ECO220Y: Homework, Lecture 3 – SOLUTIONS

(1) The sum of the heights of the bars for a frequency histogram is equal to the number of observations, n. The sum of the heights of the bars for a relative frequency histogram is 1. The total area of the bars for a density histogram is equal to 1.

(2) This sample clearly comes from non-Normal population, so we cannot apply Empirical rule. We can use Chebyshev's theorem to approximately estimate the standard deviation of these data. According to Chebyshev's theorem, at least 89% of observations should lie within 3 standard deviations from the mean. If the variance is equal to 5, the standard deviation is approximately 2.2, and at least 89% should lie within the range of (40-3*2.2, 40+382.2) or (33.4, 46.6). You can see from the histogram that far fewer than 89% of observations are within this range. Conclusion: the variance of these data cannot be equal to 5, the variance of these data must be greater than 5. (You may provide similar reasoning using 2 standard deviations' rule)

(3) Given that all three samples have the same mean and sample range, we can focus on the shape of the distribution to answer. Sample 1 has the smallest sample standard deviation, and Sample 3 has the largest sample standard deviation. Ask your TA (Inga or Cedric) during the tutorial or office hours for explanation.

(4) (a) We know that profits are normally distributed, and we can use Empirical Rule. According to Empirical Rule, about 95% of observations lie within 2 standard deviations from the mean.

Company A: (10,930 ± 2*4,196)

Company B: (13,000 ± 2*9,300)

Company C: (34,450 ± 2*4,116)

(b) The coefficient of variation is standard deviation/mean.

Company A: 4,196/10,930=0.38

Company B: 9,300/13,000=0.72

Company C: 4,116/34,450=0.12

The company with the largest CV is the most risky one – Company B. The smallest value indicates that the variation of returns relative to the mean return (profit) is low, so the company is more secure to invest in.

(5) (a) Since one-day travel expenses to Moscow are normally distributed, about 99.7% of all values are within 3 standard deviations from the mean (according to Empirical Rule). Therefore, 99.7% of all expenses are in the interval (538 ± 3*41) or (415, 661).

(b) Expenses of $720 should be considered unusually high given the shape of the distribution of all one-day expenses to Moscow. $720 is too far from the mean of the distribution and more than 4 standard deviations away from the mean which could be considered to be an outlier.

(6) This is possible IF this professor is way below the average among Yale faculty, but way above the average quality among Harvard professors. Think of Yale quality distribution as being negatively skewed when that professor was at Yale, and Harvard quality distribution as being positively skewed when he joined the faculty.

(7) No, covariance only indicates the direction of the relationship, but it does not tell us anything about the strength. We cannot compare covariance for different datasets.

(8) Compute coefficient of correlation for this sample: 12/(4*2)=1.5 But we know that correlation takes values between -1 and 1, so you must have made a mistake.