# Sample Size Estimation & Power Analysis Summary:

Warnes and Liu (2006) provide a simple method for computing sample size for microarray experiments, and reports on a series of simulations demonstrating its performance. Surprisingly, despite its simplicity, the method performs exceptionally well even for data with very high correlation between measurements.

The key component of this method is the generation of a cumulative plot of the proportion of compounds achieving a desired power as a function of sample size, based on simple gene-by-gene calculations. While this mechanism can be used to select a sample size numerically based on pre-specified conditions, its real utility is as a visual tool for understanding the trade off between sample size and power. In our consulting work, this latter use as a visual tool has been exceptionally valuable in helping scientific clients to make the difficult trade offs between experiment cost and statistical power.

Multiple comparison problem can also be taken into account with a correction of Benjamini-Hochberg Procedure. The proportion of true null hypothesis in your data can be estiamted using qvalue package.

### Input Summary:

Input Dataset: -> e.csv

Output Datasets and Files: -> post-hoc Sample Size Analysis -> power\_plot.svg & ssize\_plot.svg & ssize.csv

**- Test Type:** t-test. Compute the statistical power of the two-tailed student t-test given a sample size, and determines the sample size to obtain a target statistical power.

**- Treatment Group:** Gender. The effect size, an input parameter for power analysis, will be estimated by the Gender. Then the statistical power will be calculated based on the sample size in your dataset and the estimated sample size will be calculated based on the target statistical power.

**- Interested Number of Observations (per group):**  the given sample size. The statistical power will be estimated based on the given sample size.

**- Target Power (%):**  the target statistical power. The sample size to achieve the target power will be calculated.

**- Significance Level:**  Type I error rate (the significant criterion), the rate of falsely reject a true null hypothesis.

**- FDR Correction:**  perform the sample size estimation and power analysis considering the FDR (False Discovery Rate) correction

**- FDR Criterion:**  estimate the sample size and perform the power anlaysis while controlling the FDR at 0.05 level.

### Result Summary:

The statistical powers were estimated for each compound given a sample size of 12, and the sample size was estimated based on a target power of 0.8. The effect size of each compound was calculated from the dataset using the treatment group Gender. A significant level of 0.05 was used. Multiple comparision (or false discovery rate, FDR) problem was also taken into account with Benjamini-Hochberg procedure. The FDR was controlled at the level of 0.05. The significant level for each compounds was adjusted accordingly. See Table 1, Figure 1 and 2 for more detail

Table Explanation.

- index: the index of compounds, mainly for sorting the table.

- label: compound labels.

- power (n=12): the estimated statistical power given sample size of 12.

- n (power=0.8): the estimated sample size for the target power of 80%.

Figure 1

answers the question of What is the necessary per-group sample size for 80% powe with the observed effect size and at significant level of 0.05?.

The plot illustrates that smaple size of 5 ,7 ,10 is required to ensure that at least 10%, 20%, and 30% of compounds have a statistical power greater than 80%. It is also shown that a sample size of Inf is sufficient if 100% of the compounds need to achieve a 80% power.

Figure 2

answers the question of What is the power for 12 parients per group with the observed effect size and significant level of 0.05?.

From the plot, 35.54% of compounds achieve at 80% statistical power at the sample size of 12 and significant level of 0.05.

| index | label | power (n=12) | n (power=0.8) |
| --- | --- | --- | --- |
| 1 | xylulose NIST | 0.30343417 | 38.582701 |
| 2 | xylose | 0.46582146 | 24.290652 |
| 3 | xylitol | 0.52284167 | 21.314298 |
| 4 | valine | 0.54448880 | 20.331010 |
| 5 | uridine | 0.68430801 | 15.244885 |
| 6 | uric acid | 0.13290371 | 98.297090 |
| 7 | urea | 0.18986221 | 64.402869 |
| 8 | uracil | 0.99983687 | 4.232617 |
| 9 | tyrosine | 0.05282255 | 414.990793 |
| 10 | tryptophan | 0.99997566 | 3.840699 |

Table 1: First 10 compounds and their estimated statistial powers of having 12 samples and required sample size for 80% power.

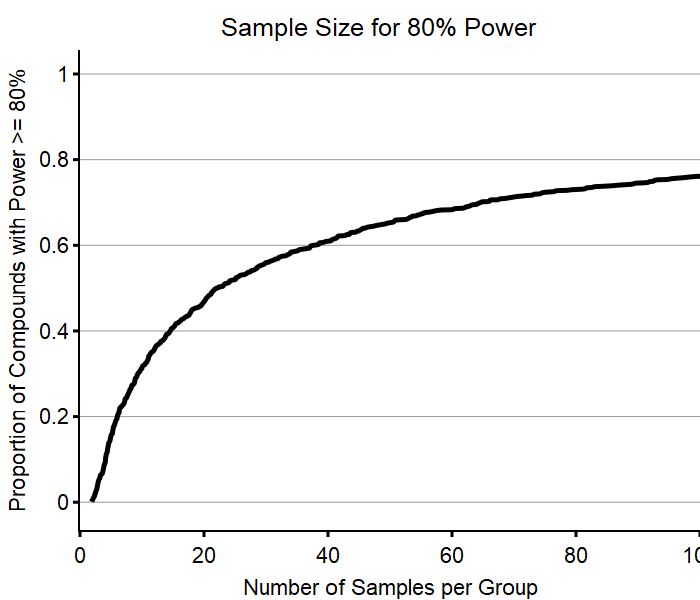


Figure 1: the proportion of compounds needing x samples to achieve a 80% statistical power.

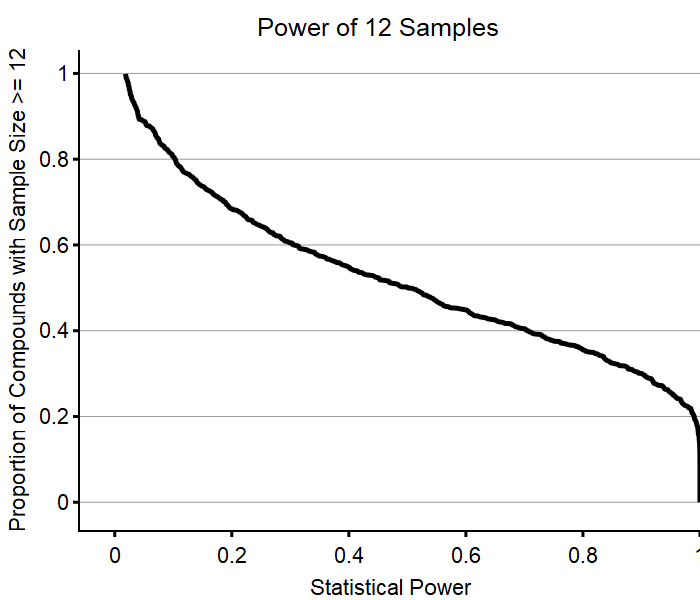


Figure 2: the proportion of compounds having x% statistical power when having 12 samples.