9/3/2009

Two types of Data

1. variable - takes on more than one value. variable varies

2 constant - takes on only one constant = does not change

- Point ~~estimation~~ Estimation - estimate the value of a particular value
- Examining Relationships - one variable is related or corolated with another. Example: if x is related to y

can have two variables that don't covary

negative slope down

Causal relationship -

- Independent variable - the cause
- dependent variable - the effect
- covariation - change in response to one another

temporal operation

Time ordering - cause of y must come first. Ex: Gender and Deliquency Age & Deliquency, Grades in School & Deliquency

Non spuriousness

- In order for X to cause Y we need to eliminate other possible of Y we need to make sure that we do not confuse other things that happen before Y.

Levels of Measurement

- Nominal-Gender
- Ordinal 15-20-21-25
- Interval = arbitrary 0 Age
- Ratio = rate frequency

Frequency - List of values of Data and the # of times it appears

Frequency Distribution - full array of values and # of times it occurs

Rate ~~F/population x standard~~

Rat = F/population x per 100,000 standard unit

Standard time size x by 12

porportion divide Frequency in total # of cases

mutiply by 100 to get %

9/8/2009

~~difference between~~ difference between 2 intervals is the same interval ratio.

Porportion (f/n x 100)

percent change
ending value - starting value)

1980 1,241 per 100,000
1990 1,789 per 100,000
2000 1,137 per 100,000

(ending ~~year~~ value - starting ~~year~~ value)/starting ~~year~~ value
x 100

1,789 - 1,241 = 548 ÷ 1,241 = .442
100 = 44.2

~~1,137 - 1,789~~

- = decrease

Number of Bad Arrest ~~from~~ in a sample of 30 police officers

person # 1	12	9
person #2	4	4
	6	10
	7	6
	13	8
	7	3
	10	4
	5	9
	8	4
	2	0
	7	5
	0	
	1	
	4	
person #15	6	

cf = cumulative frequency start at top work your way down

% vector multiply by 100

0-4 range value (class interval)

Grouped frequency distribution

Class interval = mutually exclusive
Fall into 1 and only one categories

2. Make your class intervals exhaustive
3 Make your intervals all the same width.
4. Make sure that for your first class interval contains the smallest value + the last interval the highest.

Step 1 determine the number of class intervals you want

Step 2 subtract high 89 to low 22 = 67.
8. Take the ratio of the range
67/8 = 8.375 = 8.
Class interval width of 8.
20 lower interval width
~~20-28~~
20-27
28-35
36-

Nominal 3 ordinal Level data
Pie chart
Bar chart

19.5-27.5, 27.5-37.5 = continuous

Real Class - not mutually exclusive

midpoint = exact middle of every class interval

$M_I = \frac{LL + UL}{2}$ $\frac{\text{Lower Limit} + \text{Upper Limit}}{2}$

$\frac{83.5 - 91.5}{2} = 87.5$

mode, median, mean
highest #

bimodal = two modes
trimodal = 3 modes

median interval/ratio level
50% percentile

1. Rank order the data
Find the score(s) that is in the median with formula

$x(\text{median}) = \frac{(n+1)}{2}$ not give you the value of median but the position of the median

$X = L + \left[\frac{\frac{n+1}{2} - cf}{f} \right] w_i$

$43.5 + \left[\frac{36+1}{2} \right]$

$43.5 + \left[\frac{\frac{100+1}{2} - 36}{16} \right] 8$

$43.5 + (.91)8 = 43.5 + 7.28 = 50.78$ marks mean: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{N}$

Advantages- we use all our information

The distance from the [illegible]
deviation.

$$s = \sqrt{\frac{\Sigma(x-\bar{x}^2}{n-1}}$$

$$s = \frac{\Sigma(x-\bar{x}^2)}{n-1}$$

$x = \sqrt{x^2}$ = standard deviation [illegible]
[illegible] to the square [illegible]
the [illegible]
[illegible]
to the variance

Sum $x = \sqrt{\frac{10}{10-1}} = \sqrt{\frac{10}{9}} = 1.05$ [illegible]

1.11 variance

Take each score subtract the mean [illegible]
squared each deviation
the sum overall variation

subtract
each score
square
each [illegible]
then add
to get
the sum

$$x = \sqrt{\frac{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}{n-1}}$$

$$x^2 = \frac{\Sigma(x^2) - \frac{(\Sigma x)^2}{n}}{n-1} \quad (x_1 + x_2 + x_3 \ldots)^2$$

$x_1^2 + x_2^2 + x_3^2$

rect of x^2 squares

sum

$$\frac{170 - 40^2}{10} \div 4$$

$$\frac{170 - \left(\frac{40^2}{10}\right)}{10-1}$$

$$s^2 = \frac{\ldots}{n-1}$$ NOT [illegible]

- [illegible] square
- Square the [illegible]

Multiply the square [illegible]
by the [illegible]
Sum across all [illegible]
Square

Nominal Ordinal

Mode
Median
Mean
Variation Ratio
Interquartile Range
Variance
Standard deviation

Variation Ratio

$$VR = 1 - \frac{F_{mode}}{n}$$

	f
Small (50-?)	20
Medium (?-99)	50
Modal Large (100-154)	90
Super Large (155+)	10
	N=170

$$1 - \frac{90}{170} \quad 1 - .53 = .47$$

NJ	F
Small	50
Medium	20
Large	20
Supe	100 ← Mode
	190

$1 - \frac{100}{190}$

$1 - .59 = .41\%$

Bimodal

X	F	P	%	C%
10	6	.06	6%	6%
11	7	.07	7%	13%
12	9	.09	9%	22%
Q3 13	25	.25 ×	[illegible]	47%
14	23	.23 ×	[illegible]	70%
Q1 15	14	.14	[illegible]	84%
16	6	.06	[illegible]	90%
17	3	.03	[illegible]	93%
18	7	.07	[illegible]	100%
	N = 100	1	100	

CF
6
13
22
47
70
84
90
93

Q

$\frac{50+1}{2} = 25.5$

Number of Pr...

		x^2
1.	0	0
2.	2	4
3	3	9
4	4	16
5	1	1
6	1	1
7	3	9
	$n=7$	40

Mean $\frac{14}{7} = 2$

$$s = \sqrt{\frac{\Sigma - \frac{(\Sigma x)^2}{n}}{n-1}} \qquad \frac{40 - \frac{(14)^2}{7}}{6}$$

$$\sqrt{40 - \frac{196}{7}}$$

$$\sqrt{\frac{40-28}{6}}$$

$$\frac{12}{6} \quad s = \sqrt{2}$$

$s = 1.41$

$s^2 = 2$

$S = \sqrt{\sum F m^2 - (\sum fm)^2}$

$M_i = \frac{LL + UL}{2}$

	F	Mi	Mi^2	Fm^2
0-4	6	2	4	24
5-9	11	7	49	539
10-14	27	12	144	3888
15-19	4	17	289	1156
	48			5607

Fmi
12
77
324
68
481 ÷ 48 = mean

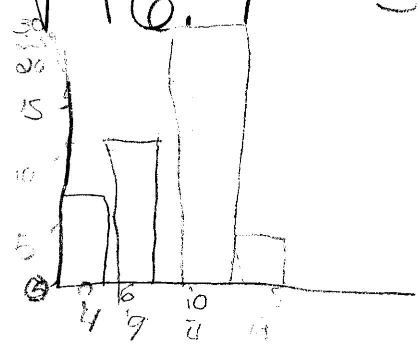
$$\frac{5607 - \frac{(481)^2}{48}}{48-1}$$

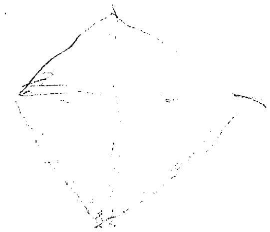
Independent = cause
dependent = effect

$$\frac{5607 - 4820}{47}$$

$\sqrt{16.7} = 4.09$

$S = \sqrt{16.7}$
$S = 4.09$
$S^2 = 16.7$





- descriptive research answer questions
- Exp[illegible] identify causes and es[illegible] of [illegible]
- evalua[illegible] of social [illegible] policy

Valu[illegible]

\- [illegible]

[illegible]

Causal/[illegible]
[illegible]
and [illegible] causes
B.

Generalization [illegible] it is used to [illegible] place [illegible]
[illegible]

\- generalization [illegible] when you [illegible] are [illegible] population from which the sample was drawn

Sampling error [illegible] difference between [illegible] characteristics of a [illegible] and the characteristics [illegible] the population [illegible] which [illegible] was taken

probability [illegible] - Allow [illegible] how likely it is that any element of a population will be [illegible]

nonprobability sample [illegible] methods [illegible] that do not let [illegible] the likelihood [illegible]

Simple random [illegible] generalize [illegible] obtained from [illegible] to a larger [illegible]

(Rate = Number in subset / Total number [illegible] × [illegible])

Proportion = Number in subset of sample / Total [illegible] sample = $\frac{f}{n}$

Percent = $\frac{f}{n}$ × 100 = Proportion [illegible])

	F	Proportion	%
[illegible]	16	.0295	2.9
[illegible]	28	.072	7.2
[illegible]	48	.088	8.8
$[illegible]	[illegible]	.159	15.9
$[illegible]	10[illegible]	.18[illegible]	18.8
[illegible]	[illegible]	.463	46.3

Mode [illegible]
then subtract that by 1

[illegible] NOT
[illegible]

Range [illegible]

[illegible]

- [illegible]
- [illegible]
- [illegible]
- [illegible]
- [illegible]
scores

- Square each of the [illegible] of the [illegible]
- Multiply each [illegible] by the [illegible] interval
- Sum [illegible] all [illegible] found
- Sum [illegible]
- [illegible]
- Divide [illegible]

Probability

- chance of an event occuring.

P(H in one flip of a coin) = .5

Suppose we flip a coin 10 times 10 times does it means we are going to get 5 heads and 5 tails? NO

expected value property of event occuring X number of events

~~[illegible]~~

$.2 \times 20 = 4$

$P(A) = \frac{f}{n}$

~~[illegible]~~

Complement of an event

the compliment of event (A) = 1 - P(A)

Odds of an event = Ratio of an event

$$\text{odds}(A) = \frac{P(A)}{1 - P(A)}$$

Mutually Exclusive - zero probability that they can occur together or at the same time.

Two events

The Bounding Rule
probability are bounded by 0 and 1.0. They can never be less than 0 or greater than 1.0.

Rule 2: The addition Rule
Restricted Addition rule
Only if events are mutually exclusive

P(A and B) is the probability of events A & B occuring together - their joint probability.

non mutually exclusive

P(A or B) = P(A) + P(B) - P(A and B)

Rule 3 multiplication Rule
Restricted Multiplication
rule for independent Rule.

If two events are independent
events then the probabity

Independent if
$P(A) = P(A|B)$

$P(A)$ is the unconditional
probability of A.
New Term: $P(A|B)$ Probability

Outcome of Trial	Retained Councel	Public Defender	total
Acquited	75	95	150
convicted	30	30	60
total	105	105	210

conditional = unconditional

	Retained Counsel	Public Defender	Total
Acquit	55	95	150
Convict	50	10	60
Total	105	105	210

If $P(A) \neq P(A|B)$ is true then the events are not independent and Rule 3a does not apply. Rule 3b applies

Rule 3b: the general Multiplication Rule if two events

Probability distribution is a theoretical distribution. It is not a distribution. It is not a distribution

P(0) Heads = P(T.T.T.T) = .5 · .5 · .5 .5 (.5)^4
.0625

P(1 Head) = P(H.T.T.T) = .5 · .5 · .5 .5 (.5)^4
= .0625

quiz thursday. calculate getting
2,3,4 heads

P(R) · q^R

p^e — probability event occuring

q — probability of q not occuring

$p^e q^{n-r}$

$$\binom{n}{r} = \left(\frac{n!}{r!(n-r)!}\right)$$

$$P(r) = \binom{n}{r} p^r q^{n-r}$$

$$P(r) = \left(\frac{n!}{r!(n-r)!}\right) p^r q^{n-r}$$

H_0 = assumption of innoscense (null hypothesis)

Non directional = Alternative hypothesis
<u>Which means</u>.

A hypothesis test begins with a null and alternative hypothesis, Lets go with a directional hypothesis.

$$P(r) = \left(\frac{n!}{n(n-r)} \right) p^r q^{n-r}$$

$p = .4$
$q = 1 - .4 = .6$
$r = 0, 1, 2 3, 4, 5$
$n = 5$

mean = $n \cdot p$ of binomial distribution

5 steps
1. State the null and the alternative hypothesis
2. Select probability distribution Binomial if 2 categories (Nominal distribution if continous

3. Select Alpha Level of significance
Balance Type 1 (False positive)
and Type II (False negative) Errors
$p = .05$ (.10; .05; .001

4. Calculate the probability of
observing 8 or More of 10 Death

$$P(r) = \left(\frac{n!}{r!(n-r)!}\right) p^r q^{n-r} = P(8)\left(\frac{10!}{8!(10-8)!}\right) .4^8 .6^{10-8}$$

Critical region defines the class
of outcomes

Step 1 - State your hypothesis
.06
D > .06

~~Choose Probability distribution~~ Choose probability distribution

$a = 0.1$ $a = 0.05$ $a = .001$

~~Normal~~ Normal distribution
- Common distribution for continuous
data