Joint Estimation of β_0 and β_1

Recall that we have assumed that

ϵ_i is independent of the model, and that the ϵ_i's are independent of each other

We have NOT assumed that the regression coefficients (β_0 and β_1) are independent of each other.

However, we place confidence intervals on each as if they were independent

Consider,
Each confidence interval is done with a 5% chance of error. If we do 2 CI, EACH has a 5% chance of error. If we do 20 CI each and every one has a 5% chance of error. As we do more CI's, the OVERALL chance of making an error increases.

Therefore, we want to obtain a confidence interval such that we are, say, 95% confident that BOTH β_0 and β_1 are contained in the interval. We start with the simple, individual confidence intervals.

$$P(b_0 - t_{1-\frac{\alpha}{2},n-2} \; s_{b_0} \; \leq \; \beta_0 \; \leq \; b_0 + t_{1-\frac{\alpha}{2},n-2} \; s_{b_0}) \;\; = \;\; 1\text{-}\alpha$$

$$P(b_1 - t_{1-\frac{\alpha}{2},n-2} \; s_{b_1} \; \leq \; \beta_1 \; \leq \; b_1 + t_{1-\frac{\alpha}{2},n-2} \; s_{b_1}) \;\; = \;\; 1\text{-}\alpha$$

Individually, each is correct with probability $1 - \alpha$
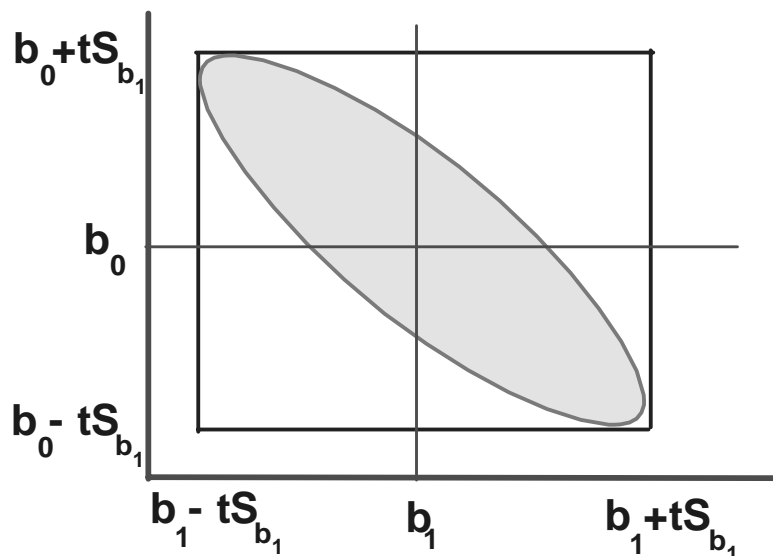
Now we wish to calculate the probability that BOTH are correct.
There is a page of probability theorems in Chapter 1, Section 1.2.

The probability statements are

$P(A_1) = \alpha$ $P(A_2) = \alpha$

where A_1 and A_2 are some events (eg. probability of error)

$$P(A_1 \cup A_2) \; = \; P(A_1) \; + \; P(A_2) \; - \; P(A_1 \cap A_2)$$

$\cup$ = the union of the events; it is the probability of EITHER event occurring
$\cap$ = the intersection of the events; it is the probability of BOTH events occurring

From this we can derive (see text) the *Bonferroni inequality*, which is the
        probability that BOTH CI ARE CORRECT, as

$$P(\bar{A}_1 \cap \bar{A}_2) \; \geq \; 1 \; - \; P(A_1) \; + \; P(A_2) \; = \; 1 - \alpha - \alpha = 1 - 2\alpha$$

This, however, as the inequality suggests is a LOWER BOUND on the probability.

Note that as a general case, $P(\prod_{i=1}^{g}\bar{A}_i) \; \geq \; 1 \; - \; g\alpha$

and if we have 20 CI, the probability is $\geq 1 - 20*0.05 \; \geq \; 0.0$, but it will not be
        zero since there is always a chance that all 20 are correct.

Important note: this effectively demonstrates that as we do more tests, the
        probability of α error deteriorates. We can calculate the lower bound of
        this deterioration using Bonferroni.

As a result of the Bonferroni inequality, we find that the JOINT CI would have a t
        value equal to $t_{1-\frac{\alpha}{2g},n-2}$. If we had chosen to do 2 CI, each with a 95%
        chance of error, then the joint probability of error is **equal to or less than**
        $1 - 2\alpha$
In order to specify a particular level of α for the joint test, we must find a t value
        with a probability of error equal to $1 - \frac{\alpha}{2g}$.

For example, suppose we wished to calculate a Joint 95% CI for β_0 and β_1.

Then $\alpha = 0.05$ and we need to find $t_{1-\frac{\alpha}{2g},n-2}$.

For the example of vial breakage, this would be $t_{1-\frac{0.05}{2*2},8\ df}$. The probability level
is then $1 - \frac{0.05}{4} = 1 - 0.0125 = 0.9975$. Most t-tables would not have
such an unlikely value. However, your textbook has this value. We
would then calculate

$$b_0 \pm t_{0.9875,8\ df} * s_{b_0}$$
$$b_1 \pm t_{0.9875,8\ df} * s_{b_1}$$

or

these are wrong $10.2 \pm 2.752 * 0.6633$
$4.0 \pm 2.752 * 0.4690$

note that the t value for the 95% CI was 2.306

Simultaneous Estimation of Mean Responses

As with the CI on a series of β estimates

There is also a problem in placing CI on a series of $\hat{Y}$ values. If we place a CI on
    estimates of 3 or 4 different values of $\hat{Y}_i$ for different X_i, there is also a
    possibility that not all are correct

    α also deteriorates

Working-Hotelling Approach : simultaneous estimates of all possible X_i (which
    will usually be accomplished by a very conservative CI).

This is given by
$$\hat{Y}_h \pm W{*}s_{\hat{Y}_h}$$

    where;
$$W^2 \;=\; 2{*}F_{1-\alpha;2,n-2\ df}$$

Bonferroni Approach
$$\hat{Y}_h \pm B{*}s_{\hat{Y}_h}$$

    where;

         g is the number of cases, and
$$B \;=\; t_{1-\frac{\alpha}{2g};n-2\ df}$$

Note that there is not "g" in the W-H interval. It is for all cases.

Since the Working-Hotelling Approach is for all possible cases, you might expect
    that it would be wider. HOWEVER, recall that the Bonferroni is a lower
    bound, and as such it may not be narrower. Which is wider will vary
    from case to case.

Use whichever is narrower.

Generally, the Bonferroni is likely to be narrower for few cases, and the Working-
    Hotelling narrower for many cases. The Working-Hotelling is probably a
    better bet overall (all possible cases).

Note that the width of the Bonferroni increases as the number of cases increases,
    while the width of the Working-Hotelling does not change (all possible)

For the VIAL BREAKAGE example

```
EXST7034 - Example 1 : NWK Example 2.19
Regression Models done with SAS REG procedure

        Dep Var    Predict   Std Err   Lower95%   Upper95%   Lower95%   Upper95%
Obs        Y        Value    Predict     Mean       Mean     Predict    Predict

  1      8.0000    10.2000    0.663     8.6704    11.7296     6.4532    13.9468
  2      9.0000    10.2000    0.663     8.6704    11.7296     6.4532    13.9468
  3     11.0000    10.2000    0.663     8.6704    11.7296     6.4532    13.9468
  4     12.0000    10.2000    0.663     8.6704    11.7296     6.4532    13.9468
  5     13.0000    14.2000    0.469    13.1184    15.2816    10.6127    17.7873
  6     15.0000    14.2000    0.469    13.1184    15.2816    10.6127    17.7873
  7     16.0000    14.2000    0.469    13.1184    15.2816    10.6127    17.7873
  8     17.0000    18.2000    0.663    16.6704    19.7296    14.4532    21.9468
  9     19.0000    18.2000    0.663    16.6704    19.7296    14.4532    21.9468
 10     22.0000    22.2000    1.049    19.7814    24.6186    18.0109    26.3891
```

These CI are calculated as $\hat{Y}_{X_i} \pm t_{\frac{\alpha}{2}} * s_{\hat{Y}}$ where $\alpha = 0.05$ and $t_{\frac{\alpha}{2}} = 2.306$.

If we wished to calculate a Joint 95% CI for 2 values of X_i, then the value used would be either

Working-Hotelling
$$W^2 \;=\; 2*F_{1-\alpha;2,n-2\ df}$$

which for 2 observations (or any number of cases actually) at 95% would be

$$W^2 = 2*F_{0.95;2,8} = 2*4.459 = 8.918$$

and

$$W = \sqrt{8.818} \;=\; 2.986$$

or Bonferroni for 2 cases is
$$B \;=\; t_{1-\frac{\alpha}{2g};n-2\ df} = t_{0.9875;8} = 2.7515$$

Then
$$\hat{Y}_{X_i} \pm W s_{\hat{Y}} \;\; \text{for the Working-Hotelling}$$

and

$$\hat{Y}_{X_i} \pm t_{\frac{\alpha}{2*2}} s_{\hat{Y}} \;\; \text{for Bonferroni}$$

which would be $t_{\frac{\alpha}{2g}}$ or $t_{\frac{\alpha}{4}}$

Simultaneous Prediction Interval for New Observations

This is the same idea as above, except that the Scheffé procedure is used instead
of the Working-Hotelling (Scheffé is more generalized)

Scheffé Procedure : provides a calculation of the CI for g cases, not all possible

This is given by

$$\hat{Y}_h \pm S*s_{\hat{Y}_{h_{new}}}$$

where;

g is the number of cases, and

$$S^2 = g*F_{1-\alpha;g,n-2 \ df}$$

Note that this is not for all possible cases, and "g" occurs in the d.f. as well as the
multiplication factor.

Bonferroni Approach is the same as with the CI for mean responses, except that
the standard error used is for individual observations

$$\hat{Y}_h \pm B*s_{\hat{Y}_{h_{new}}}$$

where;

g is the number of cases, and

$$B = t_{1-\frac{\alpha}{2g};n-2 \ df}$$

In these cases, both G and S become larger as the number of cases increases.