Basic Statistics

Theoretical Distributions including the Normal Distribution

Chapter 6

1

Outline

- Let's watch a movie…For All Practical Purposes – Place Your Bets
- Probability
- Distributions
  - rectangular
  - binomial
  - normal
- Using the normal curve to find probabilities

2

Probability: Measuring Uncertainty

- We're all familiar on some level with probability.
- Examples are all over in language: "chances are," "maybe," and "probably" all suggest an intuitive understanding of probability.

3

Probability (cont.)

- Statements: 2 parts
  - an event
  - the probability of that event
- Certain events
  - Definitely will or will not happen
  - p(event) = either 0 or 1

4

Probability (cont.)

- Uncertain events have p 0 - 1
- expresses the degree of uncertainty
  - A. P(heads with a fair coin)=.50
  - B. P(drawing ace of hearts from full deck)=.019
  - C. P(rolling a three with a fair die)=.167
- Each of these is a success, (k), other outcome is a failure (j), P(success)=k/(k+j)
- Ways of expressing probabilities
  - P(A) = .50 or ½ or 50% or 50-50 or 1 to 1
  - P(B) = .019 or 1/52 or 1.9% or 51 to 1 against

5

Calculating probabilities

- Each possible outcome is an elementary event
- Count the number of elementary events in each outcome class
- If each elementary event equally likely, probability of outcome class A can be easily computed

6

Calculating probabilities

- p(A)=# of elem. events that are A/
  Tot # of elementary events
  (same as saying k/(k+j))
- This holds only if the elementary event is a truly random event.
- Over many trials the proportion of "A" outcomes will approach p(A), with enough trials the empirical distribution will look more and more like the theoretical distribution

7

Different orders

- 5 people milling around
- What is the probability that they will line up in alphabetical order?
- N of different ways can they line up in alphabetical order?
- 2, this is the numerator
- What goes in the denominator?
  - Total number of different ways 5 people can line up (5! – what does this mean?)
- P(line up alphabetically) = .017

8

Distributions

- Theoretical distributions always have a total area of 1.0
- statements about area are equivalent to statements about probability

9

Rectangular

- pg. 122 text
- playing cards
- disregarding suit, 13 different cards, p(any particular card)=.077
- probabilities can be additive, p(7, 8, or 9)=.231 (.077*3) or 12/52=.231
- if we actually drew 52 times from a deck would deviate from theoretical rectangular dist.
- Empirical and theoretical distributions

10

Binomial

- Essential element- event(s) with two possible outcomes
- Pg. 123
- flip three coins (for independent events, we can simply multiply the probabilities of the events) .5x.5x.5 = .125
- eight possible outcomes
- hhh, hht, hth, thh, htt, tht, tth, ttt
- each has a P of .125 (1/8)
- for heads, p(0)=.125, p(1)=.375, p(2)=.375, p(3)=.125
- over enough real trials we will end up with an empirical distribution very close to the theoretical not exactly the same, but close

11

Standard Normal Curve

- $\mu = 0.0$ and $\sigma = 1.0$
- X-axis on a standard normal curve is often relabeled and called *z* scores
- page 386 Table of areas under normal curve- ***mark it somehow for quick access***

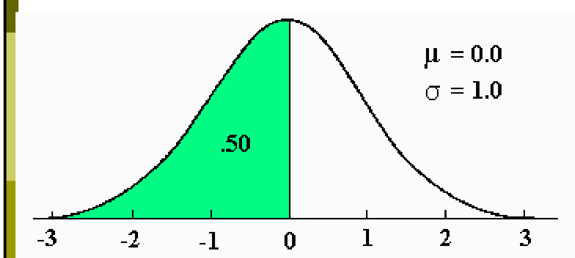
12

Normal Curve (cont.)

- total area below 0.0 is .50
- symmetrical about the mean, thus area above 0.0 is .50
- generalizes to all normal curves: total area below the value of the mean is .50 on any normal curve

13

Normal Curve (cont.)



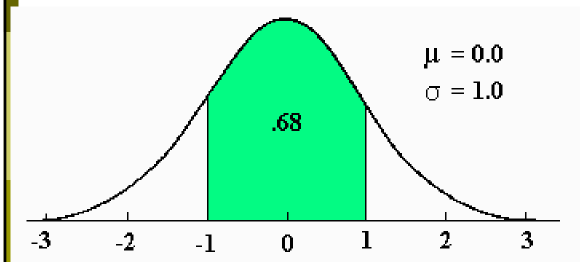
14

Normal Curve (cont.)

- area between Z-scores of -1.00 and +1.00. It is .68 or 68%
- total area between plus and minus one σ unit (1 s.d.) on any normal curve is also .68.

15

Normal Curve (cont.)



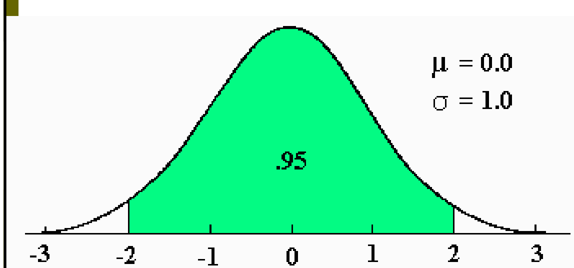
16

Normal Curve (cont.)

- area between Z-scores of -1.96 and +1.96 (round these to 2) and is .95 or 95%
- area (.95) generalizes to plus and minus two sigma units on any normal curve

17

Normal Curve (cont.)



18

Area under Normal Curve

- we can compute additional areas
- area between a Z-score of 0.0 and 1.0
- take 1/2 the area between -1.0 and 1.0
- Why?

19

Area under Normal Curve

- the distribution is symmetrical between those two points
- answer in this case is .34 or 34%
- same for the area between 0.0 and -1.0

20

Areas under Normal Curve

- area below a Z-score of 1.0?
- computed by adding .34 and .50 to get .84
- area above a Z-score of 1.0?
- subtract the area just obtained from the total area under the distribution (1.00)
- 1.00 - .84 or .16 or 16%

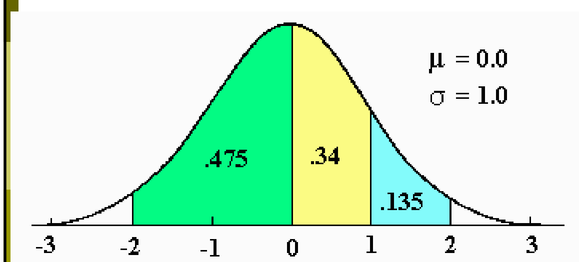
21

Areas under Normal Curve

- area between -2.0 and -1.0?
- first, the area between 0.0 and -2.0 is 1/2 of .95 or .475
- the .475 includes too much area, the area between 0.0 and -1.0 (.34) must be subtracted from this
- so, .475 - .34 or .135

22

Finding Areas (cont.)



23

Using normal curve – example

- Mean = 10, SD = 2
  - what proportion of scores is between 7.5 & 12.5?
  - what proportion of scores is between 7.5 & 10.5?
  - What score separates the lower 40% from the upper 60%?
  - If there were 250 members of population, how many would be expected to score 11 or more?
  - What proportion would be expected to score 9 or more?
  - What score separates the top 10% of scores from the rest?
  - Does this problem deal with a theoretical or empirical distribution?
- Always sketch a picture!

24