Homework 8: ECO220Y – SOLUTIONS

Required Problems:

(1) Recall the general law of variance for a linear combination of two random variables:

$$V[a + bX_1 + cX_2] = b^2V[X_1] + c^2V[X_2] + 2bc * SD[X_1] * SD[X_2] * CORR[X_1, X_2]$$

In this example, we know that $V[X_{2010} - X_{2005}] = 22.58^2$ (which is given in the title of the third histogram). Further, we know that $SD[X_{2005}] = 35.09$ and $SD[X_{2010}] = 34.37$ (which are given in the titles of the first and second histograms.) The value of $a = 0$, $b = 1$, and $c = -1$. Putting this together, we have:

$$22.58^2 = (1)^2 34.37^2 + (-1)^2 35.09^2 + 2(1)(-1) * 34.37 * 35.09 * CORR[X_1, X_2]$$

Solving for the unknown correlation, we obtain a coefficient of correlation of 0.79. We know it must have been a positive correlation because the variance of the difference is so small, relatively speaking. Notice that for a difference we subtract the term with the correlation: $V[X - Y] = V[X] + V[Y] - 2 * SD[X] * SD[Y] * CORR[X, Y]$. In words, when you are looking at a difference between two highly positively correlated variables (whenever one is big the other tends to be big and whenever one is small the other tends to be small) the difference between these two variables will not have much variation (big numbers minus big numbers are not that different from small numbers minus small numbers).

(2) Here is one: V[X] = 4, V[W] = 4, COV[W, X] = -3. This would correspond to X and W having a coefficient of correlation of -0.75. Then V[Y] = V[X] + V[W] + 2*COV[W,X] = 4 + 4 – 6 = 2. (A less creative example would be to simply point to Required Problem (1), which is another example of this circumstance.)

(3)

$$P(X = 0) = \frac{25!}{0!*25!}0.12^0 * 0.88^{25} = 0.0409; P(X = 1) = \frac{25!}{1!*24!}0.12^1 * 0.88^{24} = 0.1395$$

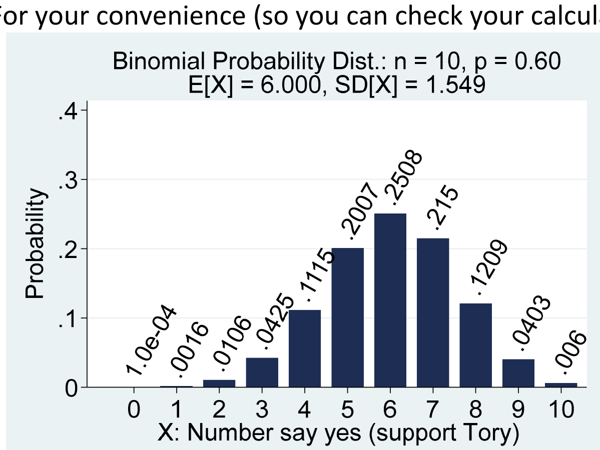
$$P(X = 2) = \frac{25!}{2!*23!}0.12^2 * 0.88^{23} = 0.2283; P(X = 3) = \frac{25!}{3!*22!}0.12^3 * 0.88^{22} = 0.2387$$

$$P(X = 4) = \frac{25!}{4!*21!}0.12^4 * 0.88^{21} = 0.1790$$

$$P(X \geq 5) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) = 0.1736$$

Yes, it is statistically plausible that our sample would have 20% or more delays caused by mechanical issues even if the claim that only 12% of the delays in the population are caused by mechanical issues is true. Our relatively high number of delays can be explained by sampling error: the probability we would see so many delayed due to pure chance is 0.1736, which is pretty high.

(4) (a) For your convenience (so you can check your calculations), the exact heights of each bar are labelled below.



(b) Given that $n = 10$, we must find: $P(5 \leq X \leq 7) = ?$ (Notice that "at least" means greater than or equal to and the "not more than" means less than or equal to.) To find: $P(5 \leq X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7) = 0.2007 + 0.2508 + 0.2150 = 0.6665$.

(5) The number of trials (n) and the probability of success (p).

(6) Mean $= E[X] = n * p = 14 * 0.5 = 7$ and standard deviation $\sqrt{n * p * (1-p)} = \sqrt{14 * 0.5 * 0.5} = 1.87$. One s.d. below the mean is 5.13 and one s.d. above the mean is 8.87. However, a Binomial random variable X is a count of successes and hence an integer. Hence, we looking for $P(6 \leq X \leq 8) = ?$ (or, equivalently, $P(5 < X < 9)$).

$$P(X=6) = \frac{14!}{6!*8!} 0.5^6 * 0.5^8 = 0.1833$$

$$P(X=7) = \frac{14!}{7!*7!} 0.5^7 * 0.5^7 = 0.2095$$

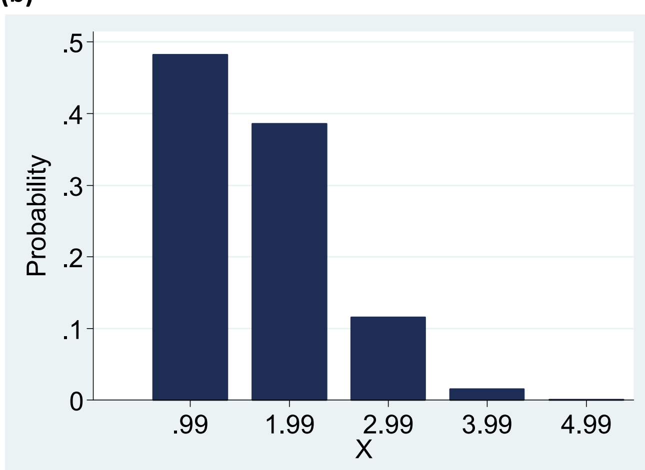
$$P(X=8) = \frac{14!}{8!*6!} 0.5^8 * 0.5^6 = 0.1833$$

$P(6 \leq X \leq 8) = 0.5761$

(7) (a) Define a random variable X to be the price in dollars of a randomly selected item from this store. X is a discrete random variable because it can take only five different values.

x	p(x)
0.99	0.4823
1.99	0.3858
2.99	0.1157
3.99	0.0154
4.99	0.0008

(b)



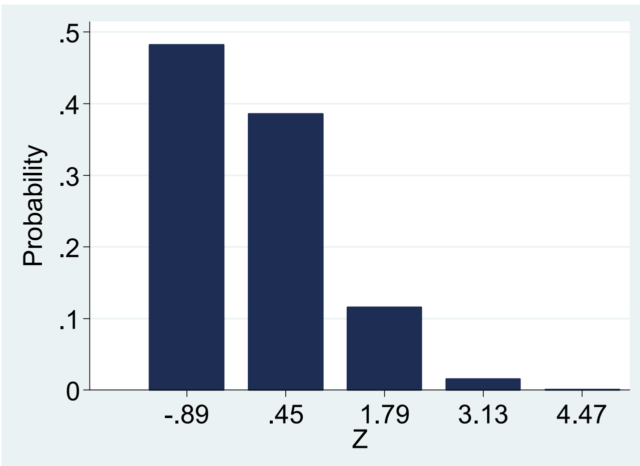
(c) The mean is $1.66. The average price of all the items in the store is $1.66.

(d) The s.d. is $0.75.

(e) [Table is on next page.] A price of $0.99 is 0.89 standard deviations below the mean price. A price of $4.99 is 4.47 standard deviations above the mean price. There are no units of measurement of z.

z	$p(z)$
-0.89	0.4823
0.45	0.3858
1.79	0.1157
3.13	0.0154
4.47	0.0008

(f) The shape is the same but the scale of the horizontal axis has changed. Recall that standardization is a linear transformation and hence will not change the shape (we still have a positively skewed discrete probability distribution).



(8) The answer is **(D)**: we have not been given any information about the correlation in daily sales across the three shifts and it is not reasonable to assume the revenues are independent in this context. For example, on holidays, weekends, and any other days that would affect restaurant traffic, we would expect to see all three shifts affected, which means they'd be positively correlated. Note that **(B)** makes the very strong assumption of independence of revenues across shifts, which makes it incorrect.

(9) (a) $Mean[M_{12} - M_{06}] = Mn[M_{12}] - Mn[M_{06}] = 72.23 - 74.43 = -2.20$ percentage points
$V[M_{12} - M_{06}] = V[M_{12}] + V[M_{06}] - 2rSD[M_{12}]SD[M_{06}] = 6.34^2 + 6.31^2 - 2 * 0.6498 * 6.34 * 6.31 = 28.0207$
Hence, the standard deviation is 5.29 percentage points.

(b)

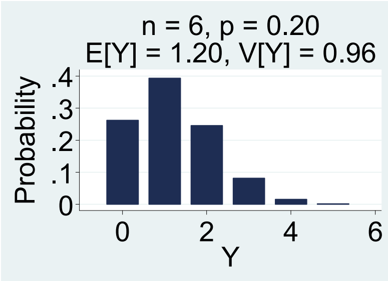
- **i)** False
- **ii)** False
- **iii)** True
- **iv)** False
- **v)** False
- **vi)** True

(10) Just like X is a discrete random variable, Y is also a discrete random variable. The fact that Y does not take on integer values is irrelevant. The definition of a discrete random variable is that it takes on a finite (countable) number of values. If, for example, $n = 5$ then X could be 0, 1, 2, 3, 4, or 5 (six possible values) and Y could be 0, 0.2, 0.4, 0.6, 0.8, 1.0 (six possible values). They are both discrete.

$$E[Y] = E\left[\frac{X}{n}\right] = \frac{1}{n}E[X] = \frac{1}{n}np = p$$

$$V[Y] = V\left[\frac{X}{n}\right] = \left(\frac{1}{n}\right)^2 V[X] = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$$

(11)



When the number of trials rises to 600 it will be close to the Bell shape with a mean of 120 and a s.d. of 9.8.

(12) (a) E[X] = 0.4*0 + 0.5*1 + 0.1*2 = 0.7 sales per day. E[X1 + X2 + X3 + X4 + X5] = E[X1] + E[X2] + E[X3] + E[X4] + E[X5] = 0.7*5 = 3.5 sales per week.

(b) V[X] = $0.4*(0 – 0.7)^2 + 0.5*(1 – 0.7)^2 + 0.1*(2 – 0.7)^2$ = 0.41 sales-squared per day. V[X1 + X2 + X3 + X4 + X5] = V[X1] + V[X2] + V[X3] + V[X4] + V[X5] = 0.41*5 = 2.05 sales-squared per week. Hence, the standard deviation of total weekly sales is 1.43 sales.