Final Exam: Statitistical Inference

Let Φ denot cdf o $\mathcal{N}$ st $\Phi(2.85) = 0.998$ and $\Phi(0.285) = 0.6725$

Chebyshev: $\mathbb{P}[\,|X-\mu| \geq \varepsilon\,] \leq \frac{\sigma^2}{\varepsilon^2}$

1) A statistical model is a distribution of a generated sample. for example, coin tossing is a Bournoulli process (two outcomes) and thus follows a Bournoulli distribution (aka Barnoulli Statistical model)

2) a) median
b) mean, std dev, variance

3) $\mu = 525$ $\sigma = 30$ $\varepsilon = \frac{1-0.889}{2} = \frac{0.111}{2} = 0.0555$ $\mathbb{P}[\,|X-\mu| \geq \varepsilon\,] \leq \frac{\sigma^2}{\varepsilon^2}$

Plugging into Chebyshev: $\mathbb{P}[\,|X-525| \geq 0.0555\,] \leq \frac{\sigma^2}{(0.0555)^2}$

We know that $[\mu-30, \mu+30]$ covers 88.9% from Chebyshev

$\Rightarrow [525-3(30), 525+3(30)] \Rightarrow \boxed{[435, 615]}$

4) $\mu = 20$ $\sigma = 7$ $n = 100$

a) we can say that sampling distripution is normal by CLT with mean $\bar{x} = 20$ and $s = \frac{\sigma}{\sqrt{n}} \Rightarrow \frac{7}{\sqrt{100}} \Rightarrow \frac{7}{10} = 0.7$ since n is sufficiently large, (100>30) ~~the~~ question our sample was randomly selected (r.v.), and it was unplanned (i.i.d.), $\mathbb{E}[X] = \mu$ $\sigma^2 < \infty$, so all conditions are met to apply Central Limit Theorem without assumptions.

b) $\mathbb{P}[18 \leq \bar{X}_n \leq 22] = \mathbb{P}[\bar{X}_n \leq 22] - \mathbb{P}[\bar{X}_n \leq 18]$

$\Rightarrow \mathbb{P}[\frac{22-20}{0.7} \geq Z] - \mathbb{P}[\frac{18-20}{0.7} \geq Z] \Rightarrow \mathbb{P}[2.86 \geq Z] - \mathbb{P}[-2.86 \geq Z]$

$\Rightarrow 0.9979 - 0.0021 = \boxed{0.09958}$ using z table in calculator

c) $\mathbb{P}[18 \leq X \leq 22] = \mathbb{P}[X \leq 22] - \mathbb{P}[X \leq 18]$

$\Rightarrow \mathbb{P}[\frac{22-20}{7} \geq Z] - \mathbb{P}[\frac{18-20}{7} \geq Z] \Rightarrow \mathbb{P}[0.29 \geq Z] - \mathbb{P}[-0.29 \geq Z]$

$\Rightarrow 0.6141 - 0.3859 = \boxed{0.2282}$

d) The reasoning is because the standard deviation is larger in X than it is in $\bar{X}_n$ ($7 > 0.7$). We know this by property of the CLT, which looks to normalize the sampling distribution by dividing by the sqrt of sample size.

e) Intuitively, this make sense because it is easier ~~to~~ predict the behavior of an average customer, through sampling, from the population & try to fit them between a CI. An individual customer has size $n=1$, whereas the average customer, under CLT, can come from a sample of size $1 < 30 \leq n \leq$ population size $< \infty$
in this case 100

5) The Fundamental Theorem of statistics is as follows:

$$\sup_x |\hat{F}(x) - F(x)| = |\hat{F} - F|_\infty$$

which states that through the Law of Large Numbers and Central Limit theorem, the emperical distribution function $\hat{F}(x)$ will almost surely converge to the true statistical distribution function.

More simply put, as our sample gets larger, we can better approximate the population (true distribution)

6) Slutsky's Theorem states that:

$$\lim_{x \to \infty} F_{X_n + Y_n}(t) = F_{X+c}(t) \qquad t \in \mathbb{R};\ c \text{ is a constant};\ X_n, Y_n \text{ are distributions}$$

Slutsky Theorem is especially helpful for establishing Normal Distribution with an estimated variance.
If we know sample variance S^2 converges in probability to σ^2, by Slutsky & CLT

$$\frac{\sqrt{n}\,(\bar{X} - \mu)}{S} \xrightarrow{\text{converges}} \mathcal{N}(0,1).$$

This can be very practicle for any sample taken in any domain.