72 + 8

Quiz 2

Adon Rosen

Date: 2020-03-27

Question 1

How fast can evolution occur in nature? Are evolutionary trajectories unique or predictable? In 1980, a European Union (EU) fly (Drosophila subobscura) was accidentally introduced into North America.
In Europe, the fly's wing size systematically varies with latitude, suggesting an evolutionary adaptation. After allowing two decades for the introduced North American flies to spread over the continent, flies were captured and the hypothesis of speedy evolution was examined by comparing the wing sizes at different latitudes between NA and EU flies.

a.

Model Restricted: $Y_{WingSize} = \beta_{intercept} + \beta_{Continent} \times X_{j,1} + \beta_{sex} \times X_{j,2}$ $+ \varepsilon$

Model Full: $Y_{WingSize} = \beta_{intercept} + \beta_{Continent} \times X_{j,1} + \beta_{sex} \times X_{j,2} + \beta_{latitude} \times X_{j,3}$ $+ \varepsilon$

b.

The number of degrees of freedom for the full model is: 38

The number of degrees of freedom for the restricted model is: 39

c. Suppose you were to see the following SAS code in your program editor window:

PROC REG;

MODEL Y = X1 X2 X3;

DEMO: Test X2=0, X3=0;

c1.

The hypothesis being tested is if the model including latitude, continent and sex explains more of the variance in wing size than a model only inclduing latitude.

c2.

model restricted: $Y_{WingSize} = \beta_{intercept} + \beta_{latitude} \boxed{+} X_{j,1}$ $+ \varepsilon$

model full: $Y_{WingSize} = \beta_{intercept} + \beta_{Continent} \times X_{j,1} + \beta_{sex} \times X_{j,2} + \beta_{latitude} \times X_{j,3}$ $+ \varepsilon$

d.

The estimated mean value for a female fly from North America captured at a latitude of 45 degrees can be written as: $\hat{Y} = \beta_{intercept} + \beta_{latitude} \times 45 + \beta_{continent}$

Question 2

To evaluate the speedy adaptional hypothesis, we need to evaluate whether or not the rates of wing change as a function of latitude vary between EU and NA flies. We may do this by including an interaction term X4 - where X4 is the interaction between latitude and continent. In your favorite program, run a multiple regression model - with wing size as the DV - that includes the linear effects of sex, continent, latitude, and the latitude by continent interaction.

a.

```
Call:
lm(formula = WingSize ~ continent * latitude + Sex, data = data.in)

Residuals:
     Min       1Q   Median       3Q      Max
-22.5537  -6.9902   0.7025   5.6303  29.9364

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          820.9865    19.8740  41.310  < 2e-16 ***
continentna           28.0224    28.3491   0.988    0.329
latitude               2.1224     0.4268   4.972 1.54e-05 ***
SexMale              -98.8571     3.3977 -29.095  < 2e-16 ***
continentna:latitude  -0.7217     0.6315  -1.143    0.260
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.01 on 37 degrees of freedom
Multiple R-squared:   0.96, Adjusted R-squared:  0.9556
F-statistic: 221.8 on 4 and 37 DF,  p-value: < 2.2e-16
```

Given the model summary results presented above, it appears that the effect which explains the most in wing size is fly sex. We can compare the relative effects of these covariates from their t values. The covariate with the largest t value is sex.

Another option to compare variable important is to observe the Type III sum of squares. These are the partial sum of sqaures, so the relative influence the variable of interest has while contorlling for all of ther variables. Because this is performed in an Anova framework, the variable which explains the greatest variance will also have the largest F value, which in this case is also the fly sex variable, with an F stat of: 846.5195501

b.

The value of the multiple correlation is: 0.9797812. This model has a poisitive relationship between the predicted values and the true values.

c.

The F value of the interaction effect is: 1.3062334

The p value for the interaction term is: 0.2604194

The squared partial correlation value is: 0.0340998

d.

The model with the correct coefficients is: $\hat{Y} = 820.9865 + 28.0224 \times X_{i,Continent} + 2.1224 \times X_{i,Latitude} - 98.8571 \times X_{i,Sex} - 0.7217 \times X_{i,Continent*Latitude}$

e.

The RMSE for this model is: 10.3338089. The root MSE is equivalent to the conditional standard deviation. This also refers to the standard deviation of the residual values.

f.

The observation with the greatest residual value has the following observed values:

	continent	latitude	WingSize	Sex	latitudeScale	WingSizeScale	PredictedValues	Residuals
39	eu	48.5	855	Male	0.6917848	-0.1821812	825.0636	29.93635

The predicted and observed values are:

	WingSize	PredictedValues	Residuals
39	855	825.0636	29.93635

g.

The coefficient of the interaction term is: -0.721734. This value indicates a modification for the slope between latitude and wing size for the observations from North America. Specifically when an observation is from the NA continent, this value will be multiplied with the latitude of the observation, this product will then be added to the products of the other main effects. The modification to slope suggests that observations from NA have a slightly lower relationship between wing size and latitude.

h.

The t value for the interaction term with standardized coefficients is: -1.1429057

i.

This question is asking for the comparison in model fit between a model with an interaction term and a model without one.

The change in residual sum of squares is: 158.34

The corresponding F stat is: 1.3062334 . Given the reported F stat, we would conclude that the interaction term does not explain a significant amount of variance.

j.

Here I will return the predicted values for a female from North America captured at 45 degrees of latitude from all trained models:

Model Reduced (Main Effects): $\hat{Y}_{WingSize} = \beta_{intercept} + \beta_{Continent} \times X_{j,1} + \beta_{sex} \times X_{j,2} = 909.6558442$

Model Full (Main Effects): $\hat{Y}_{WingSize} = \beta_{intercept} + \beta_{Continent} \times X_{j,1} + \beta_{sex} \times X_{j,2} + \beta_{latitude} \times X_{j,3} = 912.70328$

Model Full (Interaction): $\hat{Y}_{WingSize} = \beta_{intercept} + \beta_{Continent} \times X_{j,1} + \beta_{sex} \times X_{j,2} + \beta_{latitude} \times X_{j,3} + \beta_{Continent*Latitude} \times X_{j,4} = 912.0369012$

k.

Expected intercept: 849.0089209

Expected slope: 1.4006218

l.

Expected intercept: 820.9865343

Expected slope: 2.1223558

m.

Expected intercept: 750.1517781

Expected slope: 1.4006218

n.

Expected intercept: 722.1293915

Expected slope: 2.1223558

Question 3

Consider a multiple regression in which we have 4 variables: A response variable Y and 3 explanatory variables: x1, x2, and x3.

Use proper notation for all parts of the question. That is, use

r2y1.2 for (squared) partial correlations,

r2y(1.2), for (squared) semi-partial correlations, and

r2y1 for (squared) simple correlations.

R2y.12 for (squared) multiple correlations

a.

$R^2_{Y.X2X3} = r^2_{yx2} + R^2_{y(x3.x2)}$

b.

$r^2_{Y(X2.X1X3)} = \frac{r^2_{y.x2x1x3}}{1-r^2_{YX2}}$

c.

$f(R^2_{yx3.x1x2}) = \frac{R^2_{yx3.x1x2}-R^2_{x3.x1x2}-R^2_{y.x1x2}}{\sqrt{1-R^2_{x3.x1x2}\sqrt{1-R^2_{y.x1x2}}}}$

Question 4

Cor Table

Correlation x1.y: 0.2565

Analysis of Variance

Model DF: 2

Model SS: 9.7514609

Model Mean Square: 4.8757305

Error SS: 103.2632869

Error Mean Square: 1.06457

Corrected Total DF: 99

Corrected Total SS: 113.1152784

R-Square: 0.0862986

Adj R-Sq: 0.0674594

Parameter Estimates

Intercept DF: 1

Intercept t Value: 113.9284494

Intercept Standardized Estimate: 0

X1 DF: 1

X1 Standard Error: 0.120543

X1 Semi-partial Corr Type I: 0.2565

X2 DF: 1

X2 Parameter Estimate: 0.18167

X2 Semi-partial Corr Type I: 0.1432003

X2 Squared Partial Corr Type II: 0.0219505