## Testing utility of model - $F$-test

- Testing the utility of the model to predict $y$ by conducting individual $t-$tests on each of the $\beta$'s is not a good idea.
- Why? Even if we conduct each test at the $\alpha = 0.05$ level, the $overall$ probability of incorrectly rejecting $H_0$ (the probability of a Type I error) is larger than 0.05.
- Even if we begin with an $\alpha$ level of, for example, 0.05 on each individual test for $\beta$'s, the overall probability of an error will always be larger than the probability of Type I error on individual tests.
- The larger the number of predictors in the model, the higher the probability that at least one of those hypotheses tests will lead to the wrong conclusion.

## Testing utility of model - $F$-test (cont'd)

- Illustration: suppose we conduct tests on two parameters $\beta_1, \beta_2$ and we use $\alpha = 0.05$ on each test. See what may happen:

| | Correct decision for $\beta_1$ | Incorrect decision for $\beta_1$ |
|---|---|---|
| Correct decision for $\beta_2$ | 0.95 $\times$ 0.95 | 0.95 $\times$ 0.05 |
| Incorrect decision for $\beta_2$ | 0.95 $\times$ 0.05 | 0.05 $\times$ 0.05 |

- The probability of a correct decision on *both parameters* is only 0.95 $\times$ 0.95 $=$ 0.90! So our individual Type I error rate ($\alpha$) is *eroded* in multiple tests.

- In general, for an individual Type I error rate of $\alpha$, the overall error rate on $k$ tests is $(1-\alpha)^k$. For example, for $k = 4$ and $\alpha = 0.05$, the probability that we will reach the correct conclusion for all four $\beta$'s is only $0.95^4 = 0.81$, so the experiment-wise $\alpha$ is really 0.19.

## Testing utility of model - $F$-test (cont'd)

- Rather than testing each $\beta$ individually, we use a *global test* that encompasses all $\beta$'s and test the following overall hypothesis:

$$\begin{aligned} H_0 &: \quad \beta_1 = \beta_2 = \ldots = \beta_k = 0 \\ H_a &: \quad \text{at least one } \beta_j \neq 0. \end{aligned}$$

- The **test statistic** to test this hypothesis is called $F-$statistic and is calculated as:

$$F = \frac{(SS_{yy} - SSE)/k}{MSE} = \frac{R^2/k}{(1 - R^2)/[n - (k + 1)]}.$$

## Testing utility of model - $F$-test (cont'd)

- The $F$ statistic is the ratio of

  - the *explained variability* (as reflected by $R^2$) and
  - the *unexplained variability* (as reflected by $1 - R^2$),

  each divided by the corresponding degrees of freedom.

- The larger the $F$ statistic, the more useful the model.

# Testing utility of model - $F$-test (cont'd)

- Critical value for the test: $F_{\alpha,numdf,denomdf}$ (three subscripts: $\alpha$, numerator degrees of freedom and denominator degrees of freedom).
- We use an $F-$table (pages 763-770 in book).
- Each table corresponds to a different $\alpha$: 0.1, 0.05, 0.025 and 0.01.
- We search the table using the numerator degrees of freedom (column) and the denominator degrees of freedom (rows):

$$\begin{aligned} \text{Numerator d.f.} &= k \\ \text{Denominator d.f.} &= n-(k+1) \end{aligned}$$

- For example, if $\alpha = 0.1$, $n = 20$ and $k = 4$, $F_{0.1,4,15} = 2.36$.

## Testing utility of model - $F$-test (cont'd)

- Critical region: reject $H_0$ at level $\alpha$ if $F \geq F_{\alpha,numdf,denomdf}$.

- How do we interpret the result?

  – If we fail to reject $H_0$: there is no evidence that any of the predictors are linearly associated to the response.
  – If we reject $H_0$: At least one of the predictors is linearly associated to the response.

- Note that if we reject $H_0$, all we know is that one of the predictors is associated to $y$ but we do not know which.

- The next step is to conduct individual $t-$tests on the $\beta$'s, but keeping in mind the confidence erosion discussed earlier.

## Testing utility of model - $F$-test (cont'd)

- The $F$ statistic for the model is given by both SAS and JMP directly on the output. SAS calls it "'F Value"' and JMP calls it "'F Ratio"'.

- Both SAS and JMP compute the F-statistic as:

$$F = \frac{\text{Model Mean Square}}{\text{Error Mean Square}}$$

and the degrees of freedom are:

$$\begin{aligned} \text{Numerator df} &= \text{Model df } = k \\ \text{Denominator df} &= \text{Error df } = n - (k+1). \end{aligned}$$