Work all problems. Seventy points are required to pass.

1. The linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$, where $\mathbf{y}$ is a vector of length $n$, is said to be saturated if the error degrees of freedom ($n - \text{rank}(\mathbf{X})$) is equal to zero. Assume the errors are iid and mean zero.

   (a) Define BLUE. (Please do not just use four words.)

   (b) Give an example of an estimator of $\boldsymbol{\beta}$ that is not BLUE.

   (c) Show that in a saturated model, every linear unbiased estimator is the corresponding BLUE.

2. What affects the selling price of a house? Table below shows analysis results based on 100 recent home sales in Gainesville, Florida.

   (a) Report and interpret results of the global test of the hypothesis that none of the explanatory variables has an effect.

   (b) Report and interpret significance tests for the individual partial effects, adjusting for the other variables in the model.
(c) What is the conceptual difference between the tests of the effects of the independent variables in the coefficients table and in the ANOVA table?

3. Given independent random samples of sizes \(n_1\) and \(n_2\), the goal of this question is to inferentially compare \(\mu_1\) and \(\mu_2\) from \(N(\mu_1, \sigma^2)\) and \(N(\mu_2, \sigma^2)\) populations.

(a) Put the analysis in a normal linear model context, showing a model matrix and explaining how to interpret the model parameters.

(b) Find the projection matrix for the model space, and find SSR (regression sum of squares) and SSE (error sum of squares).

(c) Construct an \(F\) test statistic for testing \(H_0: \mu_1 = \mu_2\) against \(H_a: \mu_1 \neq \mu_2\). Specify a null distribution for this statistic.

(d) Relate the \(F\) test statistics in (c) to the \(t\) statistic for this test,

\[
t = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

where \(\bar{y}_i\) is the sample mean of the \(i\)th sample, and \(s_i^2\) is sample variance of the \(i\)th sample.

(e) Suppose the two independent samples had different variances. Could you use the \(F\) and \(t\) tests you used in parts (c) and (d)? Why or why not?

4. Suppose that \(y\) is \(N_n(\mu, \sigma^2 I)\) and that \(X\) is an \(n \times p\) matrix of constants with rank \(p < n\).

(a) Show that \(H = X(X'X)^{-1}X'\) and \(I - H\) are idempotent, and find the rank of each.

(b) Assuming \(\mu\) is a linear combination of the columns of \(X\), that is, \(\mu = Xb\) for some \(b\), find \(E(y'Hy)\) and \(E(y'(I - H)y)\).

(c) Find the distributions of \(y'Hy/\sigma^2\) and \(y'(I - H)y/\sigma^2\).

(d) Show that \(y'Hy\) and \(y'(I - H)y\) are independent.

(e) Find the distribution of

\[
\frac{y'Hy/p}{y'(I - H)y/(n - p)}.
\]