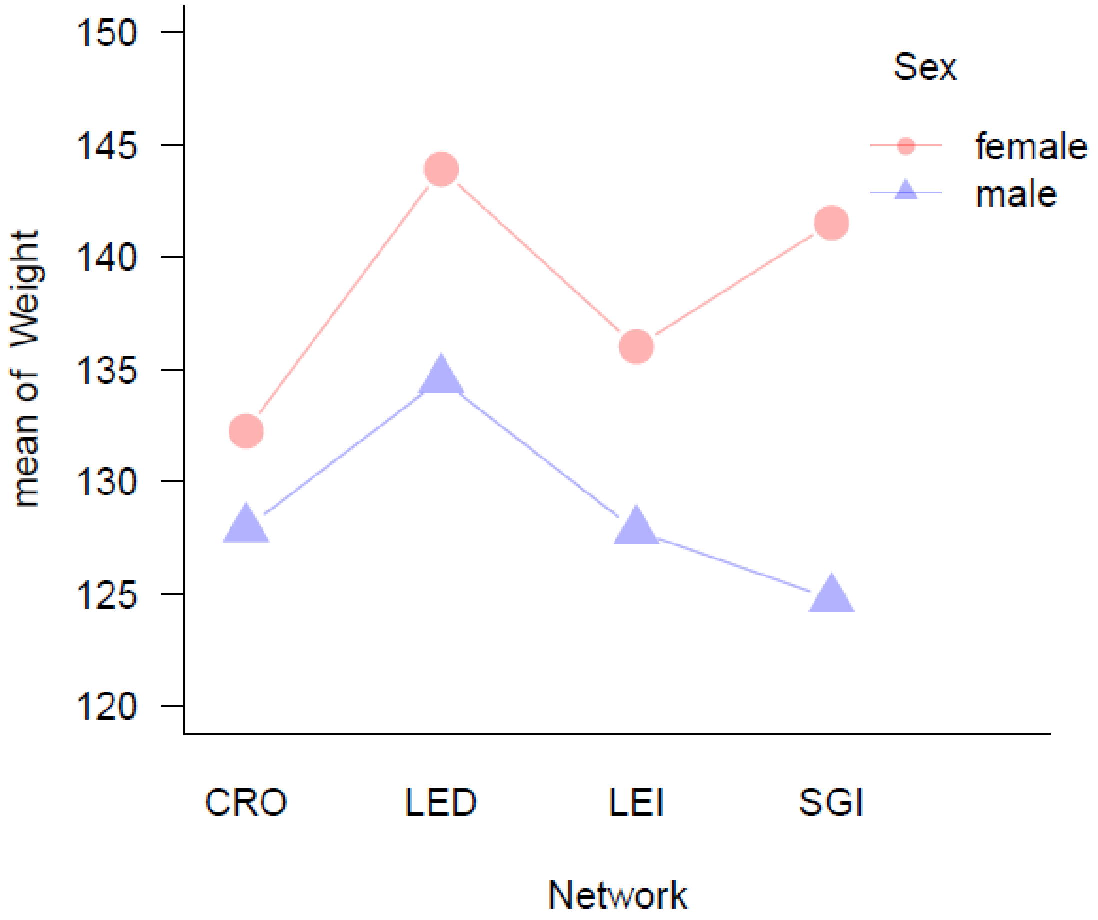
6. Interpret the results

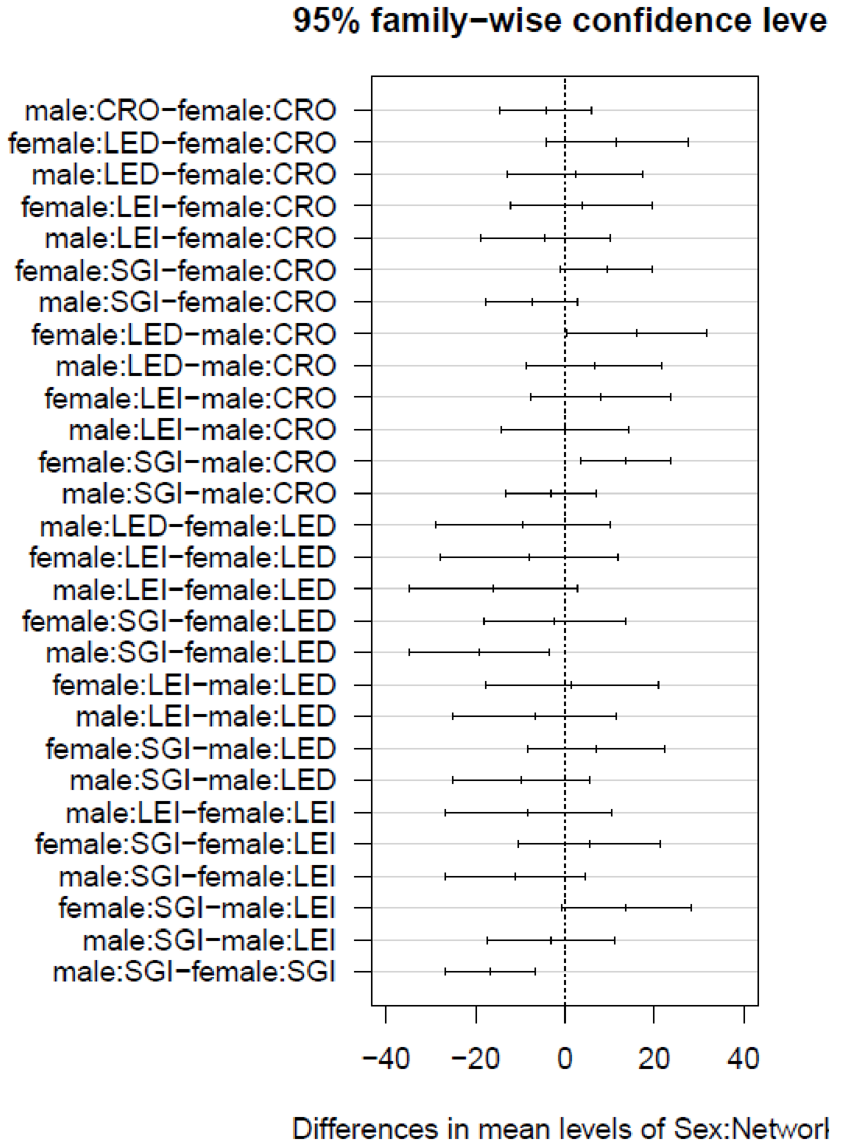
```
> summary(modList[[1]])

  Call:
  lm(formula = Weight ~ Sex * Network, data = vole)

  Residuals:
       Min        1Q    Median        3Q       Max
  -116.533   -37.937    -7.258    38.467   167.742

  Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
  (Intercept)         132.258      2.421  54.637  < 2e-16 ***
  Sexmale              -4.321      3.350  -1.290  0.19723
  NetworkLED           11.663      5.222   2.233  0.02561 *
  NetworkLEI            3.756      5.208   0.721  0.47078
  NetworkSGI            9.275      3.432   2.703  0.00692 **
  Sexmale:NetworkLED   -5.047      7.226  -0.698  0.48495
  Sexmale:NetworkLEI   -3.874      7.022  -0.552  0.58116
  Sexmale:NetworkSGI  -12.416      4.750  -2.614  0.00900 **
  ---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  Residual standard error: 54.56 on 2715 degrees of freedom
  Multiple R-squared:  0.01335, Adjusted R-squared:  0.01081
  F-statistic: 5.249 on 7 and 2715 DF,  p-value: 5.76e-06
```

What about pairwise comparisons?

Example

6. Interpret the results





Review!

So far…

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	`lm(y~1)`
Continuous	Single two-level factor	*t*-test	`lm(y~x)`
Continuous	Single multi-level factor	One-way ANOVA	`lm(y~x)`
Continuous	>1 multi-level factor (*)	Two-way ANOVA	lm($y \sim x_1 * x_2$)

Our explanatory variables were all *categorical* or factors

Inference objective was to estimate <u>sample means</u>

Review!

So far…

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	`lm(y~1)`
Continuous	Single two-level factor	*t*-test	`lm(y~x)`
Continuous	Single multi-level factor	One-way ANOVA	`lm(y~x)`
Continuous	>1 multi-level factor (*)	Two-way ANOVA	`lm(y~x1*x2)`

Null model:

$$y_i = \beta_0 + e_i$$

- β_0 is the sample mean

Review!

So far…

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	`lm(y~1)`
Continuous	Single two-level factor	*t*-test	`lm(y~x)`
Continuous	Single multi-level factor	One-way ANOVA	`lm(y~x)`
Continuous	>1 multi-level factor (*)	Two-way ANOVA	lm(y~x_1*x_2)

T-test linear model:

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

- β_0 is the reference level mean
- β_1 is the contrast

Review!

So far…

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	`lm(y~1)`
Continuous	Single two-level factor	*t*-test	`lm(y~x)`
Continuous	Single multi-level factor	One-way ANOVA	`lm(y~x)`
Continuous	>1 multi-level factor (*)	Two-way ANOVA	lm(y~x_1*x_2)

One-way ANOVA linear model:

$$y_i = \beta_0 + \beta_{1(g)} X_{1i(g)} + e_i$$

- β_0 is the reference level mean
- $\beta_{1(g)}$ is g-1 contrasts

Review!

So far…

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	`lm(y~1)`
Continuous	Single two-level factor	*t*-test	`lm(y~x)`
Continuous	Single multi-level factor	One-way ANOVA	`lm(y~x)`
Continuous	>1 multi-level factor (*)	Two-way ANOVA	$\texttt{lm(y\sim x}_1\texttt{*x}_2\texttt{)}$

Two-way ANOVA linear model (with interaction):

$$y_i = \beta_0 + \beta_{1(g)}X_{1i(g)} + \beta_{2(g)}X_{2i(g)} + \beta_{3(g)}X_{1i(g)}X_{2i(g)} + e_i$$

- β_0 is the reference level mean
- $\beta_{1(g)}$ is group 1 contrasts, $\beta_{2(g)}$ is group 2 contrasts
- $\beta_{3(g)}$ is interactions

Review!

Next!

Response (Y)	Explanatory (X)	Model	In R
Continuous	None	Intercept-only/null	`lm(y~1)`
Continuous	Single two-level factor	*t*-test	`lm(y~x)`
Continuous	Single multi-level factor	One-way ANOVA	`lm(y~x)`
Continuous	>1 multi-level factor (*)	Two-way ANOVA	$\texttt{lm(y\sim x}_1\texttt{*x}_2\texttt{)}$
Continuous	Single continuous	Simple linear regression	`lm(y~x)`

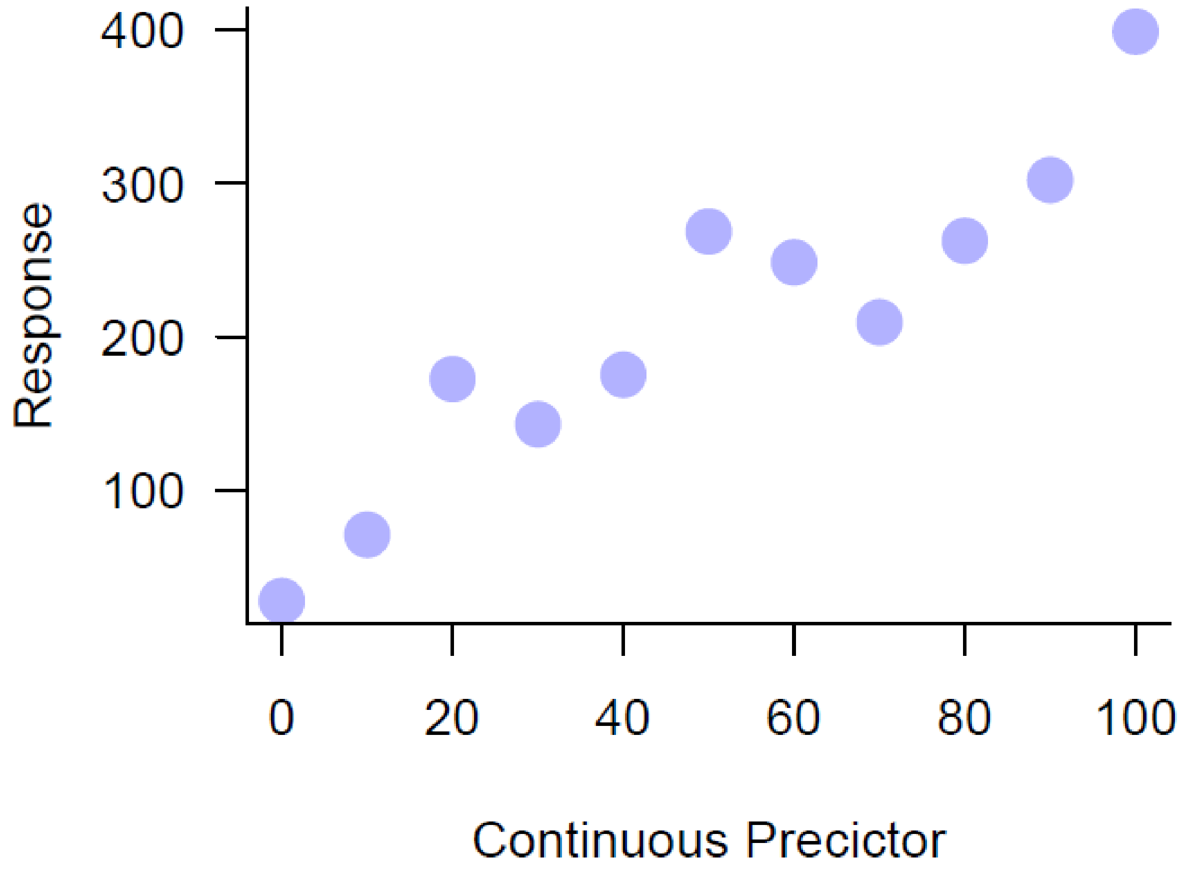
Estimating the relationship between variables!

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

Single continuous predictor

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

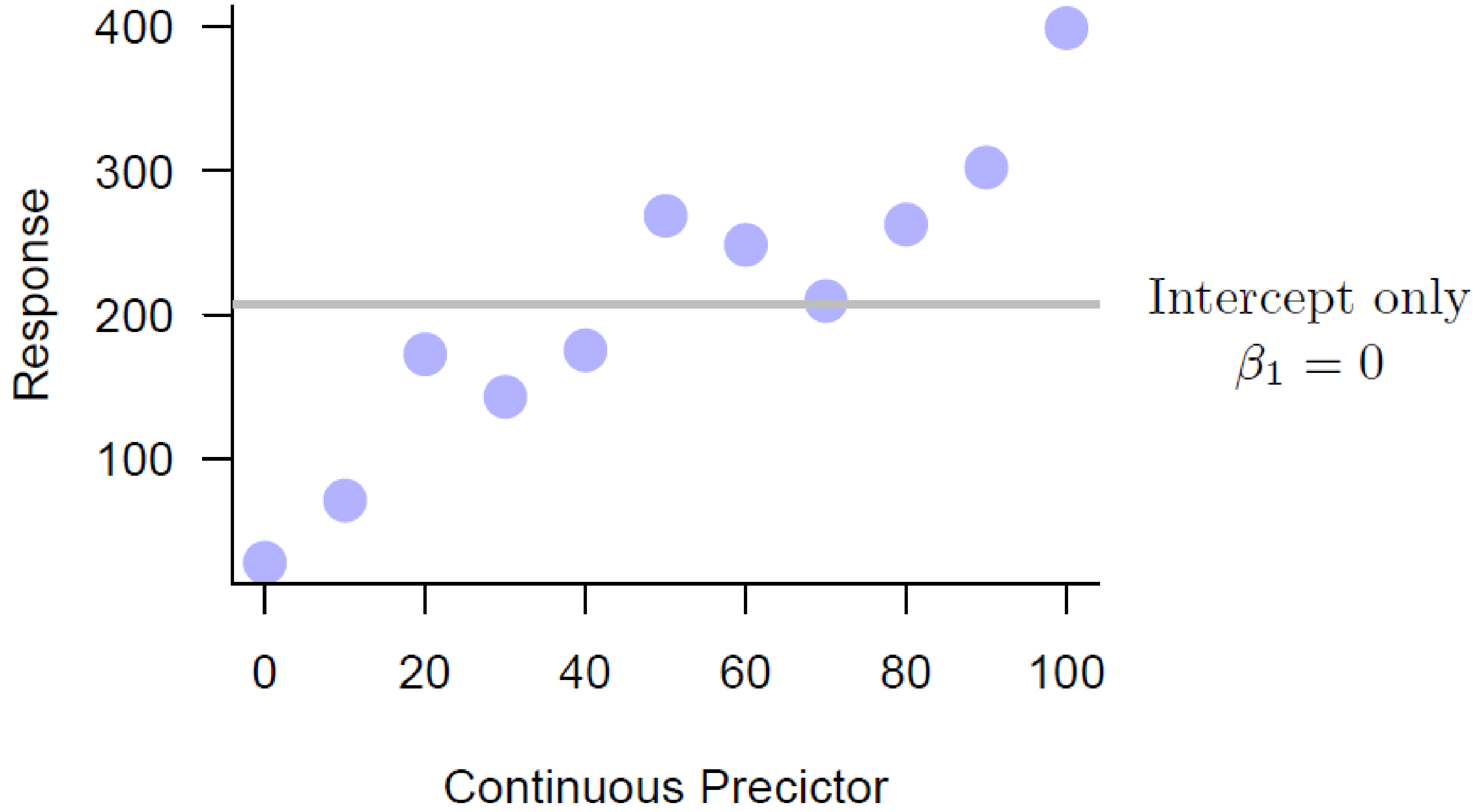
- β_0 is the intercept
- β_1 is the slope



Single continuous predictor

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

- β_0 is the intercept
- β_1 is the slope



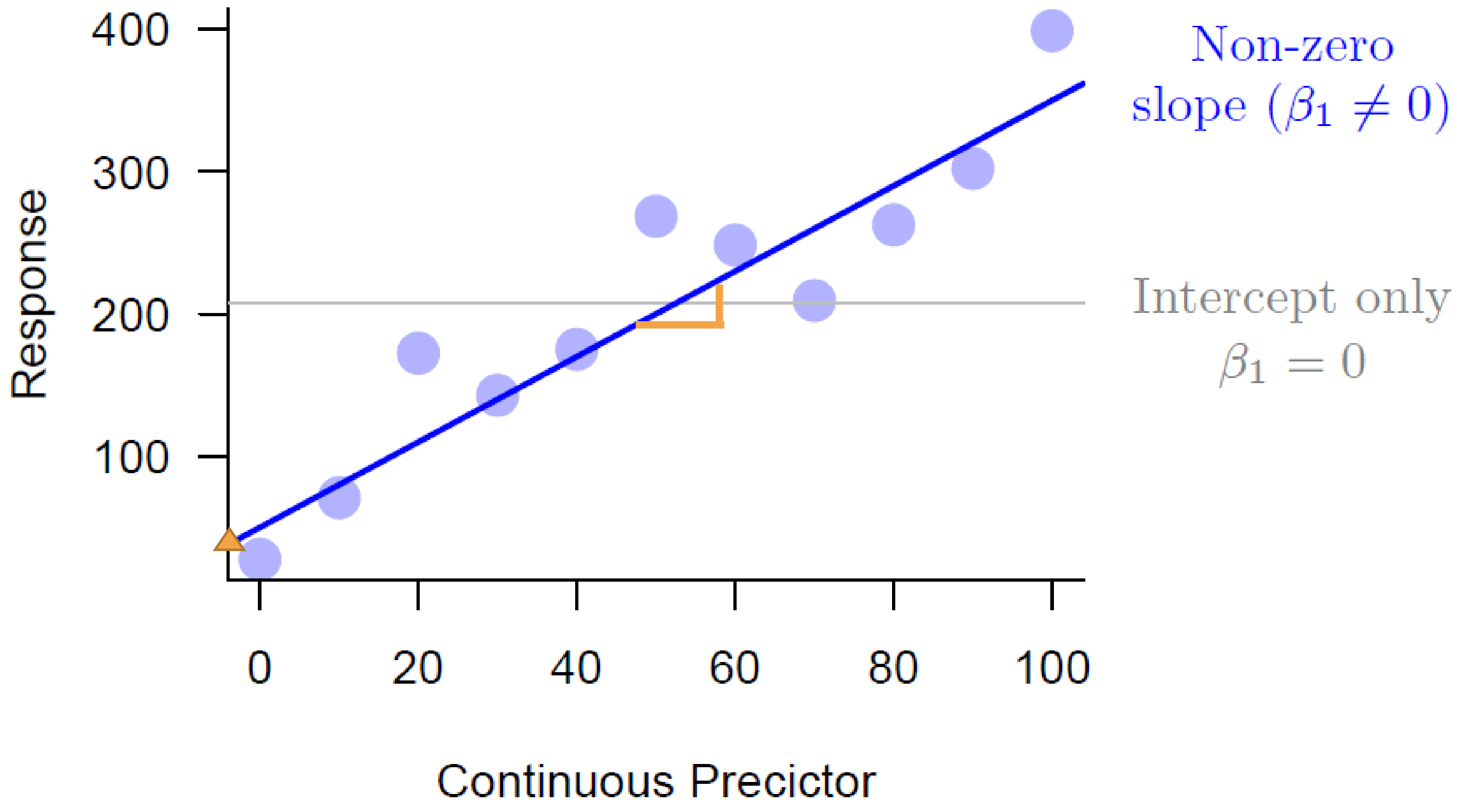
Single continuous predictor

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

- β_0 is the intercept
- β_1 is the slope

What sort of questions or studies could have a continuous predictor?

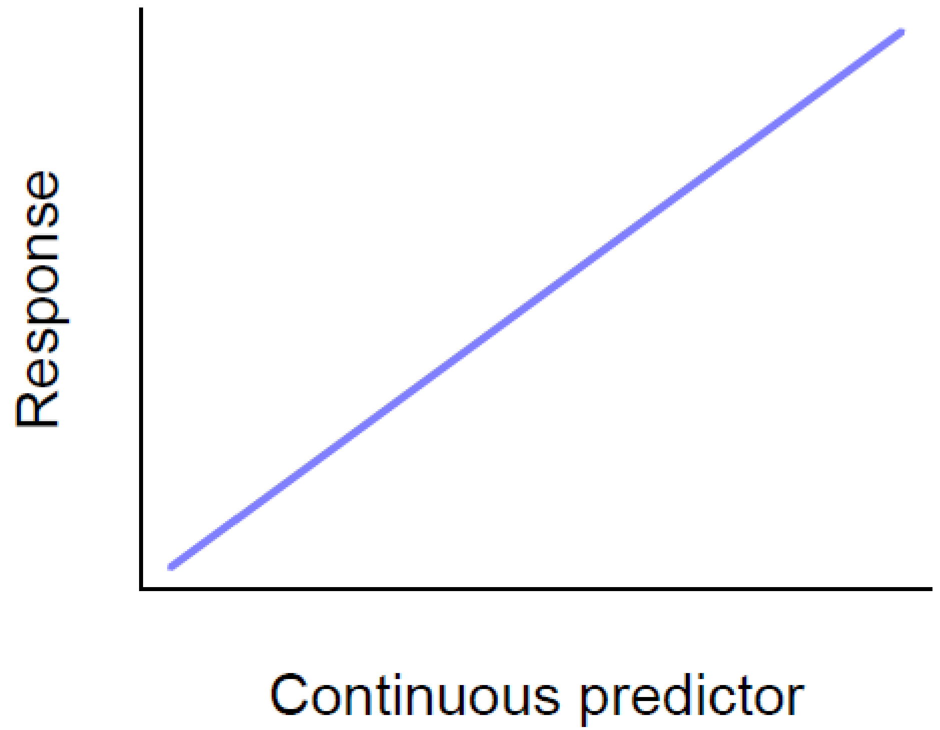
What kind of relationships might we encounter?



Single continuous predictor

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

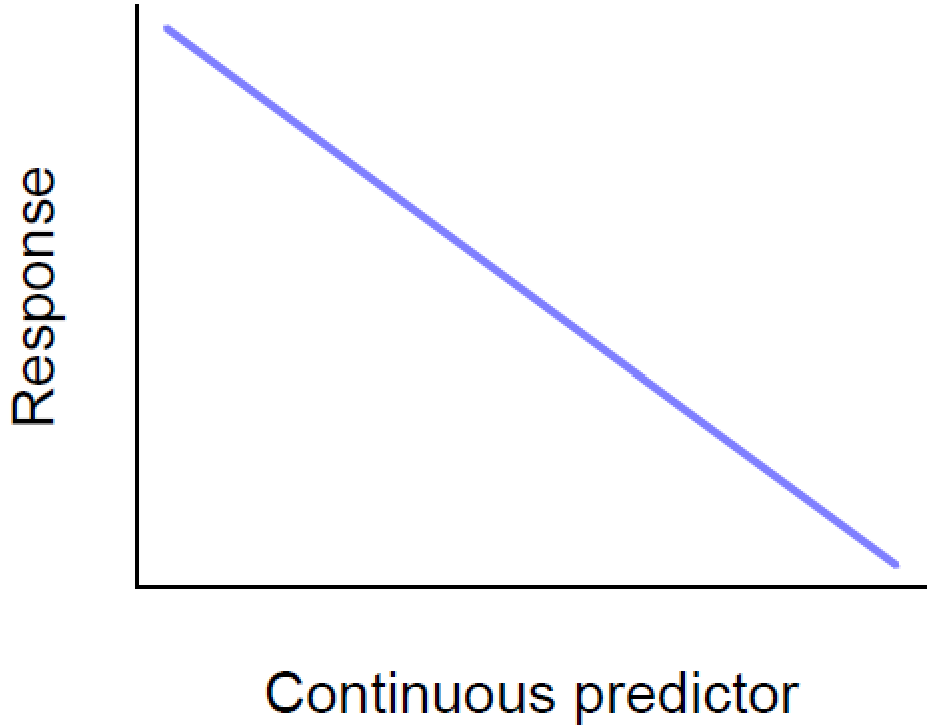
- β_0 is the intercept
- β_1 is the slope – generally 3 possible versions:
  - $\beta_1 > 0$ (positive)



Single continuous predictor

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

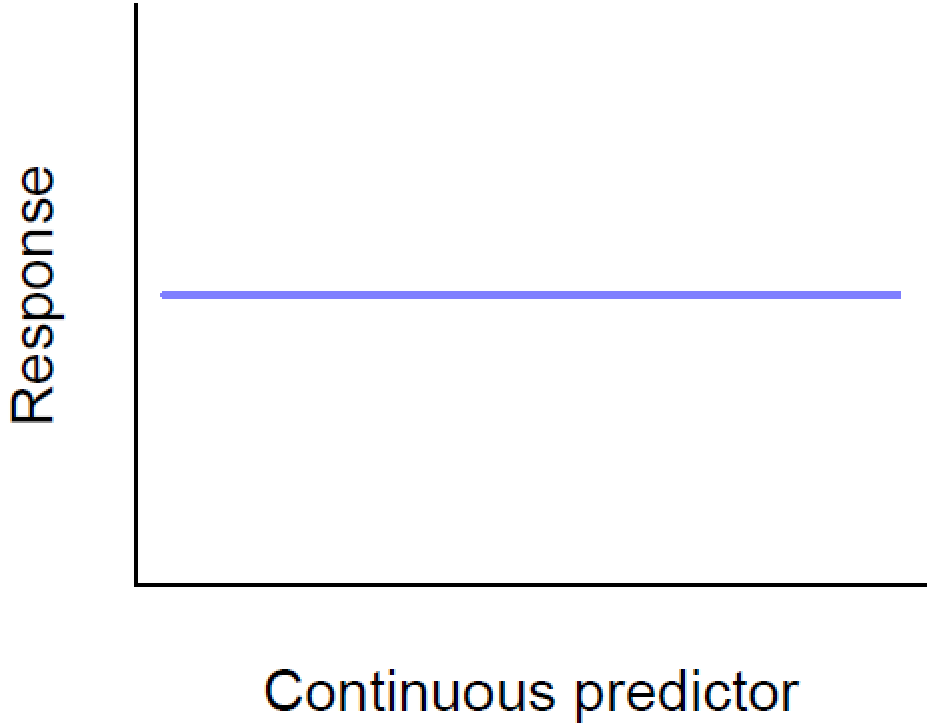
- β_0 is the intercept
- β_1 is the slope – generally 3 possible versions:
  - $\beta_1 > 0$ (positive)
  - $\beta_1 < 0$ (negative)



Single continuous predictor

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

- β_0 is the intercept
- β_1 is the slope – generally 3 possible versions:
  - $\beta_1 > 0$ (positive)
  - $\beta_1 < 0$ (negative)
  - $\beta_1 = 0$



Single continuous predictor

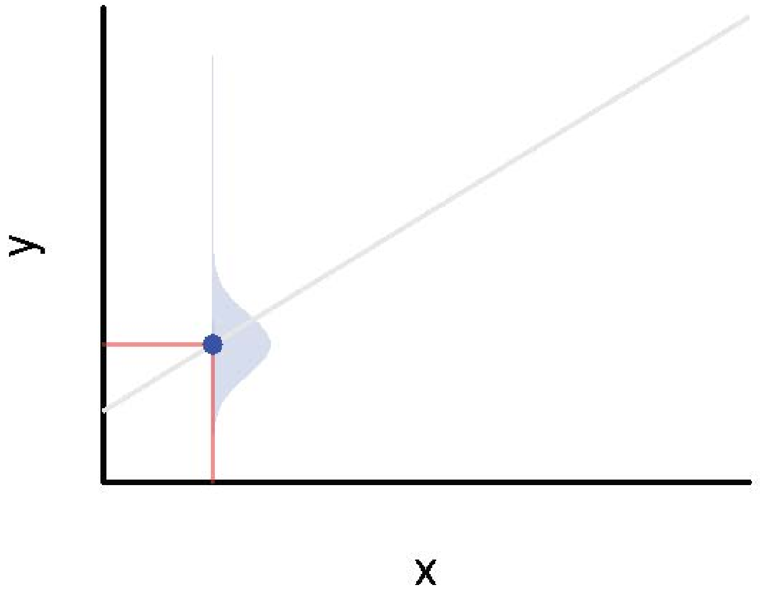
$$y_i = \beta_0 + \beta_1 X_{1i} + e_i$$

- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed
    - Should match standard normal distribution
  - Constant variance in residuals (homogeneity)
    - Random scatter of points, no shape when plotting residuals vs. estimates/fitted points
  - Observations are independent
    - No pseudo-replication, spatial/temporal autocorrelation
  - Predictors are measured without error (fixed X)
    - Avoided through training and experimental design

Single continuous predictor

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

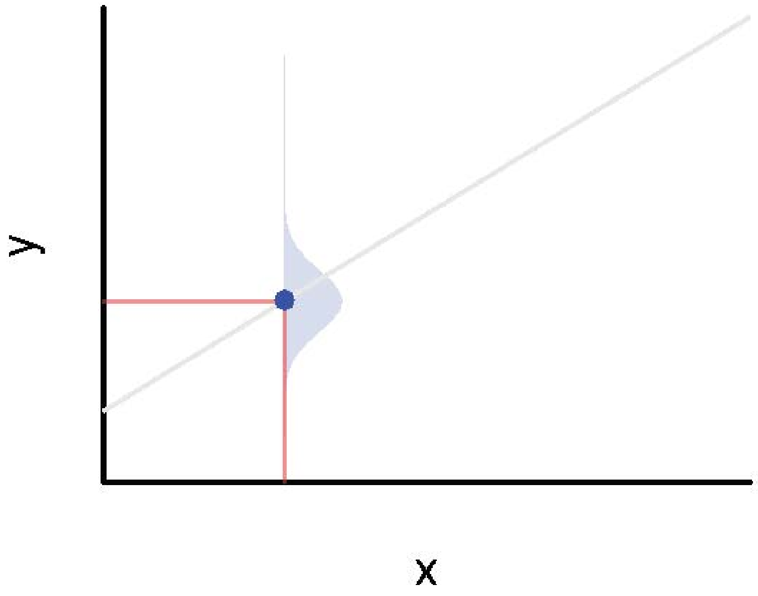
- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed



Single continuous predictor

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

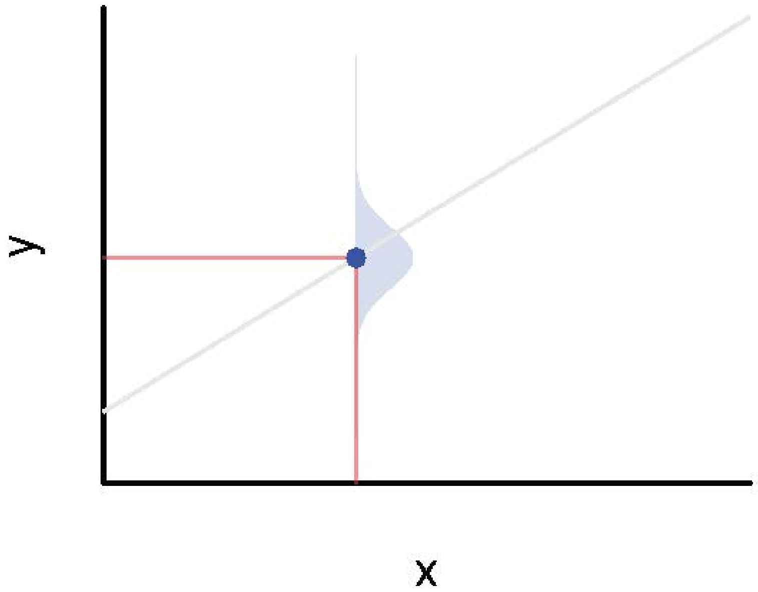
- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed



Single continuous predictor

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

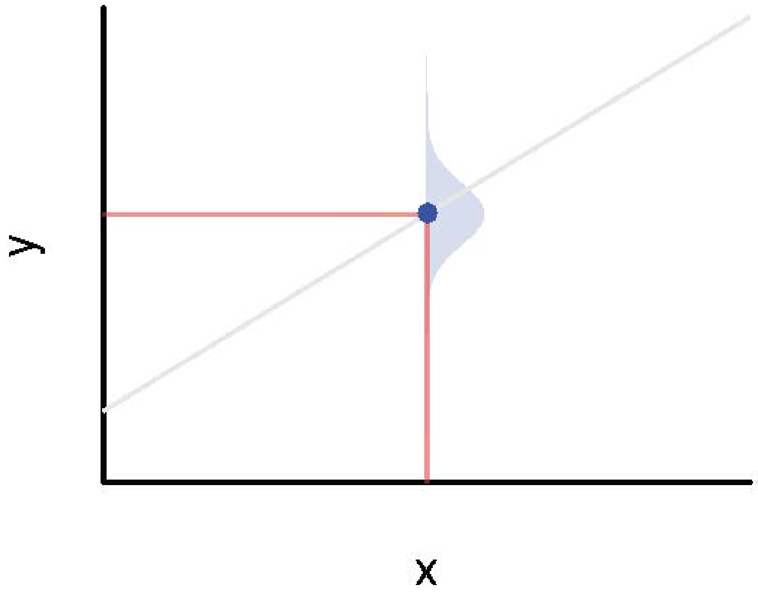
- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed



Single continuous predictor

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

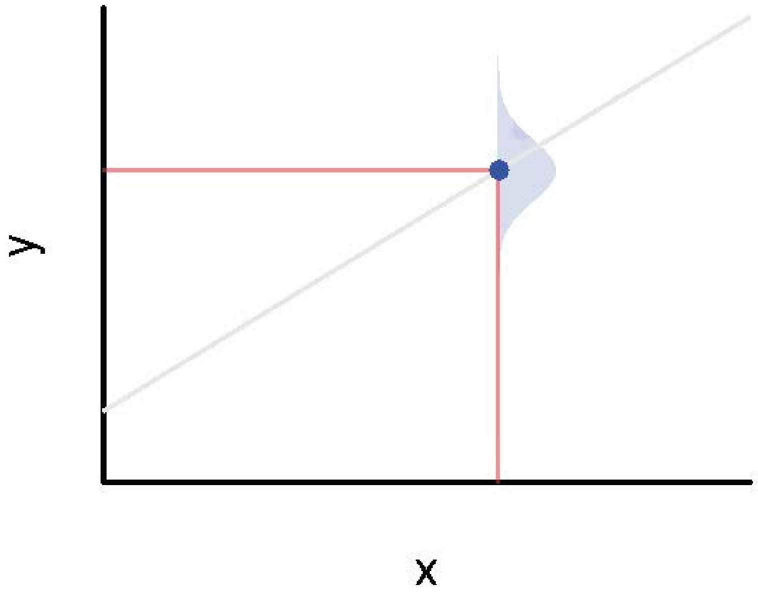
- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed



Single continuous predictor

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

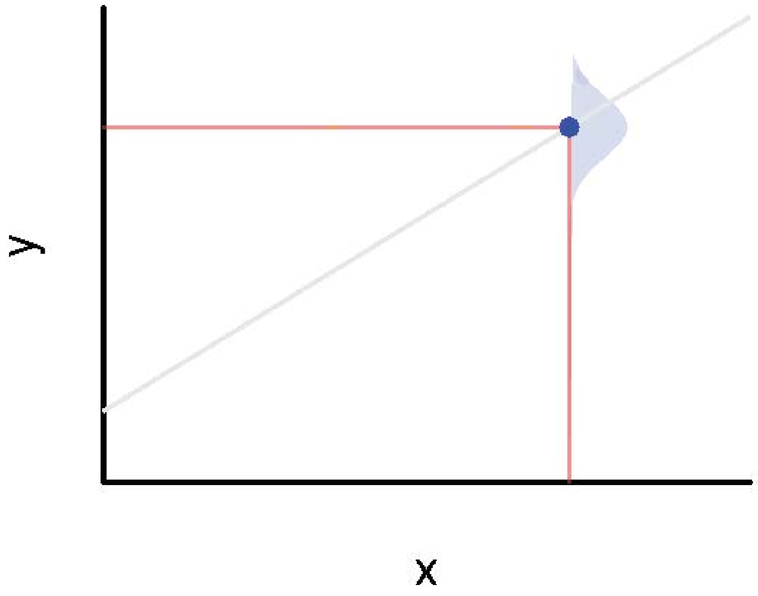
- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed



Single continuous predictor

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

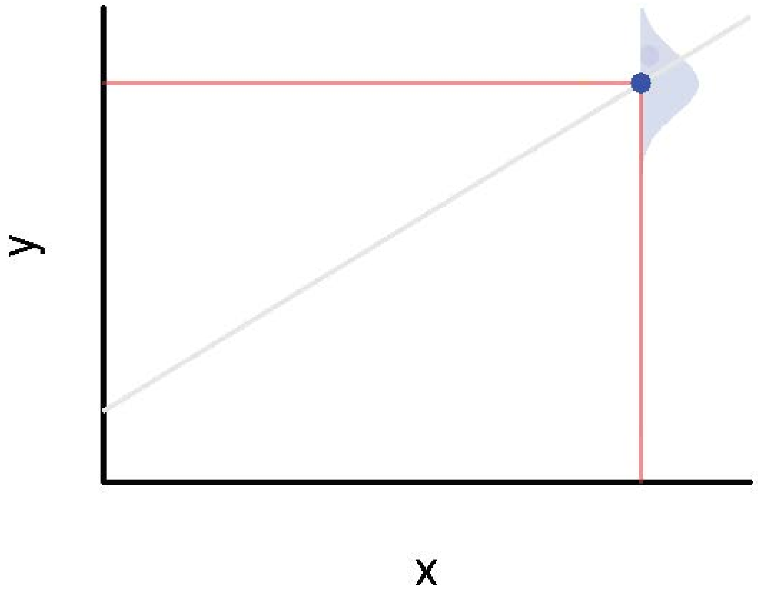
- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed



Single continuous predictor

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

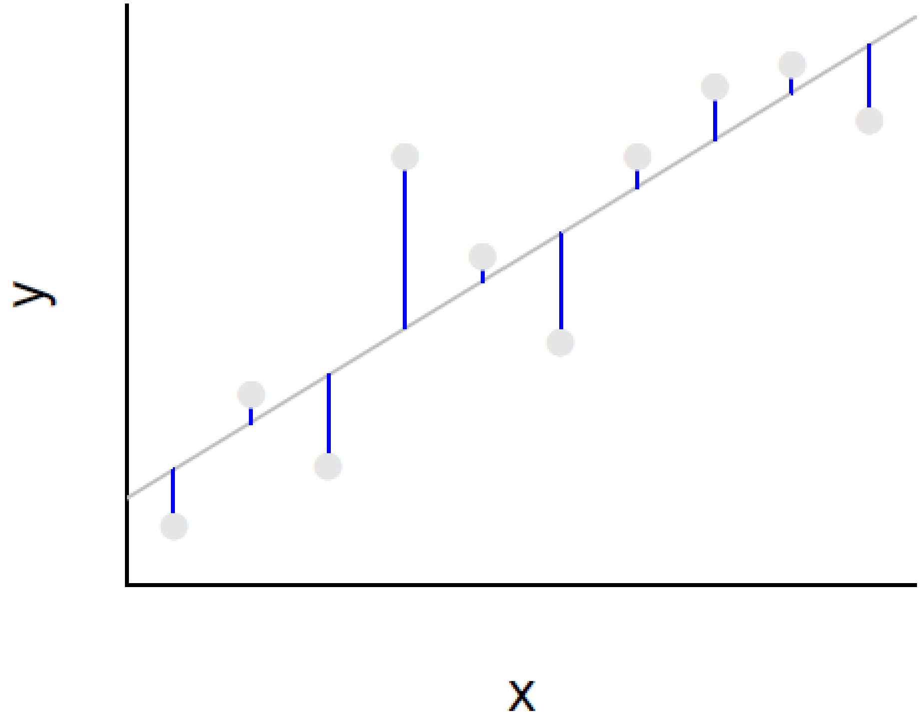
- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed



Single continuous predictor

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

- β_0 is the intercept
- β_1 is the slope
- Assumptions:
  - Residuals are normally distributed
  - Constant variance in residuals (homogeneity)



Simple linear regression example

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

Example

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

```
> head(mander)
  Season Site SVL TL Sex Cap     Ind Tail
1 Spring  P1A  43 86   U   N xxBBP1A   43
2 Spring  P1A  33 66   U   N xYxBP1A   33
3 Spring  P1A  42 84   M   N xYBxP1A   42
4 Spring  P1A  36 76   U   N xYYxP1A   40
5 Spring  P1A  44 76   M   N xxBYP1A   32
6 Spring  P1A  42 74   U   N xBxYP1A   32
```

Example

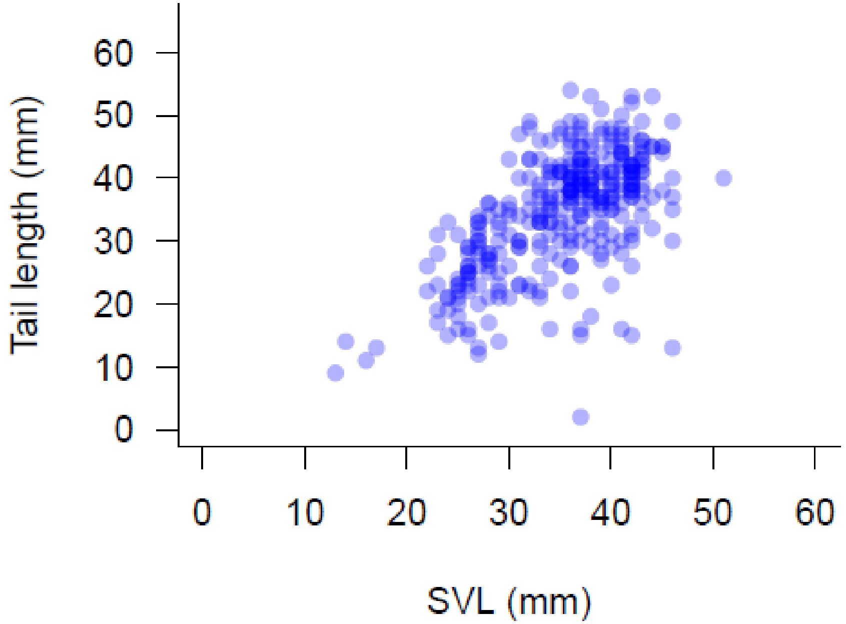
Modeling process:

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

Example

1. State the question:

- Does tail length scale predictably with SVL?
- H_0: there is no relationship between tail length and SVL
  - Response:
    - Tail length
  - Explanatory:
    - SVL



Example

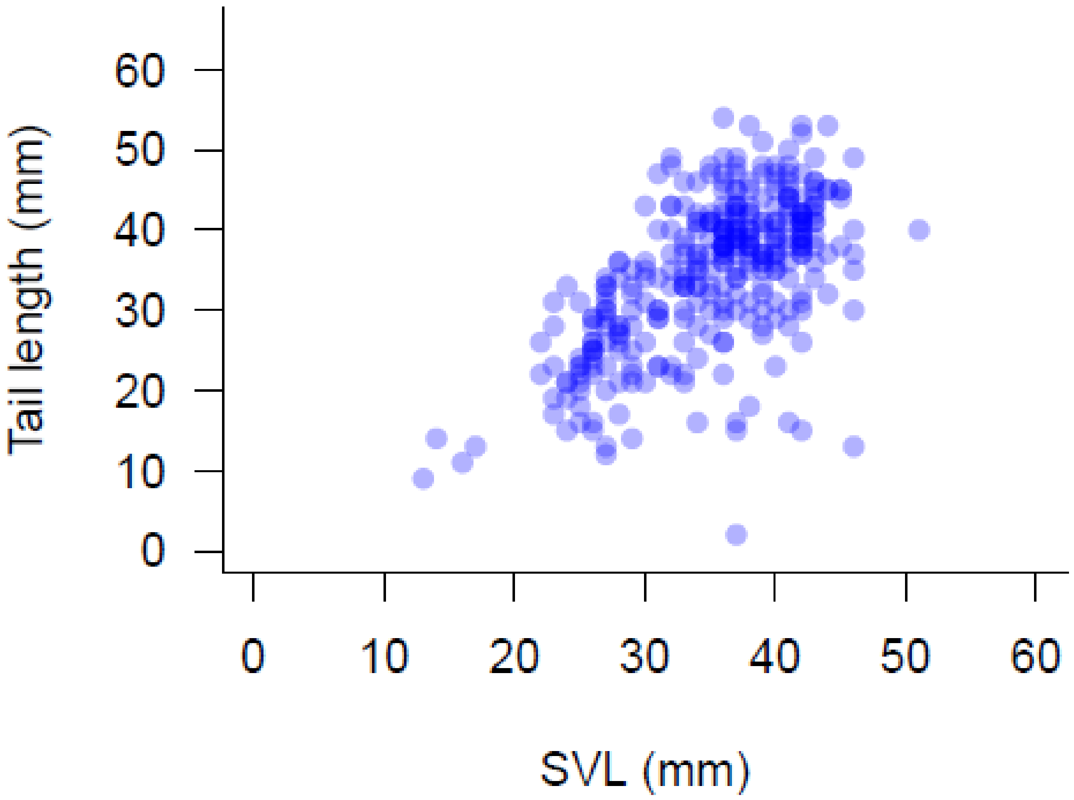
Modeling process:

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

Example

2. Data exploration:

- What kind of parameters (β_0 and β_1) do you expect?



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

Example

3. Describe the model:

- In words:
  - Is there a significant relationship between tail length and SVL?
- In mathematical form:
  - $y_i = \beta_0 + \beta_1 X_{1i} + e_i$
  - y_i is tail length, X_{1i} is SVL
  - H_0 is $\beta_1 = 0$.

Example

3. Describe the model:

- In words:
  - Is there a significant relationship between tail length and SVL?
- In mathematical form:
  - $y_i = \beta_0 + \beta_1 X_{1i} + e_i$
  - y_i is tail length, X_{1i} is SVL
  - H_0 is $\beta_1 = 0$.
- What are the model assumptions?
  - Residuals are normally distributed
  - Constant variance (homogeneity)
  - Observations are independent
  - Predictors measured without error (fixed X)

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

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
```

$$y = 3.15 + 0.89x$$

Example

Modeling process:

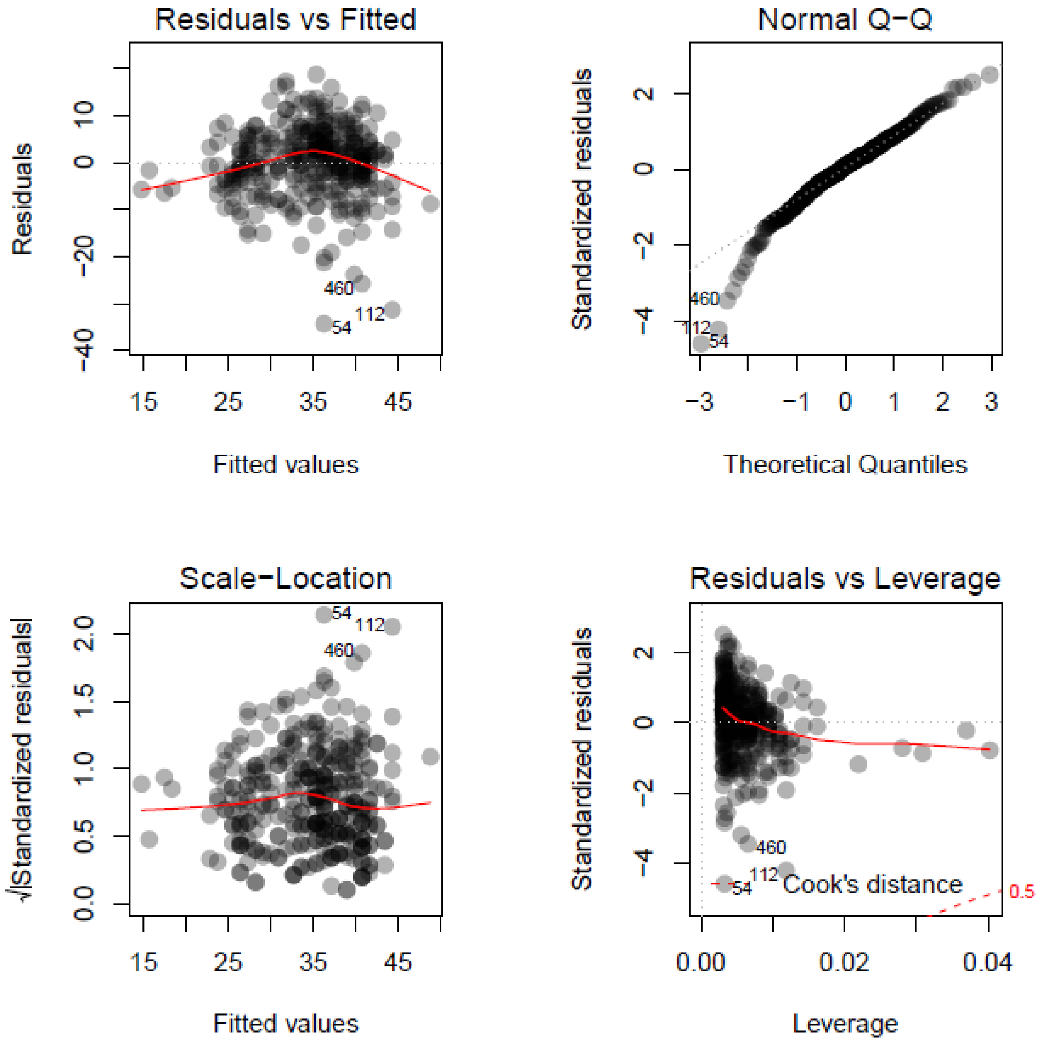
1. *State the question/ hypothesis*
   - *What is the question?*
   - *What are the variables (response and explanatory)?*
2. *Data exploration*
3. *Describe the model*
   - *In word form (should come from your question)*
   - *In mathematical form*
   - *Identify the assumptions of the model*
4. *Fit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

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)

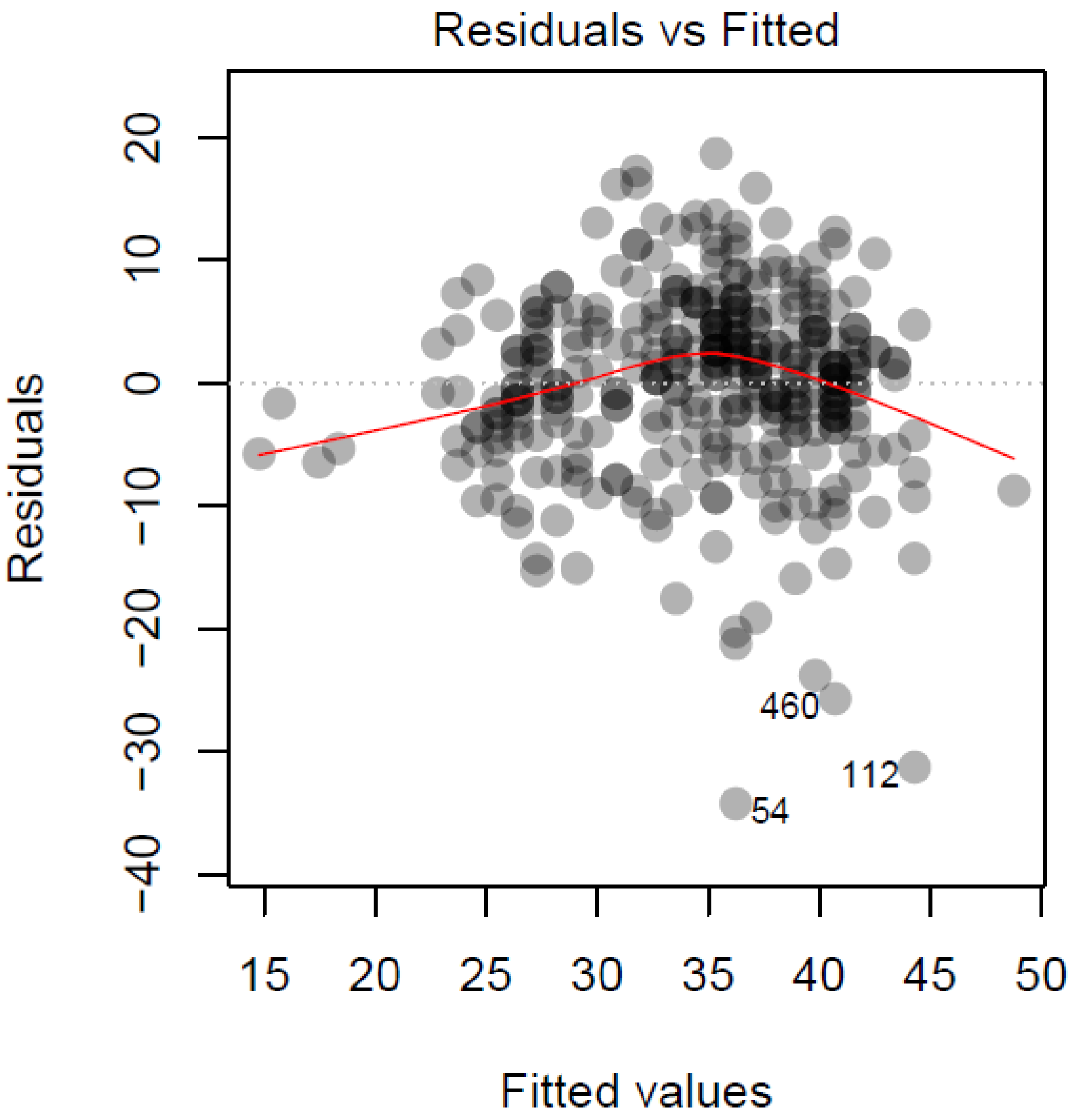
```
> par(mfrow=c(2,2), oma=c(0,0,0,0))
> plot(mAllo)
```



Example

5. Evaluate the output

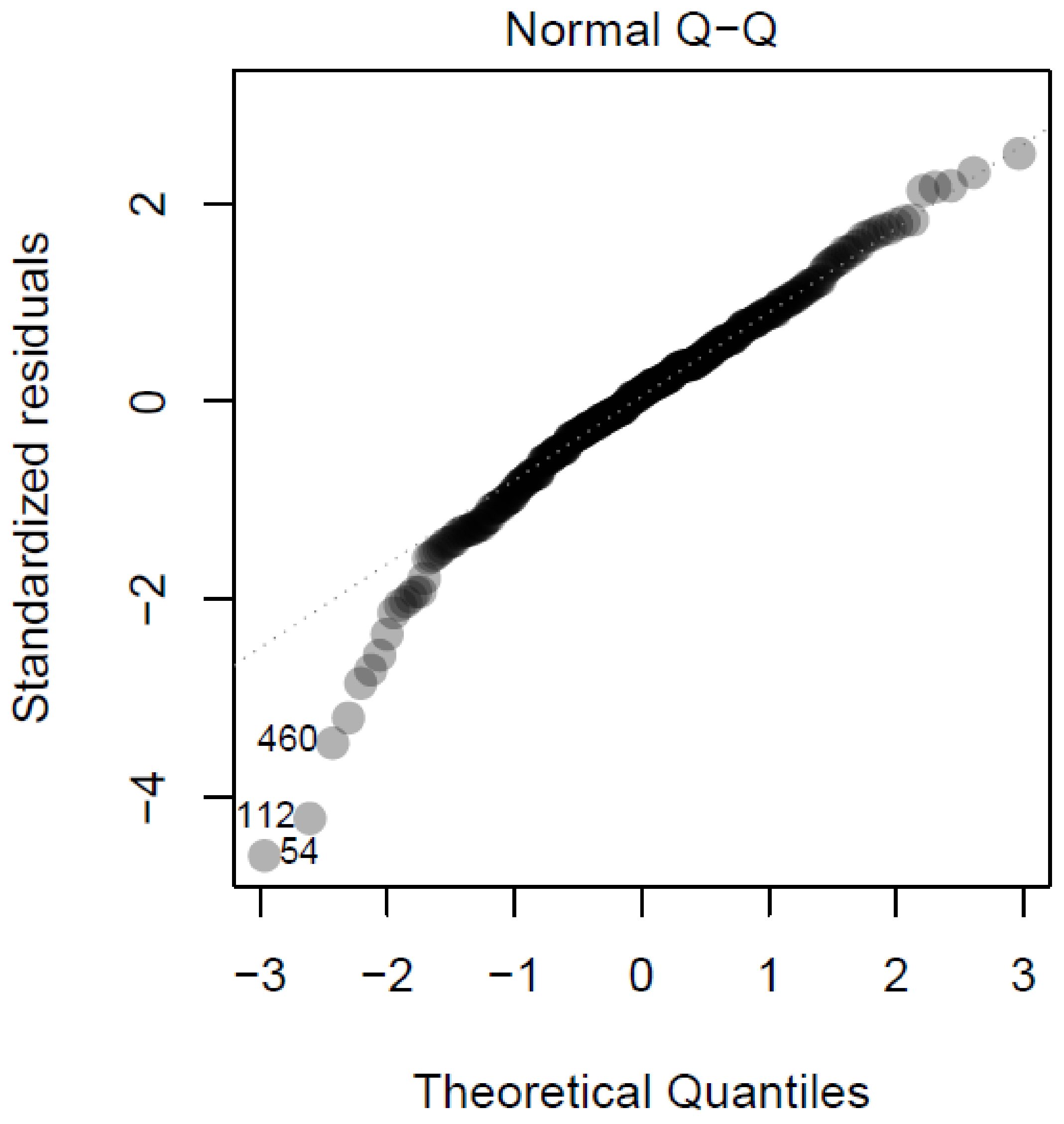
```
> # 1 = Residuals vs. fitted
> plot(mAllo,1)
```



Example

5. Evaluate the output

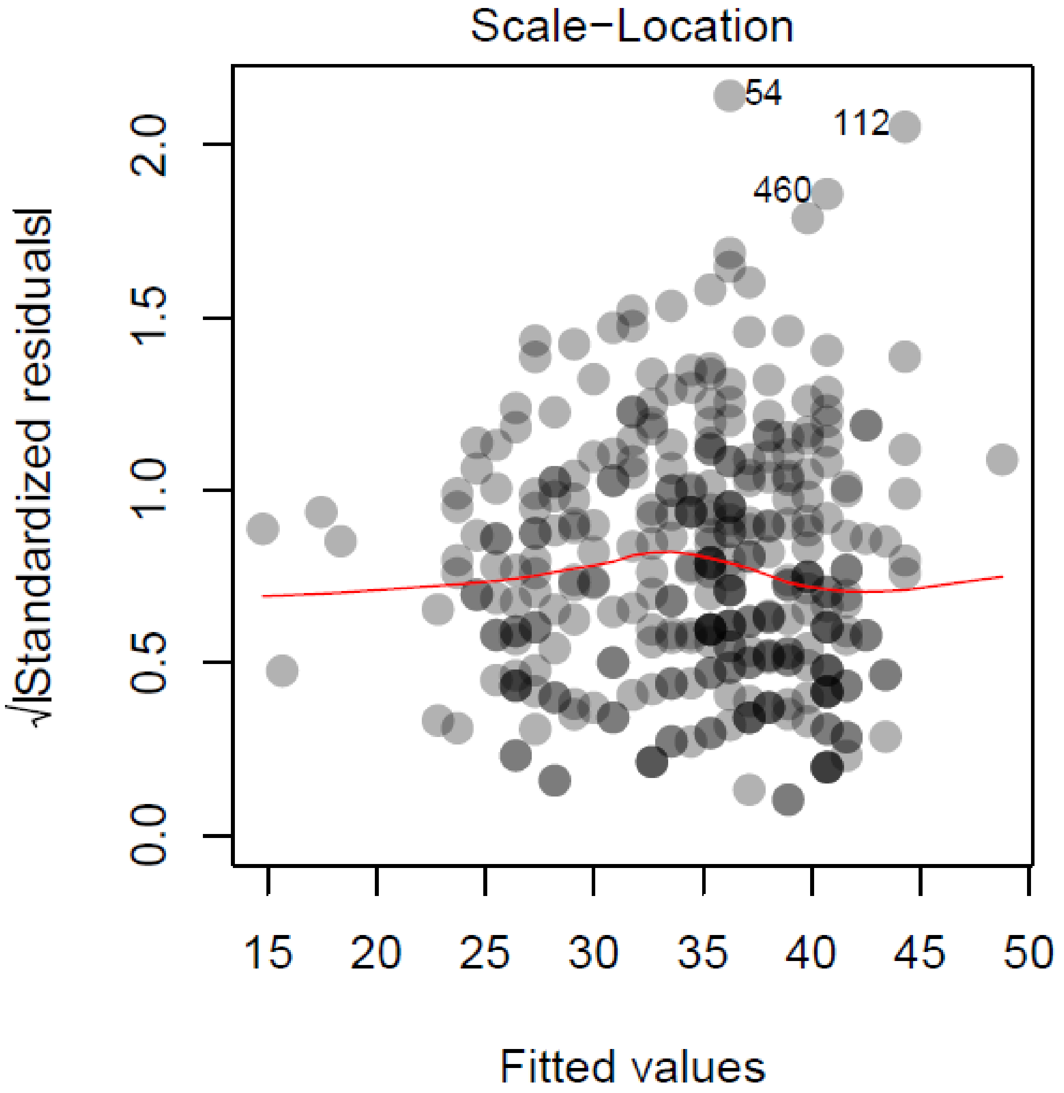
```
> # 2 = QQ-plot
> plot(mAllo,2)
```



Example

5. Evaluate the output

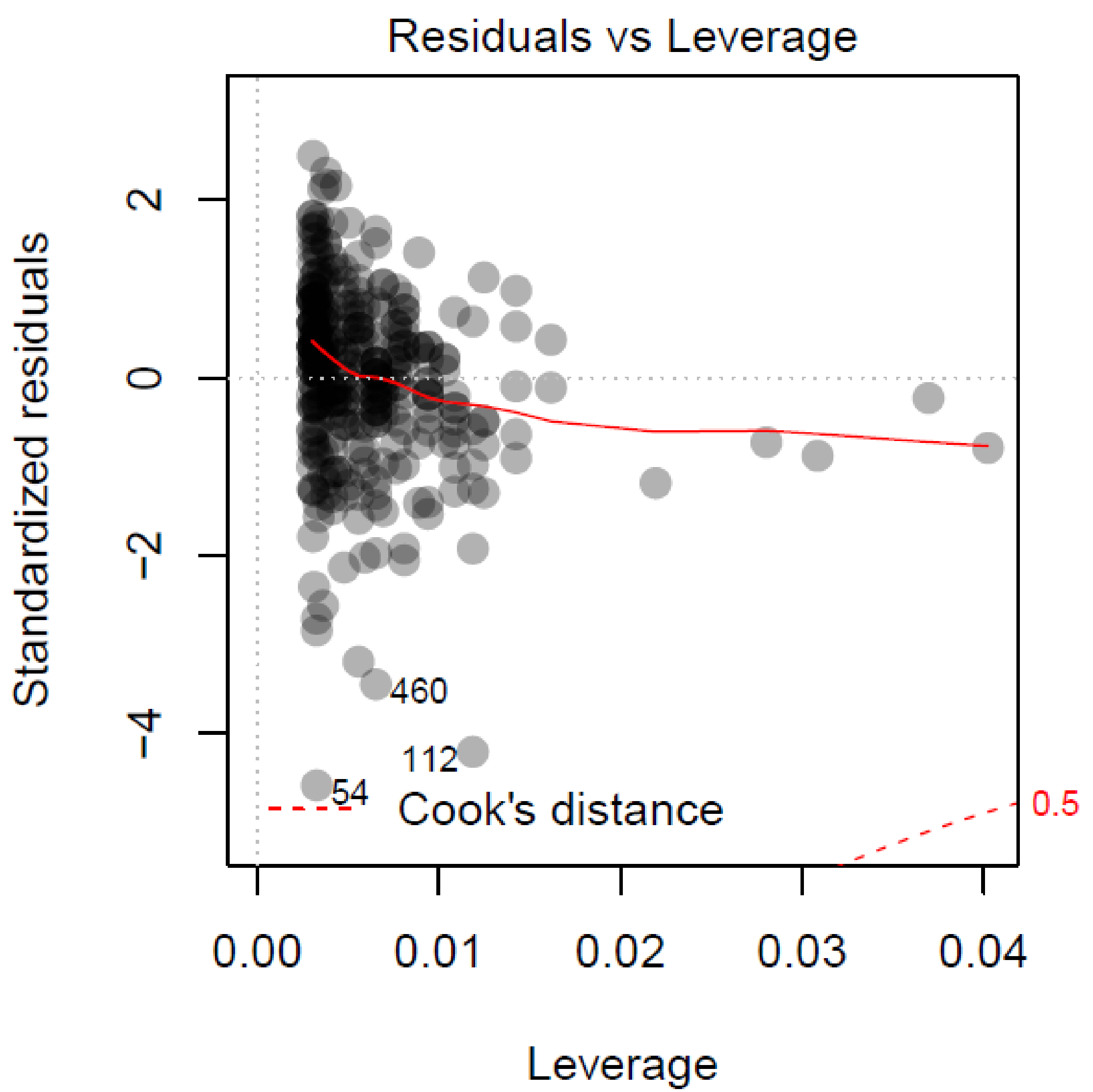
```
> # 3 = Scale-location
> plot(mAllo,3)
```



Example

5. Evaluate the output

```
> # 5 = Leverage
> plot(mAllo,5)
```



Example

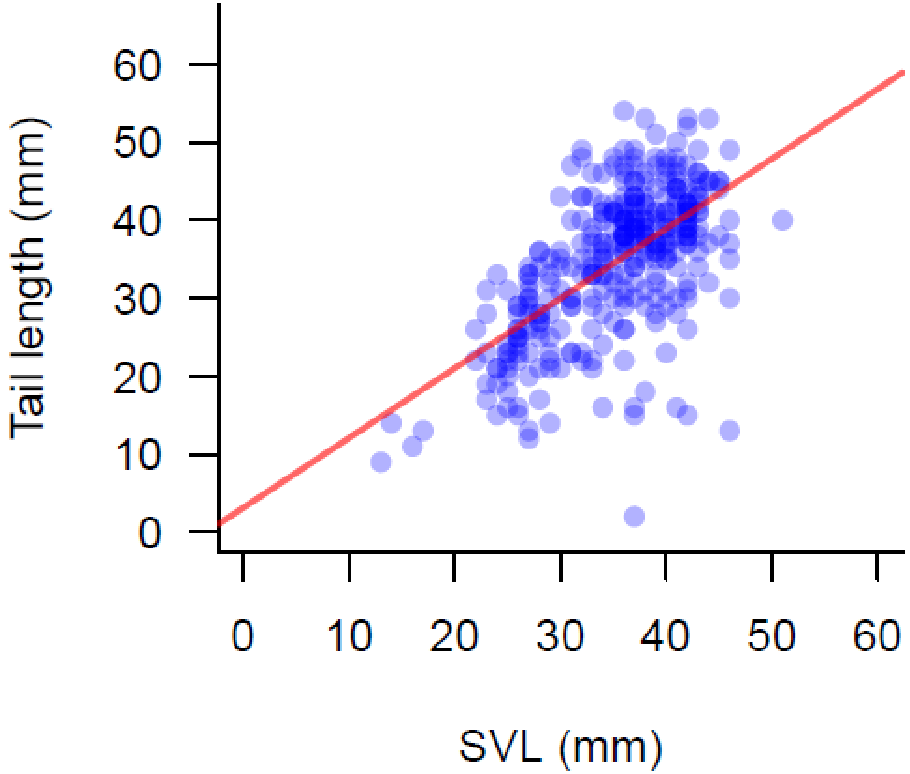
5. Evaluate the output

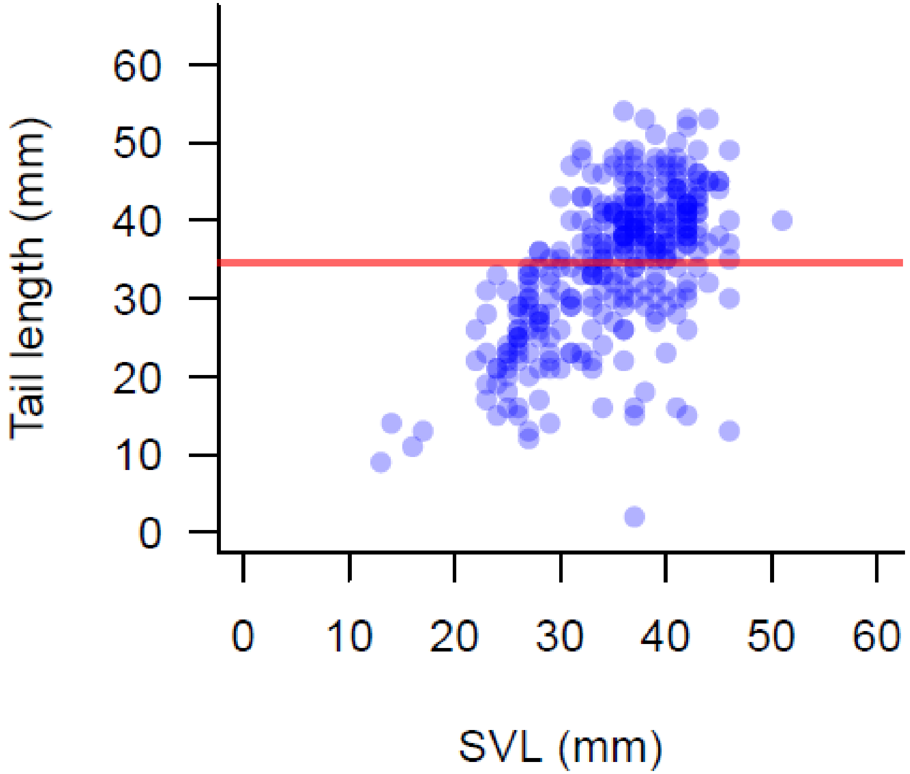
```
> summary(mAllo)

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

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

  Coefficients:
              Estimate Std. Error t value Pr(>|t|)
  (Intercept)  3.14909    2.32421   1.355    0.176
  SVL          0.89427    0.06502  13.754   <2e-16 ***
  ---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  Residual standard error: 7.477 on 329 degrees of freedom
  Multiple R-squared:  0.3651,  Adjusted R-squared:  0.3632
  F-statistic: 189.2 on 1 and 329 DF,  p-value: < 2.2e-16
```

Example

5. Evaluate the output

- Model selection
  - Two models: null and linear





Example

5. Evaluate the output

- Model selection
  - Two models: null and linear

```
> m0    <- lm(Tail ~ 1,   data = mander) # null
> mAllo <- lm(Tail ~ SVL, data = mander) # lin reg
```

```
> aictab(list(m0,mAllo),c("m0","mAllo"))

  Model selection based on AICc:

        K    AICc Delta_AICc AICcWt Cum.Wt       LL
  mAllo 3 2275.21       0.00      1      1 -1134.57
  m0    2 2423.53     148.32      0      1 -1209.75
```

Example

5. Evaluate the output

```
> 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:
              Estimate Std. Error t value Pr(>|t|)
  (Intercept)  3.14909    2.32421   1.355    0.176
  SVL          0.89427    0.06502  13.754   <2e-16 ***
  ---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  Residual standard error: 7.477 on 329 degrees of freedom
  Multiple R-squared:  0.3651,  Adjusted R-squared:  0.3632
  F-statistic: 189.2 on 1 and 329 DF,  p-value: < 2.2e-16
```

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

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:
              Estimate Std. Error t value Pr(>|t|)
  (Intercept)  3.14909    2.32421   1.355    0.176
  SVL          0.89427    0.06502  13.754   <2e-16 ***
  ---
  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  Residual standard error: 7.477 on 329 degrees of freedom
  Multiple R-squared:  0.3651,	Adjusted R-squared:  0.3632
  F-statistic: 189.2 on 1 and 329 DF,  p-value: < 2.2e-16
```

We want to show the estimated relationship… as well as the uncertainty in that relationship!

Example

6. Interpret the results – use predict to show relationship and CI

```
> #1. Fit the model
> mAllo <- lm(Tail ~ SVL, data=mander)
>
> #2. Create new X values for prediction
> new.df <- data.frame(SVL = newSVL)
>
> #3. Predict
> pred <- predict(mAllo, newdata=new.df, interval="confidence")
> head(pred, 10)
        fit      lwr      upr
1  14.77458 11.82424 17.72493
2  15.46810 12.61302 18.32318
3  16.16161 13.40152 18.92171
4  16.85513 14.18971 19.52054
5  17.54864 14.97757 20.11971
6  18.24215 15.76505 20.71926
7  18.93567 16.55210 21.31924
8  19.62918 17.33868 21.91969
9  20.32270 18.12472 22.52068
10 21.01621 18.91015 23.12227
```

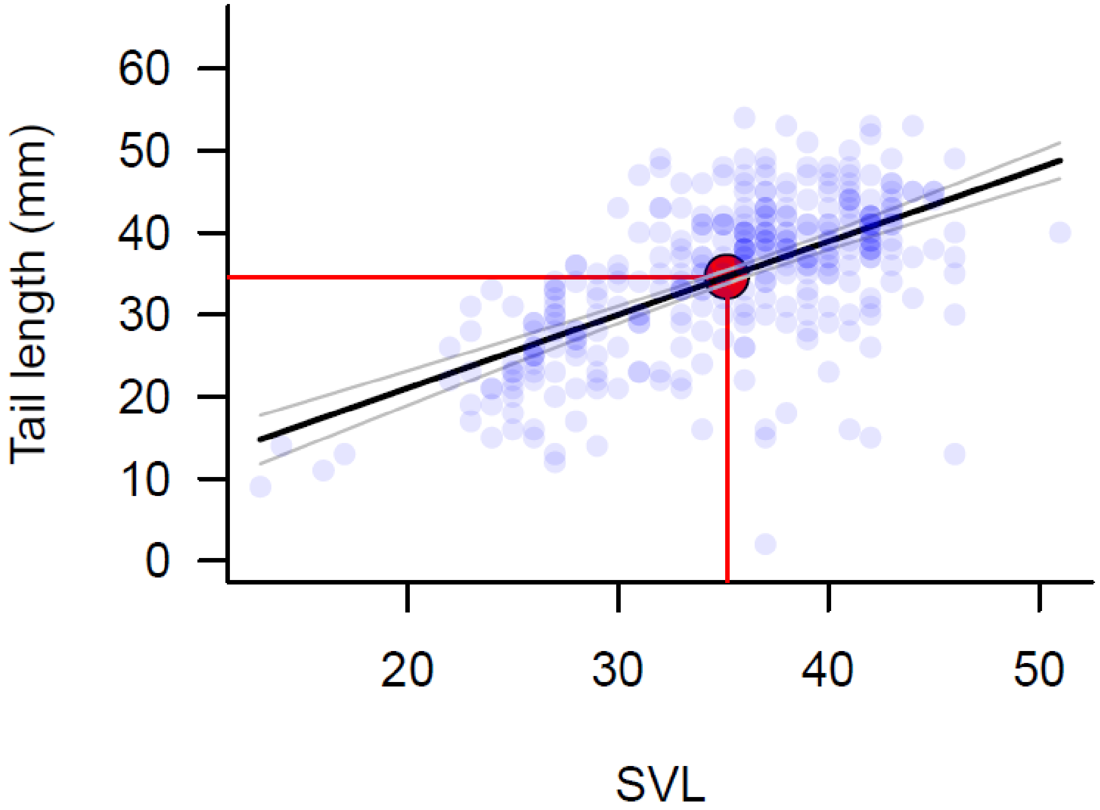
Example

6. Interpret the results – use predict to show relationship and CI

```
> #1. Fit the model
> mAllo <- lm(Tail ~ SVL, data=mander)
>
> #2. Create new X values for prediction
> new.df <- data.frame(SVL = newSVL)
>
> #3. Predict
> pred <- predict(mAllo, newdata=new.df, interval="confidence")
>
> #4. Plot
> plot(mander$Tail ~ mander$SVL)
> lines(pred[,1] ~ new.df$SVL, lwd=2)
> lines(pred[,2] ~ new.df$SVL, lty=2, col="red")
> lines(pred[,3] ~ new.df$SVL, lty=2, col="red")
```

Example

6. Interpret the results – use predict to show relationship and CI



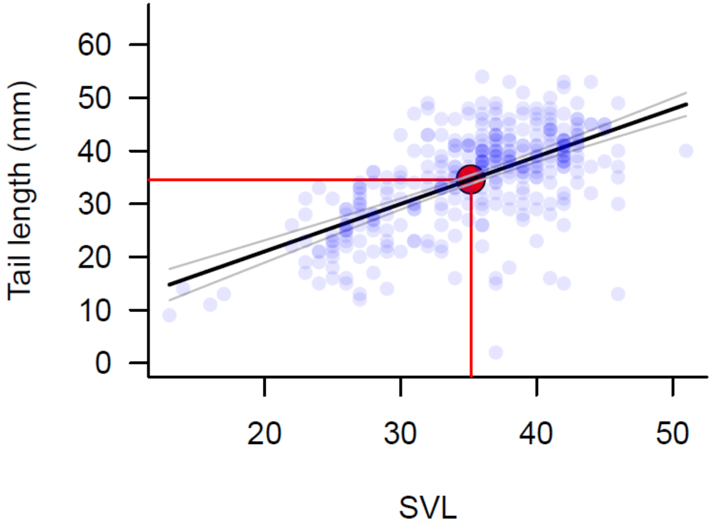
How are the confidence intervals calculated? Why aren't they the same all the way along the model line?

Example

6. Interpret the results – use predict to show relationship and CI

$$CI = \hat{y} \pm t \frac{S_E}{S_X\sqrt{n-2}}$$

- t is the t-value for a specific confidence level (usually 95%)
- S_E is the standard deviation of the residual errors
- S_X is the standard deviation of the X data
- n is the sample size



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.