BMTRY 755 HW #1 Key

1.   Point Estimate = 0.37   95% CI = [0.26, 0.49]

For every one day increase in a hospital's average patient length of stay, the risk of nosocomial infection increases by 0.37 percent.

2. There is a difference of 5 days between a LOS = 20 and LOS = 15. Since the slope gives a per-day rate of change, we need to multiply the slope by 5 to get the increased risk for a 5 day period.

$5(0.37) = 1.87$

At the hospital level, a 20-day stay risk of infection is 1.87 percent higher than a 15-day stay risk.

3. Yes, LOS is a statistically significant predictor of infection risk in this single linear regression model, since the p-value < 0.0001 for the parameter estimate for LOS. This p-value is significant at the 5% level.

4. s^2 is the sample variance, and is given by the mean square error. In this problem, $s^2 = 1.29$. If $\sigma^2 = 0$, then a plot of risk by LOS would result in all the data points assembled in a straight line – there would be no 'scatter' or 'spread' to the data. However, since $\sigma^2 \neq 0$, we have variability, which leads to a true scatterplot with observations scattered around the line. In short, σ^2 induces variability among individual observations.

5. a) Our regression model is Risk = 0.74 + 0.37 (LOS).

To determine the predicted risk of infection for a LOS = 9, we just substitute LOS = 9 into our regression model.

Risk = 0.74 + 0.37 (9) = 4.11

Thus, we predict that for a hospital with an average LOS of 9 days, there is a 4.11% risk of nosocomial infection.

b) "The average length of stay for County Hospital's patients is nine days. Therefore, we estimate the hospital's overall risk of nosocomial infection is 4%
(95% CI = **1.8 to 6.4**)"
This is a prediction interval, since we are creating a confidence interval around a single prediction, namely the infection risk for one given hospital.

 "Among US hospitals with an average LOS of 9 days, the average risk of nosocomial infection is 4% (95% CI = **3.9 to 4.3**)"
This is a CI around a mean, since we are creating a confidence interval around an estimate for an average value. Instead of an estimate for just one hospital, we have an estimate for the average across all hospitals. As a result of the additional information provided by multiple hospitals with the same average LOS, our confidence interval is narrower.