

Lecture 11

1. Adjusting: stratification, regression
2. How do LSmeans adjust?
3. Non-estimable LSmeans

1

Comparing groups while adjusting or controlling a covariate

Rat liver enzymes: we want to compare the effects of diets, adjusting for livers of different weights.

Response y_j from individuals in two or more groups (A) with additional covariate (predictor) x_j measured on each individual.

Adjusting the comparison of groups A for X means comparing groups using individuals with equal X .

Controlling for X means comparing groups using individuals with X held constant.

Adjusting = controlling

2

Adjusting 1: Strata

To adjust comparison of groups A for X :

compare levels of A using individuals with equal or nearly-equal X .

- Cut X into categories = *strata* or *blocks*, creating a new variable B where individuals have similar values of x .

This gives ANOVA adjustment.

- Match individuals drawn from different groups on their values of X .

3

Stratified or matched analysis:

1. Calculate difference between levels of A **within each stratum** or matched set
2. Test whether A differences vary with levels of B : **interaction**;
draw interaction plot (LSmeans $A*B$).
3. If no interaction, combine the A differences across X : LSmeans A
4. If there's an interaction, explain how it works.

4

Adjusting 2: Regression

To adjust comparison of groups A for X :

compare levels of A using individuals with equal or nearly-equal X .

Regression adjustment:

1. Regress response y on X for each treatment group (parallel or separate lines)
2. At selected value of X (default: overall mean \bar{x}),
compute predicted mean \hat{y}_j for each treatment group = LSmeans A
3. Compare predicted values \hat{y}_j = compare LSMeans

5

Regression-adjustment analysis:

1. Regress response y on X for each treatment group with distinct line for each level of A

$$\text{MODEL } y = A X A*X;$$

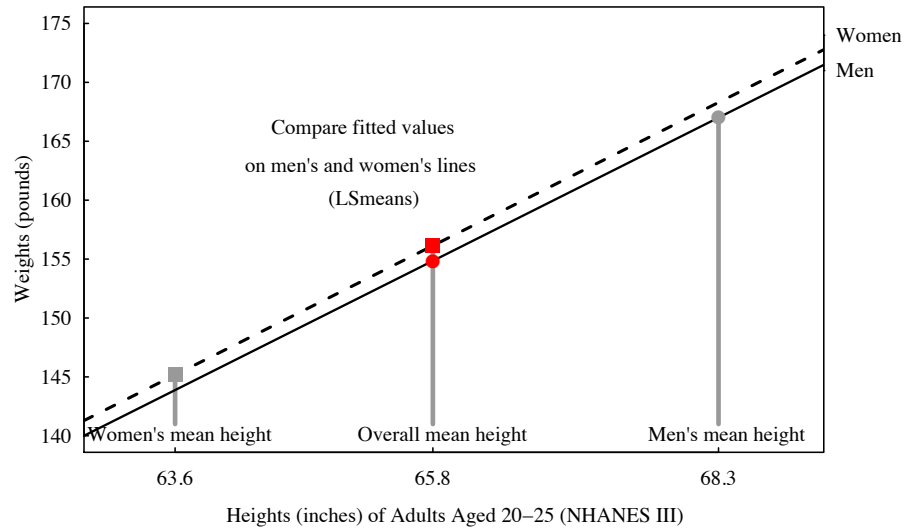
2. Test whether separate A regressions really have different slopes : **interaction**;
draw interaction plot.

6

3. If no interaction, fit the same slope for all levels of A : parallel lines

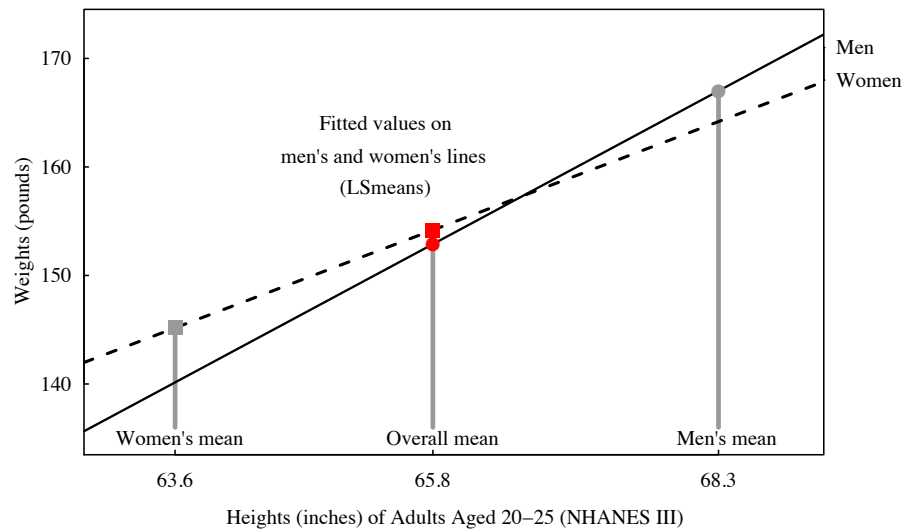
$$\text{MODEL } y = A X;$$

At selected value of X (default: overall mean \bar{x}), compute predicted mean for each treatment group = LSmeans A . Compare LSMeans.



7

4. If there's an interaction, regression adjustment will give different answers depending on where predicted means are calculated.



Regression adjustment may not make sense.

8

Dangers in regression adjustment

Often want to adjust for many continuous covariates X_1, \dots, X_p .

Lots of work to check all the interactions:

```
MODEL y = A X1 X2 X3 X4 X5 X6 A*X1 A*X2 A*X3 A*X4 A*X5 A*X6;
```

Most convenient to just add them to the model as main effects:

```
MODEL y = A X1 X2 X3 X4 X5 X6;
```

If you don't check, you assume interactions = 0.

LSMeans output doesn't reveal problems.

9

Computation of LSmeans (from Proc GLM Documentation)

Assume A has 3 levels, B has 2 levels, and C has 2 levels, and assume that every combination of levels of A and B exists in the data. Assume also that Z is a continuous variable with an average of 12.5. Then the least squares means are computed by the following linear combinations of the parameter estimates:

	A			B		A*B						C			
	μ	1	2	3	1	2	11	12	21	22	31	32	1	2	Z
LSM()	1	1/3	1/3	1/3	1/2	1/2	1/6	1/6	1/6	1/6	1/6	1/6	1/2	1/2	12.5
LSM(A1)	1	1	0	0	1/2	1/2	1/2	1/2	0	0	0	0	1/2	1/2	12.5
LSM(A2)	1	0	1	0	1/2	1/2	0	0	1/2	1/2	0	0	1/2	1/2	12.5
LSM(A3)	1	0	0	1	1/2	1/2	0	0	0	0	1/2	1/2	1/2	1/2	12.5
LSM(B1)	1	1/3	1/3	1/3	1	0	1/3	0	1/3	0	1/3	0	1/2	1/2	12.5
LSM(B2)	1	1/3	1/3	1/3	0	1	0	1/3	0	1/3	0	1/3	1/2	1/2	12.5
LSM(AB11)	1	1	0	0	1	0	1	0	0	0	0	0	1/2	1/2	12.5
LSM(AB12)	1	1	0	0	0	1	0	1	0	0	0	0	1/2	1/2	12.5

X is continuous predictor (eg liver weight)

A and B are indicator (0/1) variables (eg Api and Cru).

Fitted regression equation from Proc GLM or Proc Reg:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_X X + \hat{\beta}_1 A + \hat{\beta}_2 B + \hat{\beta}_3 A * B$$

To find LSMeans for $A, B, A * B$, substitute

11

To calculate LSmean for $A1 = 1$ and $A2 = 0$:

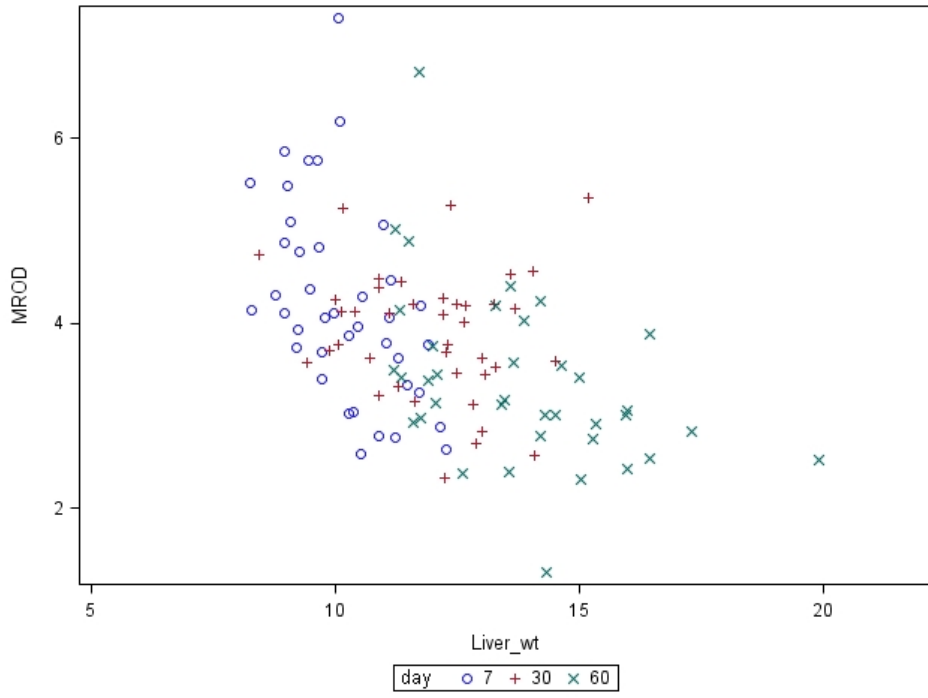
Mean value of X , over all the observations, is substituted into the fitted regression equation.

$$\begin{aligned} \text{LSmean}(A1 = 1, A2 = 0) &= \hat{\beta}_0 + \hat{\beta}_X \bar{x} + \hat{\beta}_1(1) + \hat{\beta}_2(0) + \hat{\beta}_3(1 * 0) \\ &= \hat{\beta}_0 + \hat{\beta}_X \bar{x} + \hat{\beta}_1 \\ &= (\hat{\beta}_0 + \hat{\beta}_1) + \hat{\beta}_X \bar{x} \end{aligned}$$

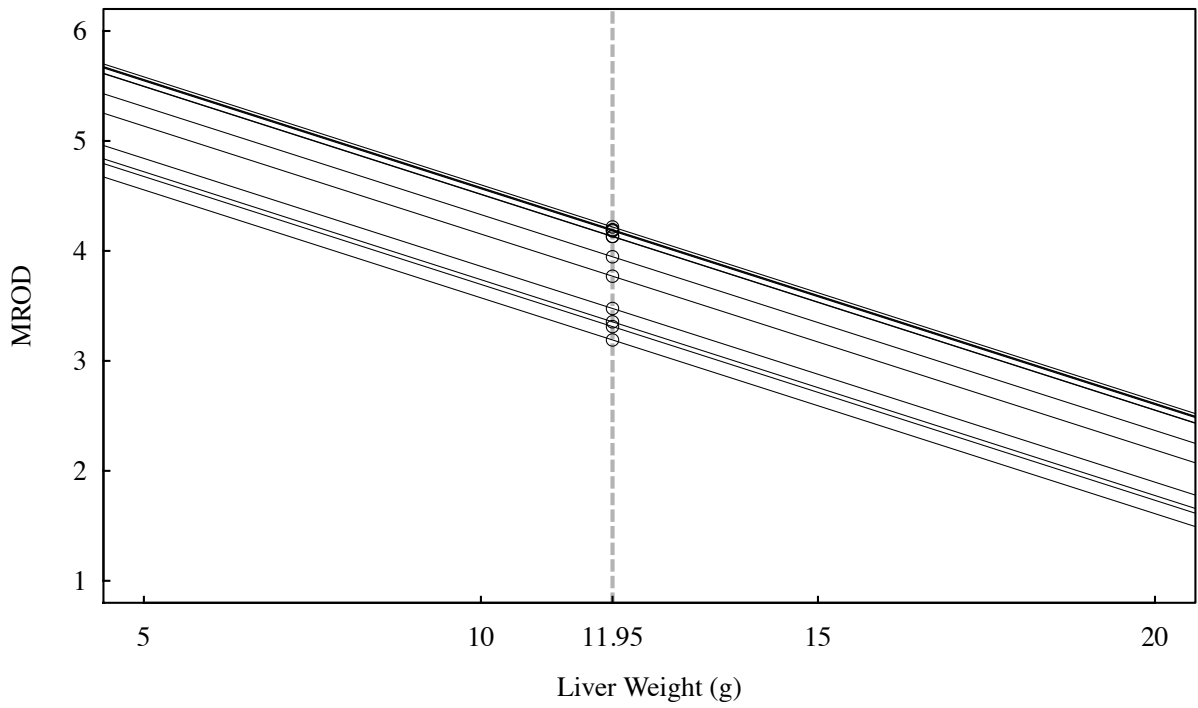
Parallel lines with slope $\hat{\beta}_X$, intercept determined by indicators.

LSmean is point on line at \bar{x} .

12

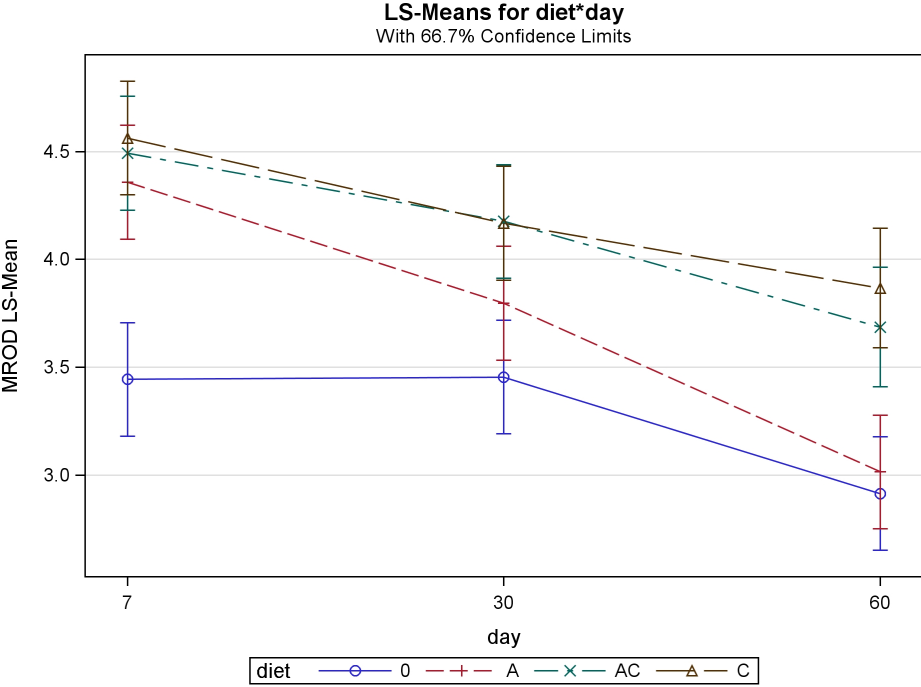


13

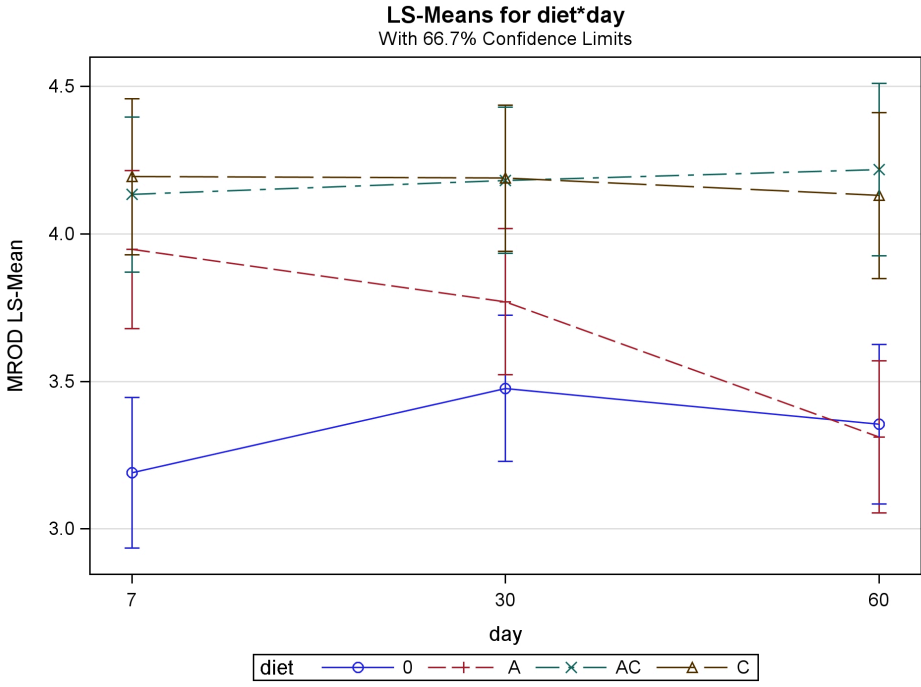


14

Unadjusted interaction plot:



Adjusted for liver weight:



To see the averaging, request LSmeans for day = 7, 30, 60 (means across diets)

```
proc glm data=flash.rat_diets;
  class day api cru ;
  model mrod = liver_wt  api cru api*cru
    day day*api day*cru day*api*cru / solution;
  lsmeans day/ stderr E;
```

17

Coefficients for day Least Square Means

Effect	day Level		
	7	30	60
Intercept	1	1	1
Liver_wt	11.9521368	11.9521368	11.9521368
Api 0	0.5	0.5	0.5
Api 1	0.5	0.5	0.5
Cru 0	0.5	0.5	0.5
Cru 1	0.5	0.5	0.5
Api*Cru 0 0	0.25	0.25	0.25
Api*Cru 0 1	0.25	0.25	0.25
Api*Cru 1 0	0.25	0.25	0.25
Api*Cru 1 1	0.25	0.25	0.25
day 7	1	0	0
day 30	0	1	0
day 60	0	0	1
day*Api 7 0	0.5	0	0
day*Api 7 1	0.5	0	0
day*Api 30 0	0	0.5	0
day*Api 30 1	0	0.5	0
day*Api 60 0	0	0	0.5
day*Api 60 1	0	0	0.5
day*Cru 7 0	0.5	0	0

18

day*Cru	7 1	0.5	0	0
day*Cru	30 0	0	0.5	0
day*Cru	30 1	0	0.5	0
day*Cru	60 0	0	0	0.5
day*Cru	60 1	0	0	0.5
day*Api*Cru	7 0 0	0.25	0	0
day*Api*Cru	7 0 1	0.25	0	0
day*Api*Cru	7 1 0	0.25	0	0
day*Api*Cru	7 1 1	0.25	0	0
day*Api*Cru	30 0 0	0	0.25	0
day*Api*Cru	30 0 1	0	0.25	0
day*Api*Cru	30 1 0	0	0.25	0
day*Api*Cru	30 1 1	0	0.25	0
day*Api*Cru	60 0 0	0	0	0.25
day*Api*Cru	60 0 1	0	0	0.25
day*Api*Cru	60 1 0	0	0	0.25
day*Api*Cru	60 1 1	0	0	0.25

Equal weight to each level, giving *balanced* LSmeans even though design is not balanced. LSMeans A / OM stderr pdiff; (observed margins) option gives arithmetic means.

19

Non-estimable LSmeans

In rat diet experiment, liver enzyme assayed in plates, 8 rats at a time.

To account for plate-to-plate variation, we should include this in model.

```
proc glm data=flash.rat_diets;
  class day api cru plate ;
  model mrod = plate liver_wt api cru api*cru
    day day*api day*cru day*api*cru / solution;
  lsmeans day*api*cru / stderr slice=day;* E;
```

20

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Plate	12	14.20216234	1.18351353	2.05	0.0284
Liver_wt	1	4.69052329	4.69052329	8.11	0.0054
Api	1	1.01404063	1.01404063	1.75	0.1887
Cru	1	12.58240943	12.58240943	21.76	<.0001
Api*Cru	1	0.76569779	0.76569779	1.32	0.2528
day	0	0.00000000	.	.	.
day*Api	2	0.46902814	0.23451407	0.41	0.6678
day*Cru	2	0.36275345	0.18137673	0.31	0.7315
day*Api*Cru	2	1.15928428	0.57964214	1.00	0.3709

So plate should be included, to reduce known variability .

Otherwise this variable goes into the error term. Consequence?

21

However, now we cannot get estimates of the means:

Least Squares Means			
day	Api	Cru	MROD LSMEAN
7	0	0	Non-est
7	0	1	Non-est
7	1	0	Non-est
7	1	1	Non-est
30	0	0	Non-est
30	0	1	Non-est
30	1	0	Non-est
30	1	1	Non-est
60	0	0	Non-est
60	0	1	Non-est
60	1	0	Non-est
60	1	1	Non-est

Non-est means non-estimable.

22