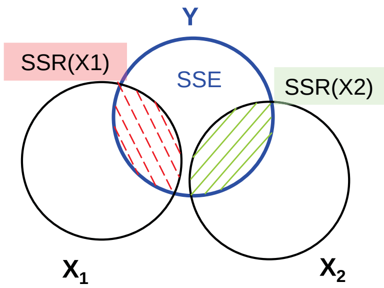
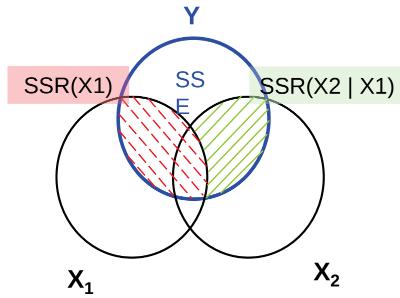
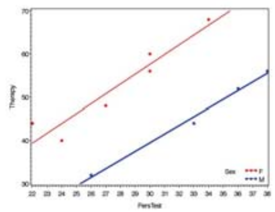
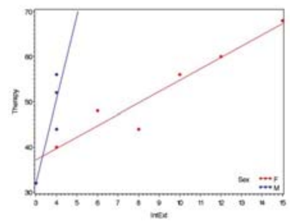




Regression: Model assessment

Psychology 6140





Topics

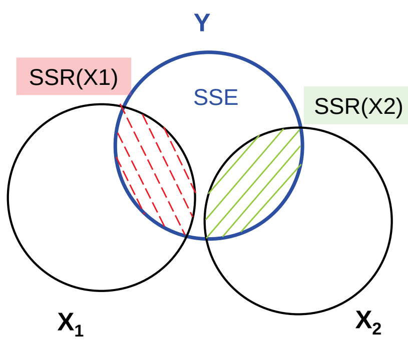
- How to assess the contributions of individual predictors to a regression model?
  - Type I (sequential) tests: added contribution of each new variable, in a given order
  - Type II (partial) tests: unique contribution of each variable, above all others
- More general methods: the general linear test: H_0: $\mathbf{L\beta=0}$
- The Marginality principle: always include low-order relatives
  - Testing hierarchical (ordered subset) models
  - Moderator variables: interaction effects

Uncorrelated predictors

- When predictors are uncorrelated, their SSR are additive

SSR(X1 X2) = SSR(X1)
+ SSR(X2)

- This makes it easy to see & test the contributions of each predictor



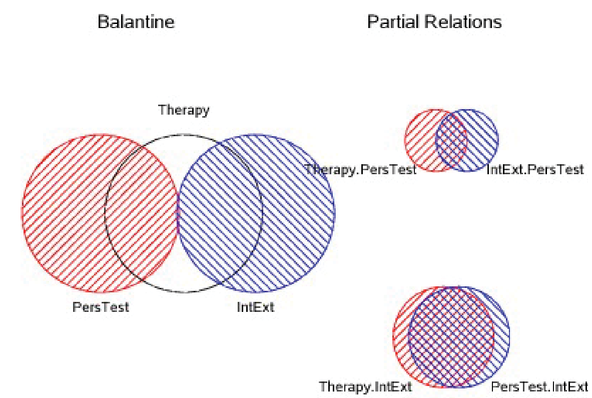
This “Ballentine” diagram shows variance or SS by areas of circles

Example: therapy data

For the therapy data, PersTest and IntExt turn out to be nearly uncorrelated.

Each have modest correlations with Therapy, but jointly account for 92%

In this example, IntExt acts as a suppressor variable for the test of PersTest, by removing effect of IntExt from error variance

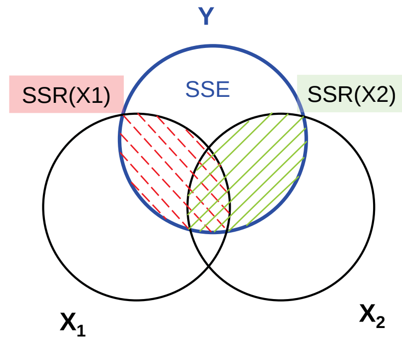


Correlated predictors

- Typically, predictors are correlated
- So, the portions of variance of Y they account for overlap

  SSR(X1 X2) < SSR(X1)
  + SSR(X2)

- How to assess contribution of each X?

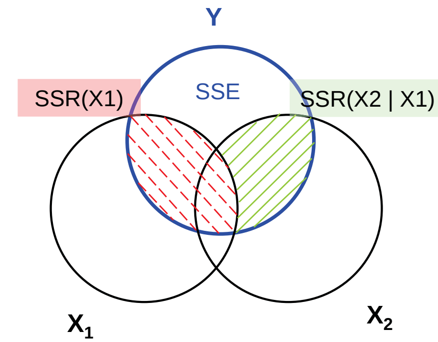


Sequential & partial SS

- Sequential SS (Type I)
  - 1st variable accts for all it can
  - Each next var: only what is left over
- ∴ contributions are additive

  SSR(X1 X2) = SSR(X1)
  + SSR(X2 | X1)

- **Only** useful if there is a reason for ordering variables
  - e.g., polynomial models
  - e.g., hierarchical models



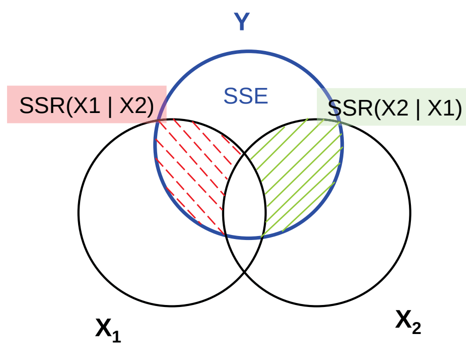
H_0: $\beta_1 = 0$ (ignoring X2, X3)

H_0: $\beta_2 = 0$ (adjusting for only X1)

H_0: $\beta_3 = 0$ (adjusting for X1, X2)

Sequential & partial SS

- Partial SS (Type II)
  - Each var accts for its unique contribution
  - Q: can we delete X_i given that all others are included?
  - t = b_i / s(b_i) is a partial test
- These are most generally useful, *except* where there is a hierarchical ordering of predictors
- In ANOVA designs there are also Type III (and IV) tests (take empty cells into account)



H_0: $\beta_1 = 0$ (adjusting for X2, X3)

H_0: $\beta_2 = 0$ (adjusting for X1, X3)

H_0: $\beta_3 = 0$ (adjusting for X1, X2)

Multiple regression: therapy data

```
proc reg data=therapy;
   model therapy = perstest intext sx;
   run;
```

Dummy (0/1) for sex
sx=1 for female here

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	982.05152	327.35051	109.43	<.0001
Error	6	17.94848	2.99141		
Corrected Total	9	1000.00000			

Root MSE	1.72957	R-Square	0.9821
Dependent Mean	50.00000	Adj R-Sq	0.9731
Coeff Var	3.45914		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-14.79157	5.22575	-2.83	0.0299
PERSTEST	1	1.71897	0.17268	9.95	<.0001
INTEXT	1	0.96956	0.25620	3.78	0.0091
SX	1	10.72600	2.40251	4.46	0.0043

Partial tests

Sequential vs. partial tests

```
proc reg data=therapy;
   model therapy = perstest intext sx / SS1 SS2;
   run;
```

```
 Parameter Estimates

                                         Parameter      Standard
 Variable    Label                  DF    Estimate         Error   t Value

 Intercept   Intercept               1   -14.79157       5.22575     -2.83
 PERSTEST    Personality Test Score  1     1.71897       0.17268      9.95
 INTEXT      Internal External scale 1     0.96956       0.25620      3.78
 SX          Sex                     1    10.72600       2.40251      4.46

                         Parameter Estimates

Variable    Label                  DF   Pr > |t|     Type I SS    Type II SS

Intercept   Intercept               1     0.0299         25000      23.96662
PERSTEST    Personality Test Score  1     <.0001     360.00000     296.42129
INTEXT      Internal External scale 1     0.0091     562.42744      42.84039
SX          Sex                     1     0.0043      59.62408      59.62408
```

F = SSR / MSE gives the test statistic for each hypothesis

Type I SS: SSR(X1), SSR(X2 | X1), SSR(X3 | X1 X2)

Type II SS: SSR(X1 | X2 X3), SSR(X2 | X1 X3), SSR(X3 | X1 X2)

Model comparison

- All statistical tests resolve to comparisons between two models
  - E.g., simple linear regression: H_0: $\beta_1 = 0$ vs H_a: $\beta_1 \neq 0$
    - Full model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$
    - Reduced model: $y_i = \beta_0 + \varepsilon_i$
    - Test:

$$F^* = \frac{(SSR_{full} - SSR_{reduced}) / (df_{full} - df_{reduced})}{MSE_{full}} = \frac{SS_{hyp} / df_{hyp}}{SS_E / df_E}$$

  - More generally, we can compare any larger model to a subset model, using the extra sum of squares, e.g.,

    H_0: $\beta_3 = \beta_4 = 0$

    $SSR(X_3\ X_4 \mid X_1\ X_2) = \underbrace{SSR(X_1\ X_2\ X_3\ X_4)}_{\text{full}} - \underbrace{SSR(X_1\ X_2)}_{\text{reduced}}$

Testing composite hypotheses

```
proc reg;
   model therapy = perstest intext sx;
   test intext, sx;
   run;
```

Test $\beta_2 = \beta_3 = 0 \mid \beta_1$

```
           Test 1 Results for Dependent Variable THERAPY

                                     Mean
     Source              DF        Square    F Value    Pr > F

     Numerator            2     311.02576     103.97    <.0001
     Denominator          6       2.99141
```

$SSR(X_2\ X_3 \mid X_1) = SSR(X_1\ X_2\ X_3) - SSR(X_1) = 982.05 - 360 = 622.05$

so

$$F^* = \frac{(SSR_{x_1,x_2,x_3} - SSR_{x_1}) / (3-1)}{MSE_{x_1,x_2,x_3}} = \frac{622.05 / 2}{2.99} = 103.97$$

General Linear Hypothesis Tests

- Even more generally, any hypothesis test can be regarded as an example of a GLH of the form

$$H_0 : \underset{q \times (p+1)}{\mathbf{L}} \ \underset{p+1}{\boldsymbol{\beta}} = \underset{q}{\mathbf{0}}$$

where the hypothesis matrix, **L**, contains specified constants and is of rank q = df for hypothesis

- e.g,

Test $\beta_2 = \beta_3 = 0$

$$\mathbf{L\beta} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

SAS syntax: **test** intext, sx;

Test $\beta_1 - \beta_2 = 0$

$$\mathbf{L\beta} = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0$$

SAS syntax: **test** perstest - intext;

General Linear Hypothesis Tests

- In all cases, the sums of squares for the hypothesis, H_0: $\mathbf{L}\ \boldsymbol{\beta} = \mathbf{0}$ has the same form,

$$SS_{hyp} = (\mathbf{Lb})^T \ \mathbf{L}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{L}^T \ (\mathbf{Lb})$$

This measures the squared distance of $\mathbf{L}\ \boldsymbol{\beta}$ from $\mathbf{0}$

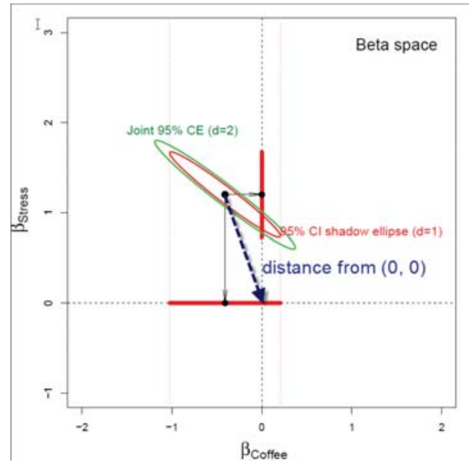
- GLH tests extend in a natural way to
  - MANOVA, MMReg: $\mathbf{Y} = \mathbf{X}\ \mathbf{B} + \mathbf{E}$

$$\mathbf{L}\ \mathbf{B} = \mathbf{0}$$

  - Repeated measures designs

$$\mathbf{L}\ \mathbf{B}\ \mathbf{M} = \mathbf{0}$$

Example: Heart disease, coffee and stress



In the model
lm(Heart ~ Coffee + Stress, data=coffee)

Test: H_0: $\beta_{Coffee} = \beta_{Stress} = 0$

$(X'X)^{-1}$ is covariance matrix of $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$

$\mathbf{L} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ selects β_1, β_2

$$(\mathbf{Lb})^T \ \mathbf{L}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{L}^T \ (\mathbf{Lb})$$

is squared distance of **b** from (0,0)

Marginality principle

- Any model including a *high-order* term should normally include all *low-order* relatives
  - Interactions: Perstest * Sex → Perstest + Sex ("main effects")
  - Polynomial models: $X^3 \rightarrow X + X^2$
- We can neither test nor interpret main effects of variables that interact
  - $X_1 * X_2 \rightarrow$ effect (β_1) of X_1 varies with X_2
- Similarly, if X^3 is important, X and X^2 must remain in the model (even if NS!)

Hierarchical testing

- Variables in regression are sometimes ordered in terms of research questions & hypotheses
  - Include necessary control variables (age, IQ)
  - Test for effects of new predictor(s) beyond old ones
- In such cases, do hierarchical (blockwise) tests

```
proc reg data=mydata;
     var y age IQ  reading math  depression anxiety;
  block1:    model y = age IQ;
  block2:    model y = age IQ  reading math;
             test reading, math;
  block3:    model y = age IQ  reading math  depression anxiety;
             test depression, anxiety;
```

Moderator variables

- A **moderator effect** occurs when the effect of one variable, x_1, depends on, or varies with variable, x_2.
  - i.e., interaction of x_1 and x_2.
  - i.e., slope (b_1), for x_1 varies with x_2.
- In regression, this is modeled by including the product, $x_1 * x_2$ in the model

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 (x_1 \times x_2) + \epsilon$$
$$= b_0 + (b_1 + b_3 x_2) x_1 + b_2 x_2 + \epsilon$$

  - In SAS proc reg & SPSS must calculate $x_1 * x_2$ explicitly
  - (often useful for interpretation to center x_1, x_2)
  - Must include x_1 and x_2 (marginality)
  - Must test $x_1 * x_2$ by partial test
  - Conclude no moderator effect if b_3 is non-significant


```
data therapy; set therapy;
  PTxSX = perstest * sx;
  IExSX = intext * sx;
*-- test moderator of Sex on Perstest;
proc reg data=therapy;
   model therapy = perstest sx PTxSX;
   run;
```

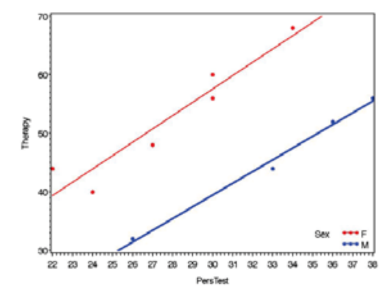
calculate products

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-20.70091	11.41755	-1.81	0.1198
PERSTEST	PersTest	1	2.00604	0.34022	5.90	0.0011
SX	Sex	1	9.94015	14.44245	0.69	0.5170
PTxSX		1	0.27279	0.46331	0.59	0.5775

There is no evidence that the slope for PersTest varies with Sex

The slope for females is only 0.27 less than that for males

Interpreting such models is easiest if you plot the fitted relationships

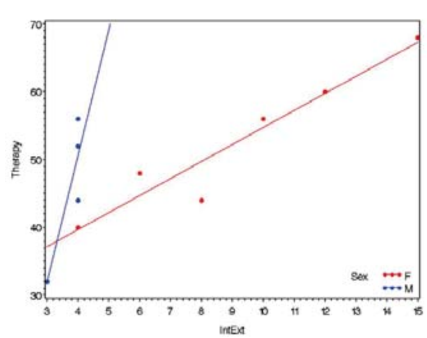



```
*-- test moderator of Sex on IntExt;
proc reg data=therapy;
   model therapy = intext sx IExSX;
   run;
```

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-24.00000	19.53597	-1.23	0.2653
INTEXT	IntExt	1	18.66667	5.17520	3.61	0.0113
SX	Sex	1	53.60825	20.14653	2.66	0.0375
IExSX		1	-16.15120	5.19916	-3.11	0.0209

There is strong evidence that the slope for IntExt varies with Sex

The slope for females is 16.15 less than for males



Arguably, it might be better to test the full model, with both interactions:

```
*-- test both moderators with sex;
proc reg data=therapy;
   model therapy = perstest intext sx PTxSX IExSX;
   test PTxSX, IExSX;
   run;
```

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-16.73684	3.48499	-4.80	0.0086
PERSTEST	PersTest	1	2.42105	0.22056	10.98	0.0004
INTEXT	IntExt	1	-4.73684	2.31673	-2.04	0.1104
SX	Sex	1	22.97128	4.53565	5.06	0.0072
PTxSX		1	-1.22461	0.26226	-4.67	0.0095
IExSX		1	6.16934	2.32193	2.66	0.0566

Joint test for both interactions: Do I need any interactions with sex?

Source	DF	Mean Square	F Value	Pr > F
Numerator	2	7.74189	12.56	0.0189
Denominator	4	0.61618		

Testing and comparing models in R

- In R, fit a model using `mod <-lm(y ~ x1+x2+ ...)`
- Test terms in that model using `summary(mod)`
- Type II *F*-tests with `car::Anova(mod)`
- Compare models using `anova(mod1, mod2, ...)`
- Linear hypotheses: `car::linearHypothesis()`

```
mod1 <- lm(therapy ~ perstest, data= therapy)
mod2 <- lm(therapy ~ perstest + intext, data=therapy)
mod3 <- lm(therapy ~ perstest + intext + sex, data=therapy)
summary(mod3)
Anova(mod3)  # F tests

# test interactions
mod4 <- lm(therapy ~ perstest*sex + intext*sex, data=therapy)

# compare models
anova(mod1, mod2, mod3, mod4)
```



```
> summary(mod3)

Call:
lm(formula = therapy ~ perstest + intext + sex, data = therapy)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.0656     3.7572  -1.082  0.32078
perstest      1.7190     0.1727   9.954 5.94e-05 ***
intext        0.9696     0.2562   3.784  0.00913 **
sexM        -10.7260     2.4025  -4.464  0.00426 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.73 on 6 degrees of freedom
Multiple R-squared: 0.9821,     Adjusted R-squared: 0.9731
F-statistic: 109.4 on 3 and 6 DF,  p-value: 1.256e-05
```

Partial *t*-tests for coefs in mod3

```
> anova(mod1, mod2, mod3, mod4)
Analysis of Variance Table

Model 1: therapy ~ perstest
Model 2: therapy ~ perstest + intext
Model 3: therapy ~ perstest + intext + sex
Model 4: therapy ~ perstest * sex + intext * sex
  Res.Df    RSS Df Sum of Sq       F    Pr(>F)
1      8 640.00
2      7  77.57  1    562.43 912.770 7.149e-06 ***
3      6  17.95  1     59.62  96.765 0.0005989 ***
4      4   2.46  2     15.48  12.564 0.0188571 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hierarchical tests of mod_i vs. mod_{i-1}

These assume nested models


```
> coef(mod3)
(Intercept)    perstest      intext        sexM
  -4.065574    1.718970    0.969555  -10.725995
> linearHypothesis(mod3, c("intext", "sexM"))
Linear hypothesis test

Hypothesis:
intext = 0
sexM = 0

Model 1: restricted model
Model 2: therapy ~ perstest + intext + sex

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1      8 640.00
2      6  17.95  2    622.05 103.97 2.206e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Linear hypotheses for 1 or more coefficients in mod3

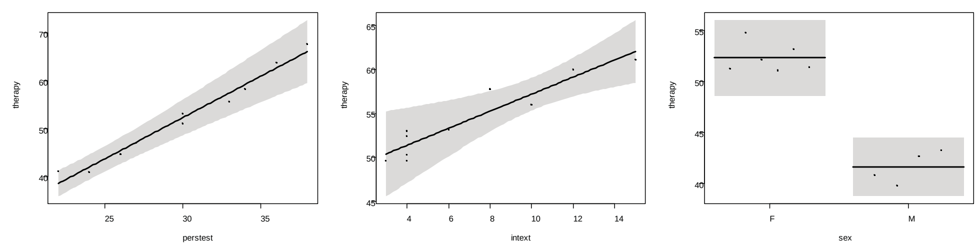
```
> tests <- matchCoefs(mod4, ":")
> linearHypothesis(mod4, tests)
Linear hypothesis test

Hypothesis:
perstest:sexM = 0
sexM:intext = 0

Model 1: restricted model
Model 2: therapy ~ perstest * sex + intext * sex

  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1      6 17.9485
2      4  2.4647  2    15.484 12.564 0.01886 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test all interactions (":" in name) in mod4

Visualizing model effects

- In R, the `effects` and `visreg` packages make it easy to visualize the effects of terms in models

```
> library(visreg)
> visreg(mod3)
```

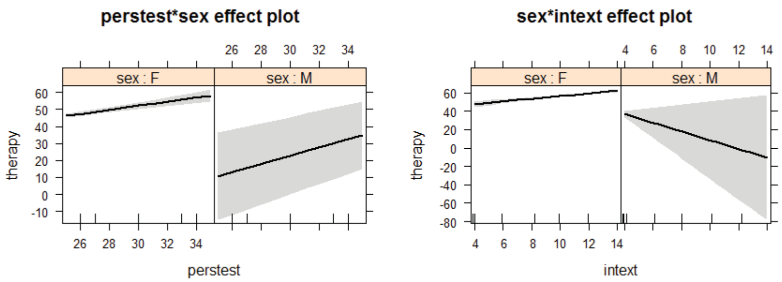


Conditional plots for each predictor, setting all others to their median value

Visualizing model effects

- The **effects** package is most useful for plotting models with interactions

```
> library(effects)
> plot(allEffects(mod4))
```



Conditional plots for each high-order term, setting others to mean value

Summary

- Sequential (Type I) and Partial (Type II) SS provide different ways of assessing the contribution of a given predictor
  - Type I: added contribution of each new variable, in order
  - Type II: added contribution of each variable above all others
- Each of these essentially give a test comparing a "full" model against a "reduced" model
- This idea extends to the General Linear Test, H_0: $\mathbf{L\beta=0}$

Summary

- Tests of complex models must respect the Marginality Principle– include low-order relatives
- Testing hierarchical models and moderator variables are examples of these ideas.
- Always important to plot model terms for interpretation
- We will consider model selection problems more generally later