Multiple Regression

Ψ320
Ainsworth

More Hypothesis Testing

- What we really want to know:
  - Is the relationship *in the population we have selected* between X & Y strong enough that we can use the relationship to make predictions about Y based on X.
- What we actually know:
  - The extent of the relationship between X and Y made on a calculations of a single sample drawn from our population.

More Hypothesis Testing

- The big question:
  - Is the relationship between X & Y in our sample SO STRONG, that it is unlikely that the relationship in the population is 0.
- We know:
  - Sampling Error will give us measures of the relationship (r and b) that will vary from sample to sample.
- BUT, if the true relationship is not zero, most of our r's and b's will be non-zero.

More Hypothesis Testing

- So, hypothesis testing on regression is of the form:

Hypotheses for correlation:

$$h_0 : \rho = 0 \quad or \quad h_0 : \rho \geq 0 \quad or \quad h_0 : \rho \leq 0$$

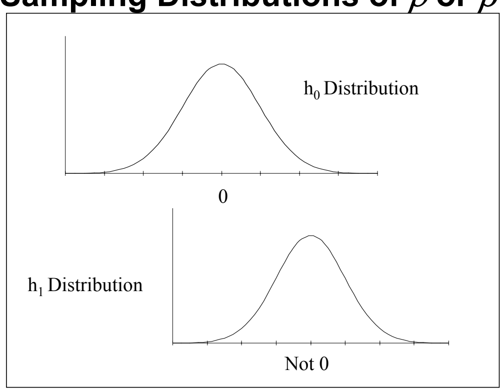
$$h_1 : \rho = 0 \quad or \quad h_1 : \rho < 0 \quad or \quad h_1 : \rho > 0$$

Hypotheses for regression:

$$h_0 : \beta = 0 \quad or \quad h_0 : \beta \geq 0 \quad or \quad h_0 : \beta \leq 0$$

$$h_1 : \beta = 0 \quad or \quad h_1 : \beta < 0 \quad or \quad h_1 : \beta > 0$$

Sampling Distributions of ρ or β



Multiple Regression

- Using several predictors to predict a single dependent variable
- Hypothesis Testing involves:
  - Finding a measure of overall fit
  - Testing each predictor

An Example

- Study by Kliewer et al. (1998) on effect of violence on internalizing behavior
  - What is internalizing behavior?
- Predictors
  - Degree of witnessing violence
  - Measure of life stress
  - Measure of social support

Violence and Internalizing

- Children 8-12 years in high-violence areas
  - Hypothesis: violence and stress lead to internalizing behavior.

Correlations

	Pearson Correlation			
	Witnessed Violence	Social Support	Stress	Internalizing
Witnessed Violence	1	.080	.048	.201*
Social Support	.080	1	-.078	-.173
Stress	.048	-.078	1	.268**
Internalizing	.201*	-.173	.268**	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

- Note: both Stress and Witnessing Violence are significantly correlated with Internalizing.
- Note: that predictors are largely independent of each other.

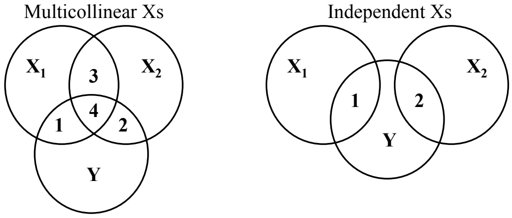
Multiple Regression Terms

- Multiple Correlation Coefficient
  - also known as the Multiple Correlation or R
  - The correlation of Y with **a set of** X variables.
- Multicollinearity
  - What happens when your set of X variables are correlated with each other?
  - Complicates analysis
  - **For this class assume Xs are <u>independent</u>**

Multiple Correlation

- Directly analogous to simple *r*
- Always capitalized (e.g. *R*)
- Always positive
  - Correlation of $\hat{Y}$ with observed *Y*
    - where $\hat{Y}$ is computed from regression equation
    - It is the correlation between all Xs and Y simultaneously
  - Often reported as R^2 instead of *R*

R^2



- X_1, X_2, and Y represent the variables. The numbers reflect variance overlap as follows:
  1. Proportion of Y uniquely predicted by X_1
  2. Proportion of Y uniquely predicted by X_2
  3. Proportion of variance shared by X_1 and X_2
  4. Proportion of Y redundantly predicted by X_1 and X_2

R^2

- When Xs are independent

$$R^2 = \sum_{k=1}^{\#\text{ of X}s} (r_{Y.X_k})^2 \text{; then R is simply } \sqrt{R^2}$$

- When Xs are multicollinear

$$R^2 \neq \sum_{k=1}^{\#\text{ of X}s} (r_{Y.X_k})^2$$

R^2 for the Example

- Assuming Independent Xs

$$R^2 = \sum_{k=1}^{\#\text{ of X}s} (r_{Y.X_k})^2 = .201^2 + -.173^2 + .268^2 =$$

$$R^2 = .040 + .030 + .072 = .143 \text{ and } \mathrm{R} = \sqrt{.143} = .378$$

- From SPSS

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.368[a]	.136	.108	2.21737

a. Predictors: (Constant), STRESS, VIOLWIT, SOCSUPP

b. Dependent Variable: INTERNAL

Regression Coefficients

- Multiple Slopes (1 for each X) and one intercept.
- With independent Xs the Bs are calculated in the same way as in simple regression
- The intercept is now

$$a = \overline{Y} - b_1\overline{X}_1 + b_2\overline{X}_2 + b_3\overline{X}_3 + \cdots$$

Slopes and Intercept

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.517	1.288		.401	.689
	VIOLWIT	.038	.018	.202	2.111	.037
	SOCSUPP	-.076	.043	-.170	-1.766	.081
	STRESS	.272	.106	.245	2.560	.012

a. Dependent Variable: INTERNAL

$$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_3 + a$$

$$= (0.038 * ViolWit) - (0.076 * SocSupp) + (0.272 * Stress) + 0.517$$

Interpretation

- Note slope for Witness and Stress are positive, but slope for Social Support is negative.
  - Does this make sense?
- If you had two subjects with identical Stress and SocSupp, a one unit increase in Witness would produce 0.038 unit increase in Internal.

Interpretation

- The interpretation holds true for other predictors as well
- R^2 has same interpretation as r^2
  - 13.6% of variability in Internal accounted for by variability in Witness, Stress, and SocSupp.

Predictions

- Assume a participant's scores were: Witness = 35, SocSupp = 20 and Stress = 5

$$\hat{Y} = (0.038 * ViolWit) - (0.076 * SocSupp) + (0.272 * Stress) + 0.517$$
$$= .038(35) - .076(20) + .272(5) + 0.517$$
$$= 1.33 - 1.52 + 1.36 + .517$$
$$= 1.69$$

- We would predict 1.69 as the level of Internal for that participant.

Residuals

- Residuals are calculated in the same way as in simple regression
- Error variance and standard error of estimate are also calculated the same way
  - Note: R^2 is used instead of r^2 in all formulas using r^2 and $1- r^2$

Partitioning Variance

- All of the SS (regression, error and total) are also calculated the same as in simple regression
  - Note: R^2 is used instead of r^2 in all formulas using r^2 and $1- r^2$
- Testing for the overall prediction makes more sense now and is not as redundant with the test of the Bs
- As before, $s^2_{regression}/s^2_{residual} = F$

Hypothesis Overall Prediction

- Test of overall prediction and R^2 given in Analyzing the of Variance as before

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	73.320	3	24.440	4.971	.003[a]
	Residual	467.090	95	4.917		
	Total	540.410	98			

a. Predictors: (Constant), STRESS, VIOLWIT, SOCSUPP

b. Dependent Variable: INTERNAL

These are the variances

Testing Slopes and Intercept

- Tests on regression coefficients given along with the coefficients.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.517	1.288		.401	.689
	VIOLWIT	.038	.018	.202	2.111	.037
	SOCSUPP	-.076	.043	-.170	-1.766	.081
	STRESS	.272	.106	.245	2.560	.012

a. Dependent Variable: INTERNAL

- *t* test on two slopes are significant

Intercept These are the Bs These are the t-values