Review of Ch1 – Ch5

Ch1 Intro to Data and Statistics

I. Key terms of a Data set

element, variable, observation

II. Types of data

Categorical

Quantitative (Discrete and Cont.)

III. Statistics and Statistical Inference

Ch2: Descriptive Statistics: Tabular and Graphical Presentation

Tables for Data Distributions :

For categorical **---** Freq, relative Freq, Percent Freq

For quantitative **---** Freq, relative Freq, Percent Freq

+ Cumulative Distribution

Graphs for Data Distributions:

1. Categorical data

   a. Pie charts b. Bar charts

2. . Quantitative data

   a. Dot plot b. Histograms c. Cumulative

3. Describing data distributions

   a. Shapes b. Outliers

Ch3 Numerical Measures

I. Measures of Location

1. Mean (Population / Sample mean)
2. Median
3. Mode

II. Measures of Variability

1. Range: R = largest – smallest
2. Variance(Population / Sample Variance)
3. Standard deviation

III. Measures of Relative Standing

***1. p*th percentile** --- *p%* of the observations are smaller, and (100 – *p*)% are larger.

index: i=(p/100)*n (Two situations: i is integer or not integer)

2. First quartile, Q_1; **i**= (25/100)**n*

3. Third quartile, Q_3; **i**= (75/100)**n*

4. Inter-quartile range: $IQR = Q_3 - Q_1$

5. Boxplot **(Min , Q_1, Median, Q_3, Max) ---- shape and outlier**

IIII. Measures of Association Between Two Variables X and Y

1. Covariance S_{xy}, Correlation Coefficient r_{xy}

3. Interpret the linear relationship between X and Y

Ch4 Intro to Prob

I. Experiments, events and the Sample Space

II. Probabilities

1. Properties of probabilities

    a. Each probability lies between 0 and 1.

    b. Sum of the possible of all possible outcomes equals 1.

2. *P(A),* the sum of the probabilities for outcomes in *A*

3.P(S)=1

III. Counting Rules

1. *mn* Rule;
2. Permutations:
3. Combinations:

$$P_r^n = \frac{n!}{(n-r)!}$$

$$C_r^n = \frac{n!}{r!(n-r)!}$$

Ch4 Intro to Prob

IV: Basic Relationships of Probability

1. Union: $A \cup B$ (either A or B)

   Intersection: $A \cap B$ (A and B)

2. Events

a. Complementary: $P(A) = 1 - P(A^C)$

b. Mutually exclusive: $P(A \cap B) = 0$

c. Independent: $P(A \cap B) = P(A)P(B)$

$$or\ P(A|B)=P(A)$$

$$or\ P(B|A)=P(B)$$

Ch4 Intro to Prob

3. Conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

4. Joint Probability Table

5. Additive Rule of Probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

6. Multiplicative Rule of Probability:

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

7. Law of Total Probability

$$P(B) = P(A \cap B) + P(A^c \cap B) \quad = P(A)P(B|A) + P(A^c)P(B|A^c)$$

8. Bayes' Rule

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

Ch5 Discrete Prob Distribution

I. Random variables (discrete and continuous)

Properties of probability function

$$0 \le f(x) \le 1 \text{ and } \sum f(x) = 1$$

II. Expected value: $E(X) = \sum x f(x)$

Variance and standard deviation

$$\text{Variance: } \sigma^2 = \sum (x-\mu)^2 f(x)$$

$$\text{Standard deviation: } \sigma = \sqrt{\sigma^2}$$

III. 1.Properties of a Binomial Experiment

2.Binomial Probability function (w/ n and p)

$$f(k) = P(X = k) = C_k^n p^k q^{(n-k)}$$

3. Expected value and variance

$$E(X) = \mu = np$$

$$Var(X) = \sigma^2 = npq$$

Example 1

A survey of job satisfaction of teachers was taken, giving the following results

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	74	43	117
	High School	224	171	395
	Elementary	126	140	266
	Total	424	354	778

If all the cells are divided by the total number surveyed, 778, the resulting table is a table of joint probabilities.

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

For convenience, let C stand for the event that the teacher teaches college, S stand for the teacher being satisfied and so on. Let's look at some probabilities and what they mean.

$P(C) = 0.150$ is the proportion of teachers who are college teachers.

$P(S) = 0.545$ is the proportion of teachers who are satisfied with their job.

$P(C \cap S) = 0.095$ is the proportion of teachers who are college teachers and who are satisfied with their job.

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.658
	Total	0.545	0.455	1.000

$P(C \cup S)$?

$P(C) = 0.150$, $P(S) = 0.545$ and

$P(C \cap S) = 0.095$, so

$$P(C \cup S) = P(C) + P(S) - P(C \cap S)$$

$$= 0.150 + 0.545 - 0.095$$

$$= 0.600$$

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
L E V E L	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

$$P(C \mid S) = \frac{P(C \cap S)}{P(S)} = \frac{0.095}{0.545} = 0.175$$

is the proportion of teachers who are college teachers given they are satisfied. Restated: This is the proportion of satisfied that are college teachers.

$$P(S \mid C) = \frac{P(S \cap C)}{P(C)} = \frac{P(C \cap S)}{P(C)} = \frac{0.095}{0.150} = 0.632$$

is the proportion of teachers who are satisfied given they are college teachers. Restated: This is the proportion of college teachers that are satisfied.

		Job Satisfaction		
		Satisfied	Unsatisfied	Total
LEVEL	College	0.095	0.055	0.150
	High School	0.288	0.220	0.508
	Elementary	0.162	0.180	0.342
	Total	0.545	0.455	1.000

Are C and S independent events?

$$P(C) = 0.150, P(S) = 0.545$$

$$P(C \mid S) = \frac{P(C \cap S)}{P(S)} = \frac{0.095}{0.545} = 0.175$$

P(C|S) ≠ P(C) so C and S are dependent events.

Or P(C∩S) ≠ P(C)P(S) so C and S are dependent events.

Example 2

Suppose a rare disease infects one out of every 1000 people in a population. And suppose that there is a good, but not perfect, test for this disease: if a person has the disease, the test comes back positive 99% of the time. On the other hand, the test also produces some false positives: 2% of uninfected people are also test positive. For someone just tested positive, what are his chances of having this disease?

Define A: has the disease   B: test positive

Information we know:

$P(A) = .001$ $P(A^c) = .999$

$P(B|A) = .99$ $P(B|A^c) = .02$

We want to know $P(A|B) = ?$

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^C)P(B|A^C)}$$

$$= \frac{.001 \times .99}{.001 \times .99 + .999 \times .02} = .0472$$

Example 3

- The following probability distribution represents the number of jobs college students
- at UCSB have:
- x 0 1 2
- P(X=x) 0.5 0.3 p
- (a) What is the probability that a randomly picked UCSB student has two jobs?
- (b) What is the probability that a randomly picked UCSB student has at least one
- job?
- (c) What is the expected value and standard deviation of the number of jobs UCSB students have?

Example 4

In San Francisco, 30% of workers take public transportation daily.

(a) In the sample of 10 people, what is the probability that exactly 3 workers take public transportations daily?

(b) In the sample of 10 people, what is the probability that at least 3 workers take public transportations daily?