Formal F-test for "lack of fit" (Section 3.7)

• Requires replicate observations at one or more levels of the *X* variables (often not applicable when one or more predictors are continuous).

• Assume there are *c* distinct levels of the *X* variable(s) in the data set. Denote the mean of the responses at the j-th *X* level as:

• We will use a "full vs. reduced model" approach.

Full model:

The least-squares (or ML) estimators of the μ_j, $j = 1, \ldots, c$, are:

The error sum of squares for the full model (here called the "pure error sum of squares" or SSPE) is:

The "pure error" degrees of freedom are

The "reduced model" is the model that holds under H_0, i.e., the model that assumes H_0: $E(Y)$ =

i.e.,

The error sum of squares for the reduced model is the usual

So the general linear test statistic is

The difference SSE – SSPE is typically called the "lack of fit sum of squares" and is denoted SSLF = SSE – SSPE.

So the test statistic is

Rejection region:

The expected mean squares are

E(MSLF) =

E(MSPE) =

Example (Bank data): For 11 bank offices, X = size of minimum deposit and Y = number of new accounts.

	X_1=75	X_2=100	X_3=125	X_4=150	X_5=175	X_6=200
Y_{ij}'s	28	112	160	152	156	124
	42	136	150		124	104
$\bar{Y}_j$	35	124	155	152	140	114

SSPE =

From a simple linear regression fit, SSE = and thus SSLF =

- **The lack-of-fit test can be carried out in SAS (the LACKFIT option) or R.**
- **The alternative H_a includes any regression relationship other than a linear one --- if H_0 is rejected, examine scatterplots and residuals to help identify an appropriate mean response function.**
- **If H_0 is not rejected, use MSE as an estimator of σ^2 as usual.**
- **If no replications exist in a data set, Reference 3.8 (p. 146) discusses other lack-of-fit tests.**