ANCOVA

- Increases precision
- Attempts to adjust for bias

- Often will include “pre” and “post” treatment to try to account for confounding individual differences

- Common: S.E.S., age

Covariate effects:

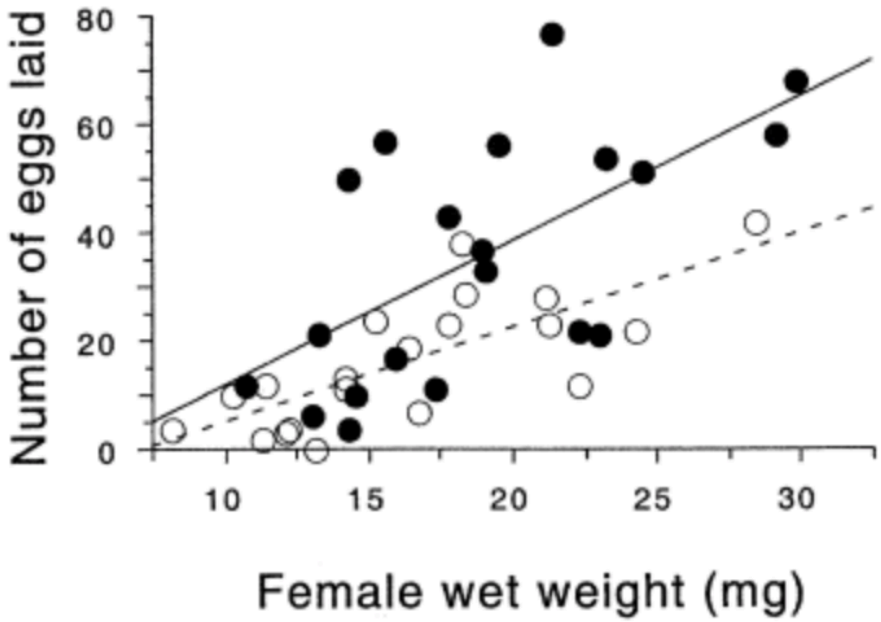
- Confounding variables bias estimates of treatment effects
- **Covariate**: a variable (or group of variables) that accounts for a portion of the variance in the dependent variable
  - Allows researchers to test for group differences while controlling for effects of the covariate

ANCOVA VS Multi-factor ANOVA?

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

- Ex. 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?)

http://www.biostathandbook.com/ancova.html



Females were mated once (white circles) or three times (black filled circles).
Since larger females are known to produce more offspring, before mating, the female fireflies were weighed.
The slopes between the once mated and thrice mated were not significantly different but the Y-intercepts are significantly different.
Is there a difference in the number of eggs laid between once and thrice mated females?

a. No
b. Yes
c. Can't tell

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**

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*

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

Example: Mole-rats are eusocial mammals with a queen, reproductive males and workers in a colony. It seems as though 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 adjusted for differences in body 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

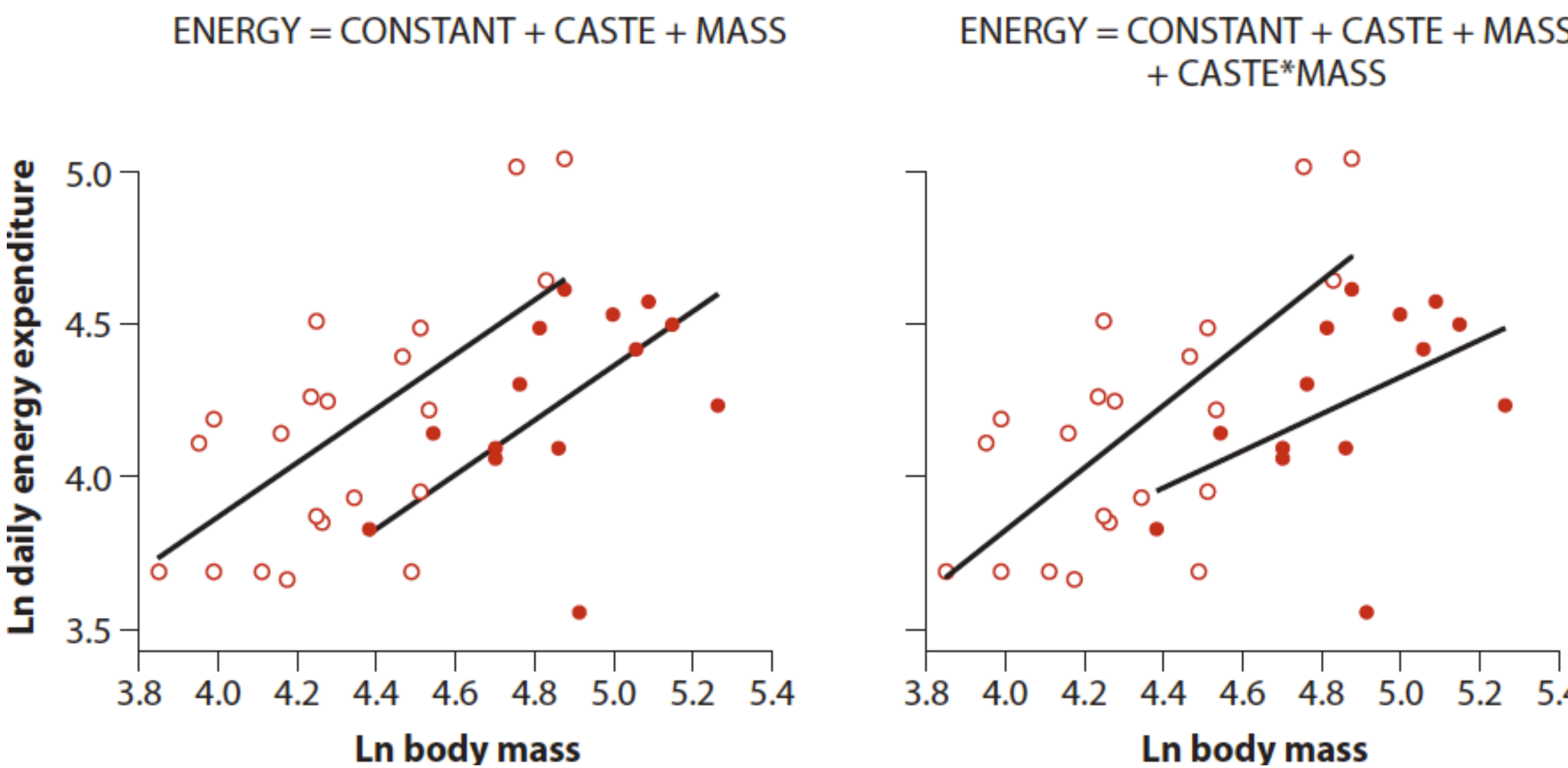
H_A: There is interaction between caste and mass

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

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

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

ANCOVA



ENERGY = CONSTANT + CASTE + MASS
ENERGY = CONSTANT + CASTE + MASS + CASTE*MASS
Ln daily energy expenditure
5.0
4.5
4.0
3.5
3.8
4.0
4.2
4.4
4.6
4.8
5.0
5.2
5.4
Ln body mass
Ln body mass

1^{st} Round Results:

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

Source of variation	Sum of squares	df	Mean Squares	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	4.2333	35			

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

The new model is:

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

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

2nd Round Results:

Source of variation	Sum of squares	df	Mean Squares	F	P
CASTE	**0.6375**	**1**	**0.6375**	**7.25**	**0.011**
MASS	**1.8815**	**1**	**1.8815**	**21.39**	**<0.001**
Residual	**2.814**	**32**	**0.0880**		
Total	5.3335	34			

Covariate effects:

- Problem 1: there might be other confounding variables that we didn't include in the study
- Problem 2: what if there **is interaction?**
  - Specify a constant value of the Factor 1

Assumptions of General Linear Models:

- Measurements at every combination (block and treatment combo) are random sample
- Measurements for every combination of values for explanatory variables have a normal distribution
- Variance of response variable is same for all combos of explanatory variables