

STAT 8200 — Design of Experiments for Research Workers
Lab 4 – Due Tuesday, Sept. 24, 2013

Example:

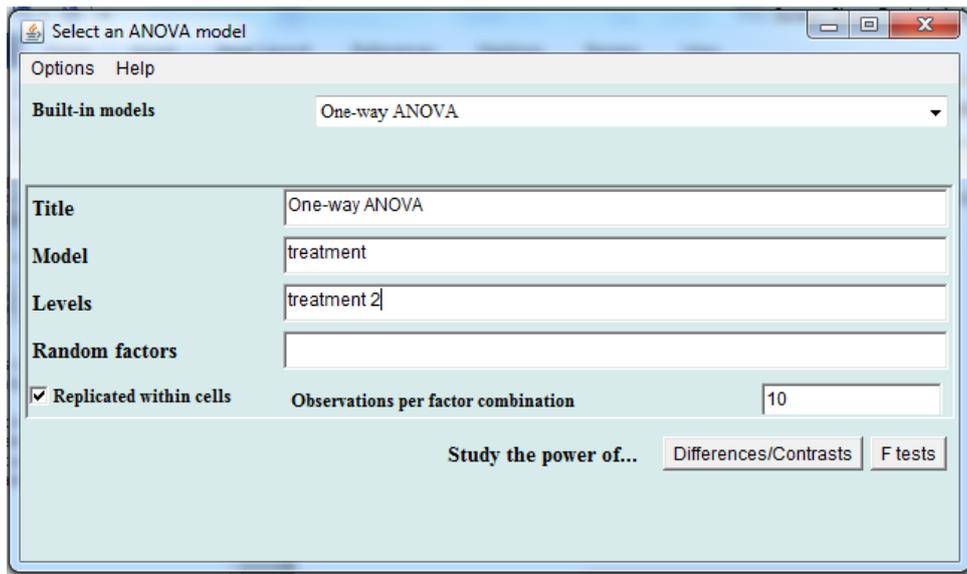
An experiment is to be conducted to investigate the effects of consuming diet soda on appetite and weight gain. Subjects who do not usually drink diet soda are to be recruited for the experiment and randomized to one of four groups that are to differ with respect to daily diet soda intake. Subjects in group A are told not to drink any diet soda, subjects in group B are told to drink one 12-ounce soda per day, subjects in group C are told to drink 24 ounces of diet soda per day, and subjects in group D are assigned to drink 36 ounces of diet soda per day. Despite the fact that diet soda contains 0 calories per 12 ounce serving and is marketed as a diet product, researchers hypothesize that consuming diet soda actually increases appetite and leads to weight gain. The experiment is to be conducted over a 16 week period with the response variable being the change in weight over the 16-week study duration. The investigators believe that consumption of diet soda could lead to substantial weight gain over time, and feel that if they could show as much as a 1.5 lb. difference in weight gain between the control group (group A) and the high consumption group (group D), they would have a clinically significant and publishable result with important public health implications. Subjects will be recruited for the study from the ranks of female college freshman. This is a population whose weight changes have been well-studied by previous researchers and it is known that this population tends to experience weight gain, and past studies that have used a 4 month follow-up period have found that the standard deviation of weight gain is about 3.5 lbs.

The researchers want to get funding for this study so as part of a grant application they must propose a sample size for the experiment that will achieve sufficient power to detect the effects they hope to find. The funding agency typically likes to see proposals for studies that will have 80% power or better.

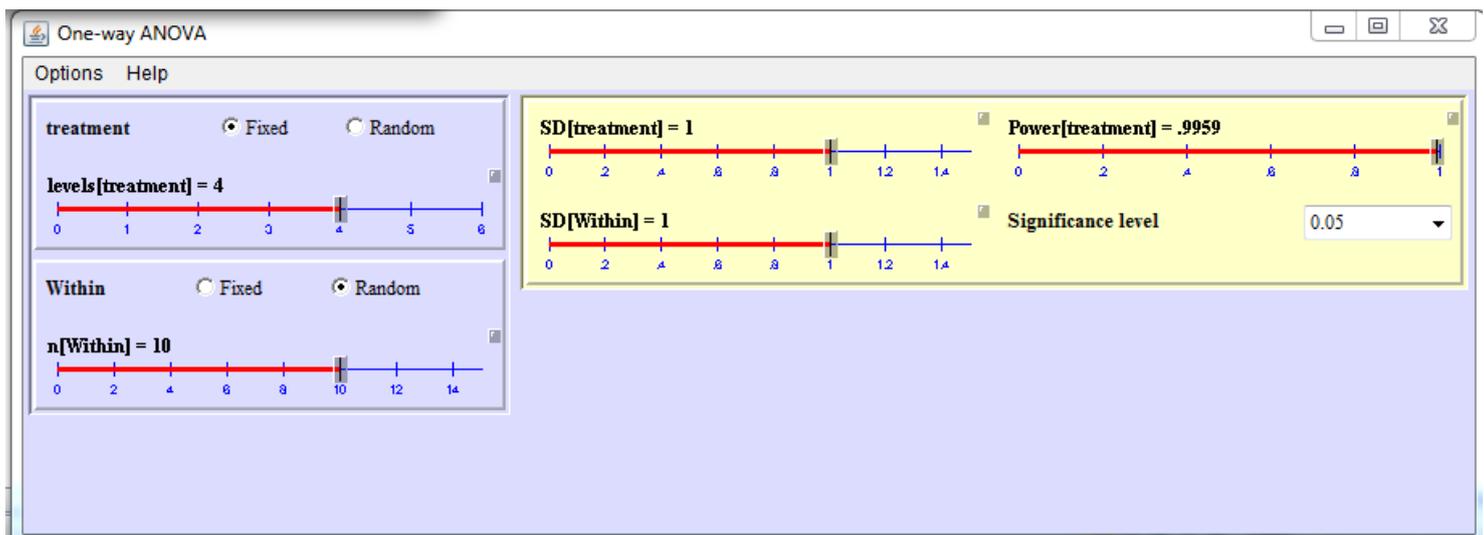
We will use Russ Lenth's Power and Sample Size applets to do the sample size determination for this problem. Go to Dr. Lenth's webpage at the following URL:

<http://homepage.stat.uiowa.edu/~rlenth/Power/>

Under the drawing of the cow a box appears that contains a list of different types of statistical analyses for which power/sample size routines are available. Select "Balanced ANOVA (any model)" and click "Run selection". This will give you a pop-up window that looks like this (you may have to agree to run the JAVA application first):



In this experiment there are four levels of the treatment factor, not 2, so change “treatment 2” on the “Levels” specification to “treatment 4”. Then, because we are going to select the sample size to get 80% power for the overall F test on the treatment factor, click “F tests”. This will bring up a second pop-up window that looks like this:



We will assume the least favorable configuration of treatment means for which the difference between the largest mean and the smallest mean is 1.5 lbs. One scenario where this would happen would be where $\mu_1 = 0$, $\mu_2 = \mu_3 = 0.75$ and $\mu_4 = 1.5$. If these were the values of the means then the standard deviation of the treatment means would be

$$\left\{ \frac{1}{a-1} \sum_{i=1}^a (\mu_i - \mu)^2 \right\}^{1/2} = \left\{ \frac{1}{3} [(0 - 0.75)^2 + (0.75 - 0.75)^2 + (0.75 - 0.75)^2 + (1.5 - 0.75)^2] \right\}^{1/2} = 0.612$$

In fact, for any set of a treatment means where the largest and smallest means are Δ units apart and the rest of the means are half-way between the largest and smallest means, the standard deviation of the treatment means will be $\frac{\Delta}{\sqrt{2(a-1)}}$.

In the second pop-up window that appeared, the number of levels of the treatment factor should be set at 4, and the number of replicates per treatment should be set at $n = 10$. Keep these values, but change SD[treatment] to 0.612 and SD[Within] (which is the error standard deviation σ) to 3.5. These values can be selected using the sliders, but it is easier to click on the little gray box above and to the right of the slider to enter the exact value in a text box. Once you enter these values, the software will immediately compute the power as .1032, or 10.32%. Obviously, $n = 10$ subjects per treatment is too few. Increase n to 50, then to 100, then to 110, and finally to 120. You will see that the power increases to 0.40, 0.72, 0.76 and finally to 0.8006. $n = 120$ is the minimum sample size necessary to get power ≥ 0.80 , so the experiment will require $N = 480$ subjects with equal replication in each of the $a = 4$ groups.

Intentionally blank

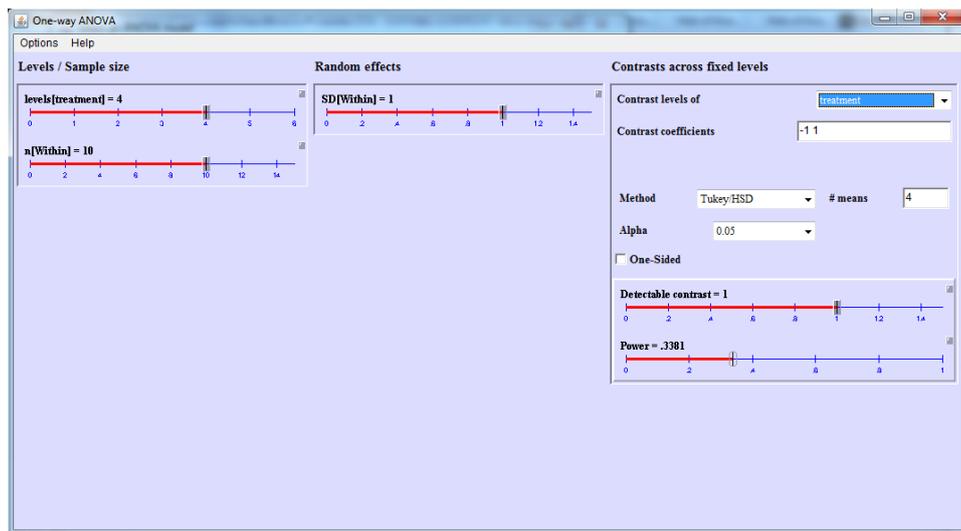
STAT 8200 — Lab 4

Name: _____

Exercise:

In the example, we computed the sample size to get 80% power in our overall F test of equal treatment means. Power does not have to be computed based on the overall F test, however. It may be that, for example, the study hypothesizes that drinking diet soda will have a dose-response effect on weight gain, such that the more diet soda one drinks, the greater the weight gain. If this were posited, then the study might instead be designed with the test of a linear effect of soda consumption as the basis of the power/sample size determination.

To determine the sample size necessary to get power $\geq .80$ for the test of linear effect of diet soda consumption on mean weight gain, close the second pop-up window. Then, in the original pop-up click on “Study the power of ... Differences/Contrasts”. The new pop-up that you get will look like this:



There are several specifications that now have to be changed. First the SD[Within] value is the error standard deviation, which we have assumed is $\sigma = 3.5$. Second, the contrast we are interested in is $\psi = -3\mu_1 - \mu_2 + \mu_3 + 3\mu_4$, so the contrast coefficients should be specified as -3, -1, 1 3 (separated by spaces, not commas). Third, we are not doing a pairwise contrast, so change the “Method” from “Tukey/HSD” to “t”. Finally, we have to specify the true value of the contrast we expect to find. Presumably, we expect a dose response relationship, and let’s stick with our assumption that the best and worst treatments would differ by no more than 1.5 lbs. Therefore, we’ll assume that the true situation is that the means are $\mu_1 = 0$, $\mu_2 = 0.5$, $\mu_3 = 1.0$ and $\mu_4 = 1.5$. This assumption puts them along a line, so we are assuming that the null hypothesis is false in such a way so that the means are linearly related to the amount of consumption. For this assumption

on the means, the value of the linear contrast is

$$\psi = -3(0) - 1(0.5) + 1(1.0) + 3(1.5) = 5.0$$

so set the “Detectable contrast” to 5.0. For the default value of $n = 10$ these values give power equal to 0.166.

1. Find the minimum n per group to get power ≥ 0.80 for the linear contrast test under these assumptions.
2. The necessary sample size is smaller than in our calculation based on the overall F test. However, for the calculations on the overall F tests, the assumptions we made about the mean were somewhat different, so the results from the “Example” portion of this lab and question 1 above are not truly comparable. What sample size would have been necessary for at least 80% power for the overall F test of equal treatment means if the means had been $\mu_1 = 0$, $\mu_2 = 0.5$, $\mu_3 = 1.0$ and $\mu_4 = 1.5$ as we assumed in question 1? Note that in this case $\text{SD}[\text{treatment}]$ would be

$$\begin{aligned} \left\{ \frac{1}{a-1} \sum_{i=1}^a (\mu_i - \mu)^2 \right\}^{1/2} &= \left\{ \frac{1}{3} [(0 - 0.75)^2 + (0.5 - 0.75)^2 + (1.0 - 0.75)^2 + (1.5 - 0.75)^2] \right\}^{1/2} \\ &= 0.645. \end{aligned}$$

3. Note that the test of the linear contrast has more power than the overall F test of equal means if the means are truly linearly arranged. Does that surprise you?

Please hand in pp.5–6, including your answers. Remember to write your name at the top. You may keep pages 1–4 for your notes.