Exercises Set 9 - Solution

1 Different tests for different situations [basic]

When deciding which test to run, there are some questions to take into account : How large are your samples ? Are the standard deviations known? If you have multiple samples, are they independent or are they "paired" (e.g. you measured the same objects before and after some operation). For two independent samples, can you assume the standard deviations ot be the same ?

To decide if your test is one-sided or two-sided, ask yourself if you want to see if your mean is greater / smaller than a reference value (one-sided), or if your mean is close to / different from a reference value (two-sided).

a) Here, we have two samples to compare you want to compare, with few points and unknown standard deviation. Comparing two small datasets is mostly done with the Welch test or 2-sample t-test. In this case, the standard deviation is unknown and we cannot assume that it is the same for the two samples (since we use 2 processes). So we choose the **Welch test**.
Since we want to see if the two processes are the different or not (similar mean), we use a **2-sided** test.

b) We compare one sample to a reference value, so we use a t-test or z-test. The standard deviation is known so we opt for a **z-test**.
We want to see if the mean is smaller than a reference value, so we use a **one-sided** test.

c) Here we have paired samples, because we are measuring the same molecule before and and after illumination. So although we would have $2 * 8000$ data points, the first thing we do is that we take the difference of the interatomic distance before and after laser illumination (leaving us with 8000 differences). So we have a **paired test**.
Then we want to know if this difference is statistically significantly different from zero or not. Technically, we would have to do a **t-test** because we do not know the standard deviation, but since N_S is so large, a **Gaussian/z-test** is also fine.
We want to detect a difference from zero, we do not care if it is positive or negative. So we use a **2-sided** test.

d) In this test, we need to detect if there is a significant deviation somewhere inside a dataset (the number of times the word is said per politician). To detect this type of deviation, the **ANOVA** test is preferred, with 100 samples and one factor ("which politician").

d) Here, you do not want to make an assumption on the mean, but on the distribution that the data follows. This is a typical use of the χ**2 test**. You will have to first find the mean and standard deviation or the number of microcracks (from your list of 100) to chose which normal distribution to compare to. Note that this removes two degrees of freedom. Then you will have to pick an appropriate set of bins, keeping in mind that you want at least 5 *expected* occurrences in each bin.

2 Three pizzaiolos [normal]

We will run an ANOVA analysis to test for differences in the mean. We will call Albertos judges "group 1", Francescos "group 2" and Paolos "group 3". $N_{Si} = 5$ for each group. First, we compute the group

means as:

$$\bar{X}_1 = 48.2, \bar{X}_2 = 35.4, \bar{X}_3 = 69.8, \bar{X}_T = 51.1$$

Next we compute the squared differences within each group from the respective group mean. For group 1, this means: $SS_1 = \sum_{j=1}^{5}(X_{1,j} - 48.2)^2 = 612.8$. The analogous calculation for the other groups yields:

$$SS_1 = 612.8, SS_2 = 515.2, SS_3 = 732.8$$

The error sum of squares is thus $SS_E = \sum_{i=1}^{k} SS_i = 1860.8$. We calculate the sum of squares between the groups as $SS_B = \sum_{i=1}^{3} N_{Si}(\bar{X}_i - \bar{X})^2 = 3022.9$. We already see that a lot of the SS is coming from variation between groups, but we still need to divide by the degrees of freedom to properly analyze the data.

We have $k = 3$ groups, and each group has $N_{Si} = 5$ participants. Hence between-group degrees of freedom are $\nu_B = k - 1 = 2$ and the error degrees of freedom $\nu_E = 3 * 5 - 3 = 12$.

This leads us to $MS_B = \frac{SS_B}{df_B} = 1511.5$ and $MSE = \frac{SSE}{df_E} = 155.1$, and finally $F = \frac{MS_B}{MS_E} = 9.75$. Now we need the critical F value at $\alpha = 0.05$, either from a table or computer. We find $qF_{2,12}(p = 95\%) = 3.885$. The observed F well exceeds the critical value, thus we have to reject the null hypothesis H_0. There is significant evidence that the pizzas are not the same.

As group 3 has the highest average score, it seems possible that Paolo is the best pizzaiolo. However, his data also have the largest SS (and hance standard deviation, as the smample size is the same for all), so we have to be careful. For example, the rejection of the null hypothesis could also mean that group 2 is particularly bad, and group 1 and group 3 are not different in a statistically significant way. To definitively settle if Paolo is the best, we should do two-sample t tests between group 3 and group 2, and then between group 3 and group 1.

We can construct what is called the ANOVA table from the values above as:

Source	df	SS	MS	F
Group	2	3022.9	1511.5	9.75
Error	12	1860.8	155.1	
Total	14	4883.7		

We have seen in the lecture that $SS_{\text{TOTAL}} = SS_E + SS_B$. Here we computed SS_E and SS_B separately. It is a good crosscheck to also compute SS_{TOTAL} and see if it adds up right.

$$SS_{\text{TOTAL}} = \sum_{k=3,i=1}^{3,5} (X_{i,j} - \bar{X}) = 4883.8$$

Indeed it does, up to a rounding error. This gives us confidence that we did this calculation correctly.

3 Radiation in Switzerland [Computational, normal]

Plot the data first, ideally as both a histogram, and a scatter plot vs time.

We find $N_S = 1084$ $\bar{X} = 0.09849$ and $s = 0.00361$ (the latter is the unbiased estimator for the standard deviation, but in fact at such a large number, divining by 1084 or 1083 does not make much of a difference).

Using the Gauss/z test is fine because N_s is so large (check the t-test table, after 120 comes infinity and the value is basically the same). Of course it would not be wrong to use the t-test, the result would be almost exactly the same.

The critical z for a one-sided z-test at 99.9% is about -3.1.

We have -13.8 so hugely beyond that, it is extremely unlikely for the *mean* daily dose to be above 0.1.

Note that the fact that the *mean* is extremely unlikely to be above 0.1, does not mean that occasional values are not above 0.1. The uncertainty about the mean decreases with more data, the standard deviation itself does not (assuming it is always the same random process generating the data).

4 Is the stockmarket a random process? [Computational, advanced]

Plot the data first, ideally as both a histogram, and a scatter plot vs time. You can just pick one of the columns (first, high, low, last). $N_S = 253$ in any case.

Taking the "first" price for example, (but you can take others), we find a mean of $\bar{X} = 12300$, and $s = 648$ using the unbiased standard deviation.

Make some number of bins (as in the tutorial) using np.histogram(). Get the expected absolute frequencies $p_i * N_s$ from the normal distribution cumulative distribution function (using the error function), and the observed values n_i from the data. Use these to compute the χ_2 statistic.

The number of degrees of freedom you have depends on the number of bins you have chosen, which is k. From that you subtract 1 as always and then the number of parameters of your PDF which you chose based on the data. In this case we determined both mean and standard deviation from the data, so we'll have to subtract another 2. We then have $\nu = k - 1 - 2$.

The resulting χ_2 is very large so we can reject the null hypothesis, the DAX is not normal-random.