CHOOSING SAMPLE SIZES (See Section 3.6)

How many observations? Consider:

- cost of each additional observation
- budget
- objectives of the experiment
- design of experiment/method of analysis
- number of observations needed to detect differences of interest

The last can be calculated (approximately); the results of the calculations may require either reducing the objectives of the experiment or increasing the budget.

One method: based on the *power* of a statistical test

*Power* of a test: The probability of rejecting $H_o$ under a specified condition.



Example: For a one-sample t-test for the mean of a population, with null hypothesis $H_o$: $\mu = 100$, you might be interested in the probability of rejecting $H_o$ when $\mu \geq 105$, or when $|\mu - 100| > 5$, etc.

Such conditions often occur naturally.

Examples:

1. If you can only measure the response to within 0.1 units, it doesn't really make sense to worry about false rejection when the actual value is within less than that amount of the value in the null hypothesis.

2. Some differences may be of no practical import -- e.g., a medical treatment that extends life by 10 minutes is probably not worth it, but if it extends life by a year, it might be.

For an Analysis of Variance test: the *power of the test at $\Delta$*, denoted $\pi(\Delta)$, is the probability of rejecting $H_o$ when at least two of the treatments differ by $\Delta$ or more.

$\pi(\Delta)$ depends on:

- sample size
- number v of treatments
- significance level $\alpha$ of the test
- (population) error variance $\sigma^2$.

$\Delta$, $\pi(\Delta)$, v, and $\alpha$: Set by the experimenter.

$\sigma^2$: Needs to be estimated from a pilot experiment.



Considerations in estimating $\sigma^2$:

- An estimate lower than the actual variance will give power lower than $\pi(\Delta)$, so it is wise to estimate on the high side.

- If the estimate is too high, the power will be higher than needed, and we will be rejecting $H_o$ with high probability in cases where the difference is less than we really care about.

The upper limit of a confidence interval for $\sigma^2$ is a reasonable choice of estimate.

***Confidence Intervals for $\sigma^2$*** (See Section 3.4.6)

It can be shown that, under the one-way ANOVA assumptions, $SSE/\sigma^2 \sim \chi^2(n-v)$.

We can use this to adapt the usual confidence interval procedure to obtain one for $\sigma^2$:

To get an upper 95% CI for $\sigma^2$:

Find the $5^{th}$ percentile $c_{0.05}$ of the $\chi^2(n-v)$ distribution -- so $P(SSE/\sigma^2 \geq c_{0.05}) = 0.95$.

Picture:

Equivalently: $P(SSE/ c_{0.05} \geq \sigma^2) = 0.95$.



Replace SSE by ssE to get a 95% upper confidence interval $(0, ssE/ c_{0.05})$ for $\sigma^2$, or a *95% upper confidence limit* $ssE/ c_{0.05}$ for $\sigma^2$.

*Example:* In the battery experiment:

$ssE = 28413$
$n = 16$
$v = 4$
$n - v = 12$
$c_{0.05} = 5.226$ ($5^{th}$ percentile for $\chi^2(12)$)

So the 95% upper confidence bound is

$$28413/5.226 = 5436.85.$$

(The estimate for $\sigma^2$ was $msE = 2368$.)

"5436.85 is a 95% upper confidence bound for $\sigma^2$" means:

***Calculating sample sizes for a given power***

Assume we are seeking *equal* sample sizes r for each treatment.

With significance level $\alpha$, we will reject

$H_0$: $\tau_1 = \tau_2 = \ldots = \tau_v$

in favor of

$H_a$: At least two of the $\tau_i$'s differ

when

msT/msE > F(v-1, n-v; $\alpha$).

Thus the power of the test is

$\pi(\Delta)$ = P(MST/MSE > F(v-1, n-v; $\alpha$)| at least one $\tau_i - \tau_j \geq \Delta$)



Recall: *If* $H_0$ is true, then the sampling distribution of the test statistic MST/MSE is F(v-1, n-v).

However, if $H_0$ is *not* true, the sampling distribution of MST/MSE has a *noncentral F-distribution*. (The distribution of the quotient of two independent "non-central chi-squared" random variables divided by their degrees of freedom.) This distribution is defined in terms of a *noncentrality parameter* $\delta^2$, which under our assumptions is

$$\delta^2 = \frac{r\sum_{i=1}^{v}(\tau_i - \bar{\tau}.)^2}{\sigma^2}$$

The hardest situation to detect (the "limiting case" on which the sample size calculations depend): when the effects of two of the factor levels differ by $\Delta$ and the rest are all equal and midway between these two.

*Exercise*: In this limiting case, the formula for $\delta^2$ reduces to

$$\delta^2 = \frac{r\Delta^2}{2\sigma^2}.$$

Thus

$$r = \frac{2\sigma^2\delta^2}{\Delta^2}.$$

Power for the noncentral F distribution is given in tables as a function of $\phi = \delta/\sqrt{v}$. In terms of $\phi$,

$$r = \frac{2v\sigma^2\phi^2}{\Delta^2}.$$

However, the tables need to be used iteratively, since the denominator degrees of freedom n - v = v(r-1) depend on r. A procedure for doing this is given on p. 52 of the text.