INTRODUCTION TO STATISTICS

PS07

Q1. Let X be a discrete random variable with probability mass function

$$P(X=x)=\begin{cases}0.3 & x=-2\\ 0.6 & x=-1\\ 0.1 & x=12\\ 0 & \text{otherwise.}\end{cases}$$

Let X1, . . . , Xn be an iid sequence of random variables with the same distribution as X. Let $\bar{X}$ be the sample mean (of X1, . . . , Xn.)

(a) Find EX.

(a) Find $E(X)$

$E(X) = \sum x_i \cdot p(x_i)$

$\Rightarrow (-2 \times 0.3) + (-1 \times 0.6) + (12 \times 0.1)$

$\Rightarrow -0.6 - 0.6 + 1.2$

$E(X) = 0$

(b) Find Var(X).

(b) $V(X) = E(X^2) - [E(X)]^2$

$E(X^2) = \sum x_i^2 \cdot p(x_i)$

$\Rightarrow (-2^2 \times 0.3) + (-1^2 \times 0.6) + (12^2 \times 0.1)$

$\Rightarrow 1.2 + 0.6 + 14.4$

$\Rightarrow E(X^2) = 16.2$

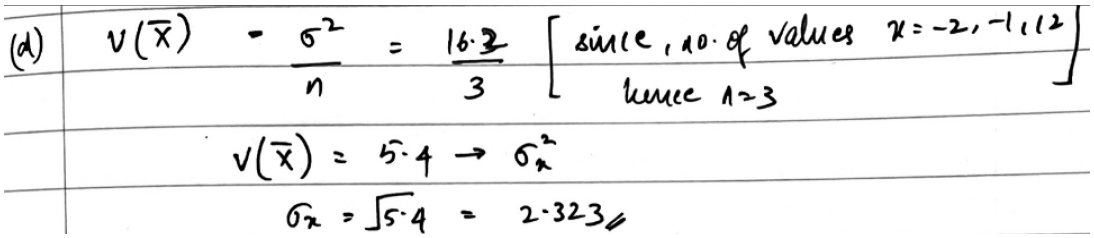
$V(X) = E(X^2) - [E(X)]^2 = 16.2 - 0$

$V(X) = 16.2$

(c) What is the expected value of $\bar{X}$?

(c) $E(\bar{X}) = E(X) = 0$

(d) What is the variance of $\bar{X}$? (Note: This will depend on n.)



(e) Suppose n = 100. Use the R function pnorm() to find the approximate probability that $\bar{X}$ is greater than 0.5

```
> mu<- 0
> sd<- 2.323
> n<- 100
> xbar<- 0.5
> Z<- (xbar-mu)/(sd/sqrt(n))
> Z
[1] 2.152389
> pnorm(Z)
[1] 0.9843166
> 1-pnorm(Z)
[1] 0.01568336
```

Q2. I downloaded data on the number of citations for a random sample of 1000 journal articles published in 1981. (The data is from the ISI Citation Indexes.) I ran some analysis on the data in R, and produced the following output:

```
> citations = scan("citations.txt")
Read 1000 items
> summary(citations)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    1.00    9.06    7.25  300.00
> var(citations)
[1] 565.2476
> # Number of articles with no citations
> sum(citations == 0)
[1] 460
```

(a) Is the distribution of the number of citations (i) exactly normal, (ii) approximately normal, or (iii) not close to normal? How do you know?
(b) Find an appropriate 95% confidence interval for the mean number of citations,
(c) Find an approximate 95% confidence interval for the proportion of journal articles with no citations.