Chapter 16 (continued): Analysis of Variance

Note: In a single-factor study, the levels are also known as treatments.

Partitioning the Total Sum of Squares

SSTO =

SSTR =

SSTR measures the variability

SSE =

SSE measures the variability

- **SSE estimates the "natural" variation in the data (variation not due to the different treatments).**

- **If the treatment means differ greatly, then SSTR will be**

Note that SSTO = SSTR + SSE.

Proof:

- **The associated degrees of freedom are also additive, and are used to obtain the mean squares.**

Expected Values of the Mean Squares (see pg. 696-698 for details):

Note: • MSE is an unbiased estimator of the error variance σ^2.
• If all treatments have the <u>same</u> population mean, then E(MSTR) =

• If the treatments have different population means, then MSTR should be _____________ than MSE, on average.

<u>F-test for Equality of Treatment Population Means</u>

• A natural test statistic to use is:

• If F* >> 1 → evidence supports

If F* near 1 → evidence supports

• Under H_0, F* has a

• We reject H_0 if

<u>Theoretical Justification of F-test</u>
(Proof of distribution of F* under H_0)

• **We need to use Cochran's Theorem:**

Suppose our observations $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$. Then if we break SSTO into *k* sums of squares $SS_1, \ldots, SS_k$ (having degrees of freedom ($df_1, \ldots, df_k$), then for *j* = 1, …, *k*,

random variables, provided that

<u>ANOVA situation</u>:

• **We have n_T observations. Under H_0, they are**

ANOVA Table

Kenton Food Example (from SAS):

• Do the four package designs have the same population mean sales?

We test:

Factor Effects Model

• **This is an alternative formulation of the ANOVA model.**

• **If $\mu_{\bullet}$ is a type of overall population mean response, then let**

• **τ_i is the effect of the *i*-th factor level, or the *i*-th treatment effect.**

• **$\mu_{\bullet}$ is often the simple average of $\mu_1, \mu_2, \ldots, \mu_r$, but it could be defined as a weighted average of $\mu_1, \mu_2, \ldots, \mu_r$.**

• **Note our model equation is:**

• **Our hypothesis of $H_0: \mu_1 = \mu_2 = \ldots = \mu_r$**
is equivalent to

• **If $\mu_{\bullet}$ is an "overall mean response" then $\tau_1, \ldots, \tau_r$ measure how much the treatment means deviate from the overall mean, and**

• **We can use a regression approach to estimate the parameters of this model.**

Regression Approach to the ANOVA Model

- **Since $\tau_r = -\tau_1 - \tau_2 - \ldots - \tau_{r-1}$, we need not estimate τ_r. We only estimate $\mu_{\bullet}$, $\tau_1, \tau_2, \ldots, \tau_{r-1}$.**

- **Recall example when $r = 3$ and $n_1 = 1$, $n_2 = 3$, $n_3 = 2$. Then let:**

- **Then the factor-effects model can be stated as**

For example:

- **If we use the indicator variables**

then this will produce the X matrix above, and we can fit the factor-effects ANOVA model via regression.

SAS Example (Kenton Foods, Table 16.1):

• To fit the cell-means model via regression, we should let

• Thus we can define indicator variables

which will produce the above X matrix.

• We can fit the cell-means model via regression (specifying NO intercept term in this case!).

<u>Kenton Foods Example</u>: