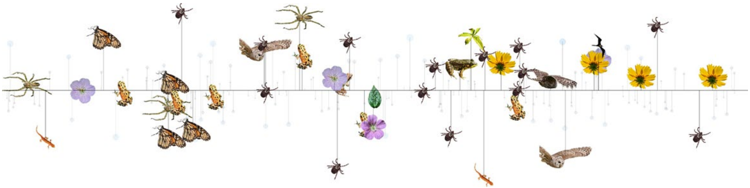
Intro to Quantitative Ecology

Deck 10A   Regression 3

Michael France Nelson

Spring 2023

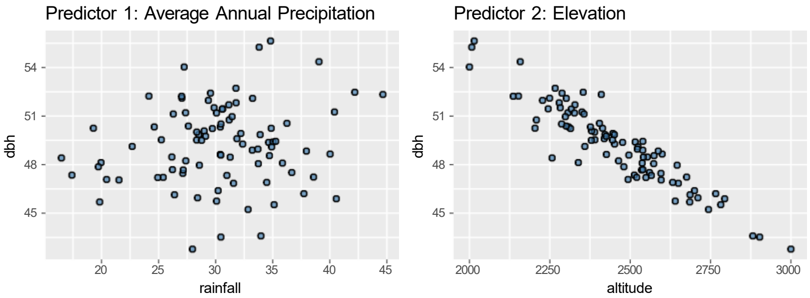


This deck's concepts

- Additive and interactive models
- Model coefficient tables
- Model coefficients and ANOVA
- Model assumptions and diagnostics

Example System: Whitebark Pine

To illustrate additive and interactive models, we'll return to the Whitebark Pine Data.



Example System: Whitebark Pine

The scenario: you think that elevation and average annual precipitation may be related to tree stem diameter (dbh).

The variables:

- Response: dbh (diameter at breast height)
- Predictor 1: Elevation
- Predictor 2: Average Annual Rainfall

Additive Models

Recall the additive multiple regression concepts:

The regression does the following:

- Simultaneously quantifies the magnitude and significance of predictor 1 and 2.
- Quantifies the relationship between predictor 1 and response, independent of the effects of predictor 2.
- Quantifies the relationship between predictor 2 and response, independent of the effects of predictor 1.

Additive Models

An additive model assumes that the effects of predictors 1 and 2 on the response are *completely independent*.

In the context of our Whitebark pine data

- There is no joint effect of changing the elevation and water *simultaneously*.

Additive Models: Coefficient Interpretation

Each coefficient is the expected change *when all other predictors are held constant*.

Additive effect coefficient 1:

- "How many units does the response change when I increase the predictor 1 by 1 unit, if I hold the value of predictor 2 constant."

Additive effect coefficient 2:

- "How many units does the response change when I increase the predictor 2 by 1 unit, if I hold the value of predictor 1 constant."

Interactive Effects: Joint Effect

Recall the additive multiple regression concepts:

The interactive regression adds one procedure

- Simultaneously quantifies the magnitude and significance of predictor 1 and 2.
- Quantifies the relationship between predictor 1 and response, independent of the effects of predictor 2.
- Quantifies the relationship between predictor 2 and response, independent of the effects of predictor 1.
- Quantifies the **joint** effect of predictor 1 and predictor 2 on the response.

Interactive Effects: Interpretation

An interaction is an extra change, above and beyond what you would expect if you increase both predictors by 1 unit.

Interactive effect coefficient:

- "How many more (or less) units does the response change, when I increase both predictors by 1 unit."

Interactive Effects: Sign of the Interaction

If the interaction coefficient is **positive**:

- Increasing both predictors by 1 unit increases the response by **more than** the sum of the two additive coefficients.

If the interaction coefficient is **zero**:

- Increasing both predictors by 1 unit increases the response by **exactly** the sum of the two additive coefficients.

If the interaction coefficient is **negative**:

- Increasing both predictors by 1 unit increases the response by **less than** the sum of the two additive coefficients.

Interactive Effects: Joint Effect

If an interaction is present

- The two predictors have a **joint** effect on the response.
- The joint effect can't be predicted from the individual effects alone.
- An interaction may mean that the two predictors somehow "work together" to influence the response.
- The predictors have a **synergistic** or **inhibiting** effect on the response.

Interactions: Whitebark Pine

We think that both altitude and rainfall influence the diameter of whitebark pine stems.

```
head(dat_whitebark)
```

```
  X rainfall altitude      dbh
1 1 20.44596 2493.433 47.08338
2 2 33.23555 2301.189 52.09458
3 3 31.15216 2328.082 51.69528
4 4 25.24166 2390.111 49.53393
5 5 29.35912 2227.444 51.96837
6 6 28.51941 2421.535 49.85675
```

We'll fit additive and interactive models.

Interactions: Whitebark Pine Models

An additive multiple regression model:

```
fit_pine_additive = lm(
  dbh ~ altitude + rainfall,
  data = dat_whitebark)
```

An interactive multiple regression model:

```
fit_pine_interact = lm(
  dbh ~ altitude * rainfall,
  data = dat_whitebark)
```

Additive Model Coefficents

```
Call:
lm(formula = dbh ~ altitude + rainfall, data = dat_whitebark)

Residuals:
     Min       1Q   Median       3Q      Max
-1.61668 -0.38439  0.06462  0.35778  1.34592

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 75.6535333  0.8418852   89.86   <2e-16 ***
altitude    -0.0123821  0.0003306  -37.46   <2e-16 ***
rainfall     0.1304770  0.0116813   11.17   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5976 on 86 degrees of freedom
Multiple R-squared:  0.9437,    Adjusted R-squared:  0.9423
F-statistic: 720.1 on 2 and 86 DF,  p-value: < 2.2e-16
```

Additive Model Coefficients: In English

How can we state the output of the model table in plain English?

Recall the coefficients:

- altitude (measured in meters): -0.012
- rainfall (measured in cm): 0.13

“If I hold the altitude constant, for every 1cm increase in precipitation, I expect dbh to be 0.13 cm larger.”

“If I hold the precipitation constant, for every 1 m increase in altitude, I expect dbh to be -0.012 cm smaller.”

Interactive Model Coefficients

```
Call:
lm(formula = dbh ~ altitude * rainfall, data = dat_whitebark)

Residuals:
     Min       1Q   Median       3Q      Max
-1.66883 -0.30129  0.05161  0.32162  1.29048

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)        57.6240619  4.6659847  12.350  < 2e-16 ***
altitude           -0.0050139  0.0019048  -2.632 0.010075 *
rainfall            0.7155606  0.1496823   4.781 7.26e-06 ***
altitude:rainfall  -0.0002387  0.0000609  -3.919 0.000179 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5532 on 85 degrees of freedom
Multiple R-squared:  0.9523,    Adjusted R-squared:  0.9506
F-statistic: 565.3 on 3 and 85 DF,  p-value: < 2.2e-16
```

Interactive Model Coefficients: In English

How can we state the output of the model table in plain English?

Recall the coefficients:

- altitude: -0.005
- rainfall: 0.7156
- rainfall/altitude interaction: -2×10^{-4}

"There is an inhibiting interaction: If I increase the altitude by 1 m and the rainfall by 1 cm, the increase in dbh is -2×10^{-4} less than what I would predict from the rainfall and altitude coefficients alone."

Additive and Interactive Coefficients

What happened to the *main effect* coefficients?

In the additive model:

- altitude: -0.012
- rainfall: 0.13

In the interactive model:

- altitude: -0.005
- rainfall: 0.7156

Interactions

What do we know about interactions?

How do we tell if an interaction is *significant*?

- We could look at an ANOVA table:

```
Analysis of Variance Table

Response: dbh
                   Df Sum Sq Mean Sq  F value    Pr(>F)
altitude            1 469.76  469.76 1535.090 < 2.2e-16 ***
rainfall            1  44.55   44.55  145.593 < 2.2e-16 ***
altitude:rainfall   1   4.70    4.70   15.359 0.0001792 ***
Residuals          85  26.01    0.31
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Relative Importance of Effects: Standardizing the Variables

Which is a more important predictor for DBH?

- The model coefficients are in terms of units, so they are hard to compare.

We can standardize (scale) the variables to make them *unitless* before we create the models. R makes this pretty easy:

```
dat_whitebark_scaled = data.frame(scale(dat_whitebark))
head(dat_whitebark_scaled)
```

```
          X    rainfall    altitude         dbh
1 -1.703049 -1.8040603  0.2066960 -0.8669988
2 -1.664343  0.5190889 -0.7814965  1.1466020
3 -1.625637  0.1406553 -0.6432565  0.9861529
4 -1.586932 -0.9329511 -0.3244115  0.1176818
5 -1.548226 -0.1850408 -1.1605679  1.0958872
6 -1.509520 -0.3375675 -0.1628828  0.2473982
```

Coefficients: Interactive Standardized Model

```
Call:
lm(formula = dbh ~ altitude * rainfall, data = dat_whitebark_scaled)

Residuals:
     Min       1Q   Median       3Q      Max
-0.67057 -0.12106  0.02074  0.12923  0.51854

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.01391    0.02383   0.584 0.561016
altitude          -0.95866    0.02404 -39.882  < 2e-16 ***
rainfall           0.28776    0.02392  12.029  < 2e-16 ***
altitude:rainfall -0.10271    0.02621  -3.919 0.000179 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2223 on 85 degrees of freedom
Multiple R-squared:  0.9523,    Adjusted R-squared:  0.9506
F-statistic: 565.3 on 3 and 85 DF,  p-value: < 2.2e-16
```

ANOVA Table: Interactive Standardized Model

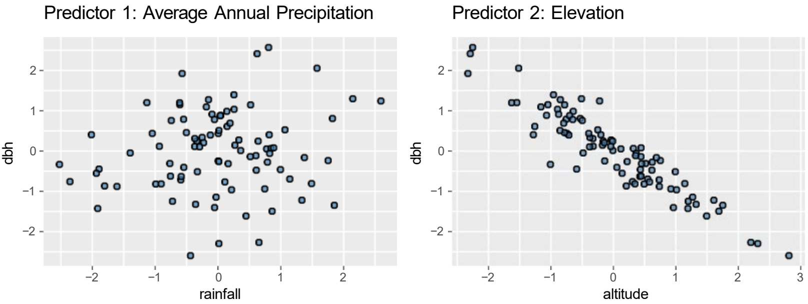
```
Analysis of Variance Table

Response: dbh
                  Df Sum Sq Mean Sq  F value    Pr(>F)
altitude           1 75.848  75.848 1535.090 < 2.2e-16 ***
rainfall           1  7.194   7.194  145.593 < 2.2e-16 ***
altitude:rainfall  1  0.759   0.759   15.359 0.0001792 ***
Residuals         85  4.200   0.049
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Scaled Plot

So... Which predictor is "more important"?

- What makes it seem more important?

Can we draw any intuition from the scaled plot?



Model Diagnostics

Remember the assumptions we made when we decided to use linear regression?

Two of the most important are:

- Normally-distributed residuals
- Independent observations

How do we know if they have been met?

Model Diagnostics: Normality of the Residuals

We'll focus on the assumption of Normally-distributed residuals.

I'll describe 3 approaches:

- Histogram of residuals
- Shapiro test of normality of the residuals
- Model plots:
  - Residuals vs. predicted values
  - **Quantile-Quantile plot**

Model Diagnostics: Cars data

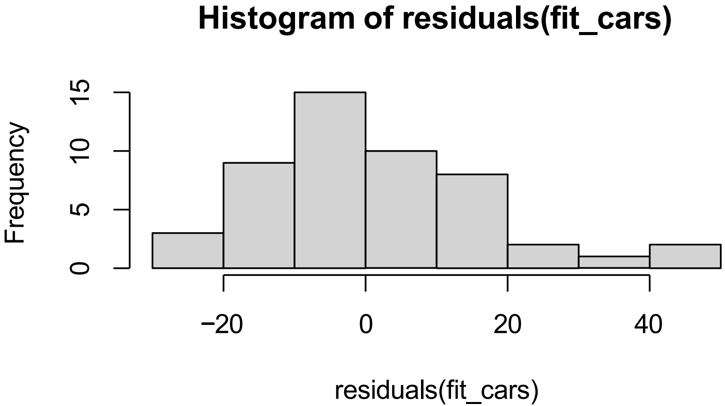
Recall our linear regression model of stopping distance and speed?

```
summary(fit_cars)
```

```

Call:
lm(formula = dist ~ speed, data = cars)

Residuals:
    Min      1Q  Median      3Q     Max 
-29.069  -9.525  -2.272   9.215  43.201 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -17.5791     6.7584  -2.601   0.0123 *  
speed         3.9324     0.4155   9.464 1.49e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Diagnostics: Cars data

```
hist(residuals(fit_cars))
```



Model Diagnostics: Cars data

```
shapiro.test(residuals(fit_cars))
```

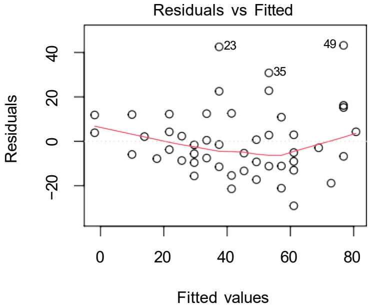
```
	Shapiro-Wilk normality test

data:  residuals(fit_cars)
W= 0.94509, p-value = 0.02152
```

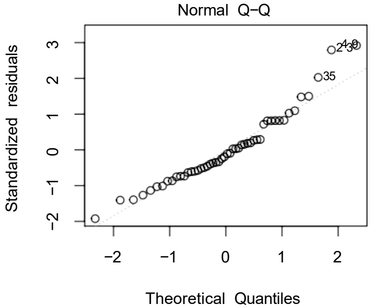
Do we think our residuals are normally distributed?

- What is the null hypothesis of the Shapiro test?

Model Diagnostics: Cars data

```
par(mfrow = c(1, 2))
plot(fit_cars, which = 1)
plot(fit_cars, which = 2)
```





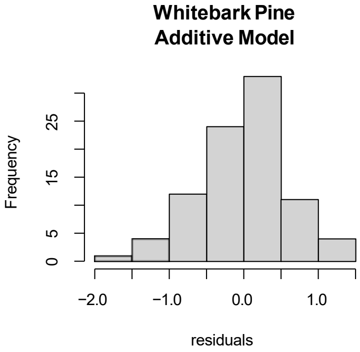
Model Diagnostics: Interpretation

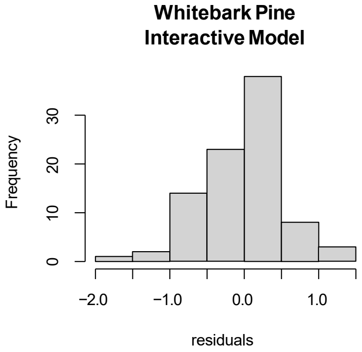
How can we interpret these diagnostics?

- Residuals should be normal:
  - histogram
  - residual plot
  - Shapiro test
- Q-Q plot
  - Normally-distributed residuals produce a straight line

Model Diagnostics: Whitebark Pine - Graphical

We created additive and interactive models of the whitebark pine. We could compare their residuals?





Model Diagnostics: Whitebark Pine - Numerical

What do the Shapiro tests tell us?

```
shapiro.test(residuals(fit_pine_additive))
```

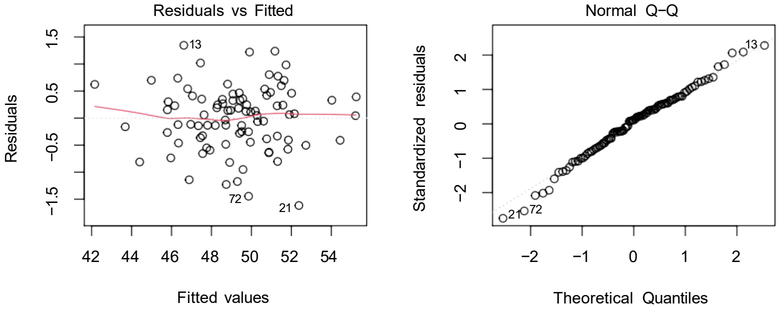
```
    Shapiro-Wilk normality test

data:  residuals(fit_pine_additive)
W = 0.98937, p-value = 0.6916
```

```
shapiro.test(residuals(fit_pine_interact))
```

```
    Shapiro-Wilk normality test

data:  residuals(fit_pine_interact)
W = 0.98506, p-value = 0.4015
```

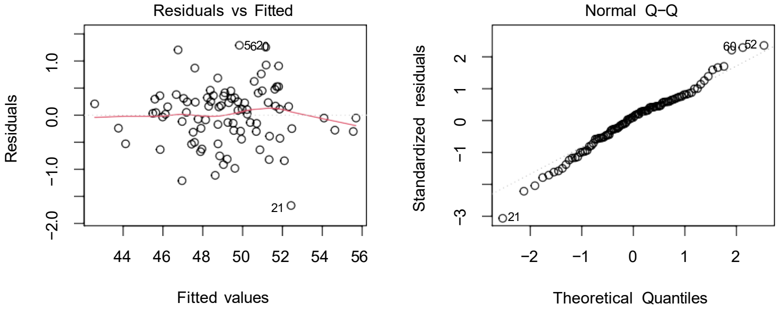
Model Diagnostics: Whitebark Pine - Additive Model Plots

Model plots for the additive model:



Model Diagnostics: Whitebark Pine - Interactive Model Plots

Model plots for the interactive model:



Model Diagnostics: Assessing Normality

What are 3 ways we can assess the normality assumption?

Model Diagnostics: Assessing Normality

What are 3 ways we can assess the normality assumption?

- Graphical: histogram of residuals (not the data!)
- Graphical: Q-Q plot
- Numerical: Shapiro test of residuals