**Exercise #3: Central limit theorem , Hypothesis testing**

1. Simulation of the central limit theorem
   1. Construct a non-normal distribution with 1,000,000 elements. For example d = exprnd(10,1,1000000) (an exponential distribution with r = 10)
   2. Compute the mean and std. Plot the histogram d (use hist(d,100)) and write the mean and std values in the title of the figure. Are the numbers close to 10 as expected?
   3. Select 100 elements at random from d and compute the mean and std. Plot the histogram of this 100 elements and write the mean and std values in the title of the figure. Are the numbers close to 10 as expected?
   4. Do this 10000 times and plot the distribution of the mean (use hist(x,100). This is an estimate of the sampling distribution and it should look normal. Plot in the title the mean and std of this distribution. Are the number as expected from the CLT?
   5. Now change the number of sampled data points used (e.g. try 10, 30, 300) and **investigative** how these affect the sampling distribution shape and std
   6. Check if instead of mean you do median or std, does the central limit theorem still works?
2. The mean height in Israel of grownup people is 177 cm and the standard deviation is 10. You sampled at random 36 people. We know that the height distribution is not normal but not very far from normal.
   1. What is the probability that the mean of this sample would be lower than 175?
   2. What is the probability the mean of this sample will be between 175 to 179 cm?
   3. How many people you should sample if you want that with probability > 0.99 that the mean of your sample will be within the above range?
3. You want to test if a coin is fair (p(tail)=p(head)= 0.5). You toss it 1000 times.
   1. What is the number of heads you should obtain so that you can reject the null hypothesis that the coin is fair with α=0.05?
   2. Is it a one sides or two sides test?
4. We know that salaries has a right skewed distribution. Assume the mean of the population salaries is 8000 NIS and that the std is 1000 NIS. We now sample 100 samples.
   1. How the distribution of the 100 salaries we sampled should look like? explain
   2. What is the probability that the mean of this sample will be higher than 8200 NIS?
   3. You want to check if the mean salary has changed. Which values you need to get in order to convince with 95% confidence that the mean salary has changed ?
   4. Can you compute the probability of a having a salary of 9000 or lower? explain
   5. Assume 1% of the people earn more than one million a year. What is the probability that at least one of the 100 samples earn more than one million a year?
5. Assume H0: u >= 75, σ=12. You sampled 36 samples and computed the mean. You reject H0 if the probability to obtain this mean or lower is less than α=0.05 (one tailed).
   1. Compute the critical value (that is, the value that below it you will reject the null hypothesis)
   2. Assume you obtained a mean of 70. What is the p value?
   3. Based on the p value you to decided to reject the null hypothesis. What is the probability of type I error?
   4. Compute the probability that you will not reject the null hypothesis when you actually should have if u is actually 72 (this probability is called type II error)?
   5. Assume now that you want to test if the mean is different from 75 with the same α. Compute the critical values.
6. The speed limit of a freeway in the United States is 120 kilometers per hour. A device is set to measure the speed of passing vehicles. Suppose that the device will conduct three measurements of the speed of a passing vehicle, recording as a random sample X1, X2, X3. The traffic police will or will not fine the drivers depending on the average speed {\displaystyle {\bar {X}}}. That is to say, the test statistic

X=(X1 +X2 +X3 )/3. We know that each measurement is normally distributed with mean=120 and variance=4. The police will not give a fine if the probability that the recorded speed is above 120 by change is less than 0.05

* 1. What is the mean and std of X?
  2. What is the actual speed limit (the one that you really get a fine)
  3. What is the probability you will not get a fine if you speed is 125 ? this is a false negative rate or Type II error