

Solution: Homework 4 Due August 10, 2020

- Exercises 4.2.1, 4.2.2, 4.2.3 (Page 98, 4th Ed).
- Data file: EXR_C04_S02_01
- Hint: Modify the Stata do file lecture05.do to complete the exercise 4.2.1 and 4.2.3. Do the calculation for 4.2.2 by hand.

4.2.1. In a study by Cross et al. (A-2), patients who were involved in problem gambling treatment were asked about co-occurring drug and alcohol addictions. Let the discrete random variable X represent the number of co-occurring addictive substances used by the subjects. Table 4.2.4 summarizes the frequency distribution for this random variable.

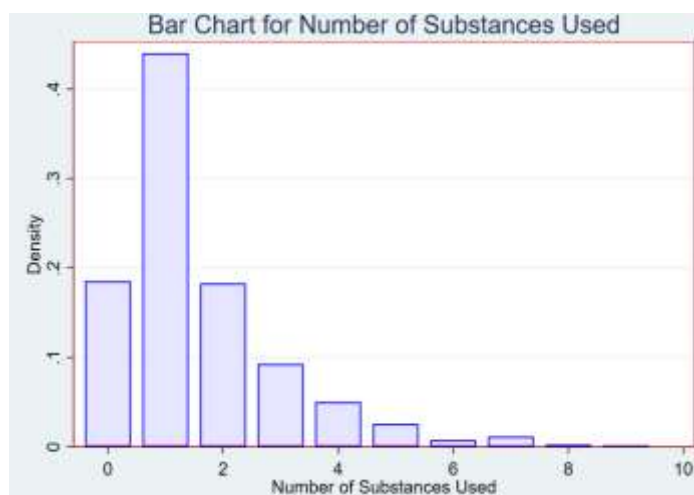
TABLE 4.2.4 Number of Co-occurring Addictive Substances Used by Patients in Selected Gambling Treatment Programs

Number of Substances Used	Frequency
0	144
1	342
2	142
3	72
4	39
5	20
6	6
7	9
8	2
9	1
Total	777

- (a) Construct a table of the relative frequency and the cumulative frequency for this discrete distribution.
- (b) Construct a graph of the probability distribution and a graph representing the cumulative probability distribution for these data.

Solution:

nstused	frequency	Frac	cum
0	144	.1853282	.1853282
1	342	.4401544	.6254826
2	142	.1827542	.8082368
3	72	.0926641	.9009009
4	39	.050193	.951094
5	20	.02574	.976834
6	6	.007722	.984556
7	9	.011583	.996139
8	2	.002574	.998713
9	1	.001287	1



***command for calculating relative percentage and cumulative percentage**

```
tabulate nstused[fweight= frequency]
```

*** Plot probability density function**

```
histogram nprg[fweight= frequency], discrete color(blue) fintensity(inten10) barwidth(0.8) xlabel(#8)
```

*** Generate new variables for relative frequency i.e., Probability density function**

```
egen totalFreq= sum(frequency)
```

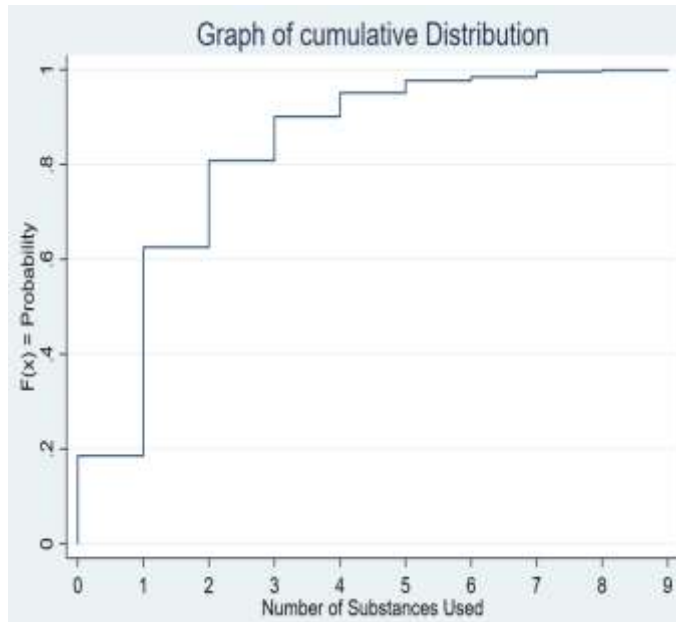
```
gen Frac= frequency/totalFreq
```

```
label variable Frac"Relative Frequency"
```

```
drop totalFreq
```

*** Generate cumulative distribution**

```
cumulnprg[fweight= frequency], generate(cum)tabulate nstused[fweight= frequency]
```



***Command for Make the plot of cumulative
*distribution**

```
gen cum1 = cum[_n -1]
replace cum1=0 if cum1==.
rename cum cum2
reshape long cum, i(nstused) j(idx)
label variable cum "Probability"
set obs 21
replace nstused= if nstused==.
replace nstused= 9 if nstused==.
replace cum = 1 if cum==.
sort nstused idx
line cum nstused, sort xlabel(#10)
```

4.2.2. Refer to Exercise 4.2.1.

(a) What is probability that an individual selected at random used five addictive substances?

Ans: $P(X = 5) = \frac{20}{771} = .0257$

(b) What is the probability that an individual selected at random used fewer than three addictive substances? Ans: $P(X < 3) = P(X \leq 2) = .808$ (see cumulative frequency Table above).

(c) What is the probability that an individual selected at random used more than six addictive substances? Ans: $P(X > 6) = 1 - P(X \leq 6) = 1 - .985 = .015$ (see cumulative frequency Table above)

(d) What is the probability that an individual selected at random used between two and five addictive substances, inclusive? Ans: $P(2 \leq X \leq 5) = P(X \leq 5) - P(X < 2) = P(X \leq 5) - P(X \leq 1) = .977 - .625 = .352$ (see cumulative frequency Table above)

4.2.3. Refer to Exercise 4.2.1. Find the mean, variance, and standard deviation of this frequency distribution.

Solution:

$$\bar{X} = \sum xP(X = x) = 0 \times .185 + 1 \times .44 + \dots + 9 \times .0013 \approx 1.6$$

$$S^2 = \sum (x - \bar{X})^2 P(X = x) = (0 - 1.6)^2 \times .185 + (1 - 1.6)^2 \times .44 + \dots + (9 - 1.6)^2 \times .0013 \approx 2.15$$

$$S = \sqrt{2.15} \approx 1.5$$

Or we can also extract mean, variance, and standard deviation from Stata output.

Command:

summarize nstused [fweight = frequency], detail