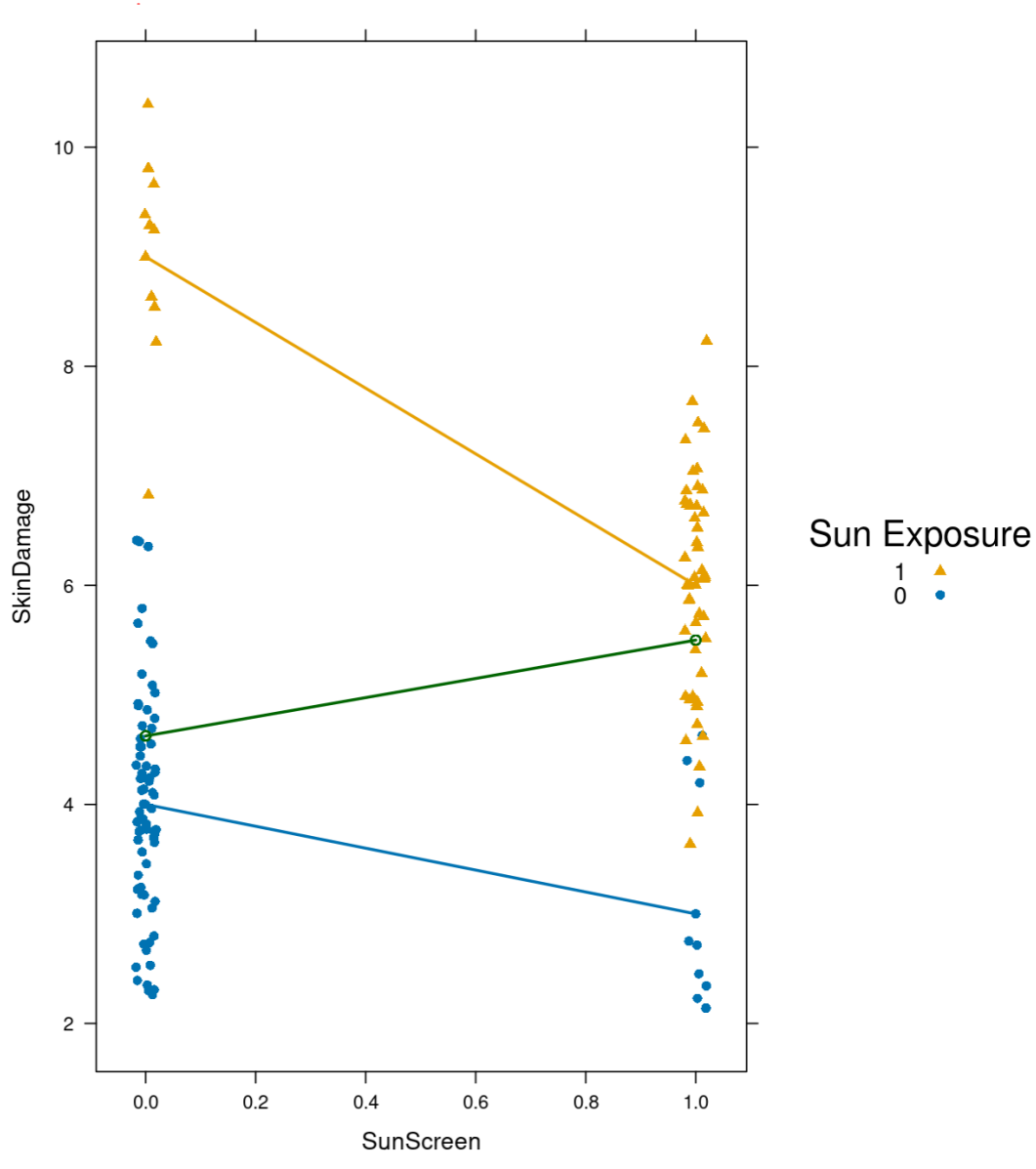
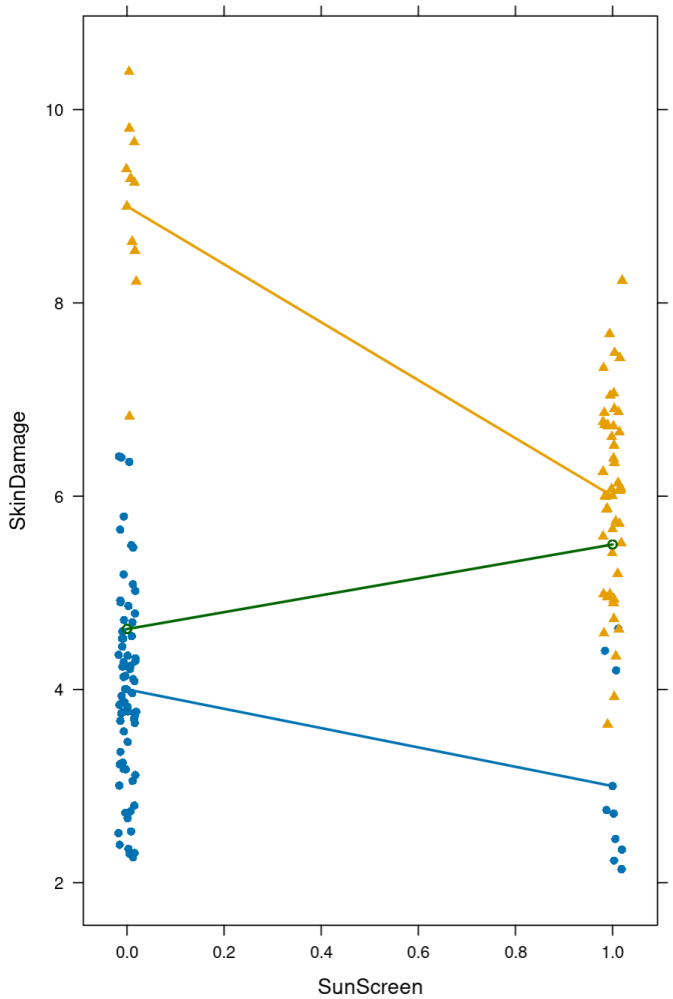
```
Call:
lm(formula = SkinDamage ~ SunScreen * SunExposure, data = dd)

                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            4.0000     0.1195  33.466  < 2e-16 ***
SunScreen             -1.0000     0.3381  -2.958  0.00365 **
SunExposure            5.0000     0.3381  14.790  < 2e-16 ***
SunScreen:SunExposure -2.0000     0.4840  -4.132 6.25e-05 ***
```

```
Call:
lm(formula = SkinDamage ~ SunScreen, data = dd)

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.6250     0.1973  23.442   <2e-16 ***
SunScreen     0.8750     0.3014   2.903   0.0043 **
```



What else would we want to estimate?



- Conditional effects of X when Z=0, Z=1
- conditional effects of Z when X=0, X=1
- difference in conditional effects of X
- difference in ... Z
- marginal effect of X
- marginal effect of Z

What does the output give us?

```
Call:
lm(formula = SkinDamage ~ SunScreen * SunExposure, data = dd)

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          4.0000     0.1195  33.466  < 2e-16 ***
SunScreen           -1.0000     0.3381  -2.958  0.00365 **
SunExposure          4.0000     0.3381  11.832  < 2e-16 ***
SunScreen:SunExposure -1.0000     0.4840  -2.066  0.04073 *
```

$\hat{\beta}_0$ $\hat{\beta}_1$ $\hat{\beta}_2$ $\hat{\beta}_3$

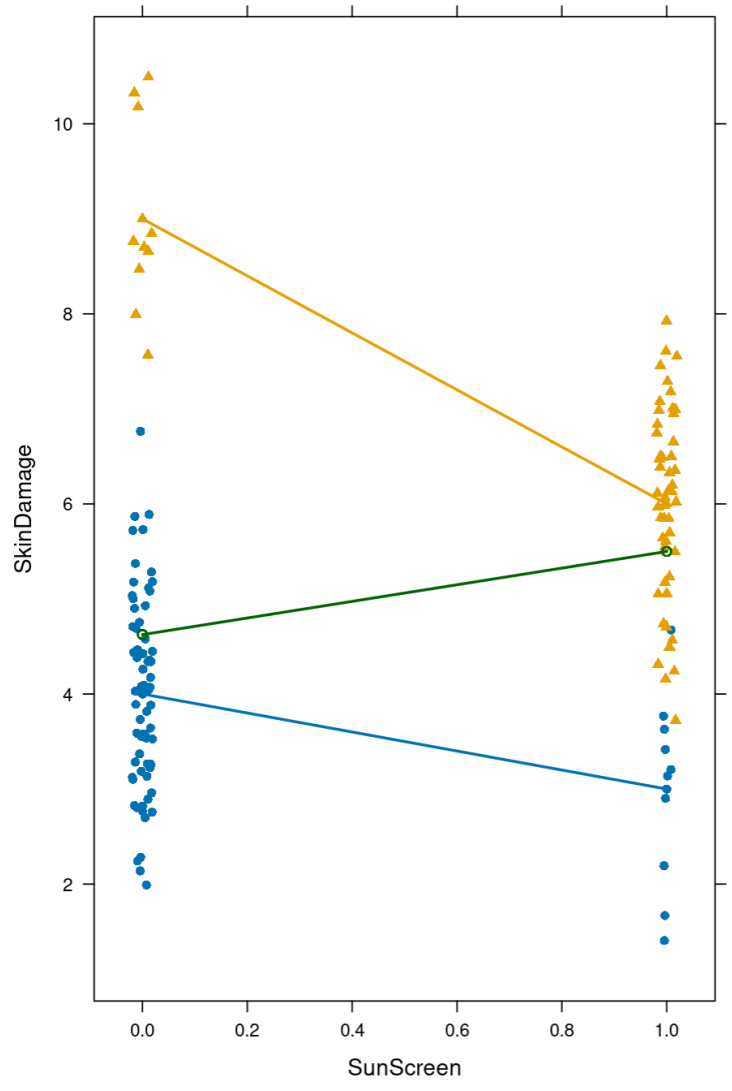
Note: We can calculate

$\beta_{x|z=0} =$

$\beta_{x|z=1} =$

$\beta_{z|x=0} =$

$\beta_{z|x=1} =$



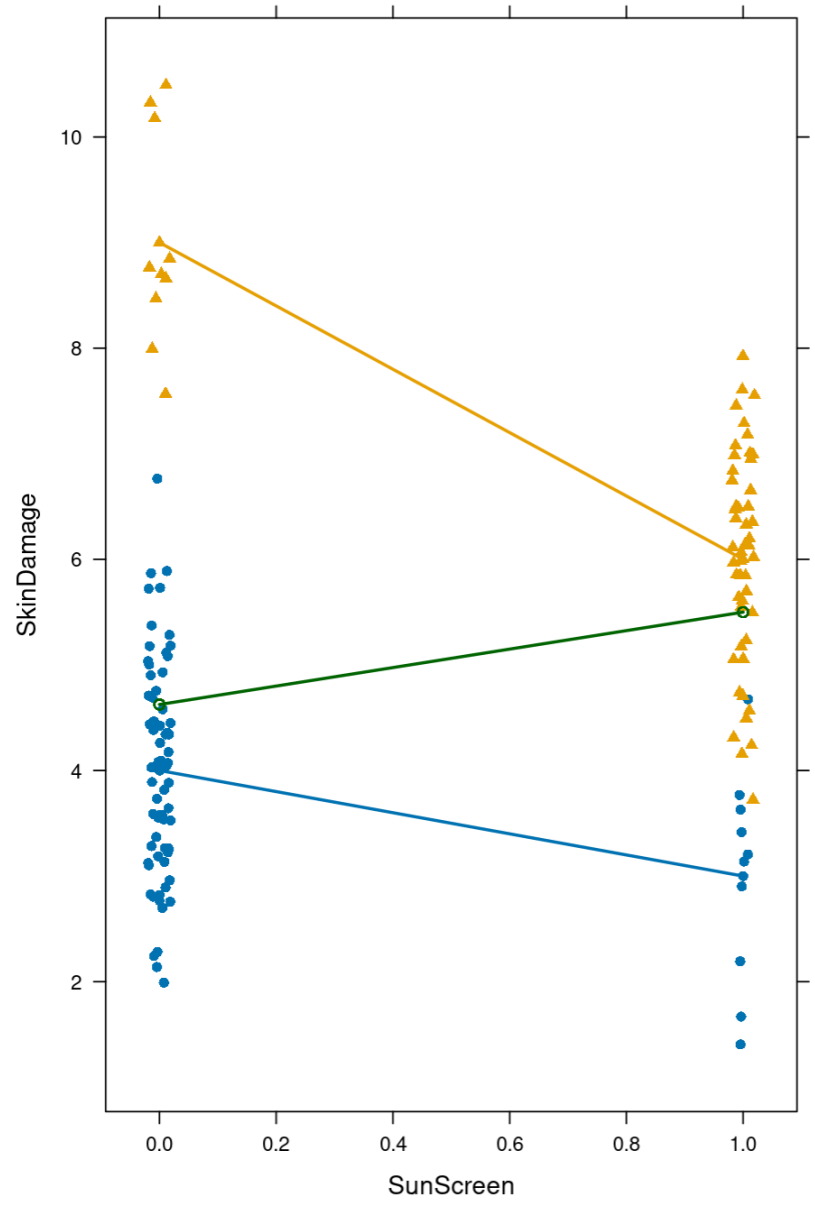
But :. Can't get CIs or p-values
. " " marginal effects

Two ways to get conditional effect CIs & p-values

1) Refit model with different - but equivalent formula
2) Use function that estimates wald tests.

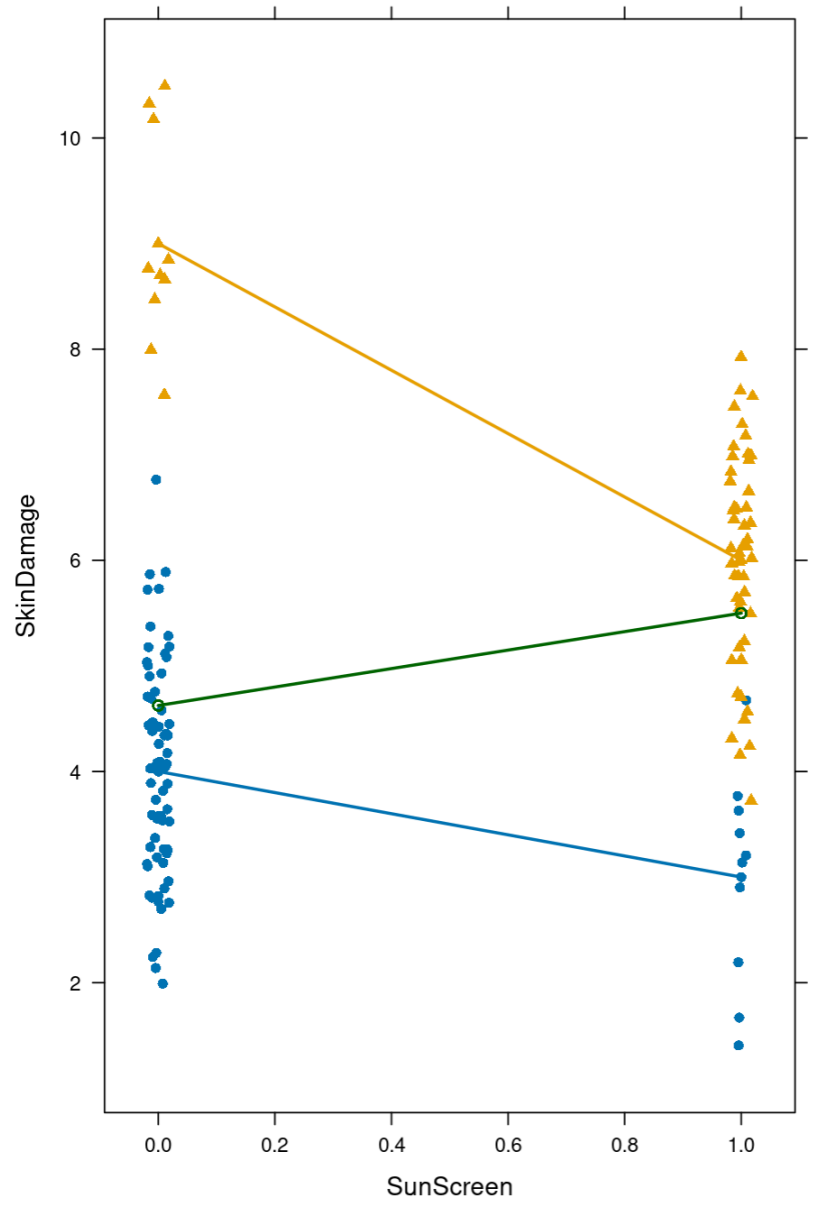
```
Call:
lm(formula = SkinDamage ~ factor(SunExposure)/SunScreen - 1,
    data = dd)

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
factor(SunExposure)0             4.0000     0.1195  33.466  < 2e-16 ***
factor(SunExposure)1             9.0000     0.3162  28.460  < 2e-16 ***
factor(SunExposure)0:SunScreen  -1.0000     0.3381  -2.958  0.00365 **
factor(SunExposure)1:SunScreen  -3.0000     0.3464  -8.660 1.21e-14 ***
```

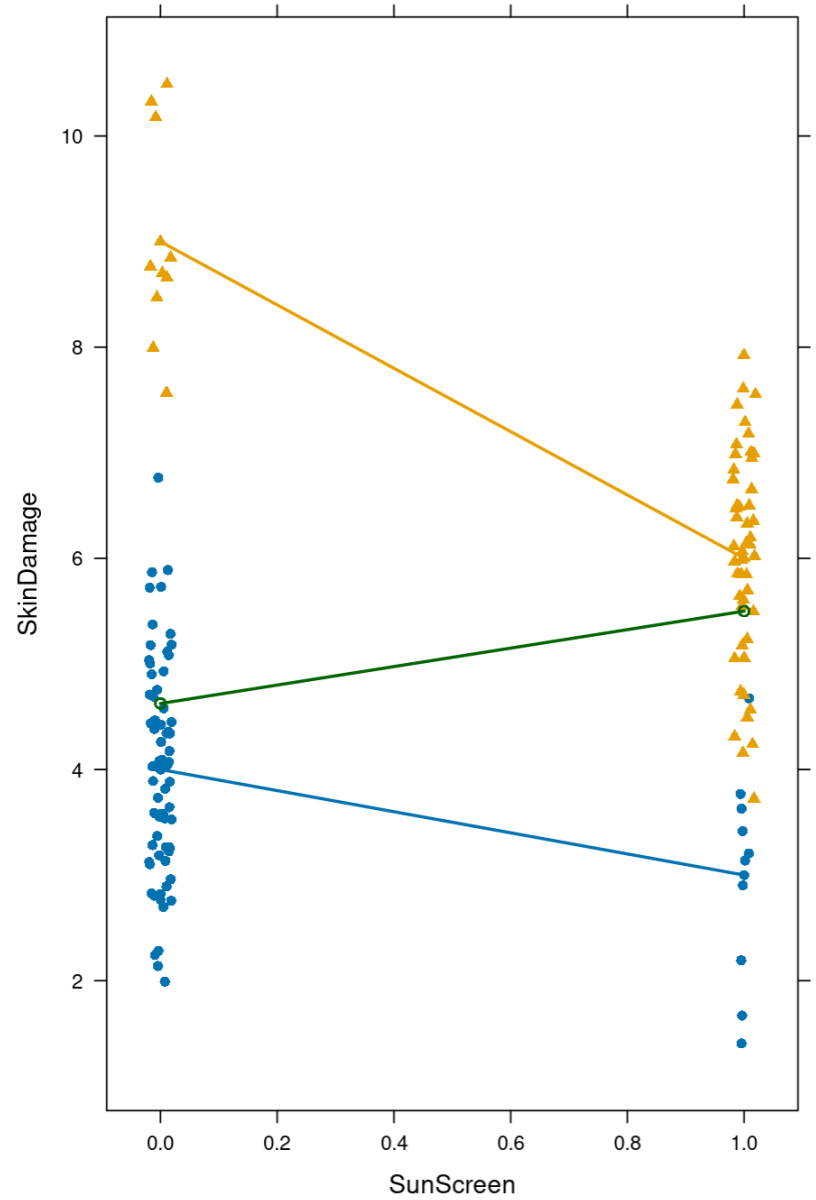
```
Call:
lm(formula = SkinDamage ~ factor(SunScreen)/SunExposure - 1,
    data = dd)

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
factor(SunScreen)0               4.0000     0.1195  33.466  < 2e-16 ***
factor(SunScreen)1               3.0000     0.3162   9.487  < 2e-16 ***
factor(SunScreen)0:SunExposure   5.0000     0.3381  14.790  < 2e-16 ***
factor(SunScreen)1:SunExposure   3.0000     0.3464   8.660 1.21e-14 ***
```



```r
fit <- lm(SkinDamage ~ SunScreen * SunExposure, dd)

summary(fit)

L <- rbind(
  "Effect of X | Z = 0" = c(0, 1, 0, 0),
  "Effect of X | Z = 1" = c(0, 1, 0, 1),
  "Effect of Z | X = 0" = c(0, 0, 1, 0),
  "Effect of X | Z = 0" = c(0, 0, 1, 1),
  "E(Y | X = 1, Z = 1)" = c(1, 1, 1, 1)  # you can keep going
)

library(spida2)
wald(fit, L)  # overall test is an F-test for all hypotheses = 0
```

```
  numDF denDF F-value p-value
1     4   136     955 <.00001
                   Estimate Std.Error DF  t-value   p-value Lower 0.95 Upper 0.95
Effect of X | Z = 0 -1       0.338062  136 -2.958040 0.00365 -1.668538  -0.331462
Effect of X | Z = 1 -3       0.346410  136 -8.660254 <.00001 -3.685047  -2.314953
Effect of Z | X = 0  5       0.338062  136 14.790199 <.00001  4.331462   5.668538
Effect of X | Z = 0  3       0.346410  136  8.660254 <.00001  2.314953   3.685047
E(Y | X = 1, Z = 1)  6       0.141421  136 42.426407 <.00001  5.720331   6.279669
```




SkinDamage
10
8
6
4
2
0.0
0.2
0.4
0.6
0.8
1.0
SunScreen

The Liu-Meng Diagram: The main effect of Z

You could always flip the roles of X and Z and draw a Paik-Agresti diagram for the main effect of Z.

But that wouldn't show the two in the same diagram.

The Liu-Meng diagram (personal communication over lunch with Xiao-Li Meng, Liu was his undergraduate student who worked on the idea with Xiao-Li. Unfortunately I don't know Liu's given name but someday I must find out.

1) Plot the points:
   - (mean of X, mean of Y)| Z = 0
     - this is a weighted mean of the
       * (mean of X, mean of Y)| Z = 0, X = 0, and
       * (mean of X, mean of Y)| Z = 0, X = 1 Thus, it lies in the convex hull of these two points, which is the line joining the two points!
   - (mean of X, mean of Y)| Z = 1
     - this is a weighted mean of the
       * (mean of X, mean of Y)| Z = 1, X = 0, and
       * (mean of X, mean of Y)| Z = 1, X = 1 Similarly, it also on the line joining these two points.
   - The vertical distance between these two points is the main effect of Z on Y.
   - The horizontal distance between these two points is the main effect of Z on X.

Small project: Write a function that produces a Liu-Meng diagram.

R makes complicated things easy .... and sometimes easy things complicated:

```
ddZ <- spida2::up(dd, ~ SunExposure, agg = ~ SkinDamage + SunScreen)
ddZ
```

```
##   SunExposure Freq SkinDamage SunScreen
## 0           0   80      3.875 0.1250000
## 1           1   60      6.500 0.8333333
```

```
pall +
  xyplot(SkinDamage ~ SunScreen, ddZ, type = 'b', lwd = 2, pch = 16, cex = 2)
```

The Liu-Meng Diagram: The main effect of Z

You could always flip the roles of X and Z and draw a Paik-Agresti diagram for the main effect of Z.

But that wouldn't show the two in the same diagram.

The Liu-Meng diagram (personal communication over lunch with Xiao-Li Meng, Liu was his undergraduate student who worked on the idea with Xiao-Li. Unfortunately I don't know Liu's given name but someday I must find out.

1) Plot the points:
   - (mean of X, mean of Y)| Z = 0
     - this is a weighted mean of the
       * (mean of X, mean of Y)| Z = 0, X = 0, and
       * (mean of X, mean of Y)| Z = 0, X = 1 Thus, it lies in the convex hull of these two points, which is the line joining the two points!
   - (mean of X, mean of Y)| Z = 1
     - this is a weighted mean of the
       * (mean of X, mean of Y)| Z = 1, X = 0, and
       * (mean of X, mean of Y)| Z = 1, X = 1 Similarly, it also on the line joining these two points.
   - The vertical distance between these two points is the main effect of Z on Y.
   - The horizontal distance between these two points is the main effect of Z on X.

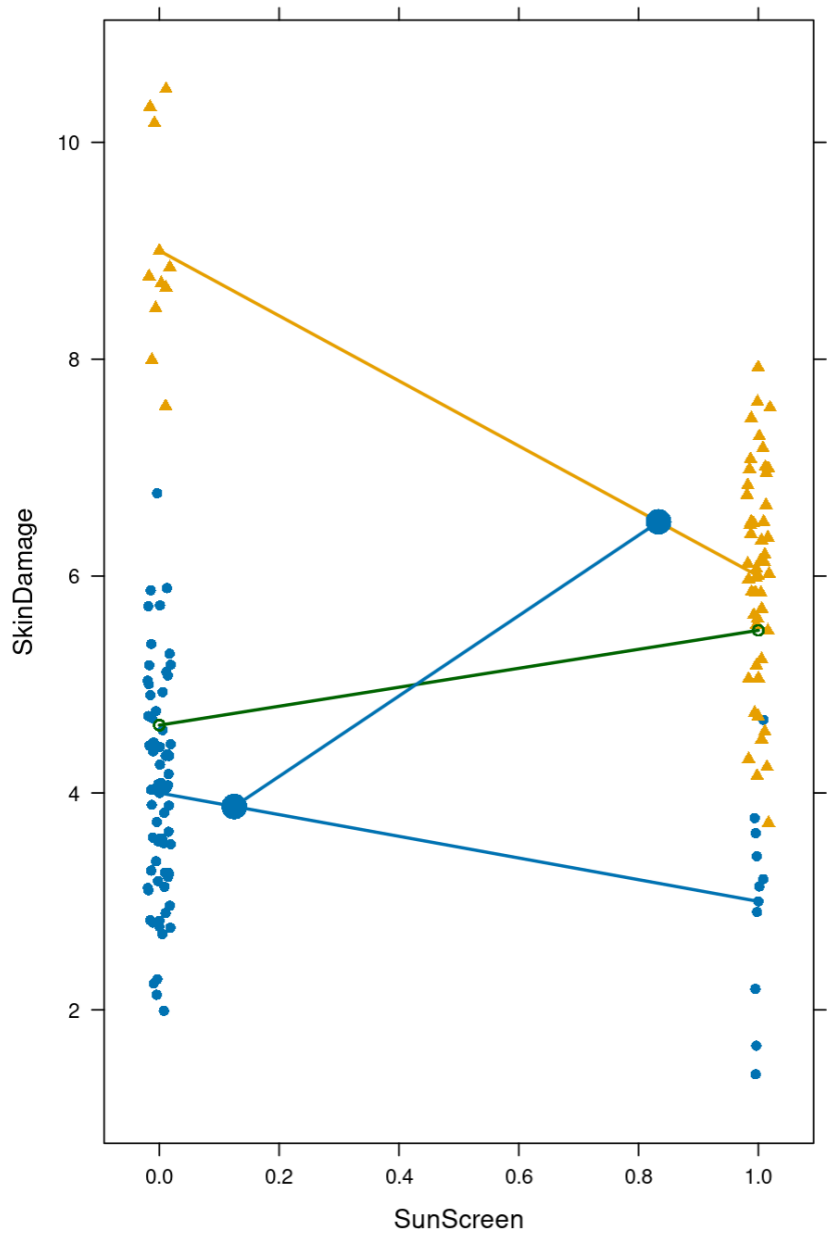
Small project: Write a function that produces a Liu-Meng diagram.

R makes complicated things easy .... and sometimes easy things complicated:

```
ddZ <- spida2::up(dd, ~ SunExposure, agg = ~ SkinDamage + SunScreen)
ddZ
```

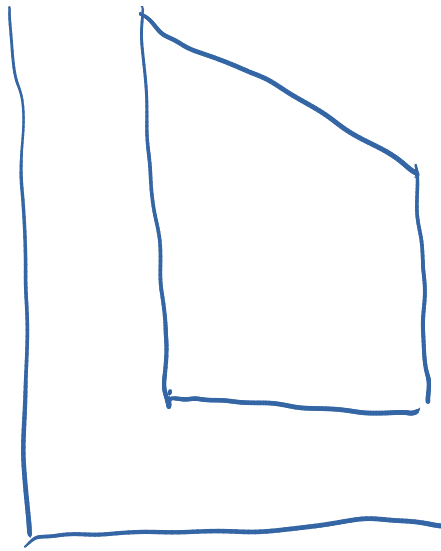
```
##   SunExposure Freq SkinDamage SunScreen
## 0           0   80      3.875 0.1250000
## 1           1   60      6.500 0.8333333
```

```
pall +
  xyplot(SkinDamage ~ SunScreen, ddZ, type = 'b', lwd = 2, pch = 16, cex = 2)
```



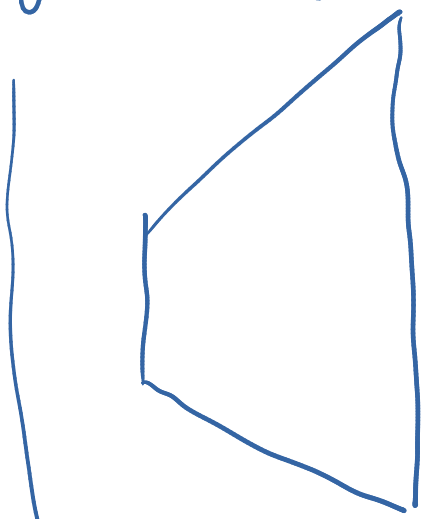
Implications of P-A & L-XX diagrams

Start with conditional effects
Draw the Trapezoid of Means

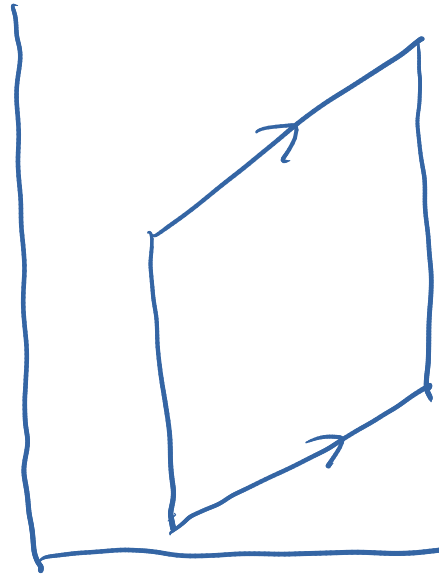


CXXs

- Same slope
- no X
- Interaction
- Removable

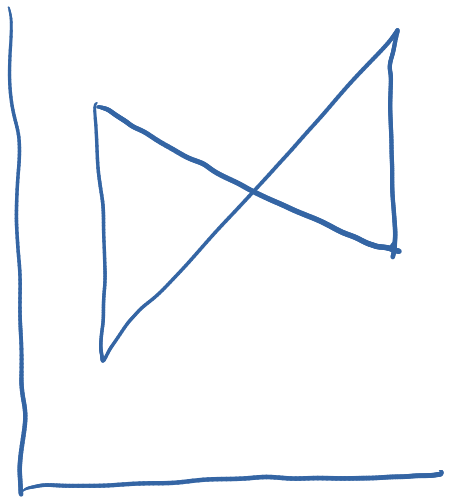


- Diff slopes
- no X
- Int.
- not removable



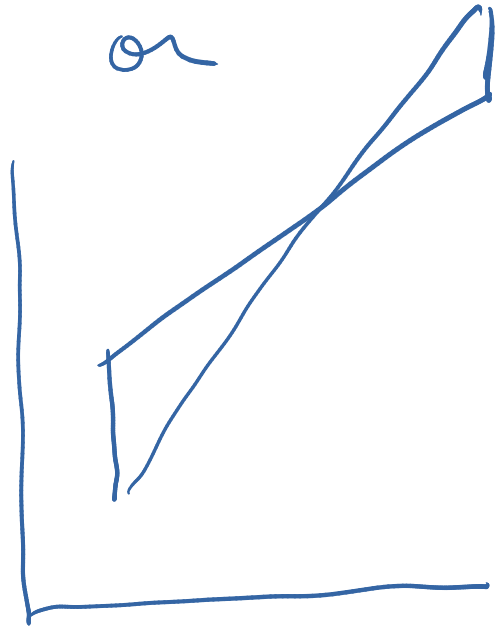
- parallel slopes
- no Int
- already removed

or



- diff slopes
- X
- Int
- not removable

or



- same slopes
- X
- Int
- not remov.

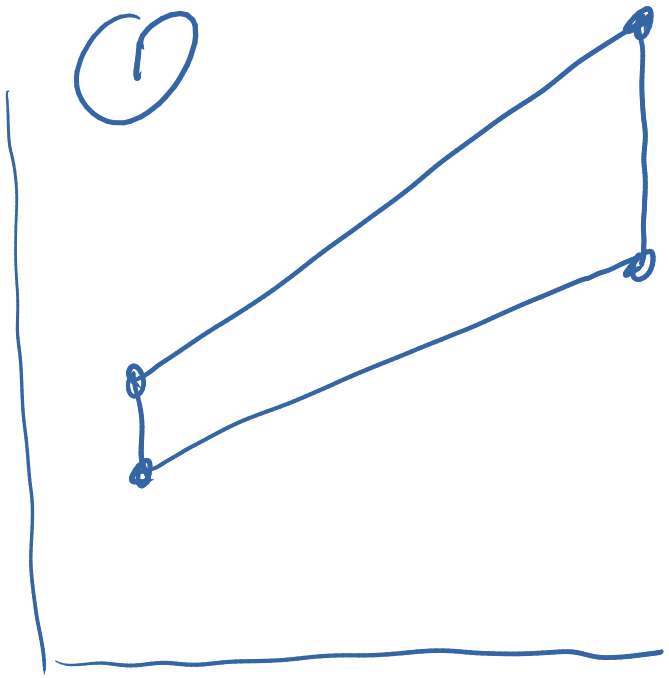
- Shows "interaction" can mean many different things
- Not enough to just say "there's interaction"
- You must display and describe the response function.

When can we have Simpson's Paradox for X?

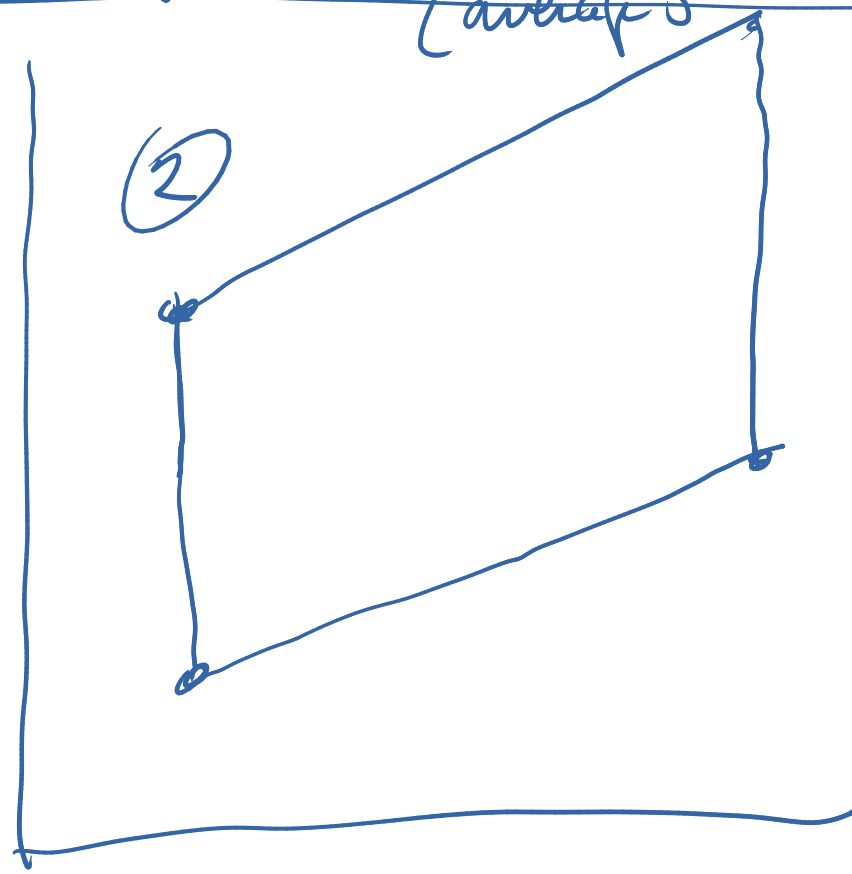
SP_X = marginal has different sign than {all / average} conditionals

what if:

(1)

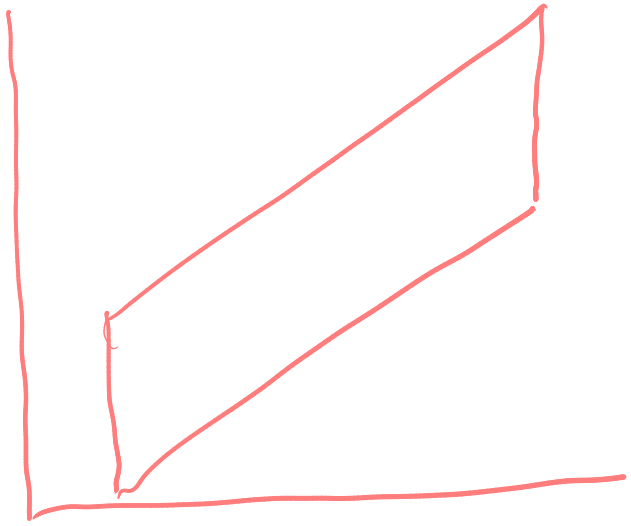


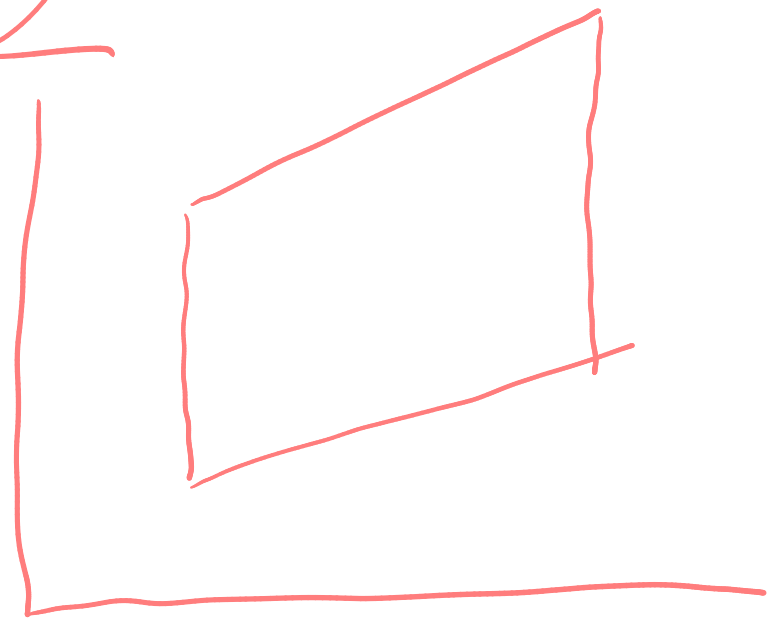
(2)



Can only have SP_x for x if there's a horizontal separation of the top and bottom edges of the ToM & slopes in some direction.

What about SP_x for Z





Take away - Can have SP_X for at most one of X & Z

But if conditionals in same direction then must have the possibility of SP_X for at least one except in special case: where diagonally opposite points have same Y.