Problem #5

a)

α_j is the population statistic. It's the treatment effect of the jth population. It is often unobservable.

$\hat{\alpha_j}$ is the sample statistic. It's $\bar{Y_j} - \bar{Y}$, the difference between jth sample mean and total sample mean. It is used to approximate α_j.

b)

The null hypothesis of the F test is,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n$$

In other words,

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

To calculate the expected value of the F score, we have:

$$\mathbb{E}(F) = \mathbb{E}(MSBG)/\,\mathbb{E}(MSWG) = \frac{\sigma_\epsilon^2 + n\sum_{j=1}^{p} \frac{\alpha_j^2}{p-1}}{\sigma_\epsilon^2}$$

Notice, when H_0 is satisfied, all α_j are 0, thus

$$n\sum_{j=1}^{p} \frac{\alpha_j^2}{p-1} = 0$$

$$\mathbb{E}(F) = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2} = 1$$

c)

$\mathbb{E}_{MS}(between)$ and $\mathbb{E}_{MS}(within)$ are population statistics that are often unobservable. They are the expected value of sample mean squares.

What we have in the summary of the ANOVA model is the computed mean square for the sample we have. Even though these sample statistics converge to the true expected values, they do not represent $\mathbb{E}_{MS}(between)$ and $\mathbb{E}_{MS}(within)$ directly.

Problem #6

a)

The question is, are the mean infant birth weights the same between populations of smoking and non-smoking mothers.

b)

The resulting p value is small, so null hypothesis of same mean is rejected. There is a statistically significant difference between the infant birth weights between smoking and non-smoking mothers.

c)

We are 95% sure that the true difference in average infant birth weights between smoking and non-smoking mothers is between 6.23 and 10.72.

d)

The plot shows that the average infant birth weight of non-smoking mothers is higher than that of smoking mothers. The non-smoking sample size is 614 while the smoking sample size is 386. The error bar shows the 95% confidence interval (which is the default in the plotmeans function) of the model.

e)

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{306566}{306566 + 17020} = .0526$$