Module 4E
Supervised Machine Learning

Different flavors of REGRESSION and General Linear Models

ANCOVA

- Increases precision
- Attempts to adjust for bias
- Often will include “pre” and “post” treatment to try to account for **confounding** individual differences
- Common: SES, age

Covariate effects:

- Confounding variables bias estimates of treatment effects

Covariate: a variable (or group of variables) that accounts for a portion of the variance in the dependent variable

- Allows researchers to test for group differences while controlling for effects of the covariate

ANCOVA VS Multi-factor ANOVA?

- ANCOVA doesn’t require that the variable we are trying to control for is ***necessarily an independent categorical variable***

Example: Amount of tv watching for girls versus boys (**independent variable** 1 = gender), in four different geographic locations (**independent variable** 2 = North, South, West, East), the interaction (gender*geography) ***and S.E.S.*** (what type of variable could this be?)

One-way ANOVA:

- 1 continuous dependent variable
- 1 categorical independent variable (≥ 2 groups)
- i.e., **Girls vs boys** in hours of tv watched

Multi-Factor ANOVA:

- 1 continuous dependent variable
- ≥ 2 categorical independent variables
- i.e., **Girls vs boys** in hours of tv watched in **four regions** of the United States

ANCOVA:

- 1 continuous dependent variable
- ≥ 1 categorical independent variables
- 1 categorical variable
- i.e., **Girls vs boys** in hours of tv watched in **four regions** of the United States and **SES**

Covariate effects:

Confounding variables bias estimates of treatment effects

Experimental - eliminate confounding variables by random assignment of treatment

Observational - include known confounding variables and correct for their distorting influence

ANCOVA: Analysis of Covariance

Two rounds of model fitting:

Response = Constant + Factor 1 + Covariate + Factor 1*Covariate

1. Interaction between covariate and treatment is tested
   Regression slopes differ among the 'groups' if interaction is present

2. If no interaction is detected, interaction term is dropped and treatment effect is tested

Response = Constant + Factor 1 + Covariate + Factor 1*Covariate

Two rounds of model fitting:

1. *Interaction between covariate and treatment is tested*

   Regression slopes differ among the 'groups' if interaction is present

   $$\text{F-test} = \frac{H_A: \text{Constant + Factor 1 + Covariate + Factor 1*Covariate}}{H_0 \quad \text{Constant + Factor 1 + Covariate}}$$

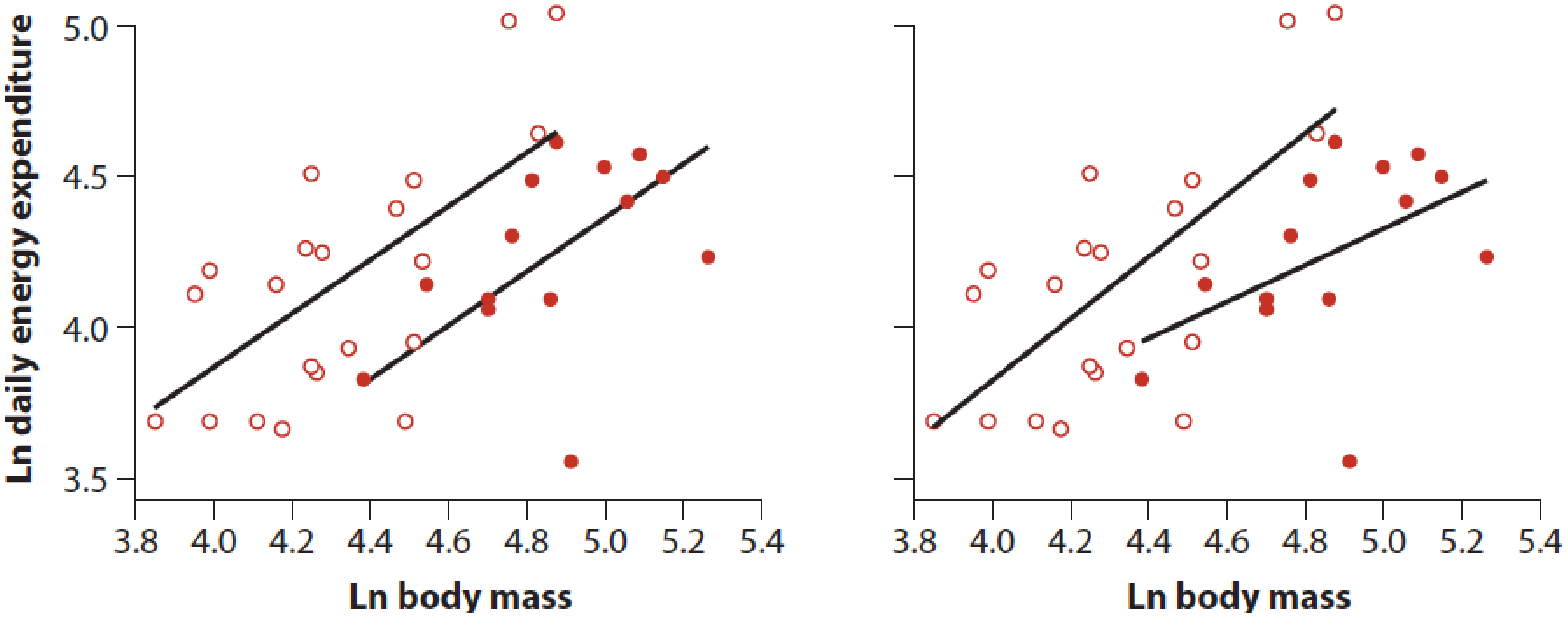
2. *If no interaction is detected, interaction term is dropped and treatment effect is tested*

   $$\text{F-test} = \frac{H_A: \text{Constant + Factor 1 + Covariate}}{H_0 \quad \text{Constant + Covariate}}$$

<u>Example:</u> Mole-rats are eusocial mammals with a queen, reproductive males and workers in a colony. It seems there might be two worker castes: “frequent workers”, who do most of the work of the colony, and “infrequent workers”, who do work after rains. Energy expenditure varies with body mass in both groups, but infrequent workers are heavier than frequent workers. **How different is mean daily energy expenditure between two groups when <u>adjusted for</u> differences in body mass?**

ENERGY = CONSTANT + CASTE + MASS

ENERGY = CONSTANT + CASTE + MASS + CASTE*MASS



Covariate effects:

Energy = Constant + Caste+ Mass + Caste*mass

Two rounds of model fitting:

1. Interaction between covariate and treatment is tested

H_0: There is no interaction between caste and mass

H_A: There is interaction between caste and mass

$$\text{F-test} = \frac{H_A: \text{Constant + Factor 1 + Covariate + Factor 1*Covariate}}{H_0 \quad \text{Constant + Factor 1 + Covariate}}$$

2. If no interaction is detected, interaction term is dropped, and treatment effect is tested

$$\text{F-test} = \frac{H_A: \text{Constant + Factor 1 + Covariate}}{H_0 \quad \text{Constant + Covariate}}$$

1^{st} Round Results:

F-test = H_A: Constant + Caste + mass + Caste*mass

H_0 Constant+Caste+Mass

Source of Variation	SS	DF	MS	F	P
Caste	**0.0570**	1	**0.0570**		
Mass	1.3618	1	1.3618		
Caste*Mass	0.0896	1	0.0896	1.02	0.321
Residual	2.7249	32	0.0879		
Total	4.233	35			

Without the interaction term, the regression lines have slopes that are not significantly different – so we can drop interaction!

Round 2: The updated model: Energy = Constant + Caste+ Mass

H_0: There is no difference in energy expenditure between different castes

H_A: There is a difference in energy expenditure between different castes

$$\text{F-test} = \frac{H_A\text{: Constant + Caste + Mass}}{H_0\quad\text{Constant + Mass}}$$

Source of Variation	SS	DF	MS	F	P
Caste	**0.6375**	1	**0.6375**	7.25	0.011
Mass	1.8815	1	1.8815	21.39	<0.011
Residual	2.72.814	32	0.0880		
Total	5.3335	34			