1. I had not heard of Stein's Paradox before, though Stein's Lemma was covered in Casella and Berger.

   Stein's Paradox may not be taught as often because a Bayesian form predated it by 200 years. If we are going to assume we know the true variance of a distribution, we might as well assume we know the true mean. I'm also not sure how his ideas extend to regression and other modeling scenarios where there are more variables to account for in the prediction. The method is not robust, so the groups chosen may not allow the paradox to hold.

2. The idea of the James-Stein estimator is that by shrinking an individual sample average to an overall grand sample average, we get a better future prediction of that individual's sample average when there are more than two groups.

   Stein's Paradox contradicts the statistical theory that "no other estimation rule is uniformly better than the observed average". Stein showed that a weighted average of the grand mean and the individual sample mean is a better predictor of the true individual group mean. It is contradictory because it is a better predictor of the true individual group mean than the observed sample average.

   Note that the individual groups are assumed independent.

3. The James-Stein estimator is $z_i = \bar{y} - c(y_i - \bar{y})$ where the shrinking factor, c = 1 - $\frac{(k-3)\sigma^2}{\Sigma(y-\bar{y})^2}$.

   The closer the individual sample mean is to the grand sample mean, the more the individual group mean is shrunk toward the grand mean.

4. The major paradoxical component that (in the case of proportions) it doesn't matter whether the "grand sample average" is the same proportional unit, the James-Stein estimator will do better than the individual sample average at predicting future averages (where here the average is a proportion). The example given in the paper was concerning batting averages and the proportion of imported cars in Chicago.

5. Admissible would mean there is no other estimator that has a smaller risk for all values of θ, that is, for all θ_i.

   Inadmissible then means there is another estimator with a smaller risk for all values of θ. The paper refers to the median as inadmissible for the normal distribution because the mean has a smaller such risk, therefore is admissible for less than three groups. For three or more groups, the James-Stein estimator is admissible.

   Note that Stein's Paradox doesn't assume the normal distribution, but because inference is about means, the CLT likely makes the normal distribution reasonable for large sample sizes.

6. Risk is the expected value of the squared error for every possible value of θ.

   $E[(y - \theta_i)]$ where θ_i is all possible values of θ.

7. In general, the James-Stein estimator is the sample version of a Bayesian version. Instead of shrinking towards the grand sample average, the individual sample means are shrunk toward a prior mean.

The idea of the James-Stein estimator is seen in hierarchical modeling. Out of hierarchical modeling we can think of the single mean model, where the different groups are not accounted for, and also the group means model where each group has a different mean. Hierarchical models balance these two extremes based on the between group variation.

In hierarchical modeling, the BLUPs are shrunk towards the grand mean if there is a lot of variation in the sample means and are more unique with less variation, which is also how the James-Stein estimator works.