Power and Sample Size Calculations

- So far: our theory has been used to compute P-values or fix critical points to get desired α levels.
- We have assumed that all our null hypotheses are True.
- I now discuss power or Type II error rates of our tests.
- **Definition**: The power *function* of a test procedure in a model with parameters θ is P_θ(Reject).

t tests

- Consider a t-test of $\beta_k = 0$.
- Test statistic is

$$\frac{\hat\beta_k}{\sqrt{MSE(X^TX)^{-1}_{kk}}}$$

- Can be rewritten as the ratio

$$\frac{\hat\beta_k / \left[\sigma\sqrt{(X^TX)^{-1}_{kk}}\right]}{\sqrt{[SSE/\sigma^2]/(n-p)}}$$

- When null hypothesis that $\beta_k = 0$ is true numerator is standard normal, the denominator is the square root of a chi-square divided by its degrees of freedom and the numerator and denominator are independent.
- When, in fact β_k is not 0 the numerator is still normal and still has variance 1 but its mean is

$$\delta = \frac{\beta_k}{\sigma\sqrt{(X^TX)^{-1}_{kk}}}.$$

- So define **non-central** t distribution as distribution of

$$\frac{N(\delta,1)}{\sqrt{\chi^2_\nu/\nu}}$$

  where the numerator and denominator are **independent**.
- The quantity δ is the **noncentrality parameter**.
- Table B.5 on page 1327 gives the probability that the absolute value of a non-central t exceeds a given level.

- If we take the level to be the critical point for a t test at some level α then the probability we look up is the corresponding **power**,
- That is, the probability of rejection.
- Notice power depends on two unknown quantities, β_k and σ and on 1 quantity which is sometimes under the experimenter's control (in a designed experiment) and sometimes not (as in an observational study.)
- Same idea applies to any linear statistic of the form $a^T\hat\beta$
- Get a non-central t distribution on the alternative.
- So, for example, if testing $a^T\beta = a_0$ but in fact $a^T\beta = a_1$ the non-centrality parameter is

$$\delta = \frac{a_1 - a_0}{\sigma\sqrt{a^T(X^TX)^{-1}a}}.$$

Sample Size determination

- Before an experiment is run.
- Sometimes experiment is costly.
- So try to work out whether or not it is worth doing.
- Only do experiment if probabilities of Type I and II errors both reasonably low.
- Simplest case arises when you prespecify a level, say $\alpha = 0.05$ and an acceptable probability of Type II error, β say 0.10.

- Then you need to specify
  - The ratio β/σ: comes from physically motivated understanding of what value of β would be important to detect and from understanding of reasonable values for σ.
  - How the design matrix would depend on the sample size.
  - Easiest: fix some small set of say j values $x_1, \ldots, x_j$; then use each member of that set say m times so that the aggregate sample size is mj.
  - This gives a non-centrality parameter of the form

$$\frac{\beta}{\sigma} \times \frac{\sqrt{m}}{\sqrt{(X^T X)^{-1}_{kk}}}$$

  - The value $n = mj$ influences both the row in table B.5 which should be used and the value of δ.
  - If the solution is large, however, then all the rows in B.5 at the bottom of the table are very similar so that effectively only δ depends on n; we can then solve for n.

Power for F tests

- Simplest example: regression through origin (no intercept).
- Model
$$Y_i = \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} + \epsilon_i$$
- Test $\beta_1 = \cdots = \beta_p = 0$
- F statistic
$$F = \frac{MSR}{MSE} = \frac{\hat{Y}^T \hat{Y}/p}{\hat{\epsilon}^T \hat{\epsilon}} = \frac{Y^T H Y/p}{Y^T(I-H)Y/(n-p)}.$$

Suppose now that the null hypothesis is false.

- Substitute $Y = X\beta + \epsilon$ in F.
- Use $HX = X$ (and so $(I-H)X = 0$).
- Denominator is
$$\frac{\epsilon^T (I-H)\epsilon}{n-p}$$

- So: even when the null hypothesis is false the denominator divided by σ^2 has the distribution of a χ^2 on $n-p$ degrees of freedom divided by its degrees of freedom.
- FACT: Numerator and denominator are independent of each other even when the null hypothesis is false.
- Numerator is

$$\frac{(\epsilon + X\beta)^T H(\epsilon + X\beta)}{p}$$

- Divide by σ^2 and rewrite this as

$$W^T HW/p$$

- $W = (\epsilon + X\beta)/\sigma$ has a multivariate normal distribution with mean $X\beta/\sigma = \mu/\sigma$ and variance the identity matrix.

- FACT: If W is a $MVN(\tau, I)$ random vector and Q is idempotent with rank p then $W^T QW$ has a **non-central** χ^2 distribution with non-centrality parameter

$$\delta^2 = E(W^T QW) - p = \tau^T Q\tau$$

and p degrees of freedom.
- This is the same distribution as that of

$$(Z_1 + \delta)^2 + Z_2^2 + \cdots + Z_p^2$$

where the Z_i are iid standard normals. An ordinary χ^2 variable is called **central** and has $\delta = 0$.
- FACT: If U and V are independent χ^2 variables with degrees of freedom ν_1 and ν_2, V is central and U is non-central with non-centrality parameter δ^2 then

$$\frac{U/\nu_1}{V/\nu_2}$$

is said to have a **non-central** F distribution with non-centrality parameter δ^2 and degrees of freedom ν_1 and ν_2.

Power Calculations

- Table B 11 gives powers of F tests for various small numerator degrees of freedom and a range of denominator degrees of freedom
- Must use $\alpha = 0.05$ or $\alpha = 0.01$.
- In table ϕ is our $\delta/\sqrt{p+1}$ (that is, the square root of what I called the non-centrality parameter divided by the square root of 1 more than the numerator degrees of freedom.)

Sample size calculations

- Sometimes done with charts and sometimes with tables; see table B 12.
- This table depends on a quantity

$$\frac{\Delta}{\sigma} = \sqrt{\frac{(p+1)\delta^2}{n}}$$

  To use the table you specify

  - α (one of 0.2, 0.1, 0.05 or 0.01)
  - Power ($= 1 - \beta$ in notation of table)– must be one of 0.7, 0.8, 0.9 or 0.95
  - Non-centrality per data point, δ^2/n.

  Then you look up n.
- Realistic specification of δ^2/n **difficult** in practice.

Example: POWER of t test: plaster example

- Consider fitting the model

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 F_i + \beta_3 F_i^2 + \epsilon_i$$

- Compute power of t test of $\beta_3 = 0$ for the alternative $\beta_3 = -0.004$.
- This is roughly the fitted value.
- In practice, however, this value needs to be specified *before* collecting data so you just have to guess or use experience with previous related data sets or work out a value which would make a difference big enough to matter compared to the straight line.)
- Need to assume a value for σ.
- I take 2.5 – a nice round number near the fitted value.
- Again, in practice, you will have to make this number up in some reasonable way.

- Finally $a^t = (0,0,0,1)$ and $a^T(X^TX)^{-1}a$ has to be computed.
- For the design actually used this is 6.4×10^{-7}. Now δ is 2.
- The power of a two-sided t test at level 0.05 and with $18 - 4 = 14$ degrees of freedom is 0.46 (from table B 5 page 1327).
- Take notice that you need to specify α, β_3/σ (or even β_3 and σ) and the design!

Sample size needed using t test: plaster example

- Now for the same assumed values of the parameters how many replicates of the basic design (using 9 combinations of sand and fibre contents) would I need to get a power of 0.95?
- The matrix $X^T X$ for m replicates of the design actually used is m times the same matrix for 1 replicate.
- This means that $a^T (X^T X)^{-1} a$ will be $1/m$ times the same quantity for 1 replicate.
- Thus the value of δ for m replicates will be $\sqrt{m}$ times the value for our design, which was 2.
- With m replicates the degrees of freedom for the t-test will be $18m - 4$.

- We now need to find a value of m so that in the row in Table B 5 across from $18m - 4$ degrees of freedom and the column corresponding to

$$\delta = 2\sqrt{m}$$

we find 0.95.

- To simplify we try just assuming that the solution m is quite large and use the last line of the table.
- We get δ between 3 and 4 – say about 3.75.
- Now set $2\sqrt{m} = 3.7$ and solve to find $m = 3.42$ which would have to be rounded to 4 meaning a total sample size of $4 \times 18 = 72$.
- For this value of m the non-centrality parameter is actually 4 (not the target of 3.75 because of rounding) and the power is 0.98.
- Notice that for this value of m the degrees of freedom for error is 66 which is so far down the table that the powers are not much different from the ∞ line.

POWER of F test: SAND and FIBRE example

- Now consider the power of the test that all the higher order terms are 0 in the model

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 F_i + \beta_3 F_i^2 + \beta_4 S_i^2 + \beta_5 S_i F_i + \epsilon_i$$

  that is the power of the F test of $\beta_3 = \beta_4 = \beta_5 = 0$.
- Need to specify the non-centrality parameter for this F test.
- In general the noncentrality parameter for a F test based on ν_1 numerator degrees of freedom is given by

$$E(\text{Extra SS})/\sigma^2 - \nu_1 \,.$$

- This quantity needs to be worked out algebraically for each separate case, however, some general points can be made.

- Write the full model as

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

and the reduced model as

$$Y = X_1\beta_1 + \epsilon$$

- Extra SS is difference between two Error sums of squares.
- One is for the full model, assumed correct, so:

$$E(\text{ErrorSS}_{\text{FULL}}) = \text{ErrorDF}_{\text{FULL}}\sigma^2$$

- The Error SS for the reduced model is

$$Y^T(I - H_1)Y$$

where $H_1 = X_1(X_1^T X_1)^{-1}X_1^T$.

- Replace Y by $X_1\beta_1 + X_2\beta_2 + \epsilon$ from full model equation; take expected value.
- The answer is

$$\sigma^2[(n - p_1) + \beta_2^T X_2^T (I - H_1) X_2 \beta_2]$$

where $p_1 =$ is the rank of X_1.
- This makes the non-centrality parameter

$$\delta^2 = \beta_2^T X_2^T (I - H_1) X_2 \beta_2 / \sigma^2.$$

- Interpretation: error sum of squares regressing $X_2\beta_2$ on X_1.

Sand and Fibre details

Assume $\beta_3 = -0.004$, $\beta_4 = -0.005$ and $\beta_5 = 0.001$. The following SAS code computes the required numerator.

```
data plaster;
infile 'plaster.dat';
input sand fibre hardness strength;
newx = -0.004*fibre*fibre -0.005*sand*sand
  +0.001*sand*fibre;
proc reg  data=plaster;
 model newx = sand fibre ;
run;
```

Output shows:

- Error sum of squares regressing newx on sand, fibre and an intercept is 31.1875.
- Taking σ^2 to be 7 we get a noncentrality parameter of roughly 4.55.
- Compute $\phi = \sqrt{4.55}/\sqrt{3+1} = 1.07$ needed for table B 11.
- For 3 numerator and 18-6=12 denominator degrees of freedom we get a power between 0.27 and 0.56 but close to 0.27.

Sample Size for F test: SAND and FIBRE example

- For same basic problem and parameter values how many times would we need to replicate the design to get a power of 0.95?
- Non-centrality parameter for m replicates is m times that for 1 replicate.
- In terms of the parameter ϕ used in the tables the value is proportional to $\sqrt{m}$.
- With m replicates have $18m - 6$ denominator degrees of freedom.
- If $18m - 6$ is reasonably large can use ∞ line and see that ϕ_m must be around 2.2 making m roughly 4 ($\phi_m = \sqrt{m}\phi_1 = 1.07\sqrt{m}$).

Using Table B 12 directly

- Table gives values of n/r where n is total sample size, $r-1$ is df in numerator of F-test, $n-r$ is df for error, non-centrality parameter δ^2 is
$$\left(\frac{\Delta}{\sigma}\right)^2 \frac{n}{2}$$
- If basic design has n_1 data points and p parameters and F test is based on ν_1 degrees of freedom then when you replicate the design m times you get mn_1 total data points, $mn_1 - p$ degrees of freedom for error and ν_1 degrees of freedom for the numerator of the F test.
- To use the table take $r = \nu_1 + 1$.
- Work out Δ/σ. Take value of ncp δ_1^2 for one replicate of basic design. Compute
$$\Delta/\sigma = \sqrt{2}\delta_1$$
.
- Look up n/r in the table and take that to be m.

- Making small mistake unless $p = \nu_1 + 1$ (which is the case for the overall F test in the basic ANOVA table).
- The problem is that you will be pretending you have $(m-1)(\nu_1+1)$ degrees of freeedom for error instead of $mn_1 - p$. As long as these are both large all is well.
- Our example: for power 0.95 and m replicates of 18 point design have $\delta_1^2 = 4.55$ as above.
- We have $r = 3 + 1 = 4$.
- We get $\Delta/\sigma = \sqrt{2}\sqrt{4.55} = 3.02$.
- For a level 0.05 test we then look on page 1342 and get $m = 5$ for a total sample size of 90.

- Degrees of freedom for error will really be 84 but table pretends that degrees of freedom for error will be $(5-1) \times 4 = 16$.
- The latter is pretty small.
- The table supposes a smaller number of error df which would decrease the power of a test.
- So $m = 5$ is probably an overestimate of the required sample size.
- A better answer can be had by looking at replicates of the 9 point design.

- For 9 data points nonecntrality parameter would be $\delta_1^2 = 4.55/2 = 2.275$.
- Gives $\Delta/\sigma = 2.13$ and m of 9 or 10.
- For $m = 10$ would have same design as before.
- For $m = 9$ we would have only 81 data points.
- At this point you go back to Table B 11 to work out the power properly for 81 or 90 data points and see if 81 is enough.