# Year and Site Comparisons

## Exploratory Analysis:

The data were collected over a period of 4 field seasons at sites in Bozeman and 2 field seasons in Moccasin. Phenotypic data were analyzed in order to determine whether conditions were similar enough at each year/site combination to consider as a single dataset, or if they differ by enough that the sites and years need to be considered as different groups.

Because lodging is considered the response of interest, we examine the Percent Lodging measured at each site/year combination. *Note that Percent Lodging is not measured in the 2013 Bozeman data; we did not include it in this analysis. Given the result of the analysis, we do not need to consider it into the analysis to determine that there are differences in the year/site combinations.* Visually, we can plot the mean percent lodging at each year/site combination to get an idea of differences between the years and sites. The plot is shown in the figure below, where the red line refers to the mean percent lodging in Bozeman and the blue line refers to the mean percent lodging in Moccasin:

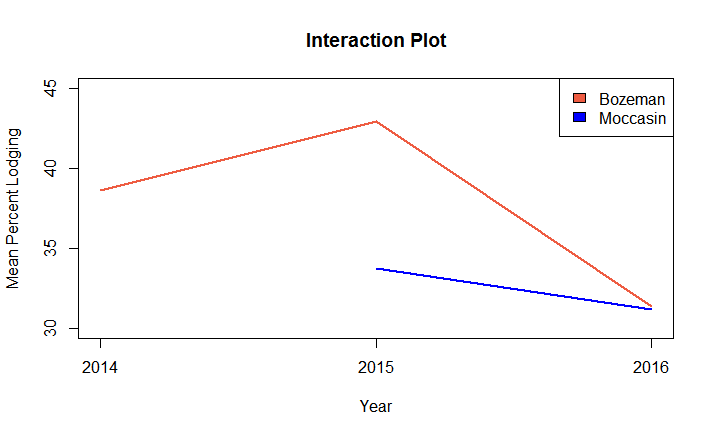


Figure 1: Mean Percent Lodging by Site/Year Combination

If the sites are reasonably similar, we would expect to see parallel lines that are either overlapping or very near each other. The red line for Bozeman indicates that average percent lodging was higher in Bozeman than in Moccasin in both 2015 and 2016; however it may also have been more variable than in Moccasin. While the blue line and red line are both decreasing from 2015 to 2016, the difference in slopes indicates that there may be some interaction between year/site combinations. More formal testing is necessary to prove that years and sites differ, but this visualization indicates that mean percent lodging for each of the years and sites differs.

## Regression Model:

To test whether year and site impact Percent Lodging, we fit a linear regression model using a 6-level factor variable as the explanatory variable and Percent Lodging as the response. We could fit a more complicated model using more of the measured covariates, but if percent lodging differs by site/year in the simple case, adding more information is not likely to change this result.

The results of the regression are displayed in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1: Regression Results | | | | |
| Coefficients | **Estimate** | **Standard Error** | **t-value** | **p-value** |
| Baseline: Bozeman 14 | 38.615 | 1.29 | 29.945 | <2E-16 |
| siteyearBozeman2015 | 4.293 | 1.761 | 2.438 | 0.0149 |
| siteyearBozeman2016 | -7.234 | 1.747 | -4.141 | 3.69E-05 |
| siteyearMoccasin2015 | -4.887 | 1.765 | -2.77 | 0.0057 |
| siteyearMoccasin2016 | -7.447 | 1.744 | -4.271 | 2.10E-05 |

The regression treats Bozeman 2014 as the baseline group. Each site-year combination is tested against the baseline; small p-values indicate that the given group differs from the baseline. The p-values for each of the other site-year combinations are all quite small – the largest, for Bozeman 2015, is only 0.015, indicating strong evidence that each of the site/year combinations differ from the baseline. Moreover, the effect size ranges from 4 to 7 points on the 1-100 percent scale on which Percent Lodging is measured. The relatively large effect size and very small p-values, when taken together, indicate not only that these differences are statistically significant but practically significant as well. The groups are likely too different to be considered as a homogenous set.

## Conclusion

This analysis indicates that the measurements taken at each site in each year should be treated separately; they should not be considered homogenous enough to be taken as a single group. The inclusion of 2013 into the data is unlikely to change the results of this study because the differences between the combinations already considered are enough to justify splitting the dataset into separate groups.