

OLS Regression Results

Dep. Variable:	domestic_adj_gross		R-squared:	0.401	
Model:	OLS		Adj. R-squared:	0.397	
Method:	Least Squares		F-statistic:	90.78	
Date:	Tue, 04 Oct 2016		Prob (F-statistic):	3.61e-101	
Time:	18:37:21		Log-Likelihood:	-5808.1	
No. Observations:	957		AIC:	1.163e+04	
Df Residuals:	949		BIC:	1.167e+04	
Df Model:	7				
Covariance Type:	nonrobust				
	coef	std err	t	P> t 	[95.0% Conf. Int.]
Intercept	1.304e+04	839.203	15.543	0.000	1.14e+04 1.47e+04
production_budget	0.7566	0.075	10.046	0.000	0.609 0.904
genre_share	3.7236	0.979	3.804	0.000	1.803 5.644
coded_mpaa	-27.6228	4.449	-6.208	0.000	-36.354 -18.891
runtime_mins	1.6685	0.193	8.628	0.000	1.289 2.048
year	-6.5059	0.418	-15.552	0.000	-7.327 -5.685
month	-0.8670	1.049	-0.827	0.409	-2.925 1.191
day	-0.5398	0.414	-1.305	0.192	-1.352 0.272
Omnibus:	571.214		Durbin-Watson:	1.973	
Prob(Omnibus):	0.000		Jarque-Bera (JB):	8046.054	
Skew:	2.462		Prob(JB):	0.00	
Kurtosis:	16.324		Cond. No.	4.97e+05	

My model focuses on predicting domestic gross (adjusted to 2016 values) as a function of production budget, genre's market share, (coded) MPAA rating (recoded to reflect level of restriction, G = 1, R = 4), runtime minutes, year, month, and day.

This is only an initial fit (given the time constraint). Tomorrow, I plan on possibly dropping day, and splitting 'month' to separate indicator variables (or grouped by season).

Below is the residual plot, which shows that there is room for improvement, even though the residuals seem to be generally random and centered about zero.

