Stat 235

Lab 3

WOOSAREE, Arun

Lab EL12

TA: Jessa Marley

October 26, 2018

Arun Woosaree 1514457

1

As λ increases, the distribution shifts to the right, since λ is the mean value of the distribution. We also see that as this distribution shifts to the right, the curve flattens as a result of the mean probability decreasing while the spread increases (Since the total area under this distribution is 1). Since Poisson distributions measure the number of successes in an interval, and in this case a "success" is a flaw in a plastic panel, we can clearly see that for manufacturing processes with higher λ, the variability of the panels produced increases. This means that we will see slightly more panels with flaws, slightly more panels with less flaws, and less panels with the expected (λ) number of flaws.

2

2.a

Assuming $\lambda = 0.5$, the probability that there are no flaws in a randomly selected panel is $P(X = 0)$. Using the *Poisson Probabilities* worksheet, we find this number to be 0.6065, or 60.65% of the panels are in perfect condition

2.b

The percentage of panels with 2 or more flaws can be found as shown below:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1)$$

Using the *Poisson Probabilities* worksheet, we find $P(X \leq 1) = 0.9098$, so

$$P(X \geq 2) = 1 - 0.9098 = 0.0902$$

So, 9.02% of the panels would need to be scrapped.

2.c

From 2.b, we know the probability of a single panel not being defective is $P(X \leq 1) = 0.9098$. For ten panels,

$$P(X \leq 1)^{10} = 0.9098^{10} = 0.38856110447...$$

Therefore, the probability that a random sample of 10 panels not containing a defective panel is approximately 0.389.

3

From in-class notes, the Central Limit Theorem states the following:

Central Limit Theorem. *When n is sufficiently large, the sampling distribution of $\bar{x}$ is well approximated by a normal curve, even when the population distribution is not itself normal. The Central Limit Theorem can safely be applied if n is at least 30.*

Note: For Question 3, $\lambda = 0.5$

3.a

Referring to the Central Limit Theorem as defined above, we cannot safely assume that the sampling distribution of the average number of flaws in a random sample of 10 panels follows approximately a normal distribution. The Central Limit Theorem states that it can only be safely applied when the sample size is greater than or equal to 30.

If we were to assume that the Central Limit Theorem does apply, however, the probability that the average number of flaws in a random sample of 10 panels does not exceed 0.3 is found as follows:

$$\sigma^2 = \lambda \Rightarrow \sigma = \sqrt{\lambda} = \sqrt{0.5} = \frac{1}{\sqrt{2}}$$

$$\Rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1/\sqrt{2}}{\sqrt{10}} = \frac{1}{\sqrt{20}} \approx 0.2236$$

Plugging in λ, and $\sigma_{\bar{x}}$ into the *Normal Probabilities Sheet*, we find that $P(X \leq 0.30) = 0.1855$.

3.b

We can safely assume that the sampling distribution of the average number of flaws in a random sample of 30 panels follows approximately a normal distribution, because a sample size of 30 satisfies the condition in the Central Limit Theorem which requires a sample size greater than or equal to 30.

Assuming that the Central Limit Theorem does applies, the probability that the average number of flaws in a random sample of 30 panels does not exceed 0.3 is found as follows:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1/\sqrt{2}}{\sqrt{30}} = \frac{1}{\sqrt{60}} \approx 0.129$$

Plugging in λ, and $\sigma_{\bar{x}}$ into the *Normal Probabilities Sheet*, we find that $P(X \leq 0.30) = 0.0607$.

4

4.a

Using the built in COUNTIF function, we find the number of perfect panels to be 362. So, the percentage of perfect panels can be found as follows:

$$\frac{362}{600} \times 100 = 60.333333333...\%$$

Therefore, about 60.33% of the panels are in perfect condition. Compared with the value obtained in 2.a (60.65%), it is pretty close. We expect these values to be close, because the random data is based on the same distribution as in 2.a. However, it would be unrealistic to expect that these two values are exactly the same, since the value obtained in 2.a is a theoretical value for an infinite number of trials, while the number obtained here is from random data representing a finite number of plastic panels.

4.b

Using the built in COUNTIF function, we find the number of samples with non defective panels to be 23. So, the percentage of non defective panels can be found as follows:

$$\frac{23}{60} \times 100 = 38.333333333...\%$$

Therefore, about 38.33% of the panels are not defective. Compared with the value obtained in 2.c (38.86%), it is pretty close. We expect these values to be close, because the random data is based on the same distribution as in 2.c. However, it would be unrealistic to expect that these two values are exactly the same, due to the randomness of the generated data. Also, the value obtained in 2.c is a theoretical value for a sample of 10 from an infinite number of panels, while we are obtaining samples from 600 randomly generated panels.

4.c

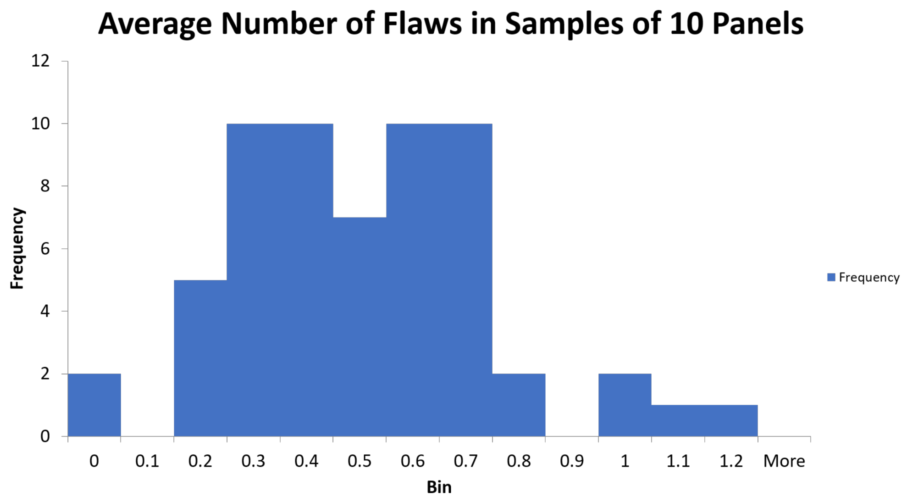

Figure 1: Average number of flaws in samples of 10 panels from 600 plastic panels.

The histogram above is a bit funky, and does not seem to resemble a normal distribution. This makes sense, because the Central Limit Theorem makes no guarantees that sample sizes less than 30 (10 in this case) will approximate a normal curve. When taking a step back and looking at the overall shape, it appears to be bimodal with the first peak being between 0.3-0.4 and the second peak being between 0.6-0.7, however there is only one bin between these two apparent peaks, so this could also be attributed to the randomness in the data.

4.d

From the automatically computed summary statistics in the *Simulation* worksheet, we find the mean and standard deviation of the average number of flaws for the 60 samples to be: $(\mu = 0.506666667, \sigma_{\bar{x}} = 0.242771195)$. In 3.a, we found the theoretical values to be $(\lambda = \mu = 0.5, \sigma_{\bar{x}} = 0.2236)$. These values are reasonably close, but we would expect them to be even closer with a larger sample size (≥ 30) due to the Central Limit Theorem. Even though we expect these values to be close, it would be unrealistic for them to be exactly the same, due to the randomness of the data, and also because we are taking samples from 600 randomly generated panels while the theoretical value is from taking samples from an infinite number of panels.

5

5.a

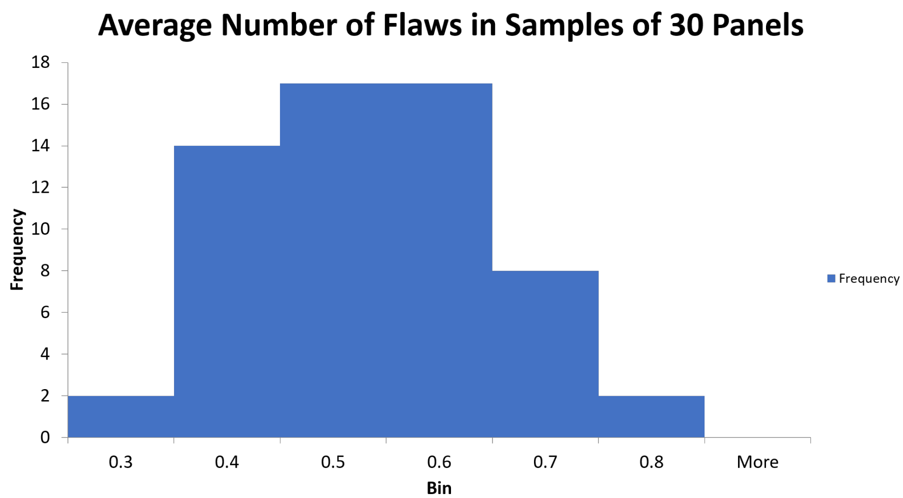

Figure 2: Average number of flaws in samples of 30 panels from 1800 plastic panels.

Unlike the histogram in 4.c, the histogram above appears to resemble a normal distribution. This makes sense, because the Central Limit Theorem guarantees that sample greater than or equal to 30 will approximate a normal curve, and we are picking samples with 30 panels. Compared to 4.c, where the sample size was 10, we saw that the resulting histogram did not look normal at all, while the histogram where the sample size was 30 is clearly unimodal, and visually symmetric like a normal distribution.

5.b

From the automatically computed summary statistics in the *Simulation* worksheet, we find the mean and standard deviation of the average number of flaws for the 60 samples to be: ($\mu = 0.502222222, \sigma_{\bar{x}} = 0.119141099$). In 3.b, we found the theoretical values to be ($\lambda = \mu = 0.5, \sigma_{\bar{x}} = 0.129$). These values are also reasonably close, and are actually even closer to each other compared to 4.d. This makes sense, because the Central Limit Theorem states that sample sizes greater than or equal to 30 should approximate a normal curve. Even though we expect these values to be close, it would be unrealistic for them to be exactly the same, due to the randomness of the data, and also because we are taking samples from 1800 randomly generated panels while the theoretical value is from taking samples from an infinite number of panels. Compared to 4.d, this estimate seems better, since the experimental values obtained are closer to the theoretical values. This is so, because according to the Central Limit Theorem, the sample distributions should approximate a normal curve with large enough sample size (≥ 30). In 4.d the sample size was smaller, and it did not satisfy the Central Limit Theorem, while here, we have a bigger sample size which does satisfy the Central Limit Theorem, so naturally the estimate here would be better than in 4.d.