201ab Quantitative methods
L.10: Multiple regression

Ed Vul | UCSD Psychology

With great illustrations from Julian Parris.

Good news!

No dealing directly with
estimation equations/calculations directly.

(it's impractical here on out)

Bad news!

From now on, getting an answer from R is much easier
than understanding what question to ask
(or which answer corresponds to which question).

Multiple regression agenda

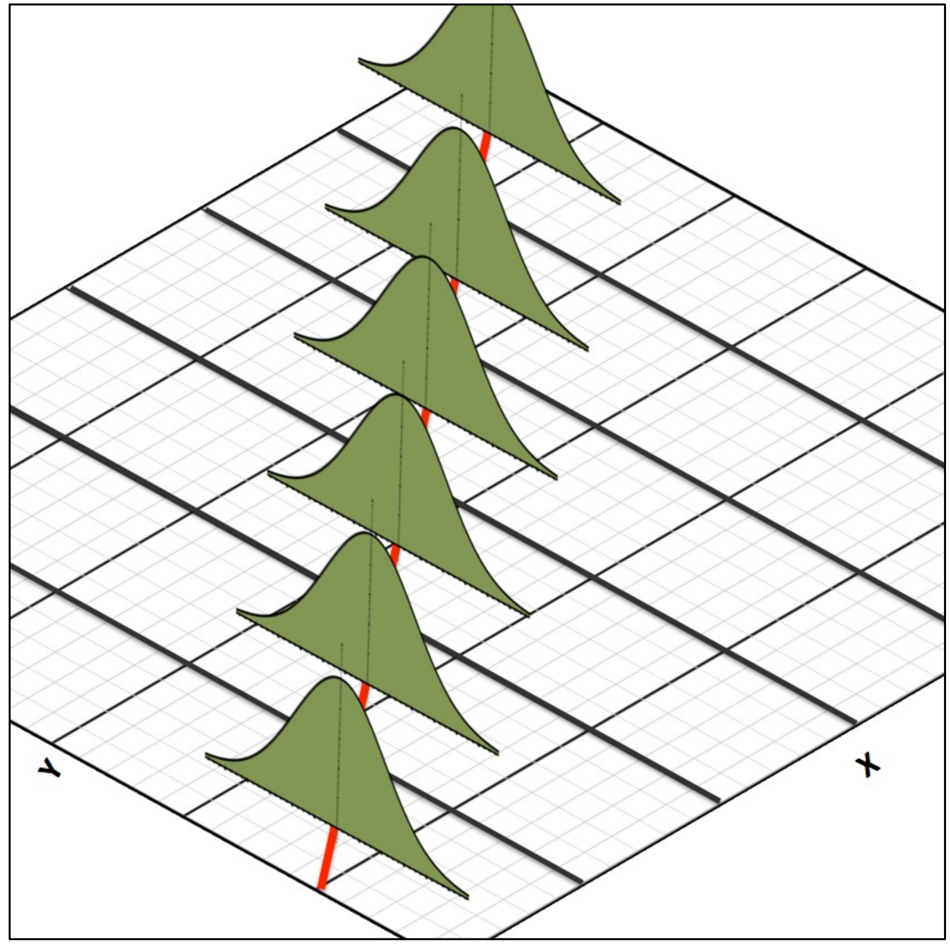
- What is it? And why do this?
- Multicolinearity & its consequences
- Sums of squares partitioning in multiple regression
- Different hypothesis tests in multiple regression
- Nested model comparison
- Non-nested models

Single predictor regression model

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_i} + \boxed{\varepsilon_i}$$

Score on Y for the ith individual = Y Intercept + (Slope *(Effect)* × Score on X for the ith individual) + Error

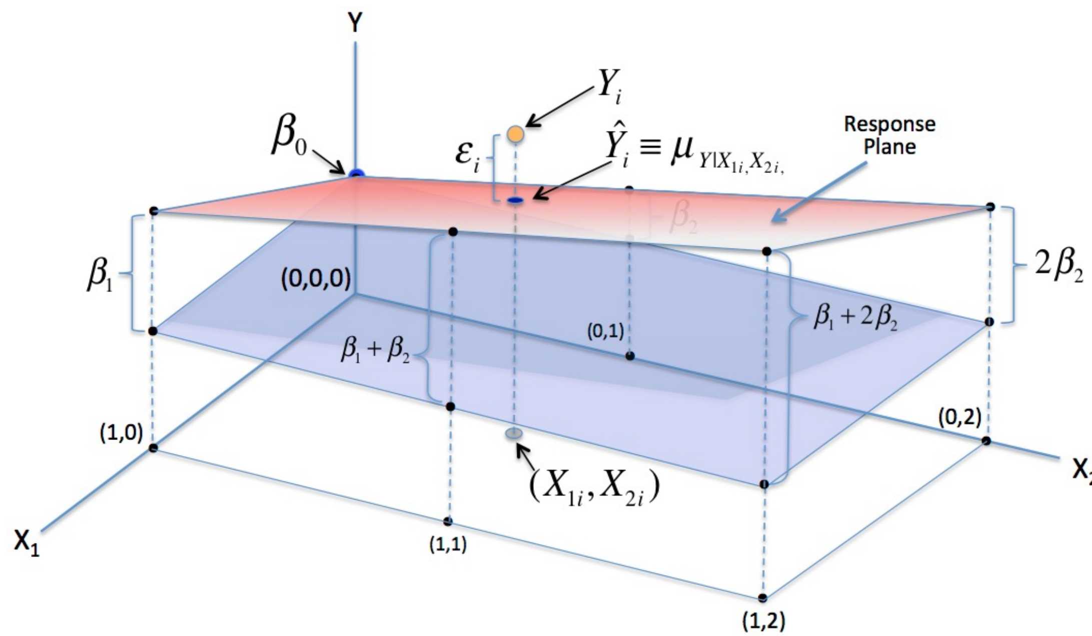
$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$



Height of daughter = Height of daughter whose mother is 0 inches tall + (Extra inches gained by daughter per inch in height of mother × Height of mother) + Other factors contributing to daughter's height plus measurement error

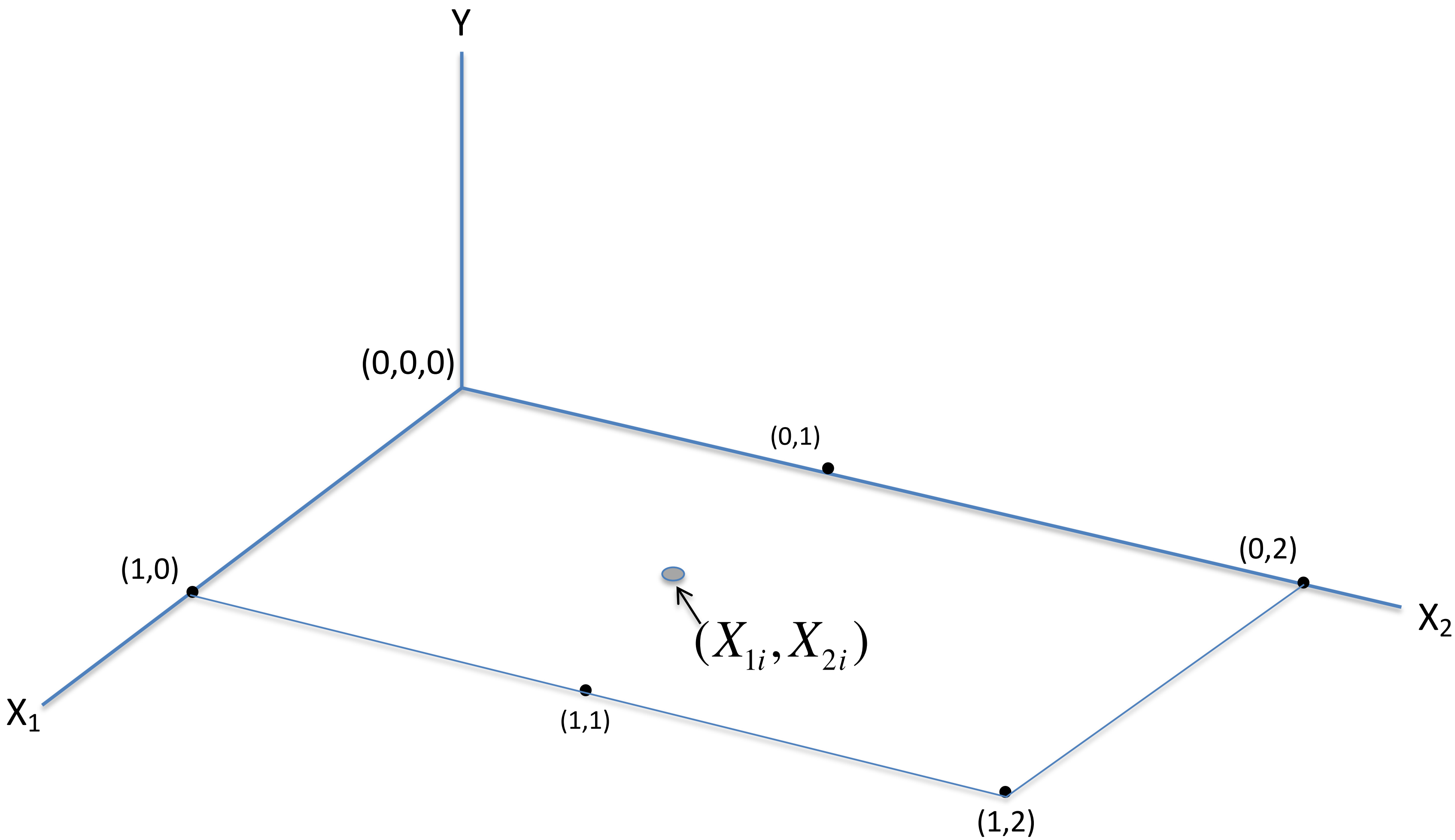
Two Predictor Regression Model

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i} \qquad \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$



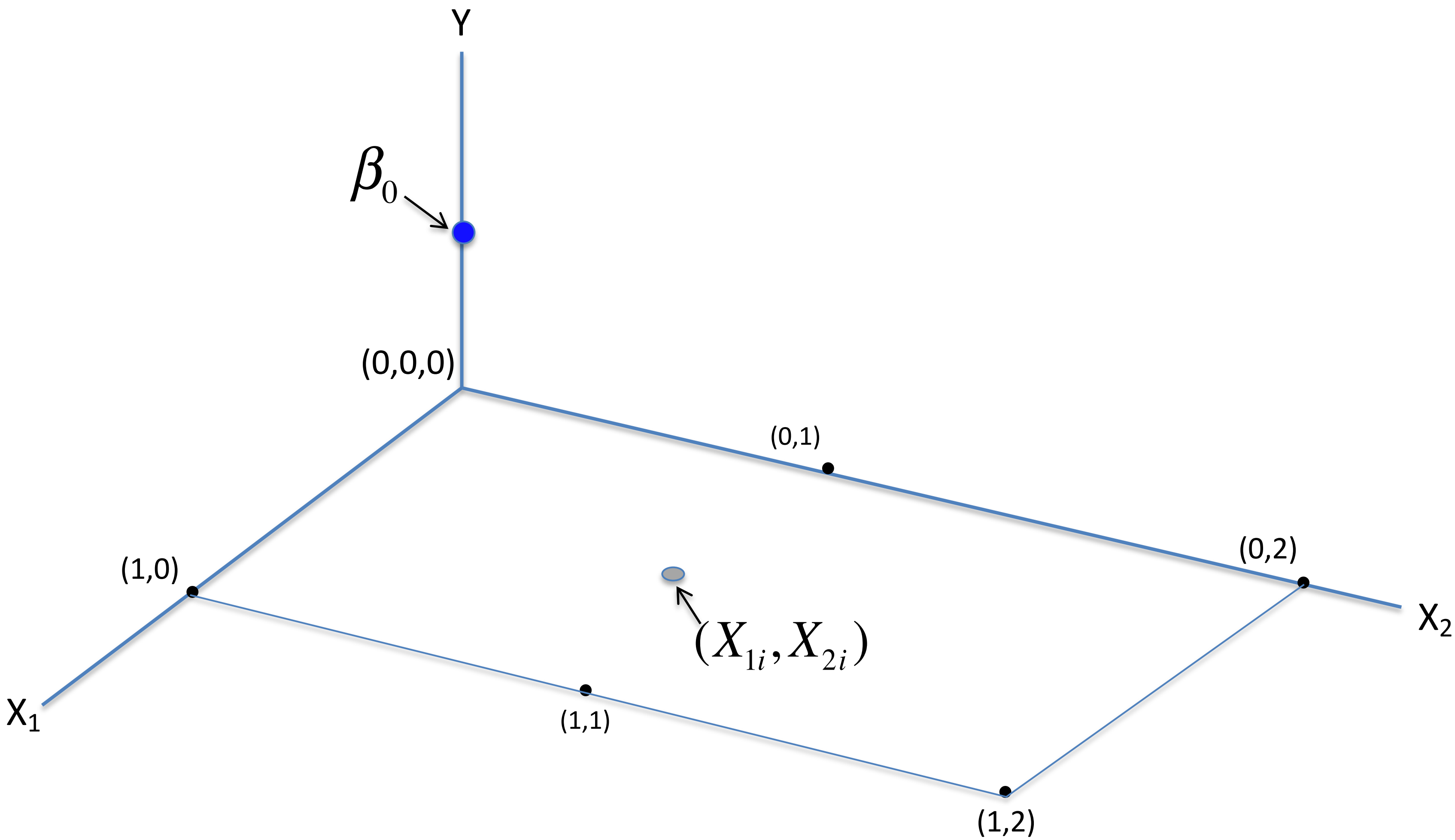
Height of daughter = Height of daughter whose parents are 0 inches tall + (Inches gained by daughter per inch in height of mother × Height of mother) + (Inches gained by daughter per inch in height of father × Height of father) + Other factors contributing to daughter's height plus measurement error

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



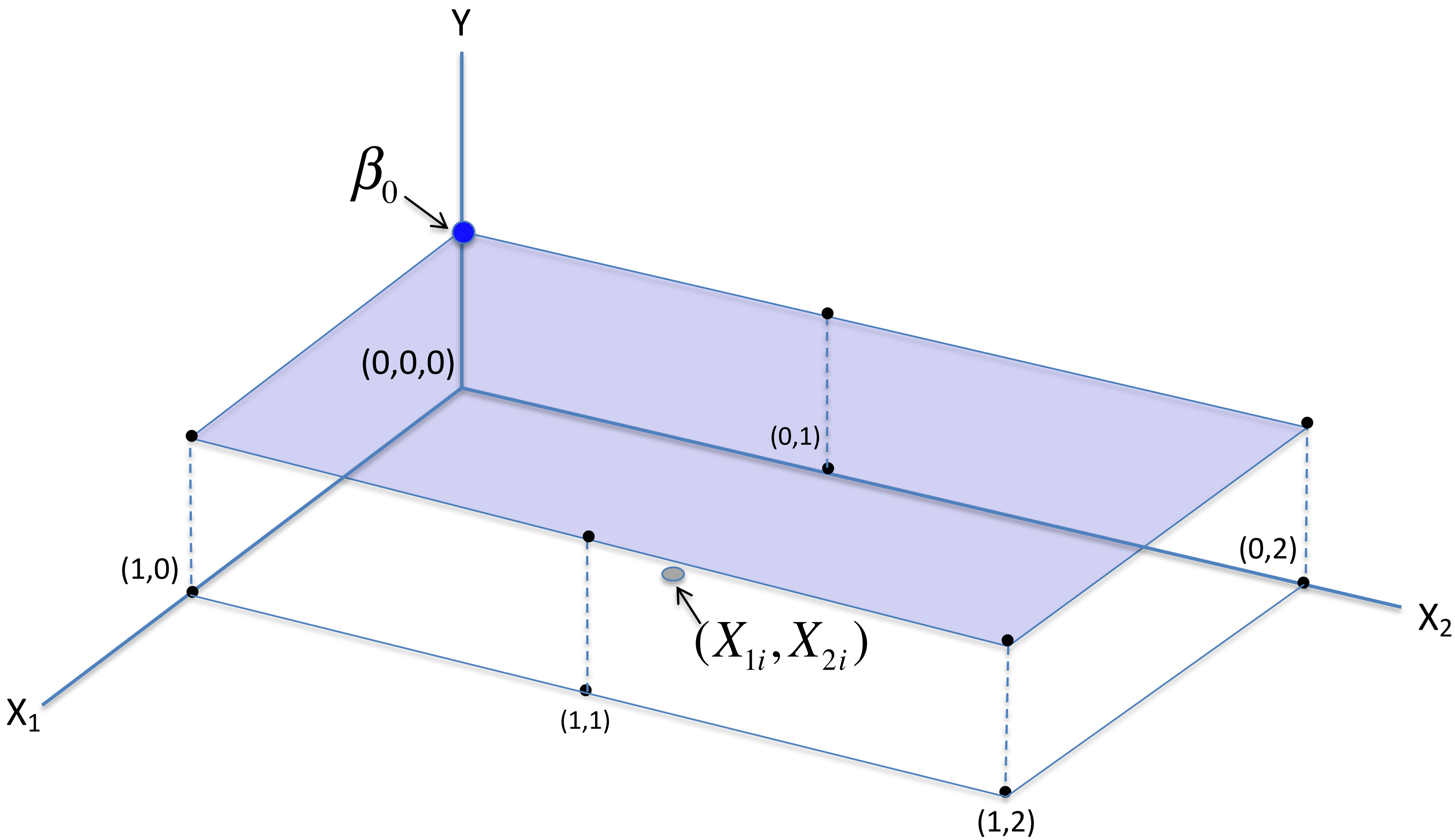
FROM JULIAN PARRIS

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



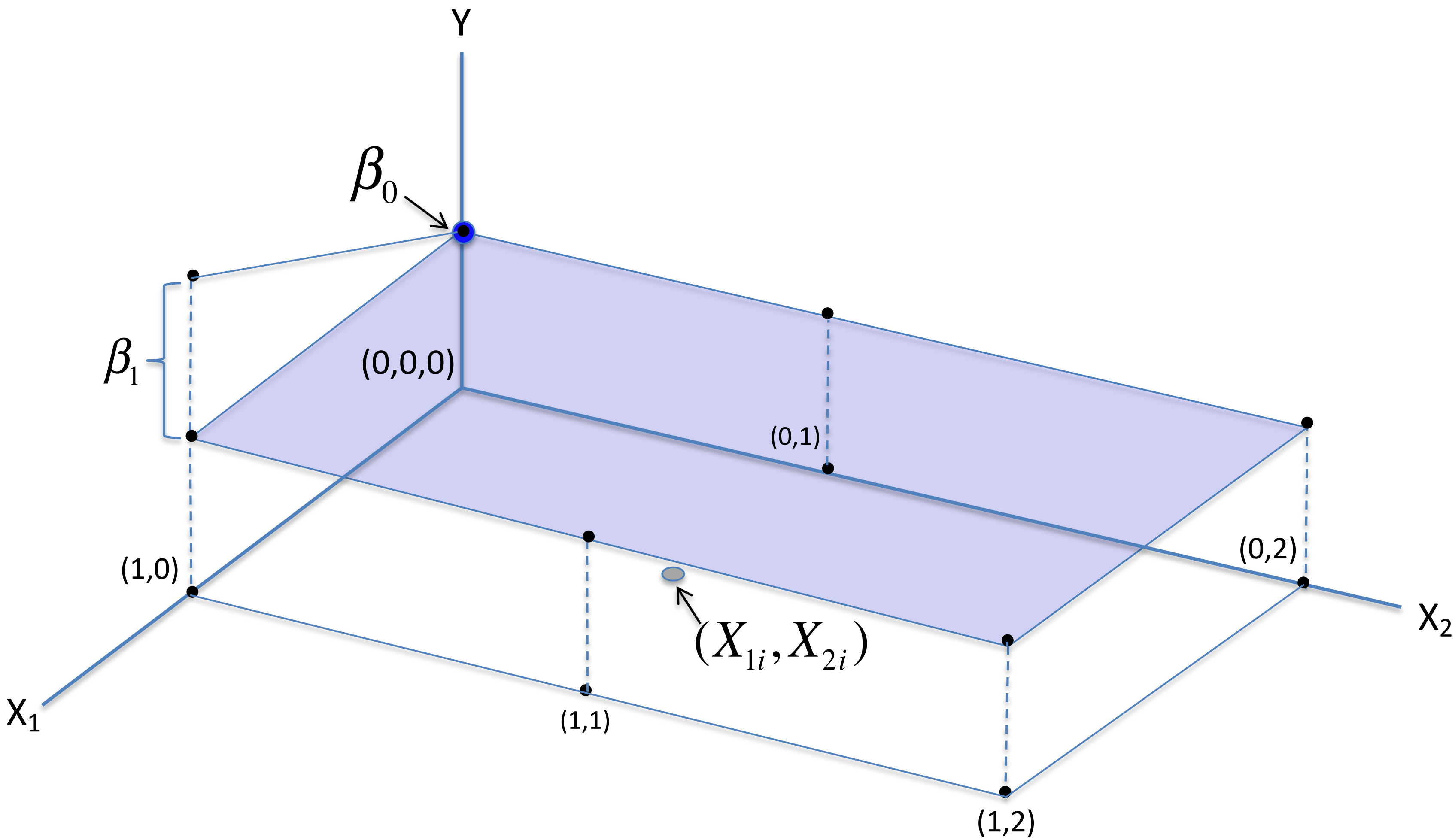
FROM JULIAN PARRIS

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



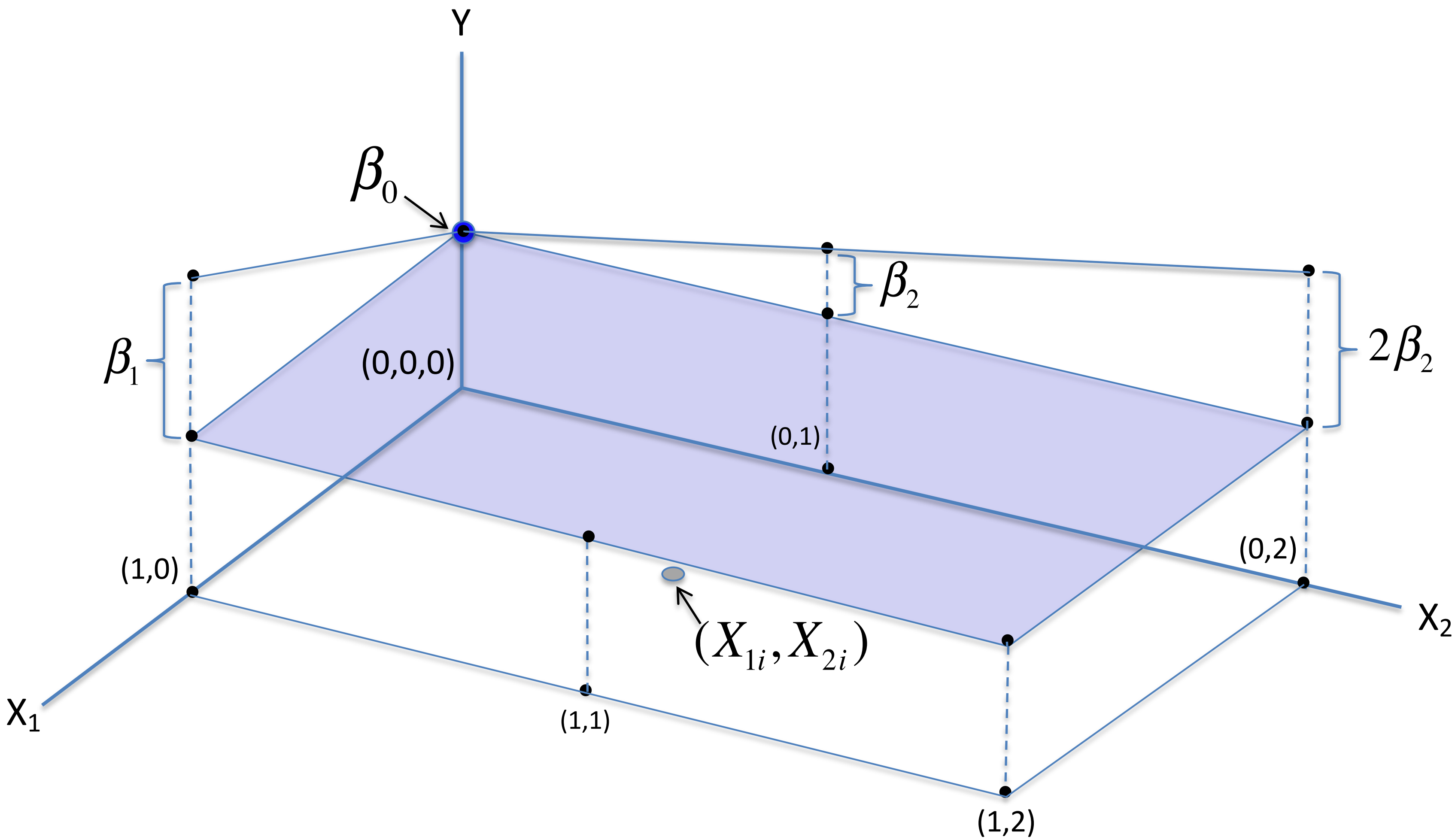
FROM JULIAN PARRIS

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



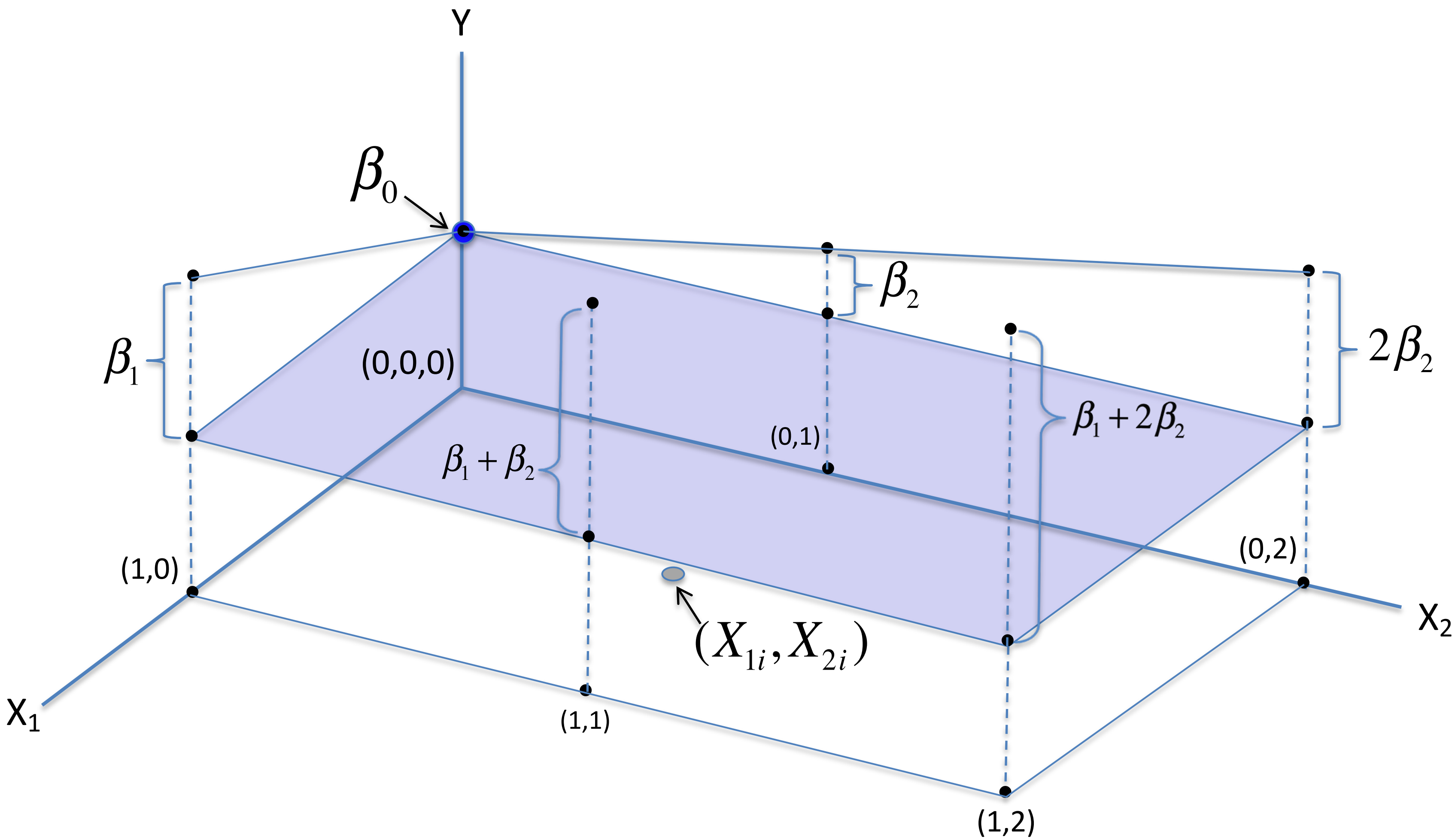
FROM JULIAN PARRIS

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



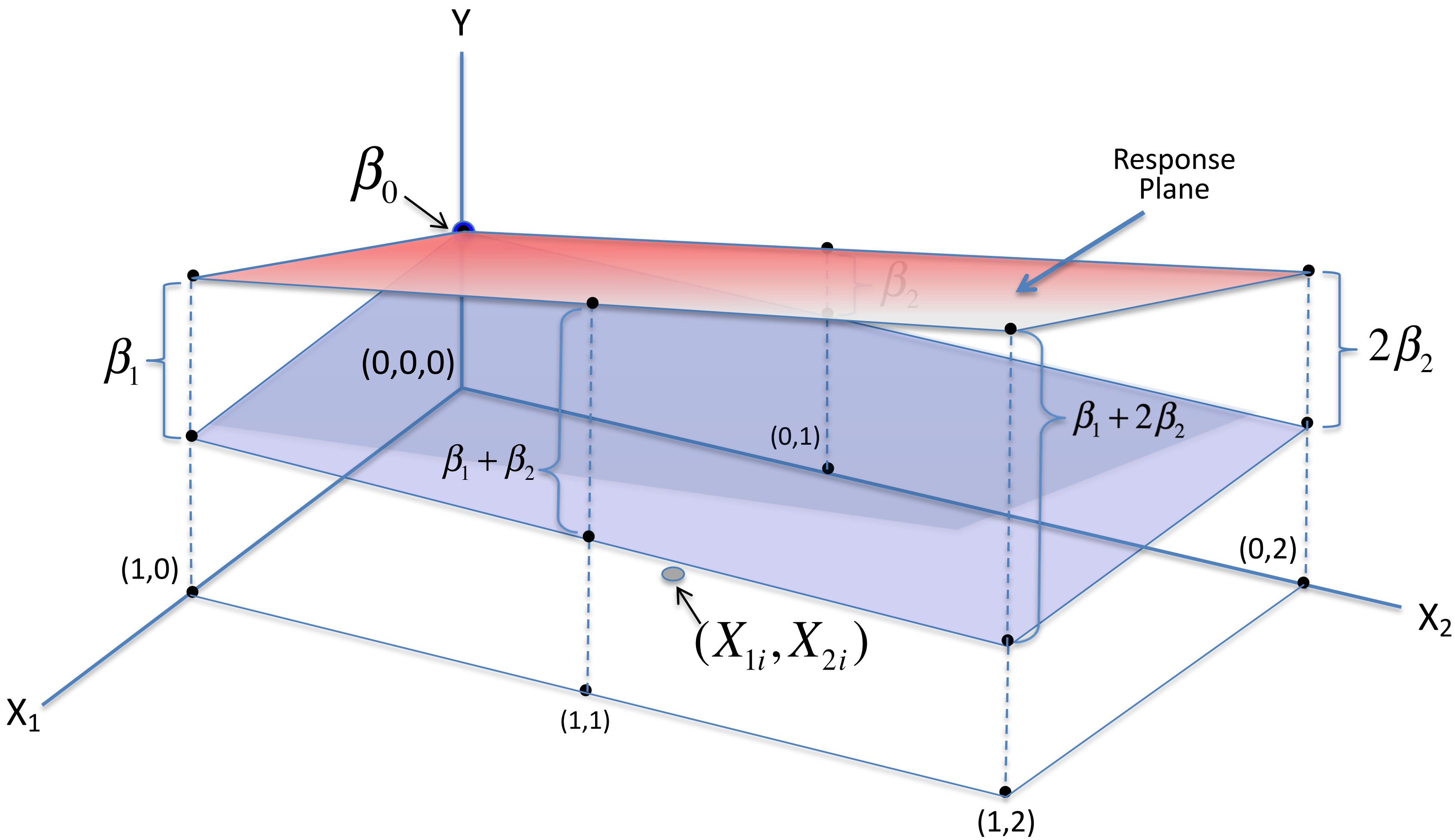
FROM JULIAN PARRIS

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$

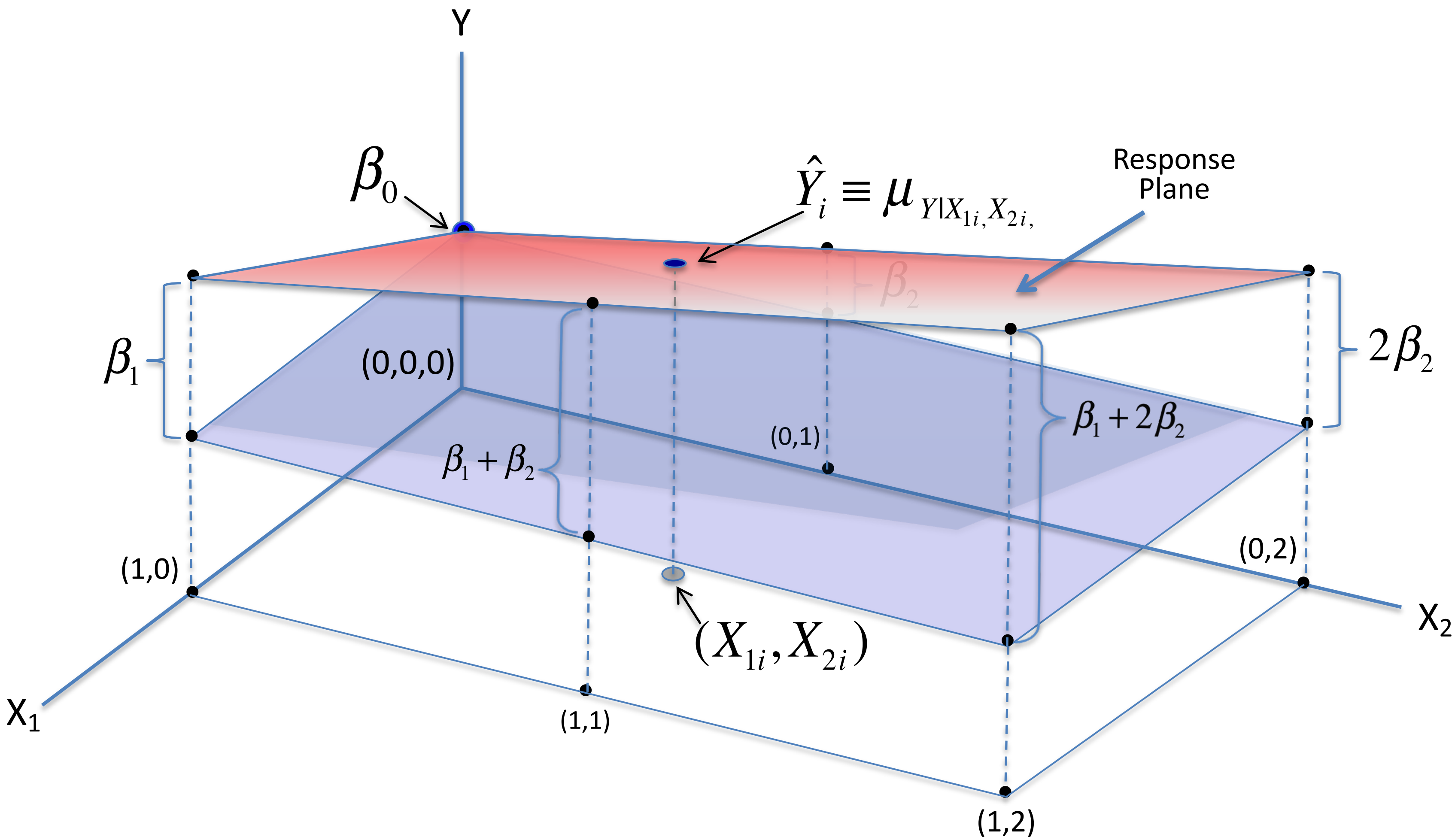


FROM JULIAN PARRIS

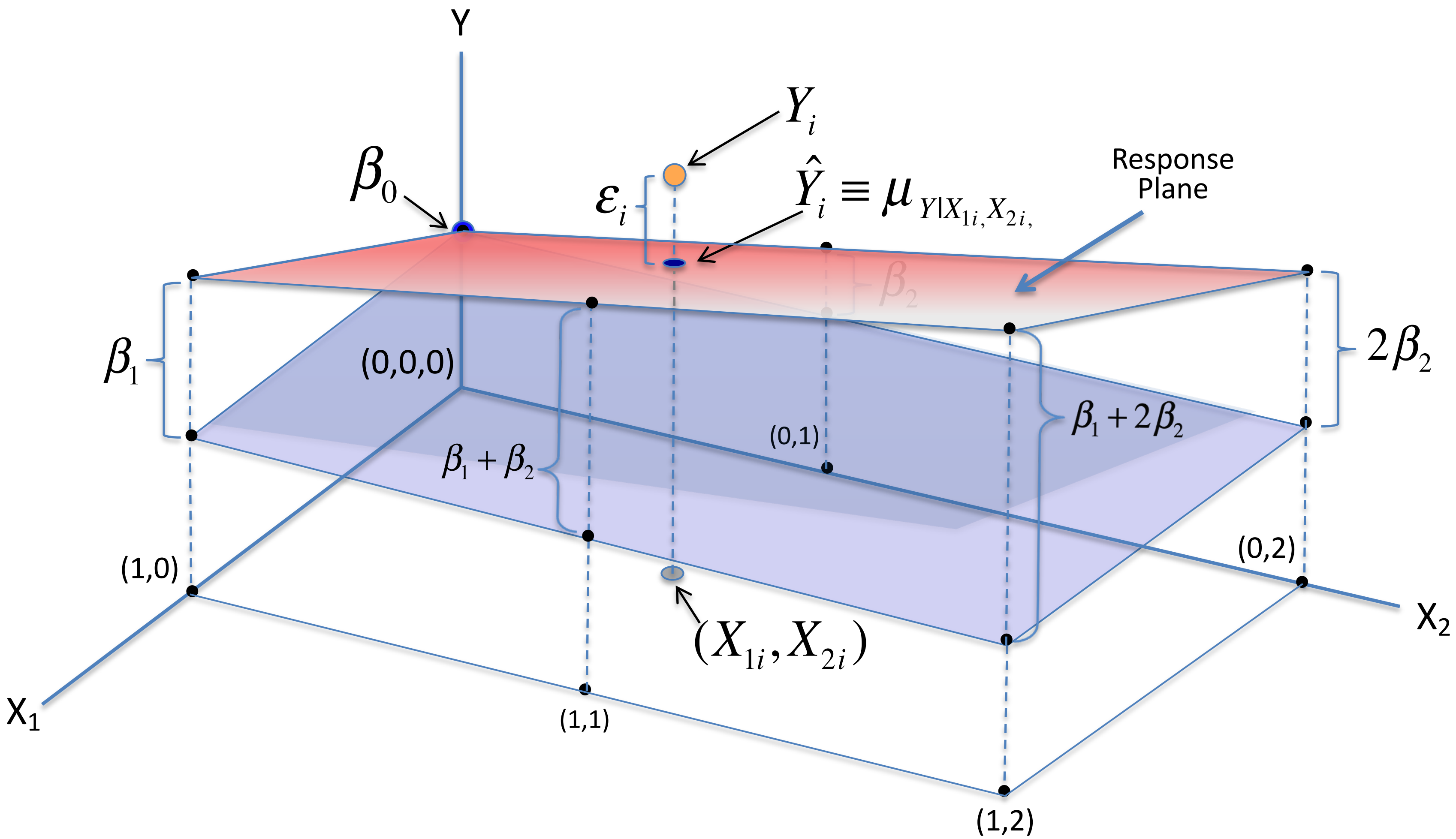
$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i}$$



Partial Regression coefficient

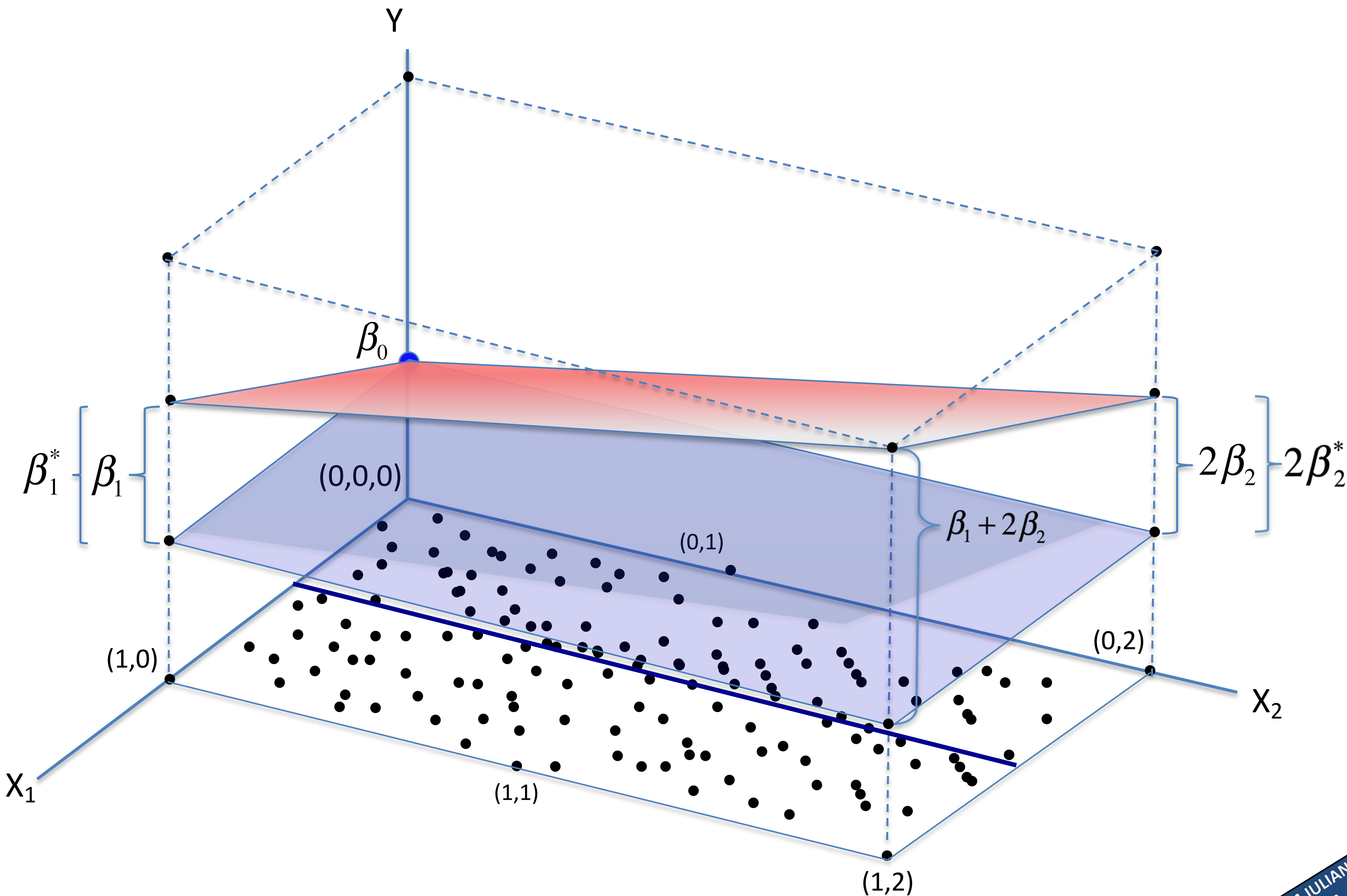
- Slope estimates are "partial regression coefficients": the partial effect of one variable with the others held constant.
  - b_1: increase in Y per unit increase in X_1, *all else constant**
  - E.g., how many inches taller will a daughter be if a mother was 1" taller, while keeping the father the same height.

  * "all else constant" is often not plausible

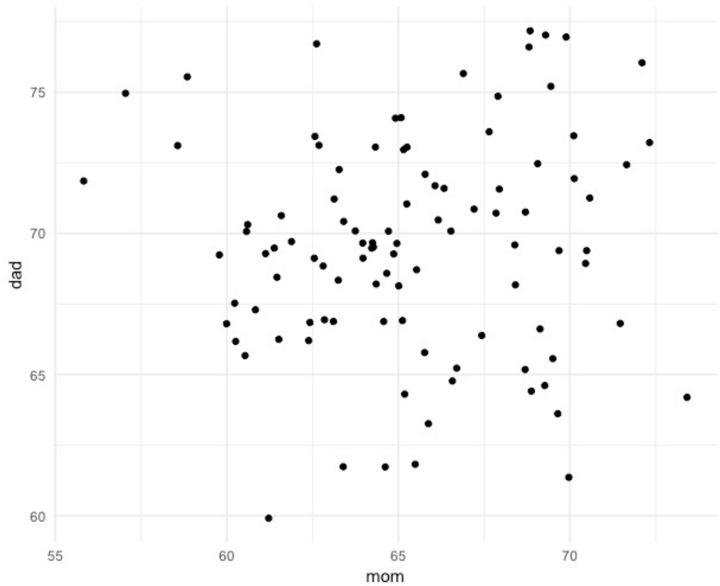
Multiple regression agenda

- What is it?
- Why do this?
  - More complete model:
    better predictions from conjunctions of variables
    less residual error
  - *Assign credit to multiple predictors*
    Estimate effect of one variable while "statistically controlling" for others

Uncorrelated predictors



FROM JULIAN PARRIS




```
summary(lm(daughter~mom))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 34.17132    6.30603   5.419  4.3e-07
mom          0.48826    0.09636   5.067  1.9e-06

Residual standard error: 3.49 on 98 degrees of freedom
Multiple R-squared:  0.2076
```

```
summary(lm(daughter~dad))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  26.8399     5.9418   4.517 1.75e-05
dad           0.5641     0.0853   6.613 1.99e-09

Residual standard error: 3.26 on 98 degrees of freedom
Multiple R-squared:  0.3086,
```

```
summary(lm(daughter~mom+dad))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.94594    7.02683  -0.135    0.893
mom          0.45337    0.07809   5.806 8.09e-08
dad          0.53768    0.07400   7.266 9.41e-11

Residual standard error: 2.823 on 97 degrees of freedom
Multiple R-squared:  0.4869,
```

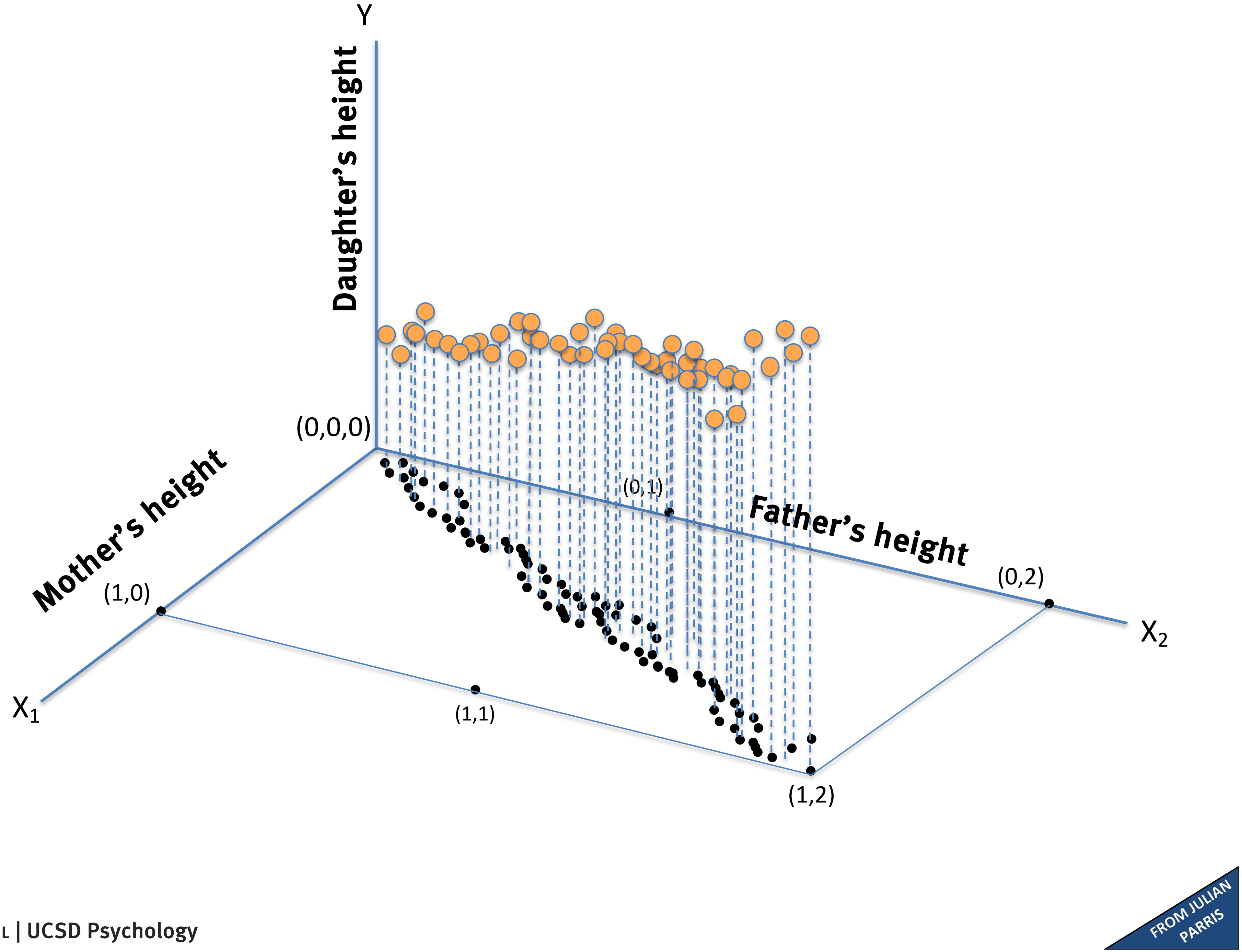
Mom and dad are uncorrelated, they explain different variability, so we lower our residual sd, increase our R^2, and get more precise estimates of the coefficients.

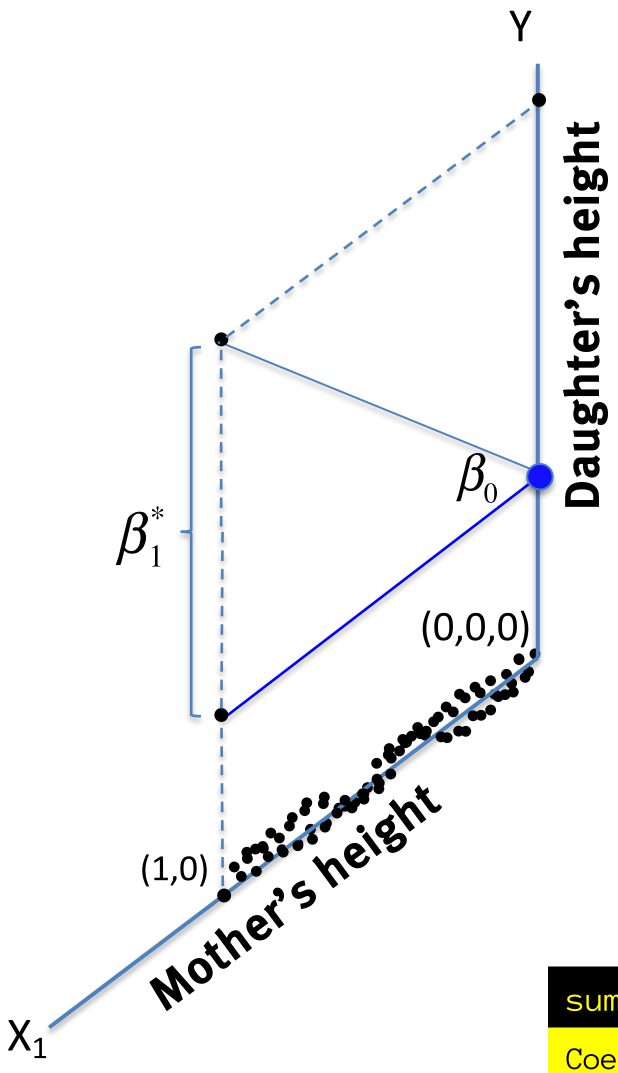
More complete model with uncorrelated predictors is a win all around.

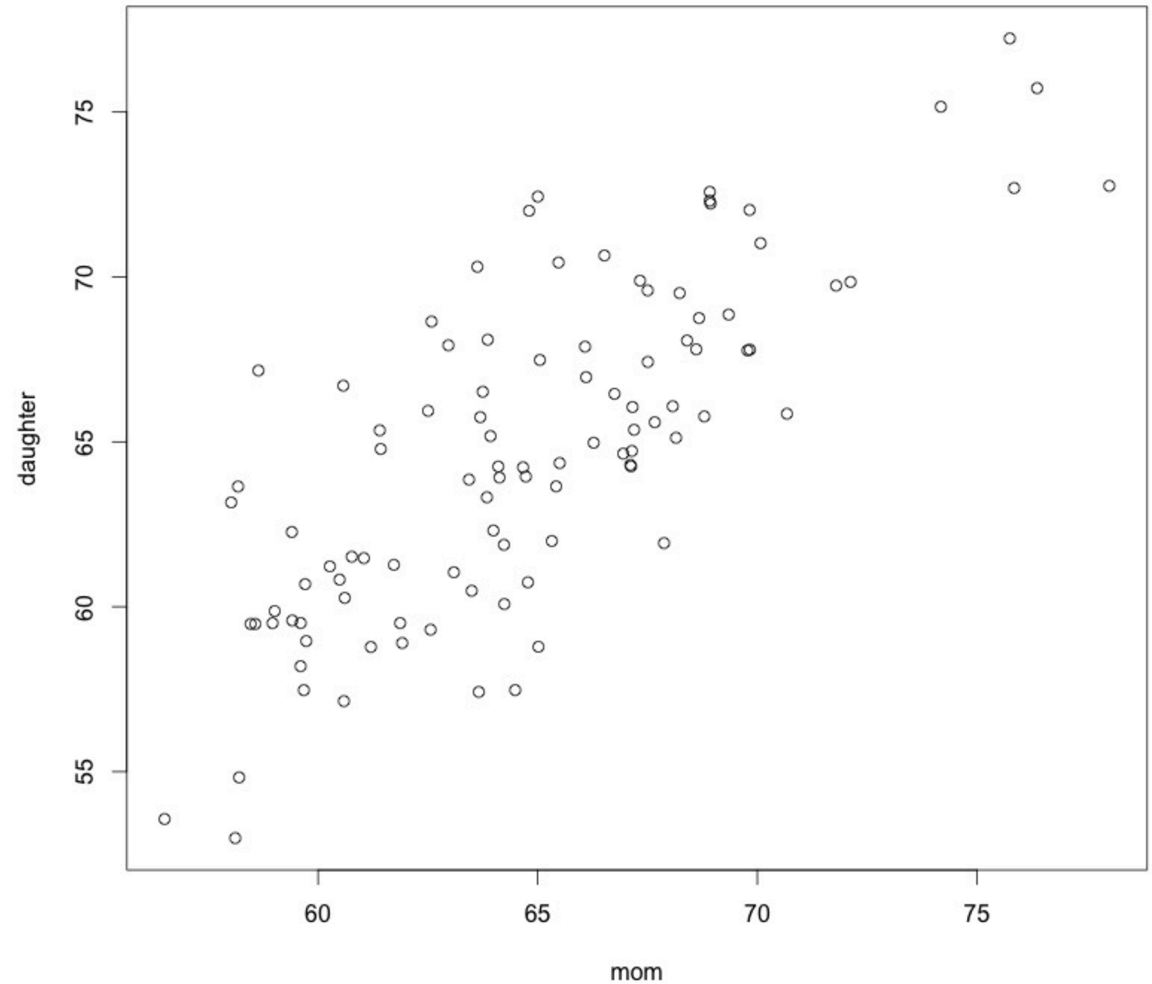
Things get trickier when predictors are correlated.

Multiple regression agenda

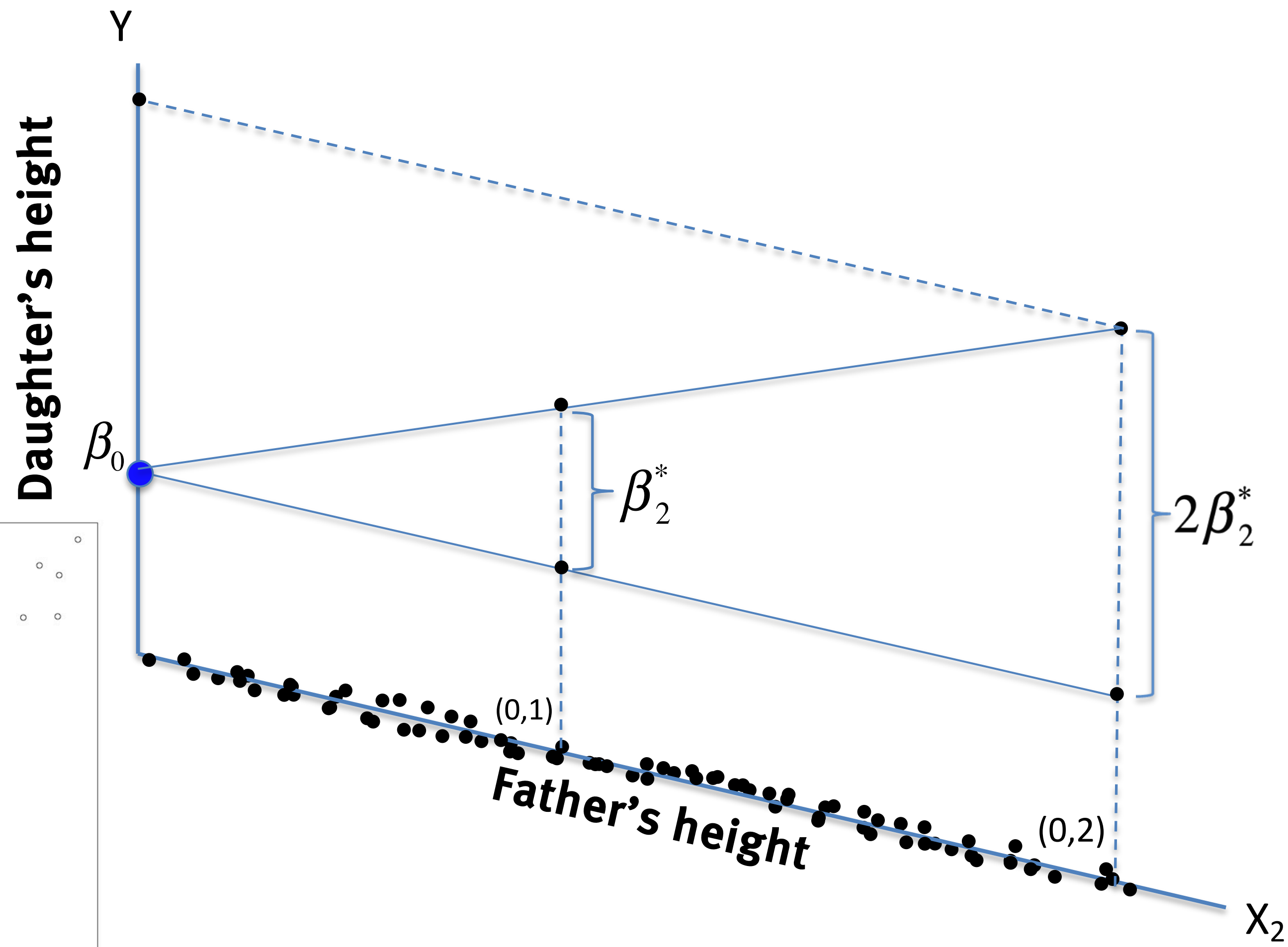
- What is it? And why do this?
- Multicolinearity & its consequences
- Sums of squares partitioning in multiple regression
- Different hypothesis tests in multiple regression
- Nested model comparison
- Non-nested models

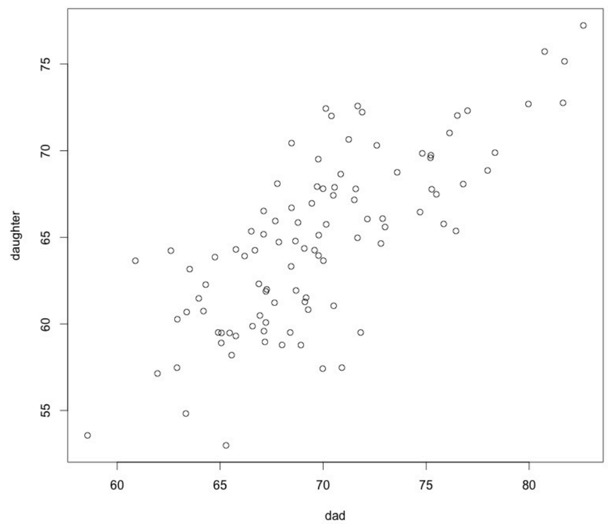

Y
Daughter's height
(0,0,0)
Mother's height
(1,0)
X1
(1,1)
(0,1)
Father's height
(0,2)
X2
(1,2)
FROM JULIAN PARRIS





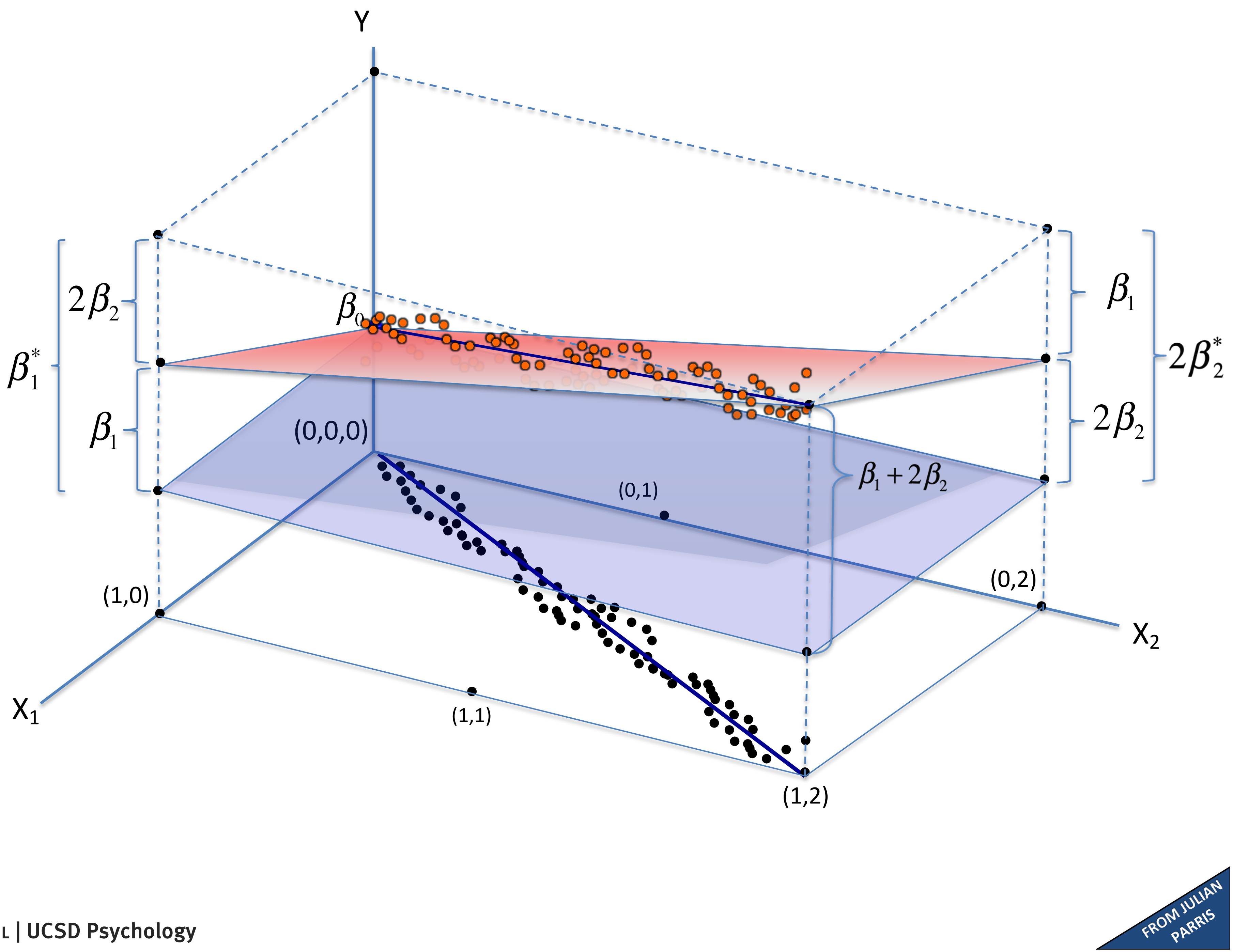

```
summary(lm(daughter~mom))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.88151    4.69354   1.892   0.0614 .
mom          0.86209    0.07223  11.936   <2e-16 ***
```



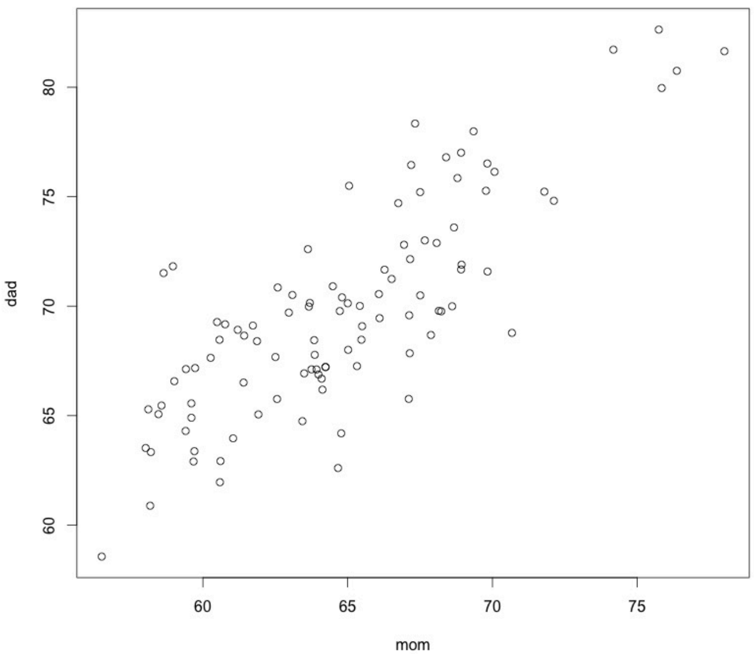



```
summary(lm(daughter~dad))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.26150    4.82647   2.126    0.036 *
dad          0.78125    0.06901  11.321   <2e-16 ***
```


FROM JULIAN PARRIS



Y
$2\beta_2$
β_1^*
β_1
β_0
β_1
$2\beta_2^*$
$2\beta_2$
$\beta_1 + 2\beta_2$
(0,0,0)
(0,1)
(1,0)
(0,2)
X_2
X_1
(1,1)
(1,2)
FROM JULIAN PARRIS



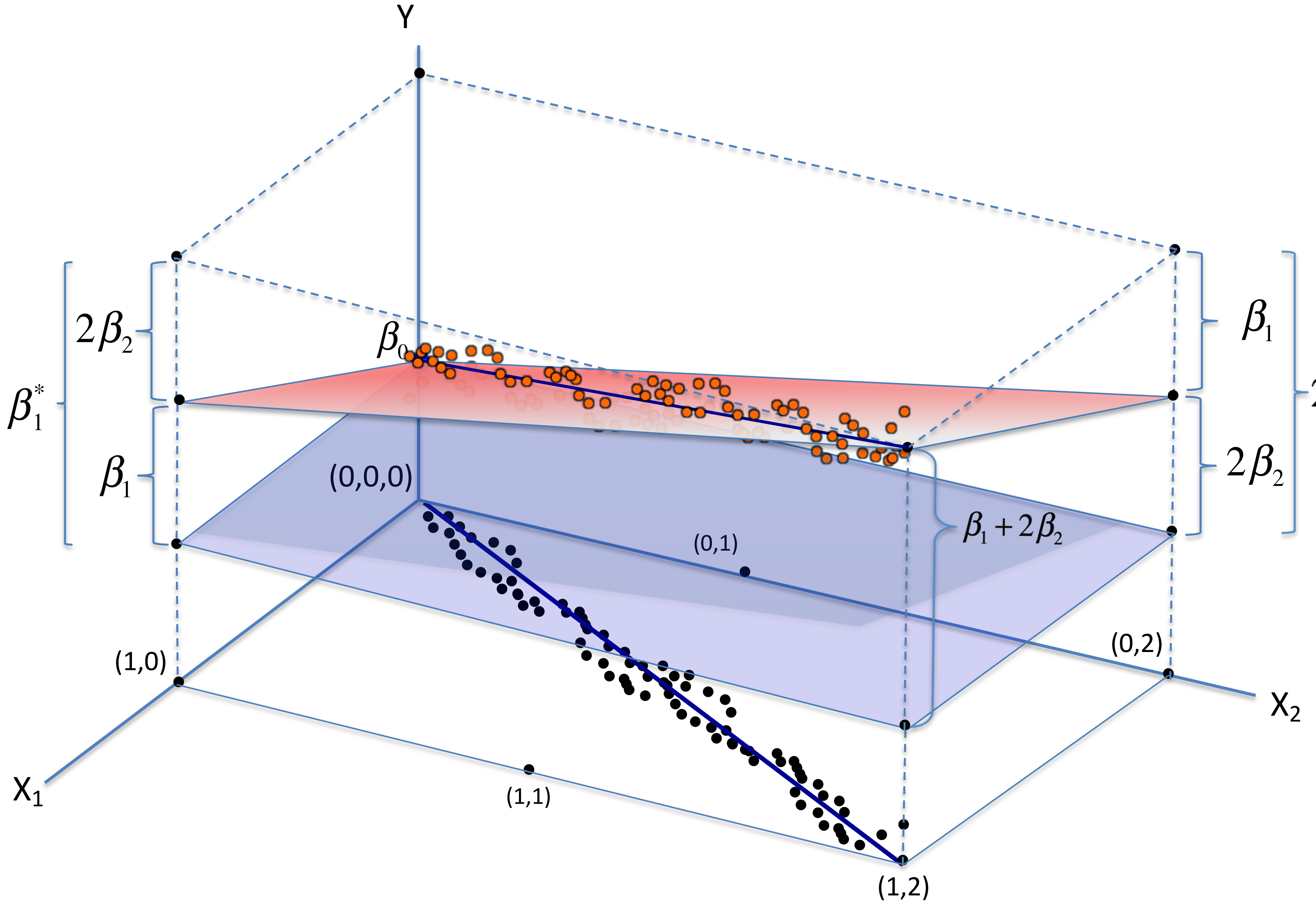

```
summary(lm(daughter~mom))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.88151    4.69354   1.892   0.0614 .
mom          0.86209    0.07223  11.936   <2e-16 ***
```

```
summary(lm(daughter~dad))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.26150    4.82647   2.126    0.036 *
dad          0.78125    0.06901  11.321   <2e-16 ***
```

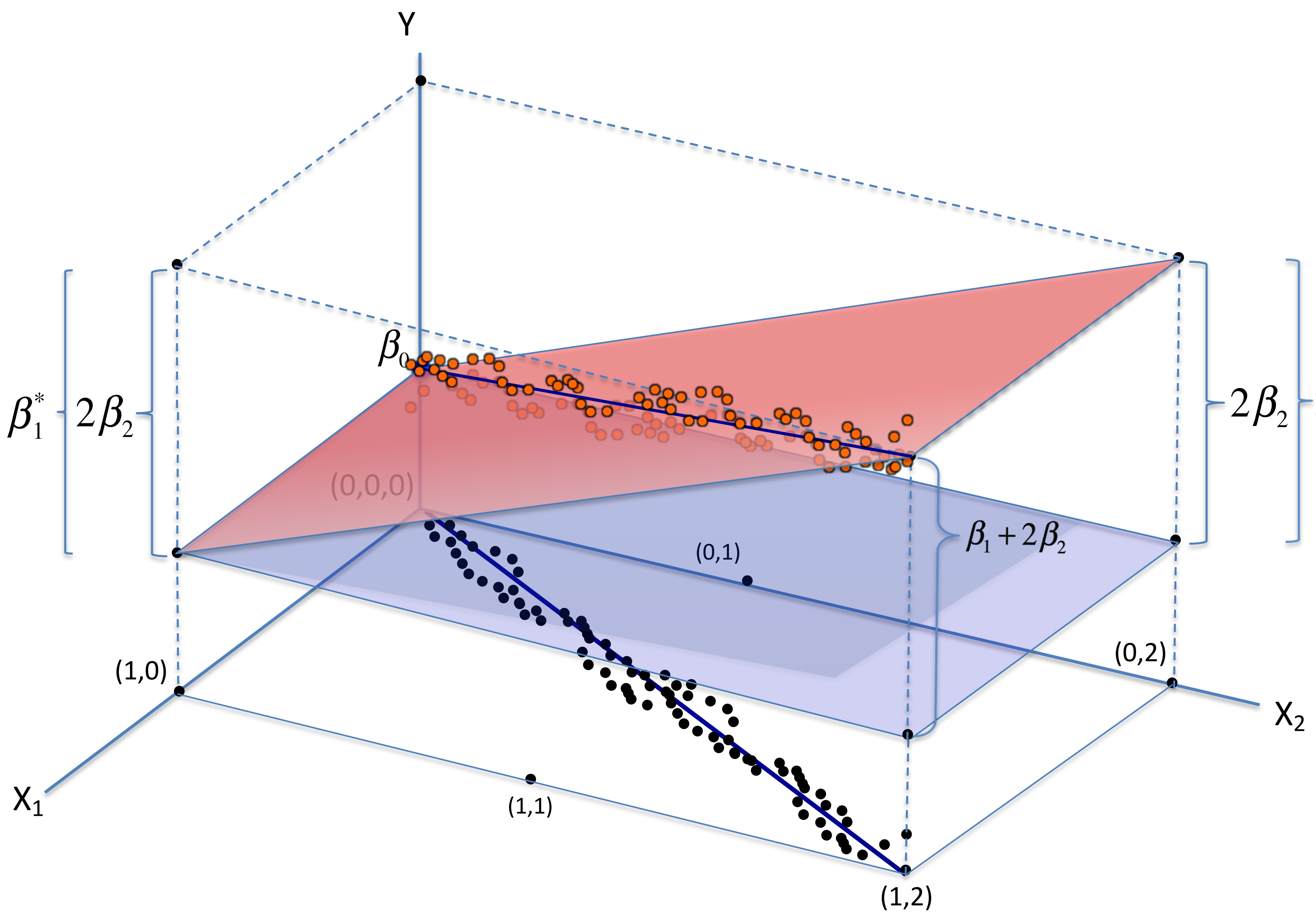
```
summary(lm(daughter~dad+mom))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.7872     4.6471   0.815 0.417082
mom           0.5210     0.1164   4.477 2.06e-05 ***
dad           0.3900     0.1078   3.617 0.000475 ***
```

Partial regression coefficients change a lot when adding correlated regressors/predictors.

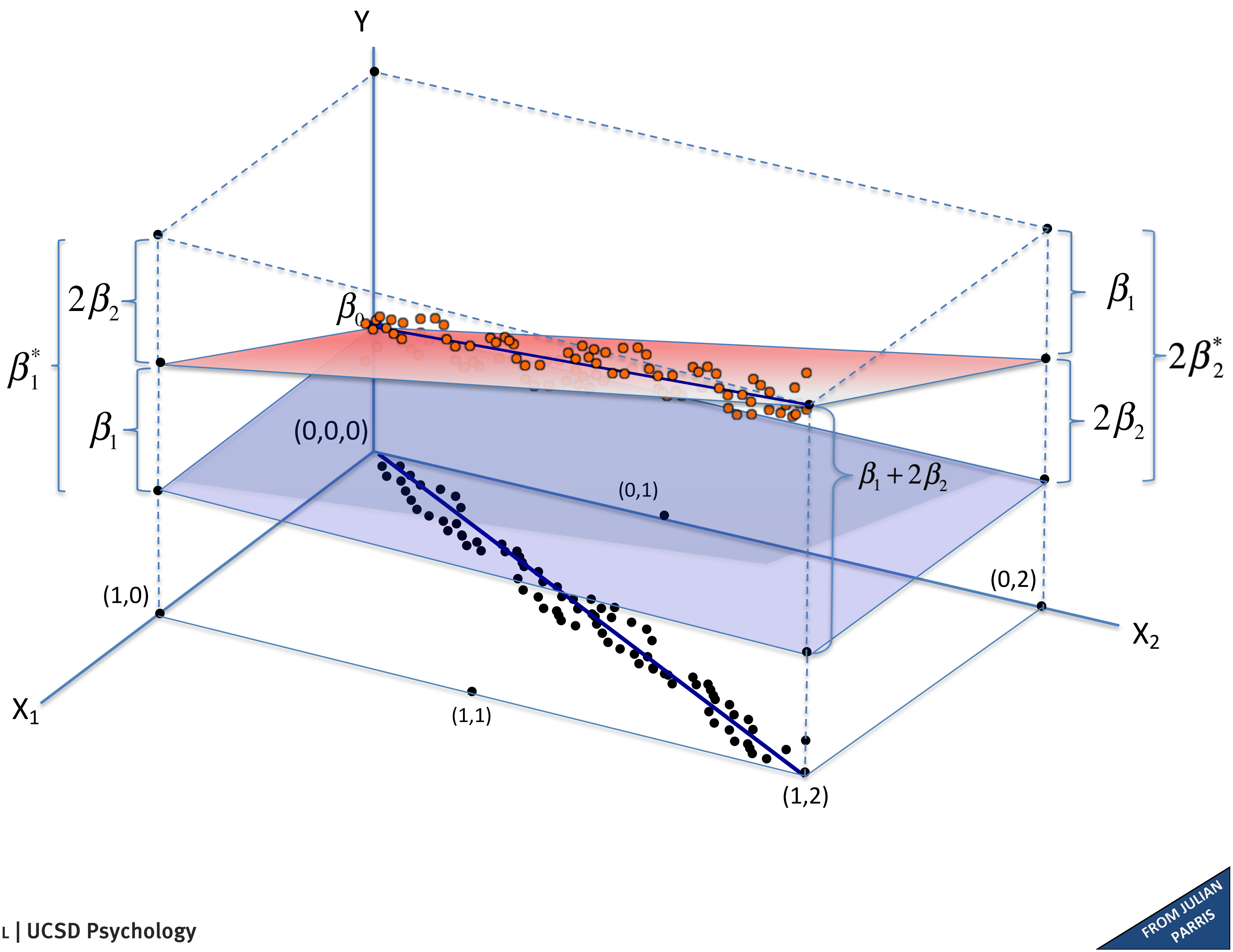
Because “credit” (for the increase in Y) is split among the different predictors.

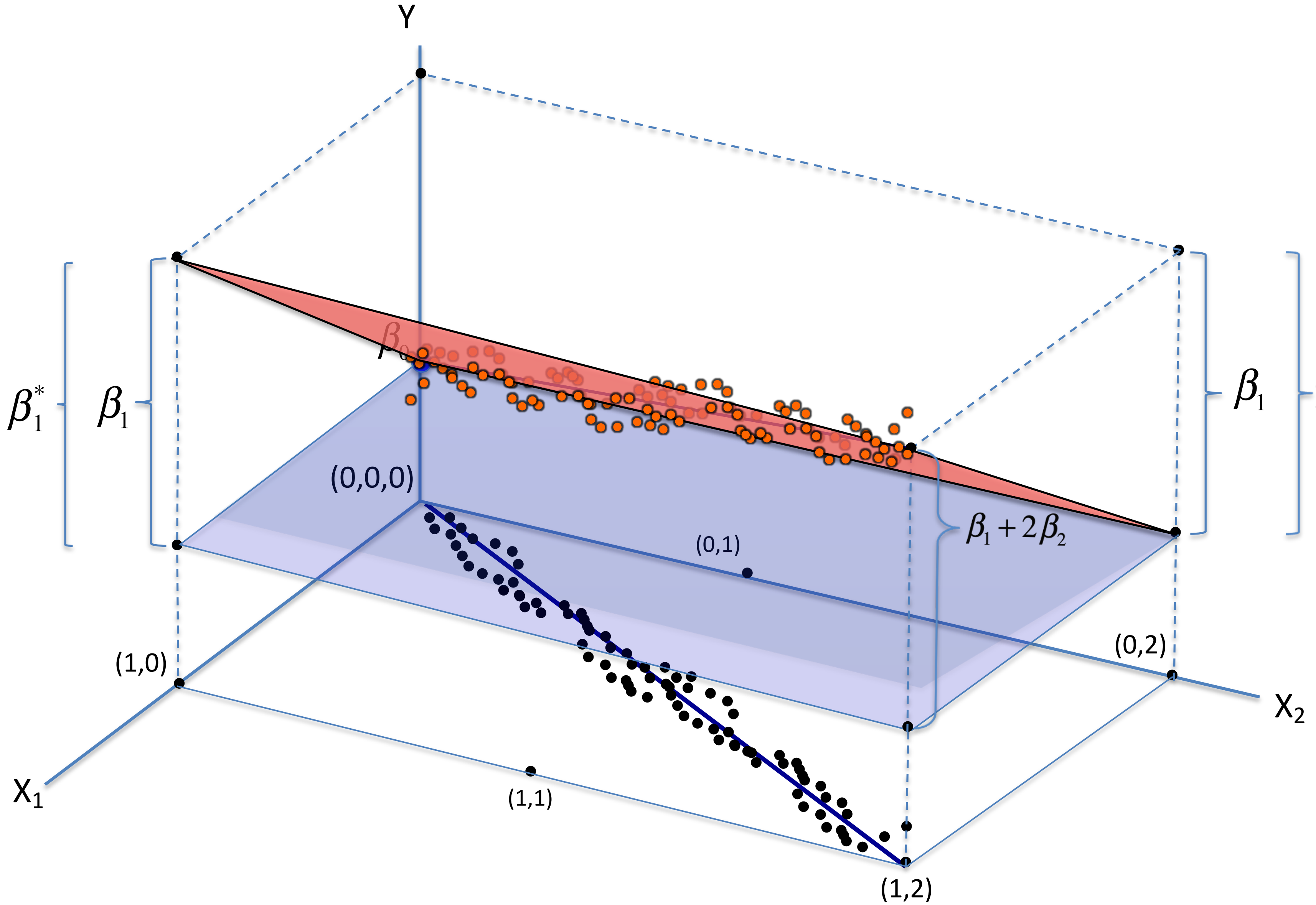


FROM JULIAN PARRIS

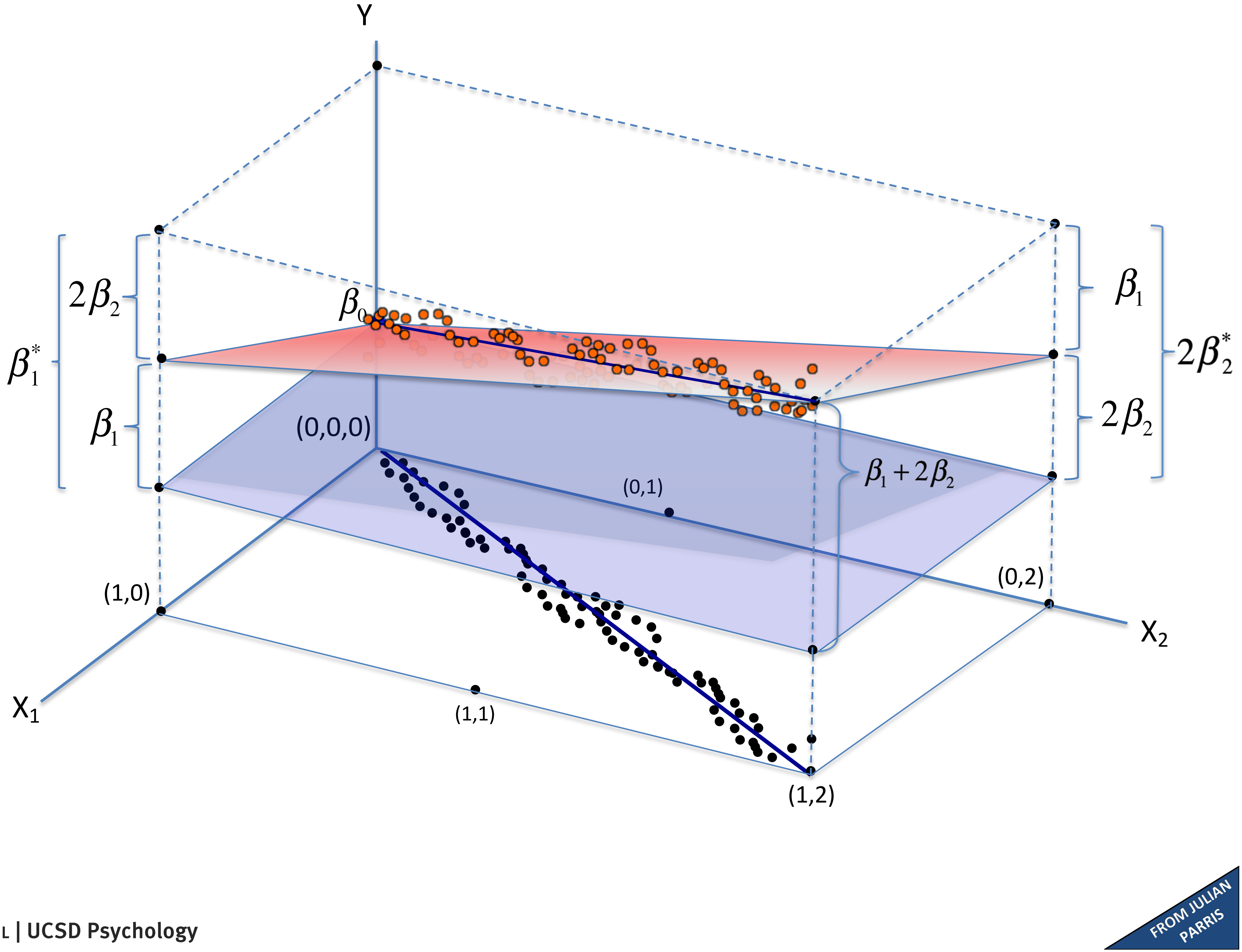


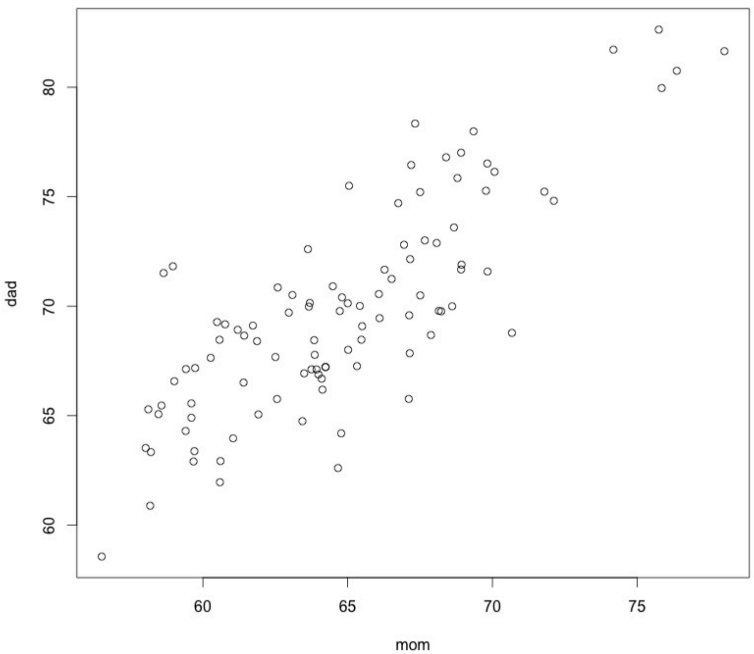
FROM JULIAN PARRIS


Y
$2\beta_2$
β_1^*
β_1
β_0
β_1
$2\beta_2^*$
$2\beta_2$
$\beta_1 + 2\beta_2$
(0,0,0)
(0,1)
(0,2)
(1,0)
(1,1)
(1,2)
X_1
X_2
FROM JULIAN PARRIS



FROM JULIAN PARRIS


Y
$2\beta_2$
β_1^*
β_1
β_0
β_1
$2\beta_2^*$
$2\beta_2$
$\beta_1 + 2\beta_2$
(0,0,0)
(0,1)
(0,2)
(1,0)
(1,1)
(1,2)
X_1
X_2
FROM JULIAN PARRIS




```
summary(lm(daughter~mom))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.88151    4.69354   1.892   0.0614 .
mom          0.86209    0.07223  11.936   <2e-16 ***
```

```
summary(lm(daughter~dad))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.26150    4.82647   2.126    0.036 *
dad          0.78125    0.06901  11.321   <2e-16 ***
```

```
summary(lm(daughter~dad+mom))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.7872     4.6471   0.815 0.417082
mom           0.5210     0.1164   4.477 2.06e-05 ***
dad           0.3900     0.1078   3.617 0.000475 ***
```

Partial regression coefficients errors tend to increase when adding correlated predictors.

Because there is ambiguity about how the credit should be split.
(Variance inflation factor)

Multicollinearity

Multicollinearity arises in multiple regression when predictors are correlated.

If this happens, we get:
(a) a credit assignment problem (which coefficients get credit for Y?)
(b) inflation of marginal standard errors for coefficients
(c) erratic changes in coefficients from small changes in the model or the data.

Measuring multicolinearity:

- How well you can account for the variance in one predictor from a linear combination of the other predictors.
- In a 2-predictor case, this boils down to their correlation:
  E.g., how well correlated are father and mother heights?
  A correlation of 1 or -1 means perfect multicolinearity.

Multiple regression agenda

- What is it? And why do this?
- Multicolinearity & its consequences
- Sums of squares partitioning in multiple regression
- Different hypothesis tests in multiple regression
- Nested model comparison
- Non-nested models

SST (SS total, also SSY)

SSR[X1] (SS regression var 1) | **SSE[X1] (SS error)**

R^2
Variability in Y accounted for by X1
e.g., Variability in daughters' heights accounted for by mothers' height

$1-R^2$
Variability unaccounted for by X1
e.g., Variability in daughters' heights not accounted for by mothers' height

1.0
Total variability in Y (around the mean)
e.g., total variability in daughter's heights

SST (SS total, also SSY)

SSR[X1] (SS regression var 1)

SSE[X1] (SS error)

R^2

Variability in Y accounted for by X1

e.g., Variability in daughters' heights accounted for by mothers' height

$1-R^2$

Variability unaccounted for by X1

e.g., Variability in daughters' heights not accounted for by mothers' height

SSR[X2] (SS regression var 2)

SSE[X2] (SS error)

R^2

Variability in Y accounted for by X2

e.g., Variability in daughters' heights accounted for by fathers' height

$1-R^2$

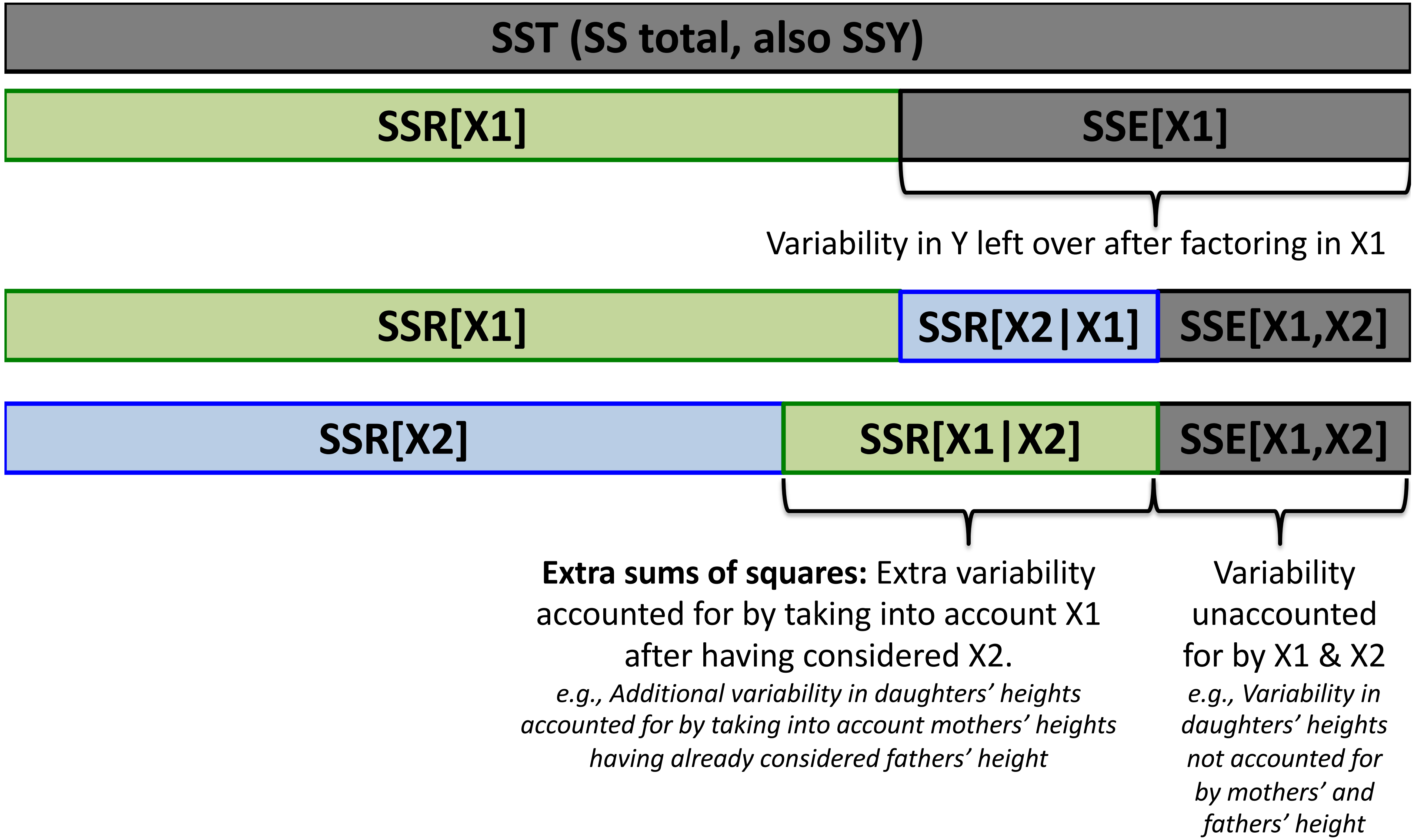
Variability unaccounted for by X2

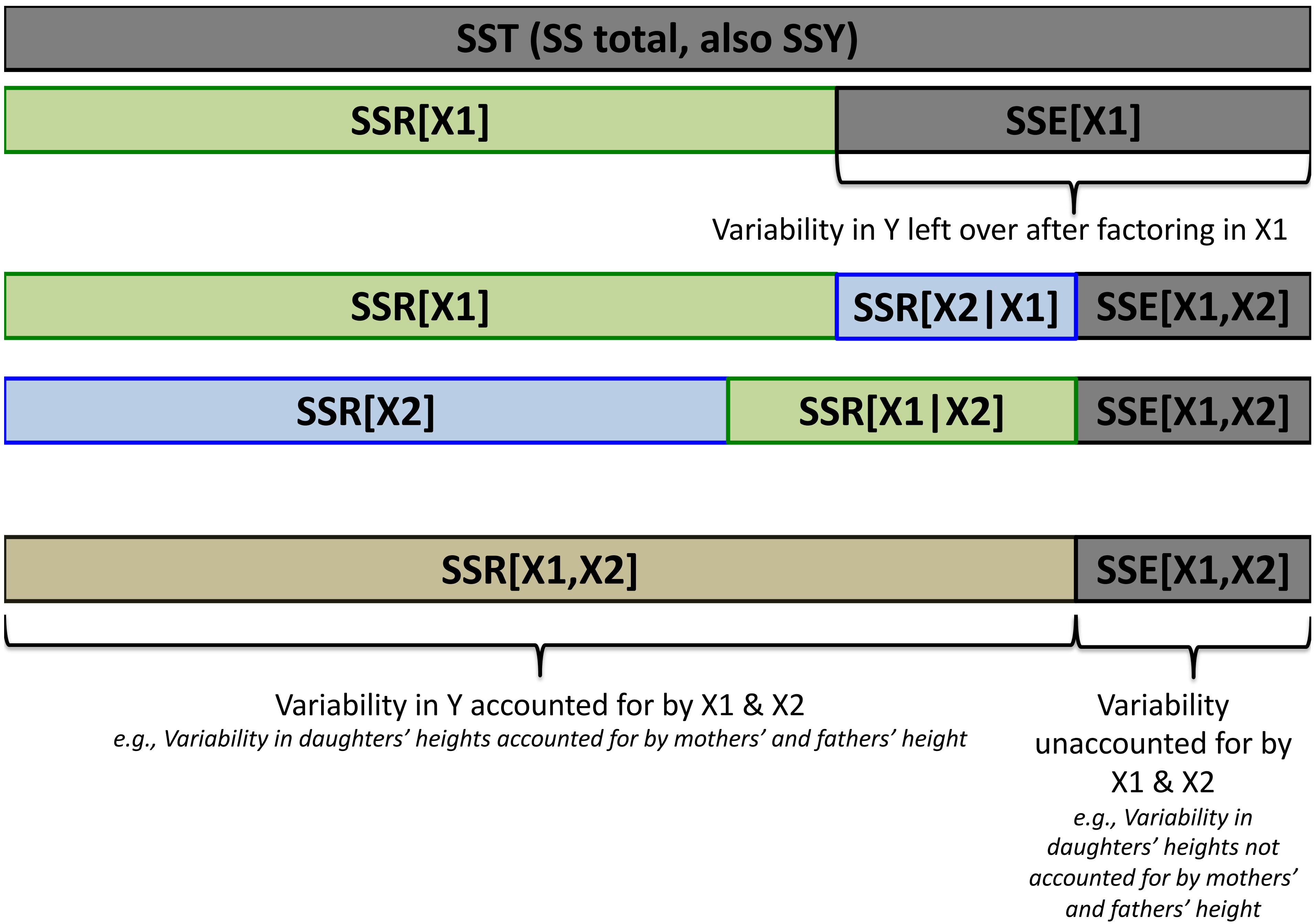
e.g., Variability in daughters' heights not accounted for by fathers' height

1.0

Total variability in Y (around the mean)

e.g., total variability in daughter's heights

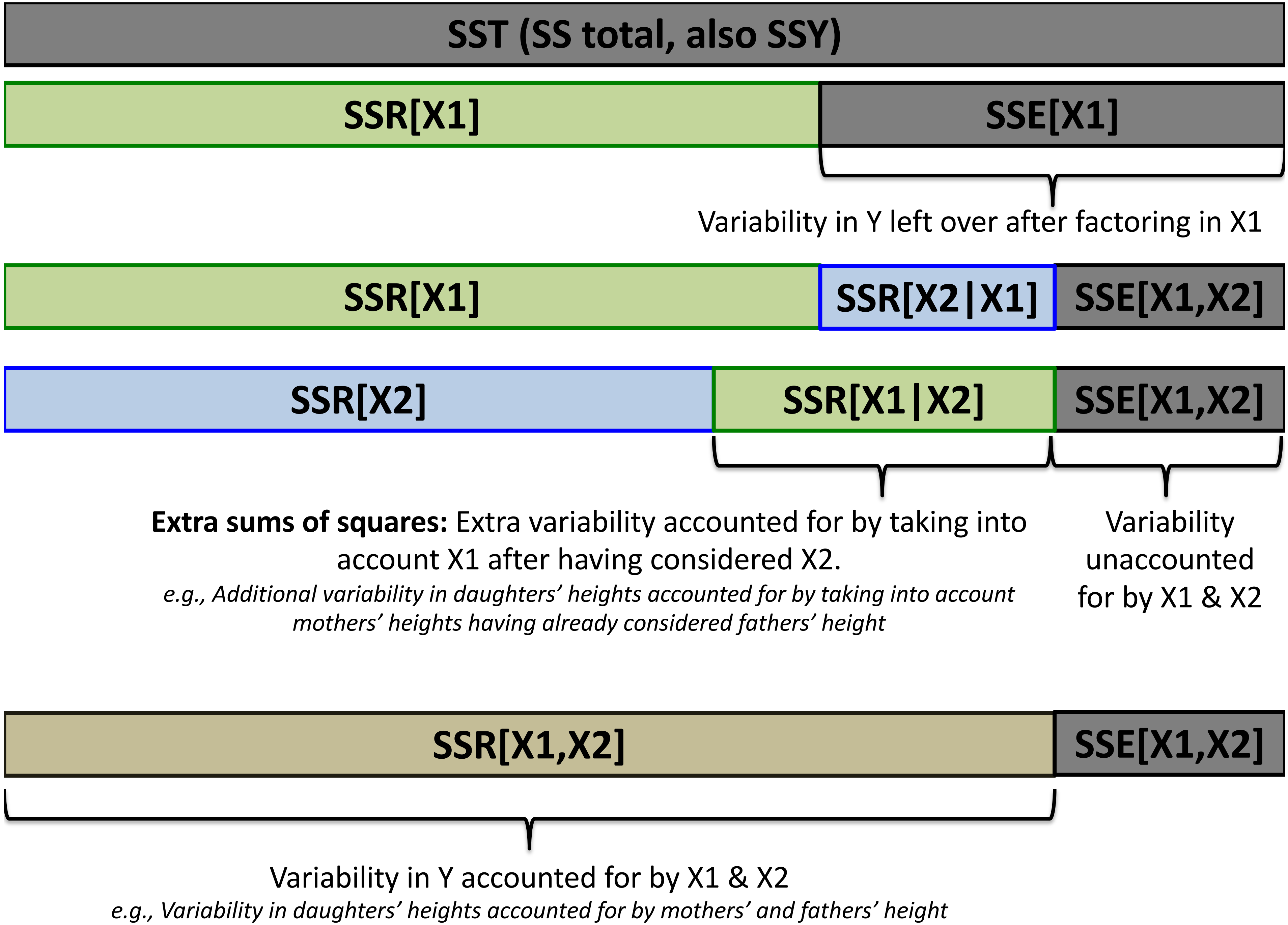

SST (SS total, also SSY)
SSR[X1]
SSE[X1]
Variability in Y left over after factoring in X1
SSR[X1]
SSR[X2|X1]
SSE[X1,X2]
SSR[X2]
SSR[X1|X2]
SSE[X1,X2]
Extra sums of squares: Extra variability accounted for by taking into account X1 after having considered X2.
e.g., Additional variability in daughters' heights accounted for by taking into account mothers' heights having already considered fathers' height
Variability unaccounted for by X1 & X2
e.g., Variability in daughters' heights not accounted for by mothers' and fathers' height

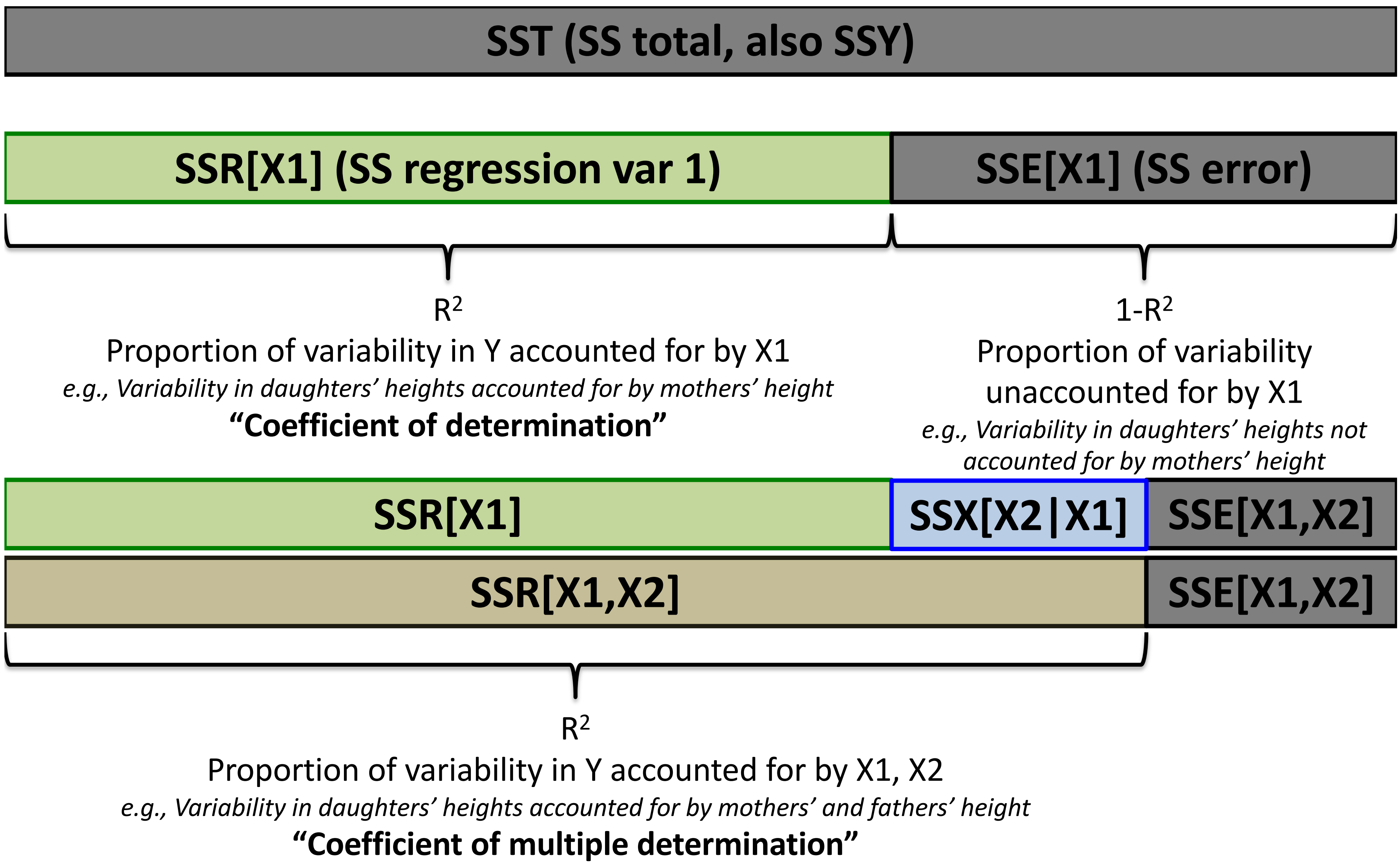

SST (SS total, also SSY)
SSR[X1]
SSE[X1]
Variability in Y left over after factoring in X1
SSR[X1]
SSR[X2|X1]
SSE[X1,X2]
SSR[X2]
SSR[X1|X2]
SSE[X1,X2]
SSR[X1,X2]
SSE[X1,X2]
Variability in Y accounted for by X1 & X2
e.g., Variability in daughters' heights accounted for by mothers' and fathers' height
Variability unaccounted for by X1 & X2
e.g., Variability in daughters' heights not accounted for by mothers' and fathers' height

Some arithmetic implications

- $SST = SSR[X_1,X_2] + SSE[X_1,X_2]$
- $SST = SSR[X_2] + SSR[X_1 \mid X_2] + SSE[X_1,X_2]$
- $SSR[X_1,X_2] = SSR[X_1] + SSR[X_2 \mid X_1]$
- $SSR[X_1|X_2] + SSE[X_1,X_2] = SSE[X_2]$

- When we do multiple regression, we have to choose how to partition the sums of squares, to test if the SS allocated to a particular variable is larger than expected by chance.


SST (SS total, also SSY)
SSR[X1]
SSE[X1]
Variability in Y left over after factoring in X1
SSR[X1]
SSR[X2|X1]
SSE[X1,X2]
SSR[X2]
SSR[X1|X2]
SSE[X1,X2]
Extra sums of squares: Extra variability accounted for by taking into account X1 after having considered X2.
e.g., Additional variability in daughters' heights accounted for by taking into account mothers' heights having already considered fathers' height
Variability unaccounted for by X1 & X2
SSR[X1,X2]
SSE[X1,X2]
Variability in Y accounted for by X1 & X2
e.g., Variability in daughters' heights accounted for by mothers' and fathers' height


SST (SS total, also SSY)
SSR[X1] (SS regression var 1)
SSE[X1] (SS error)
R²
Proportion of variability in Y accounted for by X1
e.g., Variability in daughters' heights accounted for by mothers' height
"Coefficient of determination"
1-R²
Proportion of variability unaccounted for by X1
e.g., Variability in daughters' heights not accounted for by mothers' height
SSR[X1]
SSX[X2|X1]
SSE[X1,X2]
SSR[X1,X2]
SSE[X1,X2]
R²
Proportion of variability in Y accounted for by X1, X2
e.g., Variability in daughters' heights accounted for by mothers' and fathers' height
"Coefficient of multiple determination"

SST (SS total, also SSY)

SSR[X1] (SS regression var 1) | SSE[X1] (SS error)

R^2
Proportion of variability in Y accounted for by X1
e.g., Variability in daughters' heights accounted for by mothers' height
"Coefficient of determination"

1-R^2
Proportion of Variability unaccounted for by X1
e.g., Variability in daughters' heights not accounted for by mothers' height

SSR[X1] | SSX[X2|X1] | SSE[X1,X2]

$R^2_{Y,X2|X1}$
Proportion of variability previously unaccounted for by X1 that can be accounted for by X2
"Coefficient of partial determination"

$$R^2_{YX_2|X_1} = \frac{SSX[X2 \mid X1]}{SSE[X1]}$$

Multiple regression agenda

- What is it? And why do this?
- Multicolinearity & its consequences
- Sums of squares partitioning in multiple regression
- Different hypothesis tests in multiple regression
- Nested model comparison
- Non-nested models

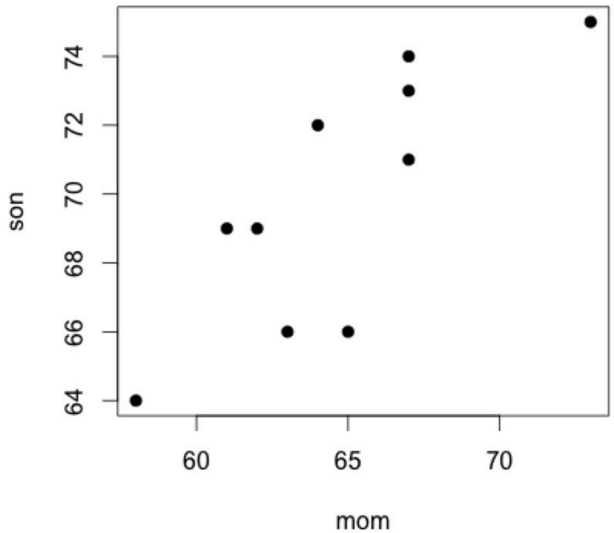
Significance of predictors.

- Pairwise correlation t-test, coefficient t-test, and variance-partitioning F-tests were the same in single variable regression, **they are all different in multivariate**.

- This is a cause for confusion – what do the different significances mean? Which ones should I care about?

- A more realistic example (less data, more noise), tenuous conclusions.

Predict: son ~ mom+dad

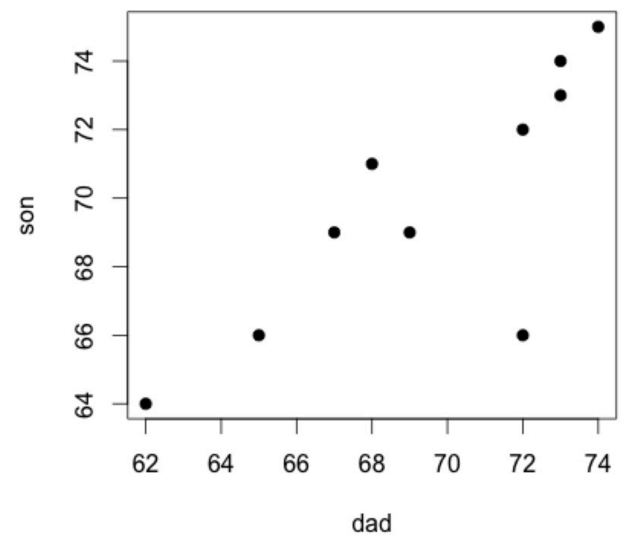



```
summary(lm(son~mom))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  23.4218    12.4358   1.883  0.09640 .
mom           0.7184     0.1919   3.744  0.00567 **
```

```
anova(lm(son~mom))
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523   14.02 0.00567 **
Residuals  8 45.377   5.672
```

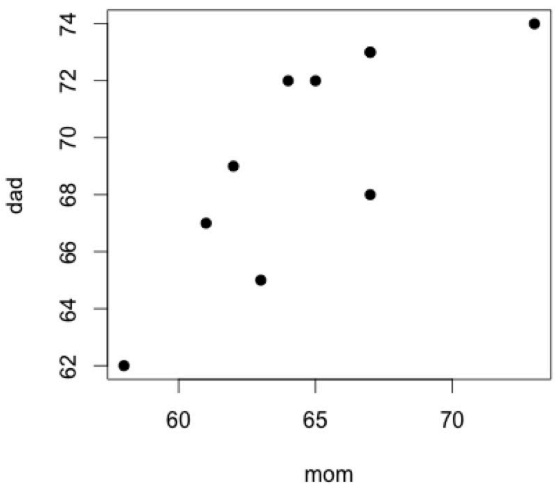
```
summary(lm(son~dad))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.9579    13.8754   1.294  0.23170
dad           0.7474     0.1994   3.749  0.00563 **
```

```
anova(lm(son~dad))
Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595  14.055 0.005632 **
Residuals  8 45.305   5.663
```



So: in single-variate regressions both mom and dad are significant predictors of son’s height.
Also, anova and regression significance are the same.

Predict: son ~ mom+dad



n	10
cor(mom,dad)	0.79

Mom and Dad height are highly correlated (colinear).

What will happen in the multiple regression?

(1) Both mom and dad coefficients will decrease (closer to 0)
(because they have same dir. Relationship w/ response , so are sharing credit)
(2) Both mom and dad coef. std. errors will go up
(because of credit assignment ambiguity)
(3) They may stop being significant!
(because t = B1/SE{B1})

Predict: son ~ mom+dad

```
summary(lm(son~mom+dad))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.0009    13.4355   1.117    0.301
mom           0.4000     0.3004   1.331    0.225
dad           0.4176     0.3124   1.336    0.223
```

(1) Both mom and dad coefficients will decrease (closer to 0 – because they are sharing credit)
(2) Both mom and dad coef. std. errors will go up (because of credit assignment ambiguity)
(3) They may stop being significant! (because t = B1/SE{B1})

```
anova(lm(son~mom+dad))
```

```
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

But the ANOVA analysis shows mom as significant, and dad as not… huh?

Predict: son ~ mom+dad

```
summary(lm(son~mom+dad))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.0009    13.4355   1.117    0.301
mom           0.4000     0.3004   1.331    0.225
dad           0.4176     0.3124   1.336    0.223
```

```
anova(lm(son~mom+dad))
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

```
summary(lm(son~dad+mom))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.0009    13.4355   1.117    0.301
dad           0.4176     0.3124   1.336    0.223
mom           0.4000     0.3004   1.331    0.225
```

```
anova(lm(son~dad+mom))
Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595 15.4116 0.005707 **
mom        1  9.153   9.153  1.7723 0.224818
Residuals  7 36.152   5.165
```

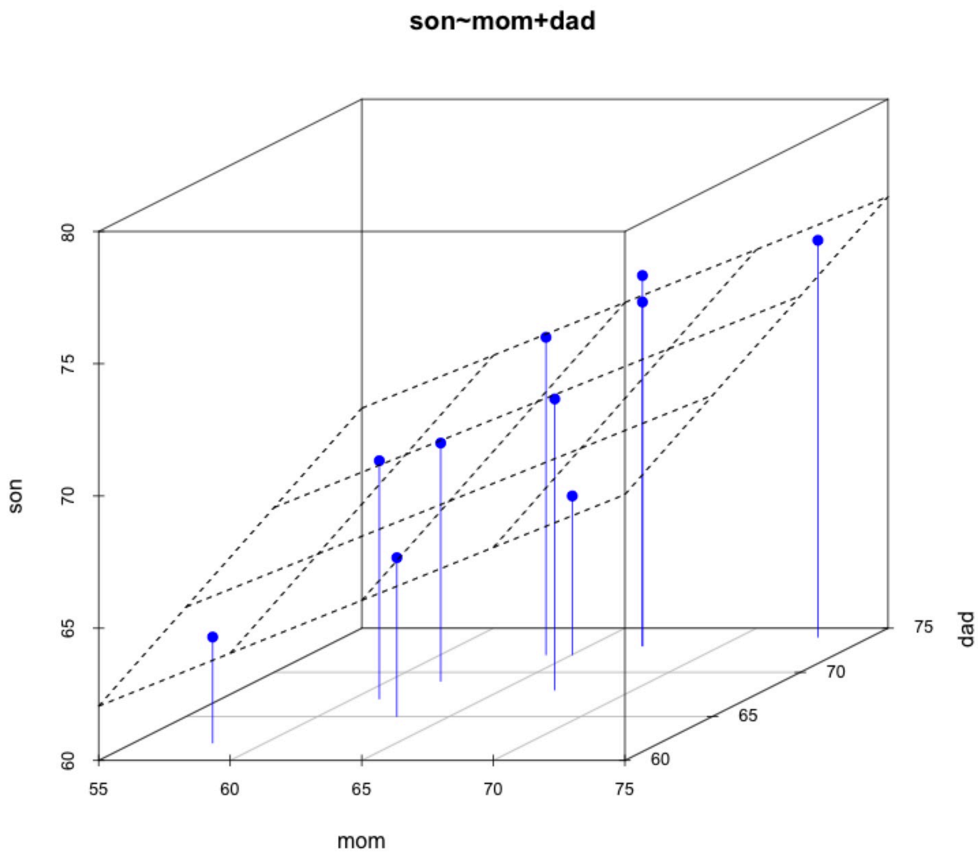
And if we change their order… coefficients stay the same, but ANOVA results change!

What is going on?

Coefficient significance.

```
summary(lm(son~dad+mom))

summary(lm(son~mom+dad))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.0009    13.4355   1.117    0.301
mom           0.4000     0.3004   1.331    0.225
dad           0.4176     0.3124   1.336    0.223
```



Significance of coefficients:
t=b1/s{b1}

s{b1} depends on s.d. of residuals (and independent variability of x1)

So: you fit the whole model (here: plane), find the residuals, then see whether the best estimated coefficient for x1 is significantly different from 0.

Formally: the partial slope on x1 is the slope of y as a function of residuals(x1~x2)

ANOVA significance.

```
anova(lm(son~mom+dad))
```

```
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

```
anova(lm(son~dad+mom))
```

```
Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595 15.4116 0.005707 **
mom        1  9.153   9.153  1.7723 0.224818
Residuals  7 36.152   5.165
```

To sort this out, we have to understand sums of squares and F statistics a bit better.

d.f. of numerator:
number of parameters for regression term

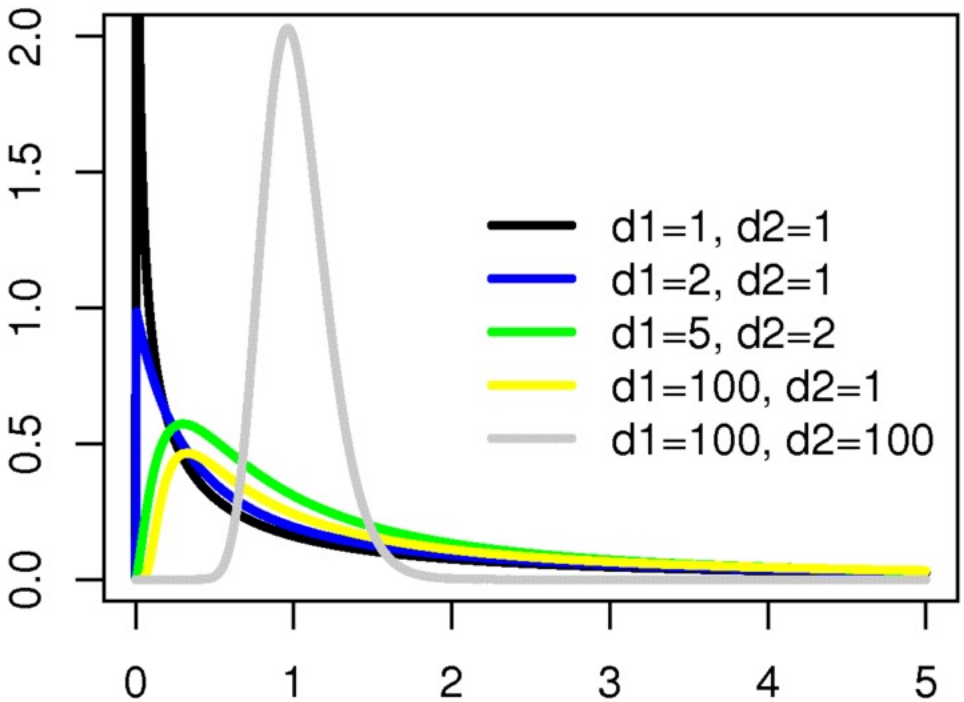
Sums of squares allocated to this regression term/component

$$F(\underbrace{df_{R-term}}_{\text{d.f. of numerator}}, \underbrace{df_{error}}_{\text{d.f. of denominator}}) = \frac{\left(\dfrac{\boxed{SSR[?\,|\,?]}}{\boxed{df_{R-term}}}\right)}{\left(\dfrac{\boxed{SSE_{FULL}}}{\boxed{df_{error}}}\right)}$$

Sum of squared residuals from full model

d.f. of denominator:
n minus number of parameters in full model

F distribution

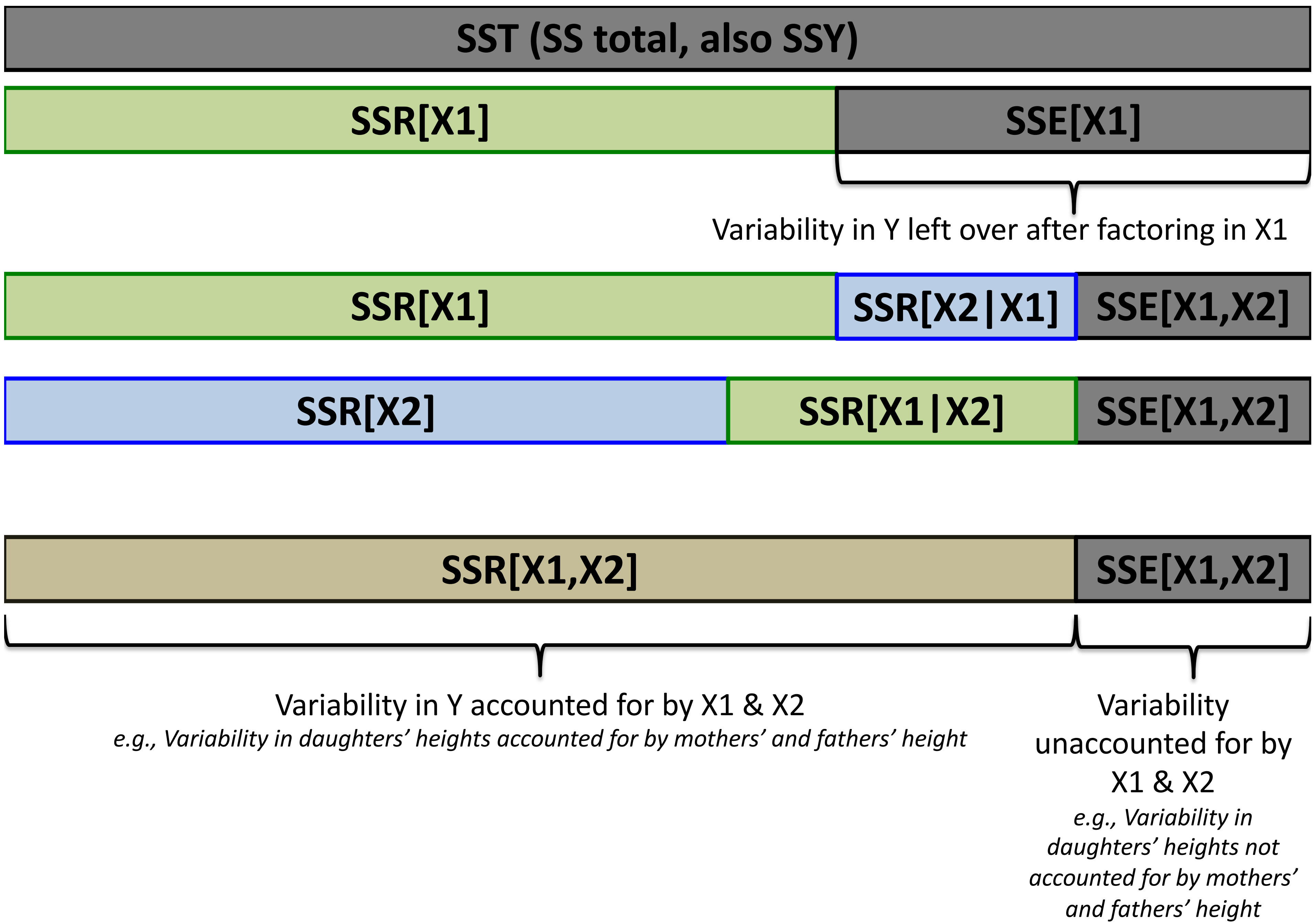


The F-statistic
The ratio of two (identical) sample variances estimated with different degrees of freedom.

Under Ho, MSR (SSR/df_R) is expected to be equal to the variance of the residuals. So numerator and denominator are two estimates of the same error variance, and the F-statistic will follow F distribution.

So, given random variation, even under Ho, we expect the regression to take up *some* variance, and our question is: does it account for *more* variance than expected by chance?

So, F-test is, like Chi-squared, one tailed (positive tail).


SST (SS total, also SSY)
SSR[X1]
SSE[X1]
Variability in Y left over after factoring in X1
SSR[X1]
SSR[X2|X1]
SSE[X1,X2]
SSR[X2]
SSR[X1|X2]
SSE[X1,X2]
SSR[X1,X2]
SSE[X1,X2]
Variability in Y accounted for by X1 & X2
e.g., Variability in daughters' heights accounted for by mothers' and fathers' height
Variability unaccounted for by X1 & X2
e.g., Variability in daughters' heights not accounted for by mothers' and fathers' height

ANOVA significance.

```
anova(lm(son~mom+dad))
```

```
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

$$F(df_n, df_d) = \frac{MSR}{MSE}$$

$$F(1,7) = \frac{79.523}{5.165}$$

```
1-pf(15.3911, 1, 7)
```

```
0.00572
```

$$F(1,7) = \frac{9.225}{5.165}$$

```
1-pf(1.7862, 1, 7)
```

```
0.2232
```

ANOVA significance.

```
anova(lm(son~mom+dad))
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

$$F(df_n, df_d) = \frac{MSR}{MSE}$$

But what are these sums of squares?

SS for mom = SSR[mom]
SS for dad = SSR[dad | mom]
The SS. Corresponds to the extra sums of squares from adding the second regressor to the first. So if we change the order of regressors, we get different results

```
anova(lm(son~dad+mom))
Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595 15.4116 0.005707 **
mom        1  9.153   9.153  1.7723 0.224818
Residuals  7 36.152   5.165
```

SS for dad = SSR[dad]
SS for mom = SSR[mom | dad]

ANOVA significance.

```
anova(lm(son~mom+dad))
```

```
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

```
anova(lm(son~dad+mom))
```

```
Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595 15.4116 0.005707 **
mom        1  9.153   9.153  1.7723 0.224818
Residuals  7 36.152   5.165
```

Which is “right”?
Neither.
They are asking different questions.
Son~mom+dad asks:
(1) is having mom in the regression better than just the mean?
(2) is adding dad to a regression with mom, worth it?

Son~dad+mom asks:
(1) Is having dad in the regression better than just the mean
(2) Is adding mom to a regression with dad worth it?

What question are you trying to ask?

ANOVA significance.

```
anova(lm(son~mom+dad))
```

```
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

```
anova(lm(son~dad+mom))
```

```
Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595 15.4116 0.005707 **
mom        1  9.153   9.153  1.7723 0.224818
Residuals  7 36.152   5.165
```

What should I do?

If your goal is to really assess the contribution of one of these predictors, you should clearly explain what the contribution is.

In this case: mom's height predicts son's height, but because it is highly correlated with dad's height, you can't tell what the causal route is. Moreover, adding mom's height to a model that includes dad's height doesn't help: mom's height accounts for the same variance in son's height as dad's height does.

Which of these things is worth emphasizing in your results depends on what your scientific question is; however, you should realize that the whole story involves understanding the full set of relationships among these variables, not just the significance assessed one way or another.

ANOVA significance.

```
anova(lm(son~mom+dad))
```

```
Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

```
anova(lm(son~dad+mom))
```

```
Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595 15.4116 0.005707 **
mom        1  9.153   9.153  1.7723 0.224818
Residuals  7 36.152   5.165
```

What should I do?
If your goal is to provide as comprehensible a model of your data as possible, consider recoding your predictors:

```
mean.mom.dad = (mom+dad)/2
diff.mom.dad = (mom-dad)
summary(lm(son~mean.mom.dad+diff.mom.dad))
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.000883  13.435464   1.117  0.30106
mean.mom.dad  0.817540   0.197538   4.139  0.00436 **
diff.mom.dad -0.008794   0.290149  -0.030  0.97667
Residual standard error: 2.273 on 7 degrees of freedom
```

Different tests for different questions

There isn't one answer to "is this predictor significant"

- F-tests: *"Does adding this predictor to some smaller model account for more variance than expected by chance?"*
  – Which "smaller model" we use depends on our question!
- T-tests for partial regression coefficients: *"Does the allocation of credit to all the predictors for variation in Y necessitate that this predictor have non-zero credit?"*
  – If another, colinear predictor *could* take credit, then the answer may well be no, but that might not matter to you
- T-tests for pairwise correlation: *"Is there a linear relationship between these two variables, disregarding relationships with all other variables?"*
  – Often useful to ask, but obscures the full picture.

Multiple regression agenda

- What is it? And why do this?
- Multicolinearity & its consequences
- Sums of squares partitioning in multiple regression
- Different hypothesis tests in multiple regression
- Nested model comparison
- Non-nested models

- **Nested Model:** A smaller model that differs only by excluding some parameters of a larger model.
  A) height ~ mom + dad + protein + exercise + milk
  B) height ~ mom + dad + protein + exercise
  C) height ~ dad + protein + exercise
  D) height ~ mom + dad
  E) height ~ dad + protein
  F) height ~ mom + dad + milk
  G) height ~ exercise + milk + beer
  H) weight ~ mom + dad + protein + exercise

B in A; C in A, B; D in A, B; E in A, B, C; F in A; A, G, H are not nested in others.

F-tests compare nested models

They ask: is a bigger model better than a smaller model?

height ~ mom + dad + protein + exercise + milk

(nested)

height ~ mom + dad + protein + exercise

(nested)

height ~ dad + protein + exercise

(nested)

height ~ protein + dad

(nested)

height ~ dad

(nested)

height ~ 1

d.f. of numerator:
number of extra parameters in full model

Extra sums of squares of full compared to reduced: Estimated by difference in SSE.

$$F(\underbrace{p_{FULL}-p_{REDUCED}}_{\text{d.f. of numerator}},\underbrace{n-p_{FULL}}_{\text{d.f. of denominator}})=\frac{\left(\dfrac{\boxed{SSE_{REDUCED}-SSE_{FULL}}}{\boxed{p_{FULL}-p_{REDUCED}}}\right)}{\left(\dfrac{\boxed{SSE_{FULL}}}{\boxed{n-p_{FULL}}}\right)}$$

Remaining sums of squares error in full model

d.f. of denominator:
n minus number of parameters in full model

$$F(p_{FULL} - p_{REDUCED}, n - p_{FULL}) = \frac{\left(\frac{SSE_{REDUCED} - SSE_{FULL}}{p_{FULL} - p_{REDUCED}}\right)}{\left(\frac{SSE_{FULL}}{n - p_{FULL}}\right)}$$

- Extra sums of squares of full compared to reduced model is the difference in sums of squares of error.
- Degrees of freedom of the extra sums of squares is the number of parameters added.
- The remaining sums of squares error from the full model is the denominator.
- Degrees of freedom of error are n minus the number of parameters in full model.

$$F(p_{FULL} - p_{REDUCED}, n - p_{FULL}) = \frac{\left(\dfrac{SSE_{REDUCED} - SSE_{FULL}}{p_{FULL} - p_{REDUCED}}\right)}{\left(\dfrac{SSE_{FULL}}{n - p_{FULL}}\right)}$$

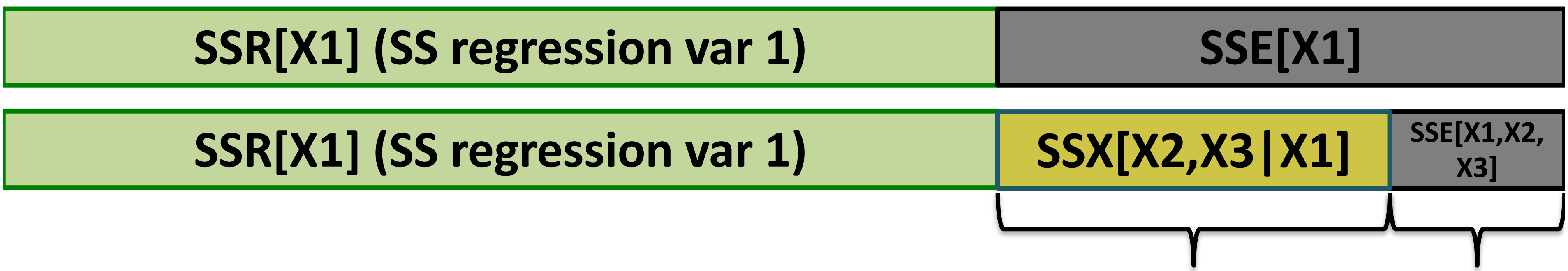
SST (SS total, also SSY)

SSR[X1] (SS regression var 1)	**SSE[X1]**

F = (SSR[x1] / (2-1)) / (SSE[x1] / (n-2))

- SSE reduced is just SST (a 1 parameter regression model considering only the mean of Y: Bo)
- SSR[X1] = SST – SSE[x1]

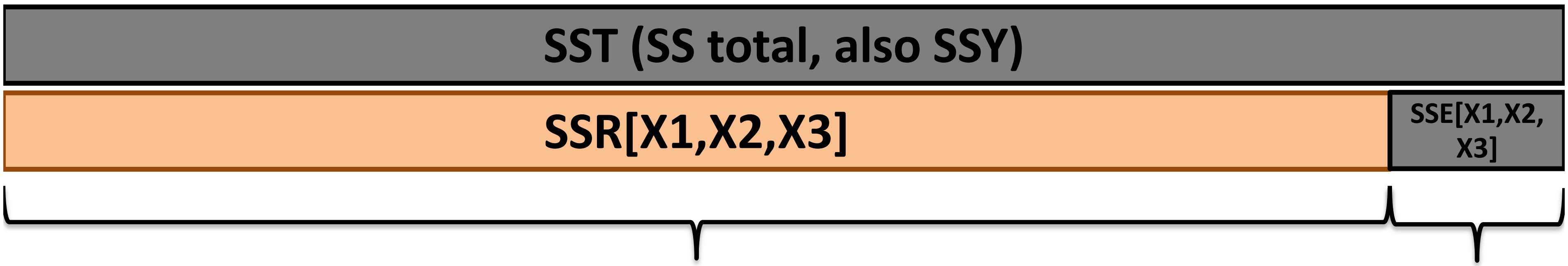
$$F(p_{FULL}-p_{REDUCED}, n-p_{FULL}) = \frac{\left(\dfrac{SSE_{REDUCED}-SSE_{FULL}}{p_{FULL}-p_{REDUCED}}\right)}{\left(\dfrac{SSE_{FULL}}{n-p_{FULL}}\right)}$$



F = (SSX[x2,x3|x1] / (2)) / (SSE[x1,x2,x3] / (n-4))

- SSE reduced is SSE[x1]. SSE full is SSE[x1,x2,x3]
- SSX[x2,x3|x1] = SSE[x1]– SSE[x1,x2,x3]
- # parameters full: 4 (b0, b1, b2, b3)
- # parameters reduced: 2 (b0, b1)

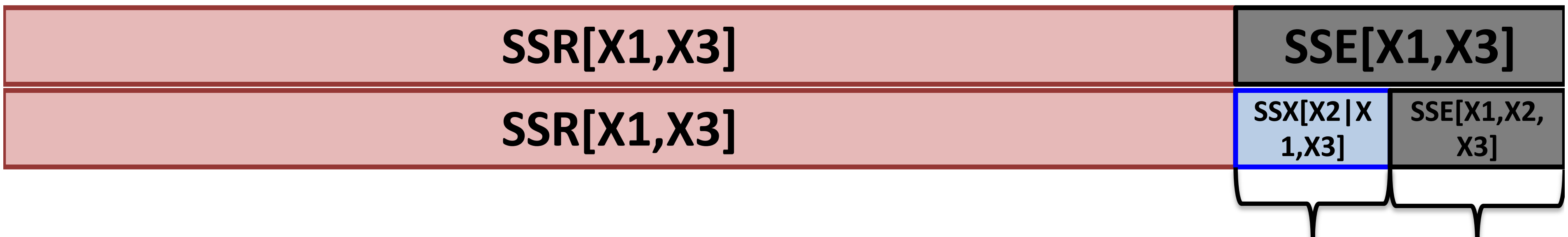
$$F(p_{FULL} - p_{REDUCED}, n - p_{FULL}) = \frac{\left(\dfrac{SSE_{REDUCED} - SSE_{FULL}}{p_{FULL} - p_{REDUCED}}\right)}{\left(\dfrac{SSE_{FULL}}{n - p_{FULL}}\right)}$$



F = (SSR[x1,x2,x3] / (4-1)) / (SSE[x1,x2,x3] / (n-4))

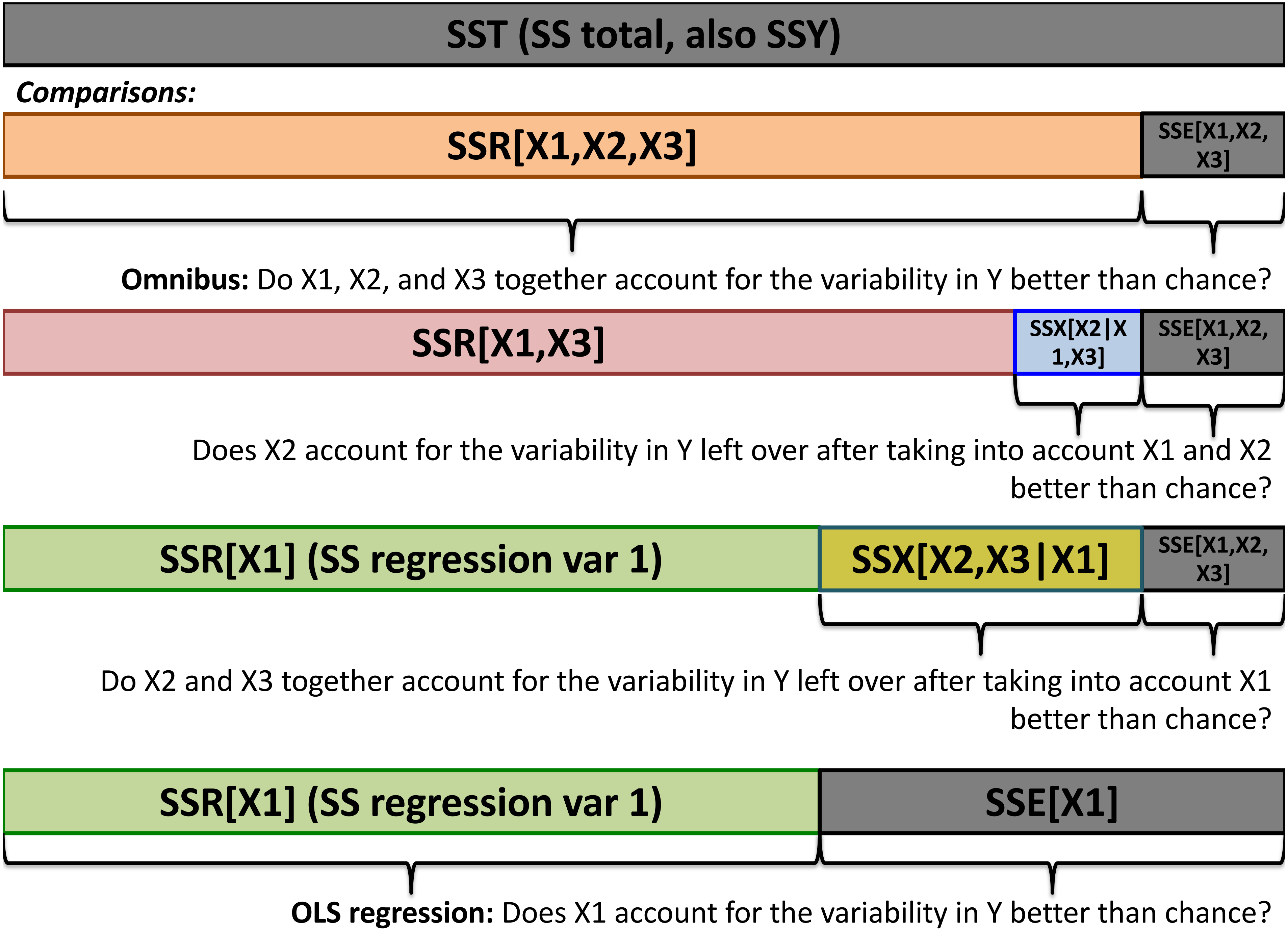
- SSE reduced is SSE[b0]. SSE full is SSE[x1,x2,x3]
- SSR[x2,x3,x1] = SSE[b0]– SSE[x1,x2,x3]
- # parameters full: 4 (b0, b1, b2, b3)
- # parameters reduced: 1 (b0)

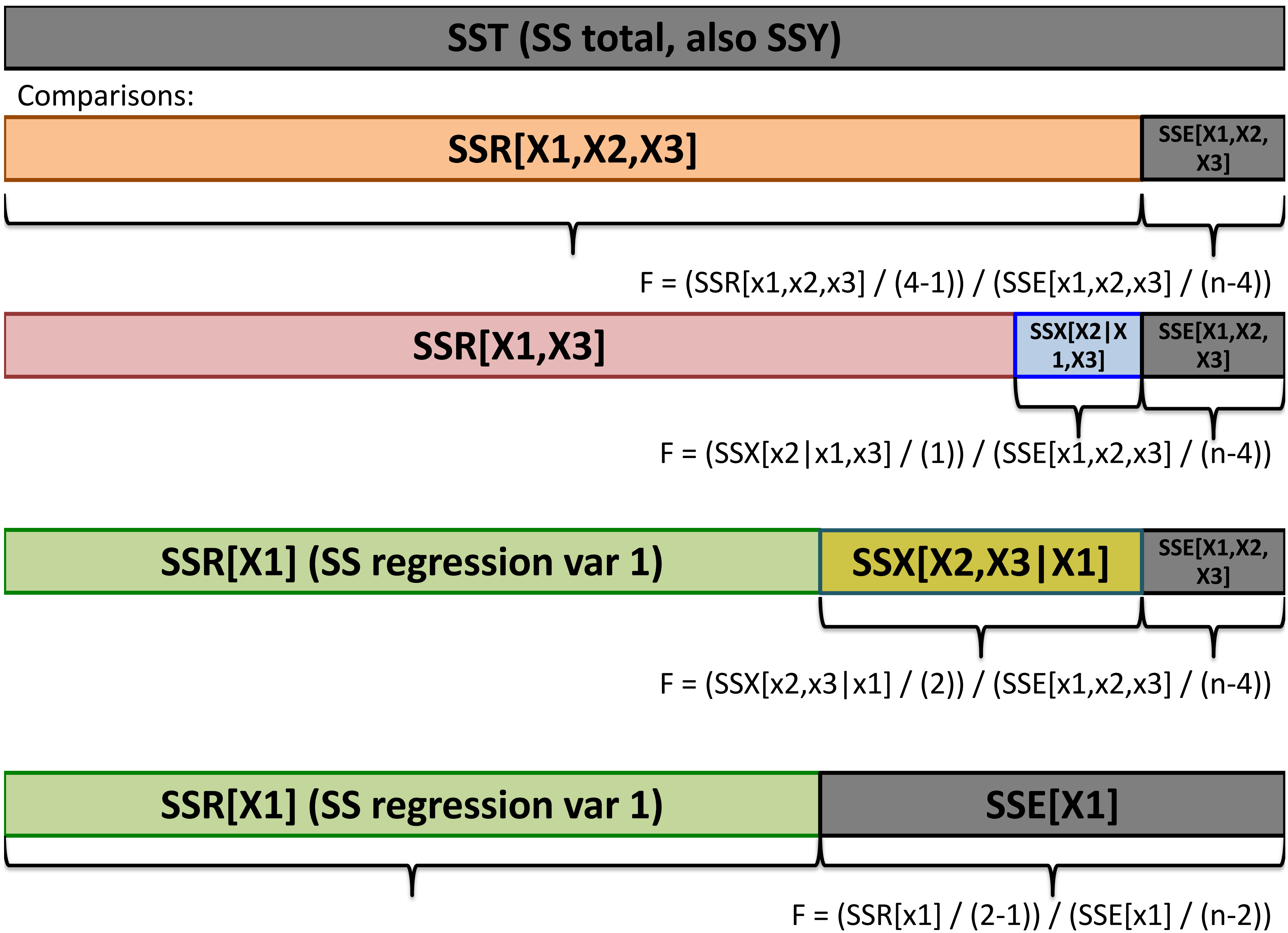
$$F(p_{FULL}-p_{REDUCED}, n-p_{FULL}) = \frac{\left(\dfrac{SSE_{REDUCED}-SSE_{FULL}}{p_{FULL}-p_{REDUCED}}\right)}{\left(\dfrac{SSE_{FULL}}{n-p_{FULL}}\right)}$$



F = (SSX[x2|x1,x3] / (1)) / (SSE[x1,x2,x3] / (n-4))

- SSE reduced is SSE[x1,x3]. SSE full is SSE[x1,x2,x3]
- SSX[x2|x1,x3] = SSE[x1,x3]– SSE[x1,x2,x3]
- # parameters full: 4 (b0, b1, b2, b3)
- # parameters reduced: 3 (b0,b1,b3)


SST (SS total, also SSY)
Comparisons:
SSR[X1,X2,X3]
SSE[X1,X2, X3]
Omnibus: Do X1, X2, and X3 together account for the variability in Y better than chance?
SSR[X1,X3]
SSX[X2|X 1,X3]
SSE[X1,X2, X3]
Does X2 account for the variability in Y left over after taking into account X1 and X2 better than chance?
SSR[X1] (SS regression var 1)
SSX[X2,X3|X1]
SSE[X1,X2, X3]
Do X2 and X3 together account for the variability in Y left over after taking into account X1 better than chance?
SSR[X1] (SS regression var 1)
SSE[X1]
OLS regression: Does X1 account for the variability in Y better than chance?


SST (SS total, also SSY)
Comparisons:
SSR[X1,X2,X3]
SSE[X1,X2,X3]
F = (SSR[x1,x2,x3] / (4-1)) / (SSE[x1,x2,x3] / (n-4))
SSR[X1,X3]
SSX[X2|X1,X3]
SSE[X1,X2,X3]
F = (SSX[x2|x1,x3] / (1)) / (SSE[x1,x2,x3] / (n-4))
SSR[X1] (SS regression var 1)
SSX[X2,X3|X1]
SSE[X1,X2,X3]
F = (SSX[x2,x3|x1] / (2)) / (SSE[x1,x2,x3] / (n-4))
SSR[X1] (SS regression var 1)
SSE[X1]
F = (SSR[x1] / (2-1)) / (SSE[x1] / (n-2))

Multiple regression agenda

- What is it? And why do this?
- Multicolinearity & its consequences
- Sums of squares partitioning in multiple regression
- Different hypothesis tests in multiple regression
- Nested model comparison
- Non-nested models

$$F(p_{FULL} - p_{REDUCED}, n - p_{FULL}) = \frac{\left(\frac{SSE_{REDUCED} - SSE_{FULL}}{p_{FULL} - p_{REDUCED}}\right)}{\left(\frac{SSE_{FULL}}{n - p_{FULL}}\right)}$$

- F test allows us to compare *nested models.*
- How do we compare non-nested models?
  - height ~ mom + dad
  - height ~ mom + protein
  - height ~ protein + exercise
  - height ~ ethnicity
  - weight ~ mom + dad

"Model building" comparison:
Is it better to add *dad* or *protein* to model that already has *mom?*
Is it better to add *mom* or *exercise* to a model that already has *protein*?

I am using these terms to describe different comparisons only for convenience, these are not really technical names for different non-nested model comparisons. In reality, all of them are 'model selection' problems.

$$F(p_{FULL} - p_{REDUCED}, n - p_{FULL}) = \frac{\left(\frac{SSE_{REDUCED} - SSE_{FULL}}{p_{FULL} - p_{REDUCED}}\right)}{\left(\frac{SSE_{FULL}}{n - p_{FULL}}\right)}$$

- F test allows us to compare *nested models.*
- How do we compare non-nested models?
  - height ~ mom + dad
  - height ~ mom + protein
  - height ~ protein + exercise
  - height ~ ethnicity
  - weight ~ mom + dad

"Model selection" comparison: Is a model with *mom* and *dad* better than a model with *protein* and *exercise?* A model with *ethnicity?*
(These can also be seen as model building problems: would it be better to add these or those regressors to null model)

I am using these terms to describe different comparisons only for convenience, these are not really technical names for different non-nested model comparisons. In reality, all of them are 'model selection' problems.

$$F(p_{FULL}-p_{REDUCED}, n-p_{FULL}) = \frac{\left(\frac{SSE_{REDUCED}-SSE_{FULL}}{p_{FULL}-p_{REDUCED}}\right)}{\left(\frac{SSE_{FULL}}{n-p_{FULL}}\right)}$$

- F test allows us to compare *nested models.*
- How do we compare non-nested models?
  - height ~ mom + dad
  - height ~ mom + protein
  - height ~ protein + exercise
  - height ~ ethnicity
  - weight ~ mom + dad

Weird (but sometimes useful) model comparison:
Is height more/less predictable by *mom* and *dad* (height?) than weight?

I am using these terms to describe different comparisons only for convenience, these are not really technical names for different non-nested model comparisons. In reality, all of them are 'model selection' problems.

- How do we compare non-nested models?
  - There isn't really a good way to test the null hypothesis that two non-nested models are equally good. Because
    (a) we don't know what 'good' means.
    Bigger models will have better fits, how do we trade off fit with model size
    (b) Even if we define 'good', the difference in goodness of two models doesn't have a definable null hypothesis distribution.
  - Consequently, we just define some goodness statistic and compare the numerical difference in goodness.
    (Bayesian methods offer ways to attach probability statements to goodness comparisons between non-nested models, but we will not be dealing with this now)

- How do we compare non-nested models?

  Goodness:

  – R^2 (no punishment for bigger models: fit is all that counts)
    - Useful for simple model building when number of parameters is constant: which parameter is a better one to add to the model I already have? Which K parameter model better fits these data?

- How do we compare non-nested models?
  - Goodness:
  - R^2_a 'Adjusted R squared'
    (like R^2, but punished for having more parameters)

$$R_A^2 = \bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p} = 1 - \frac{SSE}{SST}\frac{(n-1)}{(n-p)}$$

- How do we compare non-nested models?

  Goodness:

  – R^2

  – R^2_a 'Adjusted R squared'

  – Lots more available based on likelihood, rather than SS: AIC, BIC, WAIC, DIC, etc. (more next term)

  – Complicated ones available based on "marginal likelihood" or "model evidence" via Bayesian methods: Bayes Factor

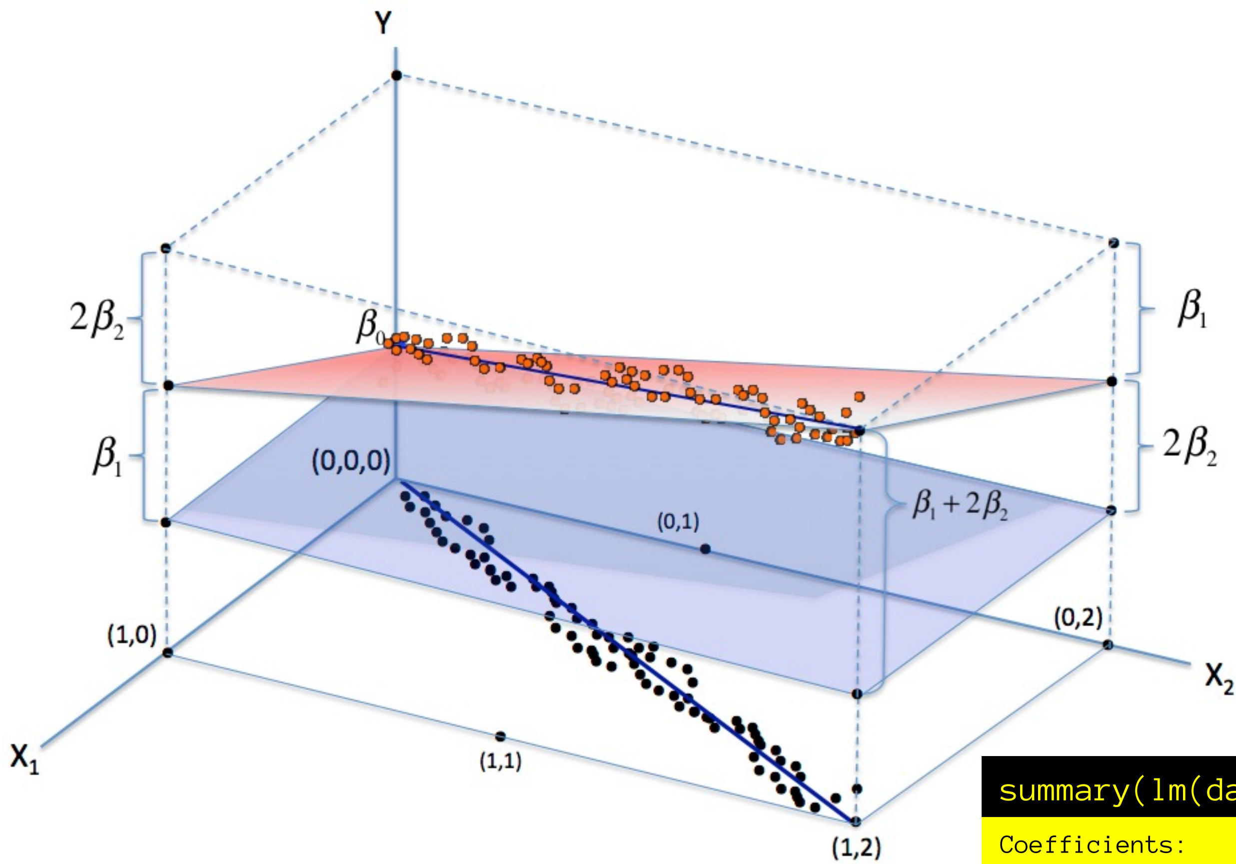
  – They all define some trade off between number of parameters and fit to the data.
    - Sometimes they will give you different answers! If so, you should be worried. A clearly better model should do better on all of these metrics. When different metrics give you different answers you should not be confident.

Multiple regression agenda

- What is it? And why do this?
- Multicolinearity & its consequences
- Sums of squares partitioning in multiple regression
- Different hypothesis tests in multiple regression
- Nested model comparison
- Non-nested models

Regression

$$\boxed{Y_i} = \boxed{\beta_0} + \boxed{\beta_1}\boxed{X_{1i}} + \boxed{\beta_2}\boxed{X_{2i}} + \boxed{\varepsilon_i} \qquad \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$



Coefficients:

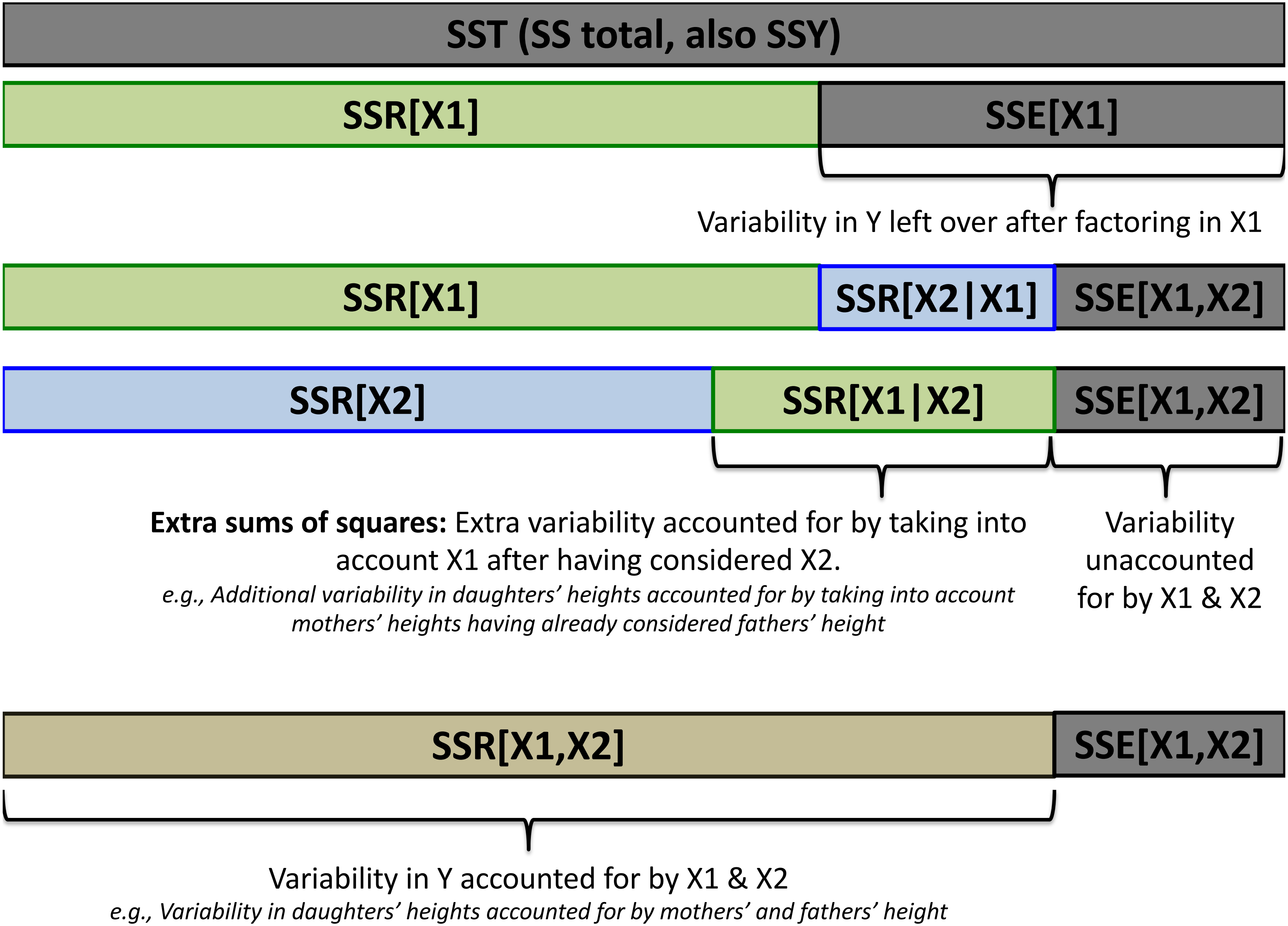
- Partial slope: dY/dX_j holding other Xs constant.

Multicolinearity:

- Correlation among predictors.
- Credit assignment is uncertain
- Coefficients change; are sensitive to model and noise; have higher marginal errors.

```
summary(lm(daughter~dad+mom))
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.7872     4.6471   0.815 0.417082
mom           0.5210     0.1164   4.477 2.06e-05 *
dad           0.3900     0.1078   3.617 0.000475 *
```


SST (SS total, also SSY)
SSR[X1]
SSE[X1]
Variability in Y left over after factoring in X1
SSR[X1]
SSR[X2|X1]
SSE[X1,X2]
SSR[X2]
SSR[X1|X2]
SSE[X1,X2]
Extra sums of squares: Extra variability accounted for by taking into account X1 after having considered X2.
e.g., Additional variability in daughters' heights accounted for by taking into account mothers' heights having already considered fathers' height
Variability unaccounted for by X1 & X2
SSR[X1,X2]
SSE[X1,X2]
Variability in Y accounted for by X1 & X2
e.g., Variability in daughters' heights accounted for by mothers' and fathers' height

SST (SS total, also SSY)

SSR[X1,X2] | SSE[X1,X2]

$$F(df_{term}, df_{error}) = \frac{\left(\frac{SS_{term}}{df_{term}}\right)}{\left(\frac{SSE_{FULL}}{df_{error}}\right)}$$

SS: Sum of squares for this term

d.f. of regression term: *# parameters of this term*

SSE: Sum of squared residuals

d.f. error: *n minus # parameters in full model*

SSR[X1] | SSR[X2|X1] | SSE[X1,X2]

```
anova(lm(son~mom+dad))

Response: son
          Df Sum Sq Mean Sq F value  Pr(>F)
mom        1 79.523  79.523 15.3977 0.00572 **
dad        1  9.225   9.225  1.7862 0.22320
Residuals  7 36.152   5.165
```

SSR[X2] | SSR[X1|X2] | SSE[X1,X2]

```
anova(lm(son~dad+mom))

Response: son
          Df Sum Sq Mean Sq F value   Pr(>F)
dad        1 79.595  79.595 15.4116 0.005707 **
mom        1  9.153   9.153  1.7723 0.224818
Residuals  7 36.152   5.165
```

Extra sums of squares of full compared to reduced

Extra parameters in full model

$$F(p_{FULL}-p_{REDUCED}, n-p_{FULL}) = \frac{\left(\dfrac{SSE_{REDUCED}-SSE_{FULL}}{p_{FULL}-p_{REDUCED}}\right)}{\left(\dfrac{SSE_{FULL}}{n-p_{FULL}}\right)}$$

Remaining sums of squares error in full model

n minus number of parameters in full model

SSR[X1] (SS regression var 1) | **SSE[X1]**

```
anova(lm(y~x1))

             Df Sum Sq Mean Sq F value    Pr(>F)
x1          1 517.18  517.18  64.373 2.263e-12 *
Residuals 98 787.34    8.03
```

SSR[X1,X2,X3] | **SSE[X1,X2,X3]**

SSX[X2,X3|X1]

```
anova(lm(y~x1+x2+x3))

            Df Sum Sq Mean Sq F value    Pr(>F)
x1          1 517.18  517.18  545.73 < 2.2e-16 *
x2          1 460.22  460.22  485.62 < 2.2e-16 *
x3          1 236.15  236.15  249.19 < 2.2e-16 *
Residuals 96  90.98    0.95
```

```
anova( lm(y~x1) , lm(y~x1+x2+x3) )

Model 1: y ~ x1
Model 2: y ~ x1 + x2 + x3

  Res.Df    RSS Df  Sum of Sq       F      Pr(>F)
1     98 787.34
2     96  90.98  2     696.37   367.4   < 2.2e-16 *
```

Significance in regression

- Pairwise correlation t-test.
  – Is there a significant linear relationship between Y and X_j ignoring other predictors?
- Coefficient t-test.
  – Does the partial slope dY/dX_j controlling for all other predictors differ significantly from zero?
- Variance-partitioning F-tests.
  – Is the sums of squares allocated to this term (depends on order, SS type) significantly greater than chance?
- Nested model comparison F-tests.
  – Does the larger model account for significantly more variance than the smaller model?

In some special cases, these end up equivalent.

Fat

```
readr::read_tsv('http://vulstats.ucsd.edu/data/bodyfat.data2.txt')
```

What variables predict bodyfat percentage?

- We have a bunch of very correlated predictors; how can we make new variables to orthogonalize them?
- What's a good model to predict bodyfat percentage?
- What would we predict is the bodyfat percentage of someone who is:
  - Height: 69
  - Weight: 175
  - Neck: 36
  - Chest: 100
  - Abdomen: 90
  - Hip: 99
  - Thigh: 58
  - Knee: 38
  - Ankle: 22
  - Bicep: 31
  - Forearm: 28
  - Wrist: 17