Session 5B: Post Class Test Solutions

1. d. Start with variables that other researchers have found statistically significant and add other variables that improve explanatory power, but only if I have a theoretical or common-sense reason for including those variables. Even this can be problematic, if prior research is data mining, but if you restrict yourself to making judgments only about the variable or variables you add, you are also minimizing the damage.
2.
    a. I would use a market pricing variable, because I am trying to explain how the market is pricing companies. While total market capitalization is an option, it is in dollar values and the range across companies can be immense, making predictions messy. I would consider a standardized pricing metric, like price earnings or an enterprise value to sales multiple.
    b. I would start with an assessment of my chosen pricing multiple, and using basic valuation model or even common sense, extract what should cause this multiple to high for some companies and low for others. Thus, with EV to sales, I would expect companies with higher margins, lower costs of capital and higher growth to trade at higher levels than companies with lower margins, higher costs of capital and lower growth. I would look at my data to see if I can find variables that measure each (margins can be measured based upon net or operating profit) or proxy for them (stable earnings should lead to lower costs of capital).
    c. I would run scatter plots of the dependent variable against each independent variable to gauge both the explanatory power of that variable as well as whether the relationship is linear. If it is not, I would consider transforming that variable (natural log, square root, square etc.) to make it more linear.
    d. I would check the F statistic and overall R-squared, hoping to see high values. I would then check the t statistics and p values for individual independent variables.
    e. Before I use the regression, I would check the residuals to make sure that there are no patterns (heteroskedasticity) and conform to normality. If they don't I would consider running a weighted or generalized least square regression.
3. e. All of the above. All of the actions listed (plus more) are part of the p-hacking process. In the hands of a skilled researcher, the hacking can be difficult to detect, even if raw data has to be made available to other researchers to try to replicate findings.
4. The regression of PE against growth is below:

Regression Statistics								
Multiple R	0.95490433							
R Square	0.91184227							
Adjusted R S	0.90506091							
Standard Err	4.28714599							
Observations	15							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	2471.3783	2471.3783	134.462965	3.1484E-08			
Residual	13	238.93507	18.3796208					
Total	14	2710.31337						
	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4.00072159	2.60216266	1.53746023	0.14815748	-1.6209091	9.62235223	-1.6209091	9.62235223
Expected Gr	217.685697	18.7727784	11.5958167	3.1484E-08	177.129575	258.241819	177.129575	258.241819

The regression passes the statistical significant test, and PE ratios are clearly higher for higher growth companies. Every 1% increase in expected growth, on average, increases the PE by 2.177.

To bring in the effect of the founder/professional manager, I created a founder dummy:

	PE	*Expected Growth*	*Founder Dummy*	*Founder/Professional Mgt*
Tech Co 1	12.33	5.35%	0	Professional
Tech Co 2	14.60	3.80%	1	Founder
Tech Co 3	17.50	4.22%	1	Founder
Tech Co 4	19.00	8.05%	0	Professional
Tech Co 5	22.56	7.68%	1	Founder
Tech Co 6	25.48	9.35%	1	Founder
Tech Co 7	27.98	12.22%	0	Professional
Tech Co 8	29.04	10.90%	1	Founder
Tech Co 9	32.12	16.50%	0	Professional
Tech Co 10	34.77	14.44%	0	Professional
Tech Co 11	37.65	15.02%	0	Professional
Tech Co 12	39.18	19.19%	0	Professional
Tech Co 13	45.27	18.50%	1	Founder
Tech Co 14	50.35	20.33%	0	Professional
Tech Co 15	61.80	22.62%	1	Founder

The regression of PE against growth and the founder dummy is below:

Regression Statistics								
Multiple R	0.97769244							
R Square	0.95588251							
Adjusted R Square	0.94852959							
Standard Error	3.15663622							
Observations	15							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	2590.74115	1295.37057	130.00048	7.3734E-09			
Residual	12	119.572227	9.96435222					
Total	14	2710.31337						
	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.226274	2.27212331	-0.099587	0.92231657	-5.1768054	4.72425744	-5.1768054	4.72425744
Expected Growth	229.694142	14.2512425	16.1174818	1.7043E-09	198.643352	260.744932	198.643352	260.744932
Founder Dummy	5.82980608	1.68439547	3.46106729	0.00470788	2.15982361	9.49978855	2.15982361	9.49978855

The regression remains statistically significant, and the founder dummy is adding to that significant; the t statistics and p values back up that significance.

For tech company 1, with an expected growth rate of 5.3% and a professional money manager, the regression yields a predicted PE ratio of 12.06.
PE = -0.2263 + 229.69 (.0535) + 5.82 (0) = 12.06
The stock, trading at 12.33, is mildly over priced, but not by enough to be significantly so.

5. Using the regression, I estimate predicted PEG ratios for companies:

	Growth rate	*ln(Growth)*	*Predicted PEG*
Company A	25%	-1.3863	1.68
Company B	15%	-1.8971	1.62
Company C	2%	-3.9120	1.38

For a company that is fairly priced, with a PEG ratio of 1.60, you can back out the expected growth rate:

PEG ratio	1.60	
Ln(Expected growth Rate) =	-2.0833333	Solve for (1.60- 1l85)/.12
Expected growth rate	12.45%	Take exp(-2.08333)