<table>
<tr><th colspan="3">Assumption</th><th colspan="3">Implications</th><th colspan="3">Assessment</th></tr>
<tr><td>Name</td><td>Mathematical Representation</td><td>Description</td><td colspan="3"></td><td>Evidence – Tool</td><td>How to Use</td><td>Mitigation</td></tr>
<tr><td>MLR1<br>(Gauss-Markov 1)</td><td>$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$</td><td>Linearity in Coefficients.<br>Weak assumption.</td><td rowspan="5">OLS Estimators Consistent. For large samples, $\beta \to \beta$</td><td rowspan="6">OLS Estimators Best Linear Unbiased Estimators (BLUE). $E\left(\hat{\sigma}^2\right) = \sigma^2$</td><td rowspan="7">$\beta_j \sim N(\beta_j, Var(\beta_j))$ , $\beta_j \sim N(\beta_j, Var(\beta_j))$ $t_{df} = (\beta_j - a_{H_0})/se(\beta_j) \sim t_{n-k-1}$<br>The OLS coefficients are normally distributed, so you can conduct t-tests, etc. on the values of the coefficients.</td><td colspan="3">The model is linear. Input variables can be of any form: log, polynomial, etc. only coefficients are assumed to be linear.</td></tr>
<tr><td>MLR2<br>(Gauss-Markov 2)</td><td></td><td>Random sample of data. Independent and Identically Distributed (iid). Strong.</td><td>Discern from sampling technique as described by researcher.<br>Check resid_v_leverage</td><td>Look for outliers with influence and leverage. Cook's distance lines.</td><td>Look hard at those points. Exclude carefully with strong evidence of errant data only.</td></tr>
<tr><td>MLR3<br>(Gauss-Markov 3)</td><td>There is no $x_i = \delta_0 + \delta_1 x_j$</td><td>No perfect co-linearity.<br>Weak assumption.</td><td>VIF Variance Inflation Factor<br>$VIF_j = \frac{1}{1 - R^2}$<br>vif(model)</td><td>If VIF is close to 1, you have multi-colinearity.</td><td>Remove one of the variables from the model.</td></tr>
<tr><td>MLR4<br>(Gauss-Markov 4)</td><td>$E(u \mid x_1, x_2, \ldots, x_k) = 0$</td><td>Zero conditional mean for errors. A strong assumption.</td><td></td><td></td><td></td></tr>
<tr><td>MLR4'</td><td>$Cov(u_i, x_i) = 0, all\ x$</td><td>All X uncorrelated with error.</td><td>Residuals v. Fitted values.<br>Scale-location plot.<br>plot(model1) gives both</td><td>Show that exogeneity is realistic. This preserves consistency of estimators.<br>$\lim_{n \to \infty} \hat{\beta} = \beta$</td><td>Claim an assumption of exogeneity and that model is associative and not causal. Just looking gor best fit line.</td></tr>
<tr><td>MLR5<br>(Gauss-Markov 5)</td><td>$var(u \mid x_1, x_2, \ldots, x_k) = \sigma^2$</td><td>Homoskedasticity of errors. Constant error variance across all x. A strong assumption.</td><td></td><td>Residuals v. Fitted values.<br>Scale-location plot.<br>plot(model1) gives both<br><br>Breusch-Pagan Test<br>bptest(model1)<br><br><br>White Test</td><td>R v FV: look for different width of bands as a function of x.<br><br>BP: Regresses residuals on Xs and does F-test on model.<br><br><br>White: regresses error residuals on fitted values.</td><td>Use White standard errors, which are robust to homoskedasticity.<br><br>coeftest(model3,<br>vcov = vcovHC)<br><br>(se.model1 = sqrt(diag(vcovHC(model1))))</td></tr>
<tr><td>MLR6 CLM<br>(Classical Linear Model)</td><td>$u_i \mid x \sim N(0, \sigma^2)$</td><td>Errors are normally distributed.</td><td></td><td></td><td>Hist(residuals)<br>Q-Q plot – look for perfect diagonal line, no tails<br>Shapiro-Wilk test of normality of residuals</td><td>hist(model3$residuals)<br>shapiro.test(model1$residuals)</td><td>Large data set: rely on asymptotics of OLS est. Need 1-5.<br>Small data sets: transform skewed variable, try x^2. If resi_v_fitted shows curvature have zcm violation too.<br>Bootstrap errors.</td></tr>
</table>