Assignment #5 MWW test and al other tests

We have now learned several methods (with their variants)

1. Exact probability (Binomial, Poisson, Normal, Hypergeometric)
2. t-test and z-test, paired and non-paired
3. MWW, 2 samples paired and non-paired
4. Sign test.

In each of the below questions decide which method is most applicable. Use it and explain what exactly you can state. When applicable try the other methods and compare the results. It is possible that none of methods we learned are applicable so be careful…

If you need to assume normality support it by plotting the sample distribution (e.g. using boxplot)

State your null and alternative hypotheses. Use Matlab\Python functions when possible

1. Assume you have two samples: A = [10 20 30], B=[25 35 40 41]
   1. How many different pairs can be constructed if you sample one number from A and one from B? mark this value as U
   2. In how many of these pairs the number from B is higher from the number in A? mark this number UB
   3. In how many of these pairs the number from A is higher from the number in B? mark this number as UA
   4. Check that UA + UB gives U
   5. How many groups of size 3 and 4 can one construct out of 7 different numbers?
   6. In how many of these groups all the numbers from group A are smaller than those in group B?”
   7. Use these value to compute the probability of obtaining UA or less
   8. Verify your result by running a ranksum test in matlab\python (note the one or two sides issue)
2. We are testing the effect of background music on student’s mental performance. Two groups of 15 students each were randomly selected. One group was tested with music A in the background and the other with music B. The scores are:

X = [18 23 21 20 20 29 20 16 20 26 21 25 17 18 19];

Y= [26 21 22 26 19 22 26 25 24 21 23 23 18 29 22];

Are the scores of mental performance significantly different between the two groups?

1. In a clinical trial, survival time (weeks) is collected for 10 subjects with non-Hodgkin's lymphoma. The exact survival time was not known for one subject who was still alive after 362 weeks, when the study ended. The subjects' survival times were

49, 58, 75, 110, 112, 132, 151, 276, 281, 362+

The plus sign indicates the subject still alive at the end of the study. The researcher wished to determine if the median survival time was less than or greater than 70 weeks.

1. State the null and alternative hypotheses
2. Run the appropriate test and explain the result
3. Can you say something about the mean?
4. We count the number of cars in a boarder road block every day. The results are shown below.
   1. Can we claim that the mean number of cars is higher than the previous year that was 60?
   2. Can we claim that the median is higher than the previous year median that was 60?
   3. **Estimate** the probability that the number of cars arriving tomorrow will be less than 20

Number of cars = [45 69 77 77 75 55 83 82 90 101 65 67 68 95 64 110 30 55 50 52 58]

1. Everitt (1994) compared several different therapies as treatments for anorexia. One condition was cognitive behavior therapy, and he collected data on weights before and after therapy. These data are shown below.
   1. Is the treatment significantly better? What can you conclude from this study?
   2. Plot the data in a graph that best demonstrate the effect you found

Before = [ 80.50, 84.90, 81.50, 82.60, 79.90, 88.70, 94.90, 76.30, 81.00, 80.50 85.00, 89.20, 81.30, 76.50, 70.00, 80.40, 83.30, 83.00, 87.70, 84.20 86.40, 76.50, 80.20, 87.80, 83.30, 79.70, 84.50, 80.80, 87.40];

After=[ 82.20, 85.60, 81.40, 81.90, 76.40, 103.6, 98.40, 93.40, 73.40, 82.10, 96.70, 95.30, 82.40, 72.50, 90.90, 71.30, 85.40, 81.60, 89.10, 83.90, 82.70, 75.70, 82.60, 100.4, 85.20, 83.60, 84.60, 96.20, 86.70];

1. We want to compare the mean salary in socialist countries (which is known to be normally distributed) versus capitalist ones (which is known to be highly skewed). We sample 22 salaries from Sweden (e.g. socialist) and 23 from Israel (e.g. semi-socialist). Our hypothesis is that the mean salary is lower in socialist countries.

Sweden = [20 30 30 40 42 45 47 48 50 52 55 56 57 60 60 60 64 65 70 75 80 90]

Israel = [20 20 20 22 23 23 22 25 26 30 30 30 32 33 35 40 52 60 80 90 110 120 350 250]

* 1. Plot the two distributions (use [a,b]=hist(x); plot(b,a))
  2. Does the mean salary in Sweden significantly lower than in Israel?
  3. Does Swedes tend to earn more than Israelis?

1. Does added calcium intake reduce the blood pressure of African American men? In a randomized comparative double-blind trial, 10 men were given a calcium supplement for twelve weeks and 11 others received a placebo. For each subject the researchers recorded whether or not blood pressure dropped. Here are the data:

Subjects Success Proportions

Calcium 10 6 0.60

Placebo 11 4 0.36

Total 21 10 0.48

* 1. Does the proportion of success with calcium is significantly higher than with placebo?

1. Can the full moon influence behavior?

A study observed 15 nursing-home patients with dementia. The number of incidents of aggressive behavior was recorded each day for 12 weeks. Call a day a “moon day” if it is the day of a full moon or the day before or after a full moon. The Table below gives the average number of aggressive incidents for moon days and other days for each subject.

* 1. Is there significant evidence to claim that nursing-home patients are more aggressive on Moon days?
  2. Is this test has an equal or unequal variance? Is the difference in variance is irrelevant?
  3. Any idea why there is more aggression in moon days ????

Patient, Moon days, Other days

|  |  |  |
| --- | --- | --- |
| 1 | 3.33 | 0.27 |
| 2 | 3.67 | 0.59 |
| 3 | 2.67 | 0.32 |
| 4 | 3.33 | 0.19 |
| 5 | 3.33 | 1.26 |
| 6 | 3.67 | 0.11 |
| 7 | 4.67 | 0.3 |
| 8 | 2.67 | 0.4 |
| 9 | 6 | 1.59 |
| 10 | 4.33 | 0.6 |
| 11 | 3.33 | 0.65 |
| 12 | 0.67 | 0.69 |
| 13 | 1.33 | 1.26 |
| 14 | 0.33 | 0.23 |
| 15 | 2 | 0.38 |