## STATISTICAL ANALYSIS

**Pearson's r correlation between text length and MT errors:**

| | |
|---|---|
| Text Length - DL MQM errors: | r(50) = -0.128, p = 0.366 |
| Text Length - GT MQM errors: | r(50) = -0.116, p = 0.411 |
| Text Length - SY MQM errors: | r(50) = -0.081, p = 0.568 |
| Text Length - DL Bert scores: | r(50) = -0.128, p = 0.366 |
| Text Length - GT Bert scores: | r(50) = -0.078, p = 0.583 |
| Text Length - SY Bert scores: | r(50) = -0.155, p = 0.273 |

--> No statistically significant correlation between text length and MT error for any of the three MT engines

**Pearson's r correlation between readability and MT errors:**

| | |
|---|---|
| Readability - DL MQM errors: | r(50) = 0.359, p = 0.09 |
| Readability - GT MQM errors: | r(50) = 0.331, p = 0.164 |
| Readability - SY MQM errors: | r(50) = 0.306, p = 0.271 |
| Readability - DL Bert scores: | r(50) = 0.055, p = 0.699 |
| Readability - GT Bert scores: | r(50) = -0.094, p = 0.508 |
| Readability - SY Bert scores: | r(50) =-0.061, p = 0.670 |

--> No statistically significant correlation between readability and MT error for any of the three MT engines

**T-Test for statistical significance of the difference (superiority) of DL over GT and SY**

DL - GT MQM errors: Two sample t-test (left-tailed) p = 0.2647

DL - SY MQM errors: Two sample t-test (left-tailed) p = 0.2634

--> No statistically significant difference: the MQM error average of DL is smaller than the sample average of both GT and SY, but not small enough to be statistically significant