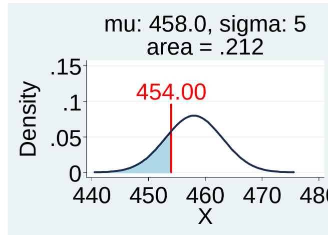
Homework 9: ECO220Y – SOLUTIONS

Note: For Exercises 1 – 8 in "The Normal Table: Read it, Use it" (posted on Quercus), the answers are on pages 6 – 7.

Required Problems:

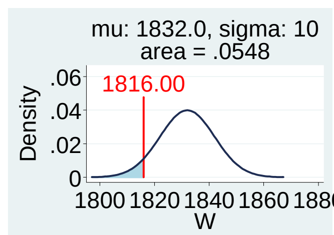
(1) (a) The chance that a single box has less than 454 grams is 0.212.



(b) Each box either is or is not less than 454 grams and each box's weight is independent. Hence it is Binomial with n = 4 and p = 0.212. Let X be the number of boxes that are underweight. P(X >= 1) = 1 – P(X = 0) = 1 - 0.3856 = 0.6144.

(c) The probability that all four are underweight is 0.0020 (= 0.212^4).

(d) Let X1, X2, X3, X4 be the random variables that record the number of grams contained in boxes one through four respectively. P((X1 + X2 + X3 + X4) < 454*4) = ? Because X1 through X4 are all Normal and independent we know that the sum is Normal (see pages 282 – 285 of our textbook) and we can use the laws of expectation and variance to find the mean and variance of the sum.



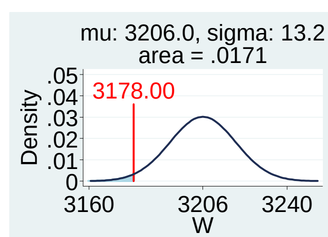
E[X1 + X2 + X3 + X4] = 458 + 458 + 458 + 458 = 1832
V[X1 + X2 + X3 + X4] = V[X1] + V[X2] + V[X3] + V[X4] = 25 + 25 + 25 + 25 = 100

Define W = X1 + X2 + X3 + X4. P(W < 1816) = 0.0548.

(e) Let X1, X2, …, X7 be the random variables that record the number of grams contained in boxes one through seven respectively. P((X1 + X2 + X3 + X4 + X5 + X6 + X7) < 454*7) = ? X1 through X7 are all Normal and independent.

E[X1 + X2 + X3 + X4 + X5 + X6 + X7] = 3206
V[X1 + X2 + X3 + X4 + X5 + X6 + X7] = V[X1] + V[X2] + V[X3] + V[X4] + V[X5] + V[X6] + V[X7] = 175

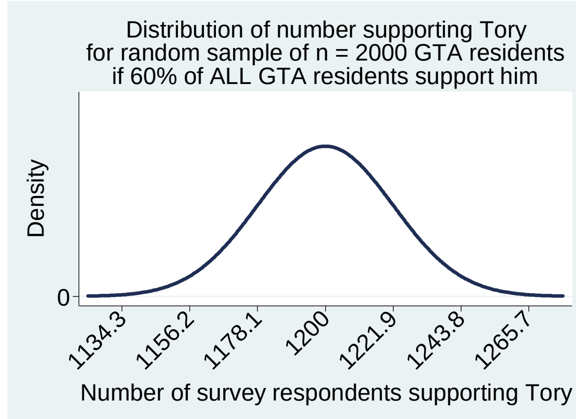
Define W = X1 + X2 + X3 + X4 + X5 + X6 + X7. P(W < 3178) = 0.0171.



(f) P(X1 < 454) = 0.212
P((X1 + X2 + X3 + X4) < 454*4) = 0.0548
P((X1 + X2 + X3 + X4 + X5 + X6 + X7) < 454*7) = 0.0171

The probabilities go down as we combine more and more boxes. This is because when we combine boxes some are likely to be above and some below weight and would cancel as we combine them. (Please remember TK71: random sampling is NOT a self-correcting process. However, as we draw more and more boxes there will inevitably be some cancelling of above average boxes with below average boxes.) The independence assumption is very important to make this argument work: if instead of combining seven independent boxes we multiple one box by seven we would only amplify the variability of a single box. Instead by combining independent boxes there is the chance for random positive draws to offset random negative draws. This is an example of an important concept we will study further and is very much related to the idea that as we increase sample sizes we reduce sampling error.

(2) (a) $X \sim B(n = 2{,}000, p = 0.60)$. We can definitely use the Normal approximation to the Binomial (we pass either of the standard rules of thumb discussed in Lecture 9 with flying colors). $\mu = E[X] = n * p = 1{,}200$ and $\sigma = SD[X] = \sqrt{n * p * (1-p)} = 21.91$. We use the Empirical Rule to tick off values on the x-axis. (You must label the vertical axis "Density" but you do not need to use the Normal density function to find numeric values to tick on the y-axis.)



(b) That probability is, for all practical purposes, 1. Notice in the graph above: 1134.3/2000 = 0.567 and 1265.7/2000 = 0.633. Hence, it is nearly certain that a random sample of 2,000 GTA residents will have somewhere between 56.7% and 63.3% of GTA residents supporting Mayor John Tory so long as it is true that 60% of *all* GTA residents support him (i.e. $p = 0.60$). Obviously, the range 50% to 70% is even wider, so the probability is 1 (even machine precision will round the answer to 1, even though in mathematical theory it would technically be infinitesimally smaller than 1).

(c) The primary difference between this question and required problem (4) in HW 8 is that here you use the Normal approximation to the Binomial and in the earlier homework you had to use the Binomial distribution. Notice the difference in sample sizes across the two homework questions.

(3) About 25% of observations are within ± 0.32 s.d. of mean; About 50% of observations are within ± 0.675 s.d. of mean; About 75% of observations are within ± 1.15 s.d. of mean

(4) (a) The probability that a non-ALDC applicant has an academic rating of 2 or better is $0.4228 = \frac{612+59{,}731}{142{,}728}$. Define the random variable X to be the number of non-ALDC applicants with a rating of 2 or better. In this case it is *not* appropriate to use the Normal approximation to the Binomial so we use the Binomial formula.

$$
\begin{aligned}
&P(X > 1 \mid p = 0.4228, n = 3) \\
&\quad = P(X = 2) + P(X = 3) \\
&\quad = \tfrac{3!}{2!(3-2)!} 0.4228^2 (1 - 0.4228)^{3-2} + \tfrac{3!}{3!(3-3)!} 0.4228^3 (1 - 0.4228)^{3-3} \\
&\quad = 0.3095 + 0.0756 = 0.3851
\end{aligned}
$$

(b) In this case it *is* appropriate to use the Normal approximation to the Binomial. The calculation below also includes the continuity correction.

$$
\begin{aligned}
&P(X > 100 \mid p = 0.4228, n = 300) \\
&\quad = P\left(Z > \tfrac{100.5 - 300*0.4228}{\sqrt{300*0.4228*(1-0.4228)}}\right) = P(Z > -3.08) = 0.4990 + 0.5 = 0.9990
\end{aligned}
$$

(c) The probability of being LDC conditional on being an admitted, non-athelete student is $0.2077 = \frac{2{,}041}{7{,}784+2{,}041}$. Define the random variable X to be the number of LDC students in the sample. In this case it *is* appropriate to use the Normal approximation to the Binomial. The calculation below also includes the continuity correction.

$$
\begin{aligned}
&P(X < 50 \mid p = 0.2077, n = 200) \\
&\quad = P\left(Z < \tfrac{49.5 - 200*0.2077}{\sqrt{200*0.2077*(1-0.2077)}}\right) = P(Z < 1.39) = 0.4177 + 0.5 = 0.9177
\end{aligned}
$$

(5) (a) Define the random variable X to be the number of invitees that accept the invitation. In this case it is *not* appropriate to use the Normal approximation to the Binomial so we use the Binomial formula.

$$P(X > 18 \mid n = 20, p = 0.85) = P(X = 19) + P(X = 20)$$

$$P(X = 19) = \frac{20!}{19!(20-19)!} 0.85^{19}(1 - 0.85)^{20-19} = 0.1368$$

$$P(X = 20) = \frac{20!}{20!(20-20)!} 0.85^{20}(1 - 0.85)^{20-20} = 0.0388$$

$$P(X > 18) = 0.1368 + 0.0388 = 0.1756$$

$$E[Cost] = (1 - 0.1756) * 1000 + (0.1756) * 2000 = 1175.6$$

$$E[Revenue] = E[60 * X] = 60 * E[X] = 60 * 20 * 0.85 = 1020$$

$$E[Profit] = 1020 - 1175.6 = -155.6$$

(b) Define the random variable X to be the number of invitees that accept the invitation. In this case it *is* appropriate to use the Normal approximation to the Binomial. The calculation below also includes the continuity correction.

$$P(X > 1500 \mid n = 2000, p = 0.70) = P\left(Z > \frac{1500.5-(2000*0.70)}{\sqrt{2000*0.70*0.30}}\right) = P\left(Z > \frac{100.5}{20.4939}\right) = P(Z > 4.90) \approx 0$$

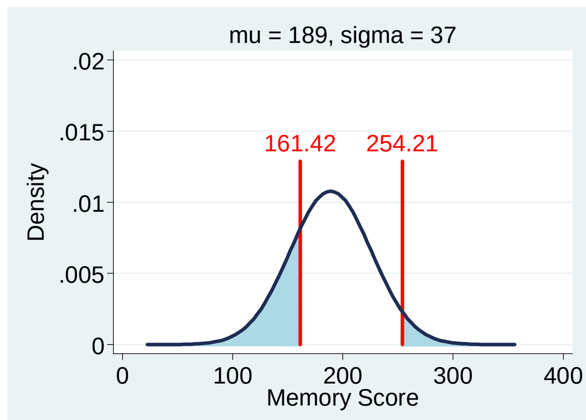
$$E[Cost] \approx 50{,}000 * 1 + 55{,}000 * 0 = 50{,}000$$

(6) We are given that:
$P(X < 254.21) = 0.961$ and $P(X < 161.42) = 0.228$.

Using the Normal table, we know that:
$P(Z < 1.76) = 0.961$ and $P(Z < -0.745) = 0.228$.

Recalling that $Z = \frac{X-\mu}{\sigma}$, we have that:
$1.76 = \frac{254.21-\mu}{\sigma}$ and $-0.745 = \frac{161.42-\mu}{\sigma}$.

Given two equations with two unknowns, solve for μ and σ to obtain: $\mu \approx 189$ and $\sigma \approx 37$.



(7) The standard deviation would be about 16. To get this approximation, you must notice the shape is Triangle. (The sample size is very large, $n = 2{,}500$, so you can see the subtle difference between the Normal shape and Triangle shape in the histogram. You wouldn't be able to see the difference with a small n.) The standard deviation of the Triangle distribution, where $X \sim T[2a, 2b]$, is $SD[X] = \frac{b-a}{\sqrt{6}}$, we see that $2a \approx 0$ and $2b \approx 80$ so $SD[X] \approx \frac{40-0}{\sqrt{6}} = 16.3$.

(8) First have to find the equation of the blue line (density function): in particular, what is y when x is 1? We know that the area under the density function must be 1 and that this is a triangle: A = ½ b*h. Hence, 1 = ½ 4*h and solving for h we find h = 0.5. Thus, when x = 1, y = 0.5. Find equation of the line (we have two points and two points determine a line): $y - y_1 = m(x - x_1)$. We find that the slope = m = (0.5 – 0)/(1 – 5) = -0.125.
y – 0 = -0.125(x – 5)
y = 0.625 – 0.125x
Hence, f(x) = 0.625 – 0.125x if x is between 1 and 5 and f(x) = 0 otherwise.

(9) U[0.5, 6.5]. Keep in mind that we are approximating a discrete probability distribution (rolling a die) with a continuous distribution (the Uniform). Hence, we have to consider the "continuity correction." Note: U[1,6] is *not* a good approximation because according to that the probability of getting a 1 (0.5,1.5) is only 0.5*0.2 = 0.1, which is substantially less than 1/6 (=0.167).