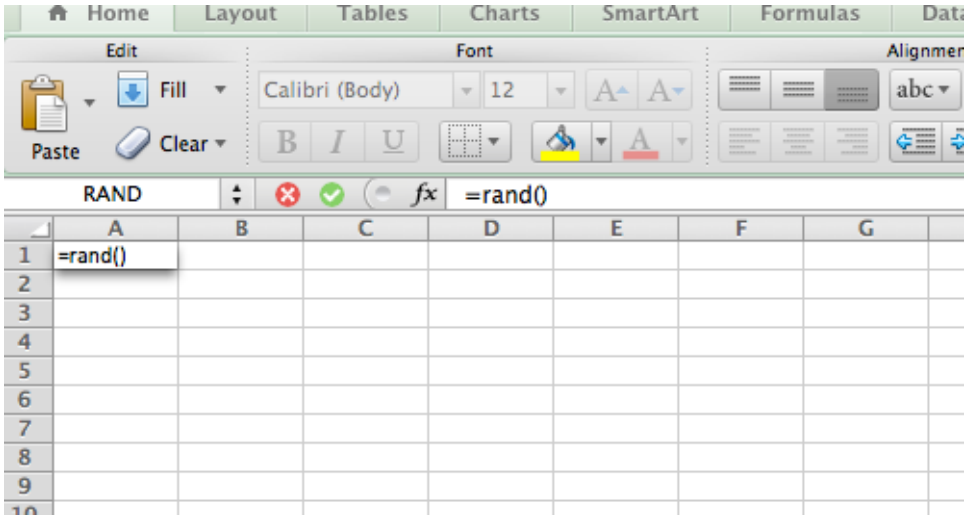
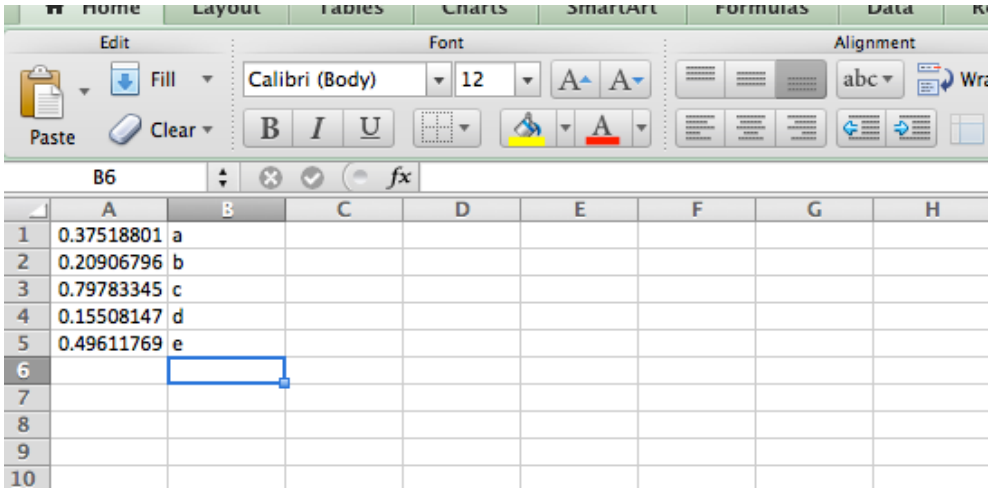
I generated 500 random numbers in excel using the =rand() function.



I then made five pretend treatment levels (a-e)



and copied these 100 times each, to make a file containing 500 datapoints, with 5 treatment levels (i.e., n = 100 for each treatment level).

I then ran various analyses on these random data (null-hypothesis = no difference between the treatment levels)

The results of a 1-way ANOVA were:

$F(4, 495) = 1.85$, $p = 0.12$

This is, unsurprisingly, not significant. As we would expect, the null hypothesis is TRUE (and therefor not rejected) that there is NO DIFFERENCE between these 5-treatment levels (of course: they were based on random numbers ☺).

However, I then ran a series of t-tests to compare the treatment levels. Here are the results (remember this is the SAME DATA as I just ran the ANOVA on):

	t-statisic	p value
a vs b	-0.78	0.44
a vs c	0.38	0.71
a vs d	0.73	0.47
a vs e	1.82	0.07 .
b vs c	1.16	0.25
b vs d	1.48	0.14
b vs e	2.53	0.01**
c vs d	0.36	0.72
c vs e	1.48	0.14
d vs e	1.1	0.27

Hmmm. According to this, there is a significant difference between levels 'b' and 'e' at the 0.01 level (i.e., pretty high), even though this was random numbers! There is also a 'trend' for a difference between levels 'a' and 'e' (p = .07).

For anyone who thinks that non-parametric tests will sort this out, here are the results of a Mann-Whitney run on groups 'b' vs 'e':

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of dv is the same across categories of iv.	Independent-Samples Mann-Whitney U Test	.010	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

This illustrates the problem with choosing which parts of the data to look at rather than looking at all the data together. You introduce the problem of **familywise error** (i.e., the probability of falsely rejecting the null-hypothesis [a.k.a. Type-I error]). In other words, by running an ANOVA, you REDUCE THE FAMILYWISE ERROR probability, as illustrated by the example above. If these were real data, as the ANOVA was non-significant, you would assume that any variation in your dependent variable was due to chance, not due to any systematic effects of your treatment.

So, the moral of the story is, under NO circumstances is it EVER ok to run multiple tests when one test is possible. It's just good science folks ☺