

7963 Review (Exam is scheduled for March 20th)

1. What assumptions are made in OLS regression? What are the consequences of violating each assumption?
2. How is each parameter estimate interpreted in multiple regression? What is meant by the "true" regression equation? The estimated regression equation?
3. What criterion does OLS use for fitting a line?
4. How and why would one 'center' variables? What is the consequence of centering variables?
5. What does the variance (or standard error) of the slope estimate measure?
If the data points are more tightly compacted about the OLS line, what is the effect on the variance of the slope estimate?
If the X's are more spread out about the mean of X, what is the effect of the variance of the slope estimate?
As the sample size increases, what is (generally) the effect on the variance of the slope estimate?
6. What does it mean to say that the slope estimate is 'biased'? 'Inefficient'?
7. How is the R^2 computed? What does the R^2 measure? How is the R^2 sometimes misused? What are some of the limitations of the R^2 ?
8. How are standardized estimates (sometimes called beta weights) interpreted? What are some of the limitations of standardized estimates?
9. Why do some researchers force the intercept to 0? What are some reasons that you might advise a researcher *not* to force the intercept to 0? How can you tell if your estimated intercept is significantly different from zero?
10. Why do we occasionally use 'functional transformations' of the independent variable? What, for instance, is the purpose of the log transformation – that is, what relationship between the dependent and independent variable does it capture? (You should also be familiar with the X^2 and square root of X transformations). In general, how can one interpret the slope coefficients of transformed variables in order to maximize a reader's understanding of the relationship?
11. Describe the meaning of confidence intervals for β . How would you compute (i.e., what is the formula) a two-tailed CI for β ? How are confidence intervals interpreted?
12. Describe the meaning of confidence intervals for $E(Y_0)$. Describe the meaning of predictive intervals for Y_0 . Define extrapolation, and explain why extrapolation is risky.
13. You should be able to identify such problems as non-linearity, autocorrelation, heteroskedasticity, etc. from various plots.
14. How many parameters are estimated in a multivariate regression equation? How many degrees of freedom are there when estimating a multivariate regression equation?

15. How are t-statistics computed? When should a researcher use a one-tailed test of significance? Using p-values, how would one test the hypothesis that $\beta \neq 0$? Using p-values, how would one test the hypothesis that $\beta > 0$?

16. What is the purpose of the adjusted R^2 ? What is the standard error of regression (or the standard error of the estimate, or the root mean square error)? What is the relationship between the two measures?

17. Consider the following model of support for Clinton versus (George Herbert Walker) Bush in the 1992 election. The data were gathered through the 1992 National Election Study survey. The dependent variable is based on two thermometer scores: one for Clinton (coded 0-100, 0 indicating 'very cold' toward Clinton), and one for Bush (coded 0-100, 0 indicating 'very cold' toward Bush). The dependent variable itself is calculated by subtracting the Bush thermometer score from the Clinton thermometer score, and ranges from -100 to 100 (-100 indicating that the respondent liked Bush *much* more than Clinton; 100 indicating that the respondent liked Clinton *much* more than Bush).

The following independent variables were used to predict the dependent variable:

- Conserve: Ranges from 1-7; higher values indicate more conservative.
- Repub: Ranges from 1-7; high values indicate more Republican.
- Female: Coded 1 if female, 0 otherwise.
- Black: Coded 1 if African-American, 0 otherwise.
- Married: Coded 1 if married, 0 otherwise.
- Age: Ranges from 17 to 91.
- Income: Ranges from 1-24; higher values indicate higher income.
- Educate: ranges from 1-7; higher values indicate more highly educated.
- Femrepub: Multiplicative term between female and Republican.

Variable	Parameter Estimate	Std. Error	P-Value
Conserve	-6.580	.607	.000
Repub	-11.784	.566	.000
Female	1.505	3.170	.635
Black	2.942	2.587	.256
Married	-3.730	1.562	.017
Age	.067	.046	.143
Income	-.238	.135	.078
Educate	.615	.508	.226
Femrepub	-1.392	.728	.056
Constant	76.226	4.194	.000

- A researcher expects a negative relationship between Republicanism and the dependent variable; that is, the more Republican a respondent is, the more they like Bush relative to Clinton. She also expects that women will be significantly more likely than men to like Clinton more than Bush. She expects this difference between the sexes to be particularly marked among Republicans than among Democrats. Are her expectations upheld? Explain why or why not.

- What is the predicted value of difference between thermometer scores for a female Republican? Female Democrat? Male Republican? Male Democrat?

(Constant=1. Hold other variables at their mean:

Black=.1280

Married=.5441

Age=45.7553

Income=10.2482

Educate=3.8136)

- Graph the researcher's expectations regarding the variables female, Repub, and femrepub; graph the results.

18. What is multicollinearity? Why might one suspect multicollinearity? How can one diagnose multicollinearity? Assess some solutions to multicollinearity. What are the consequences of multicollinearity?

19. Suppose one wanted to test the effect of religion on feeling toward George W. Bush. All of the cases (i.e., voters) have been divided up into four mutually exclusive categories: Catholic, Protestant, Jewish, other. What is problematic with the following estimated equation:

$X_1 = 1$ if Catholic, 0 otherwise $X_2 = 1$ if Protestant, 0 otherwise
 $X_3 = 1$ if Jewish, 0 otherwise $X_4 = 1$ if other, 0 otherwise

$Y_i = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_i$

How would one solve this problem? How would one then interpret each slope coefficient?