Hi all,  
  
here are comments on some typical errors in Assignment 0. Some of the  
question were already answered or discussed in slack.  
  
General comment: the whole statistics and econometrics is about  
explanatory modelling. That means that we try not only to compute sth  
but also try to find some economic explanation what these numbers mean.  
  
Problem 1.2c: "What can we learn from the corresponding location and  
dispersion (!) measures?(1.2.с)" Mathematically it is an elementary  
question: two samples -> two means and two variances. But we can also  
learn that the average income of younger CEO is lower and more  
importantly their variance is lower. This Indicates that younger ceo  
tend to have the same level of salary, whereas there is larger variation  
for older CEOs, implying their income can heavily depend on other factors.  
  
Problem 1.3.b: Interpretation on n12, h12m n1. and h1.  n12 gives simply  
the number of CEOs in the sample who are younger than 50 AND have salary  
between 3000 and 5000. h12 gives the relative value, i.e. the percentage  
of CEOs satisfying these conditions. n1. is the marginal frequency or  
row sum for the 1st row: the number of CEOs younger than 50 (regardless  
of the salary). h1. is then the % of such CEOs.  
  
Problem 2.1.b: the densities look similar but the t\_df distribution has  
heavier tails, i.e. it allows for more extreme observations. These  
extreme observations are particularly important in practice, since these  
are typically extreme losses, extreme demand for some product, large  
sales, large waiting time, etc. Thus it is not correct just to look at  
the histogram and say "well, looks like normal, lets take normal". It is  
necessary to test the distribution for normality and look for a suitable  
alternative if it is not normal (or at least use the results with care).  
The strange numbers in the t-distribution guarantee that the sample will  
have the same expectation and the same variance as the sample from the  
normal. Note that if the number of degrees of freedom is large than say  
30 then normal and t\_df are almost indistinguishable.  
  
Problem 2.3.a: It looks complicated, but one simply needs to apply the  
definition of the covariance = E(U\*-E(U\*))(V\*-E(V\*)), open the  
parentheses and then divide everything on the standard deviations. So no  
difficult math, just the basic properties and definitions.  
  
Problem 2.3.b: Should give you the feeling how imprecise the estimator  
of correlation is. In fact one believes a lot in rho's, but one needs  
really a large sample to get it precisely. This precision also depends  
on the true correlation: if variables are in fact uncorrelated, then it  
is more difficult to estimate this zero correlation precisely.  
  
Problem 2.3.c: the idea of the task is to show that the Spearman  
measures a non-linear but monotone dependence. Thus if the point lie on  
a monotone curve, than the Spearman is large. So in the simulations if  
you transform the sample with exp, then the Pearson gets smaller (the  
dependence is less linear), but Spearman remains almost the same.

https://ssl.gstatic.com/ui/v1/icons/mail/images/cleardot.gif