CEE518
Introduction

Dorothy A. Reed
Civil and Environmental Engineering

In this lecture

- Course Schedule and Logistics
  - Website is courses.washington.edu/cee518
  - Lectures and assignments posted
    - Assignments (usually weekly) 25%
    - Quizzes 50%
    - Project 25%
  - Office hours to be determined today
  - My e-mail is reed@u.washington.edu
  - Feedback always appreciated
  - Over the course of the quarter, we will follow the text [Haldar and Mahadevan] topics fairly closely.
- Topics
  - Basics of probability and statistics
  - Applications to structural engineering problems
- Introduction to Probability Theory
  - Bayes Theorem and Applications
  - Subjective Probability Assessment Considerations
  - Frequentist Approach
- Applications

Terms and Definitions

- **What is reliability?**
  - New World Dictionary:
    - *Reliable* is an adjective meaning "that can be relied on; dependable; trustworthy".
    - *Rely* :[M.E. *relien*, to rally O.Fr. *relier*, L *religare*: relates to the modern word *religion*]to have confidence in a person or thing ; trust
  - Kapur (1977): "The reliability of a system is the *probability* that when operating under stated environmental conditions, the system will perform its intended function adequately for a specified interval of *time*."
  - Dai and Wang (1992): "Reliability is the *probability* that a component, equipment or system will perform a required function under the operating conditions encountered for a stated period of *time*."
  - Kale (1998): "*Reliability* is the branch of quality assurance that deals specifically with functionability upon demand…"
  - Halder and Mahadevan (2000): "…the probabilistic assurance of performance is referred to as *reliability*." In this context, *performance* is defined as a structural system (or component) whereby the capacity of the system (or component) exceeds the demand placed on it. Capacity can be represented by a strength value and demand by a load-induced stress value.

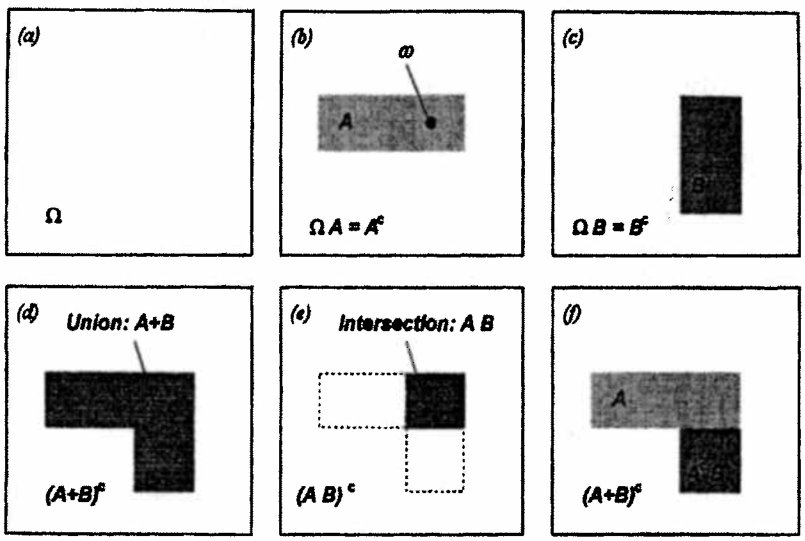
- Common Elements
  - All definitions involve probability.
    - Capacities and demands are random variables.
    - Uncertainty plays an important role.
  - Time intervals vary from discipline to discipline; IE and ME: explicit time; CEE more implicit.
  - In this class, structural reliability means design for a low or prescribed probability of failure: "Safety with economy"
- How does reliability differ from risk?
  - Haldar and Mahadevan (2000): "…*risk* is not just the probability of failure but includes the consequence of failure." Example: risk is the probability of failure [reliability] multiplied by the cost of failure.
  - Hyman (1998): "The potential for something of value …to be adversely impacted by an event." Risk analysis consists of assessment, perception and communication, and management.
  - Reiter (1990) in his discussion of Earthquake Hazard Analysis:
    - "*Seismic hazard* describes the potential for dangerous, earthquake-related natural phenomena such a ground shaking, fault rupture or soil liquefaction. These phenomena could result in adverse consequences to society such as the destruction of buildings or the loss of life.
    - *Seismic risk* is the probability of occurrence of these consequences."

Probability Theory

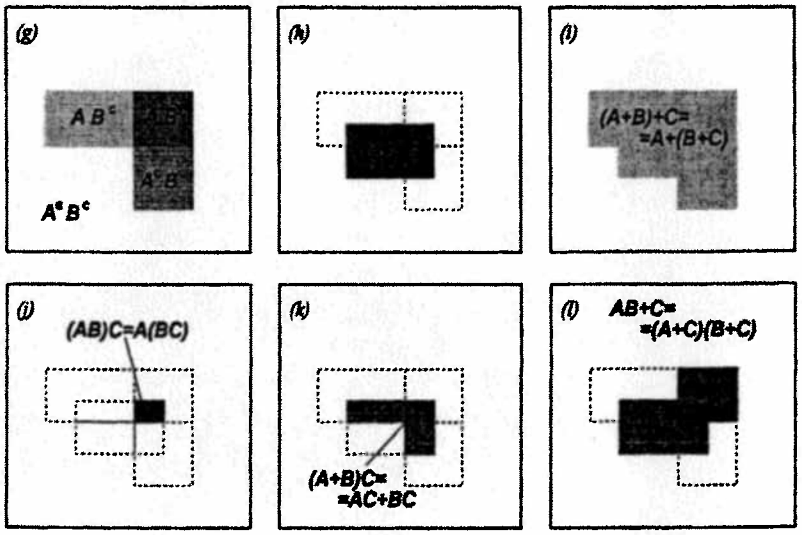
- Classical
  - aka frequentist or objective
  - Probabilities are obtained through long-term repeated trials or observations
- Bayesian
  - aka subjective
  - Prior knowledge is incorporated via Bayes Theorem
  - Statements usually in terms of degrees of belief
- Axioms of Probability Theory
  - Probabilities must be between 0 and 1.
  - Probabilities must add up.
  - Total probability must equal 1.
- Set theory
  - Venn diagrams useful in presenting probability concepts.
  - Following figures from Kottegoda and Rosso (1997).

Ω is the sample space.

9/26/2011 week1.ppt 5



9/26/2011 week1.ppt 6



Notation

- P(A) is the probability of event A
- The diagram can be used to identify the complement of A.

$$P(\overline{A}) = 1 - P(A)$$

sometimes A^C [A complement]

is used to denote $\overline{A}$ ["not A"].

Civils typically refer to probability of failure...

$$P(F) + P(\overline{F}) = 1$$

$P(F) \equiv$ probability of structural failure, aka p_f

$p_f \approx 10^{-5}$ to 10^{-3}

$P(\overline{F}) \equiv$ probability of not failing; reliability; R

R typically denoted in IE, ME applications as 0.9^4, etc.

- Joint probabilities can also be assessed: the probability of A and B.
- Conditional probability: P(A|B)= probability of A given B

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A,B)}{P(B)}$$

Two events A and B are said to be independent if either

$$P(A \mid B) = P(A) \quad \text{if} \quad P(B) > 0$$

or

$$P(B \mid A) = P(B) \quad \text{if } P(A) > 0$$

Further,

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

9/26/2011 week1.ppt 9

Bayes Theorem

$$P(B \mid A)P(A) = P(A \mid B)P(B)$$

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$$

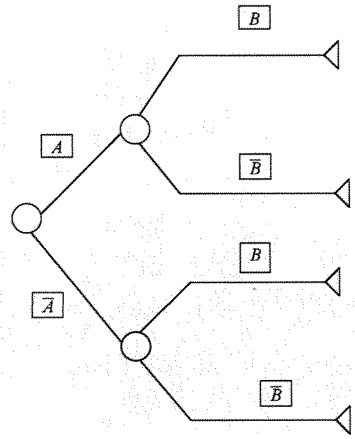
$$= \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})}$$

Application

The easiest way to remember these relationships is through the use of probability trees.
These are commonly part of decision trees.
Basically, a decision tree has probability trees leading to a square terminal node that represents a decision.

9/26/2011 week1.ppt 10

Probability tree

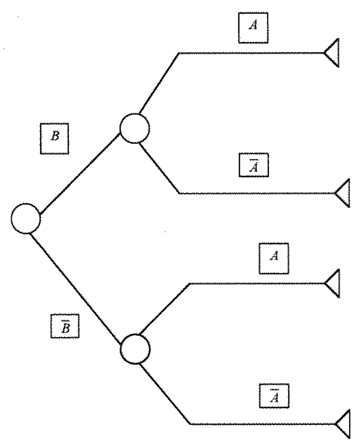


$$P(A,B)=P(A)P(B|A)$$
$$P(A,\bar{B})=P(A)P(\bar{B}|A)$$
$$P(\bar{A},B)=P(\bar{A})P(B|\bar{A})$$
$$P(\bar{A},\bar{B})=P(\bar{A})P(\bar{B}|\bar{A})$$

$$P(B)=P(B,A)+P(B,\bar{A})$$
$$P(\bar{B})=1-P(B)=P(\bar{B},A)+P(\bar{B},\bar{A})$$

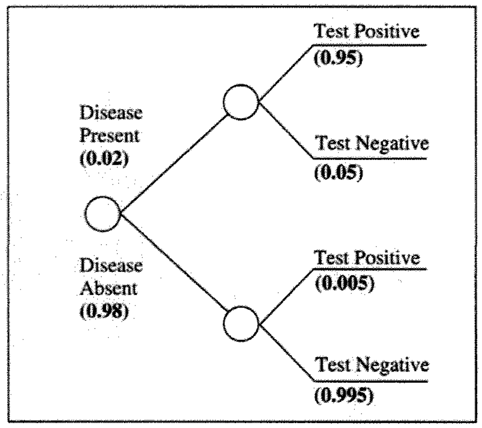
9/26/2011 week1.ppt 11

Flipped tree



9/26/2011 week1.ppt 12

Example from Clemen and Reilly (2001)



Let's flip this tree

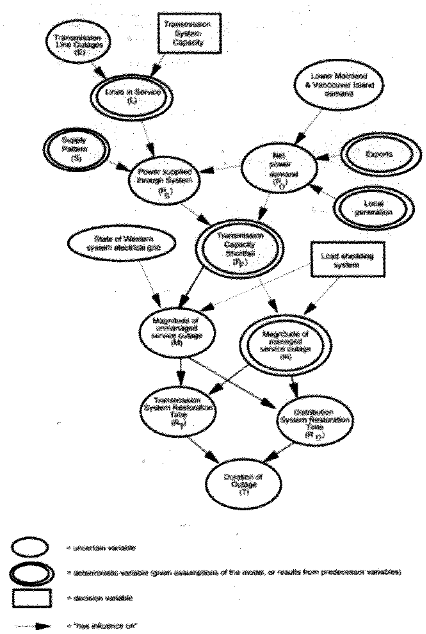
Subjective Assessment

- Difference between causality and correlation: Baurind p. 19 of Glymour et al.
  - "The number of never-married persons in certain British villages is highly inversely correlated with the number of field mice in the surrounding meadows. Marital status of humans was considered an established cause of field mice by the village elders until the mechanisms of transmission were finally surmised: Never-married persons bring with them a disproportionate number of cats relative to the rest of the village populace and cats consume field mice. With the … mechanisms understood, village elders could concentrate their attention on increasing the population of cats rather than the proportion of never-married persons.

Influence Diagrams

- Directed graphs with nodes connected via arcs or arrows in a causal manner
- Four main node types:
  - Calculation – probabilities typically
  - Relevance
  - Decision
  - Consequence
- Assessment of conditionals
  - Easier cognitively to assess in the causal direction; that is
  - Typical Usage:
    - Observable →Unobservable
  - Assessment:
    - Effect ← Cause
  - Example: Shachter
    - Let A= Sore throat; B= Flu
    - P(B|A) common usage
    - P(A|B) easier assessment due to causal direction

9/26/2011 week1.ppt 15



Keeney et al. (1995)

9/26/2011 week1.ppt 16

Fragility Curve Assessment

- ATC-13 and ATC-25 Applied Technology Council Reports
- Fragility is the probability of damage or failure given a level of demand or load
- P(damage|earthquake demand) typically means P(damage|PGA) where PGA is peak ground acceleration
- Is this easier or harder to assess than P(PGA|damage)?
- Parametric models exist for fitting fragility curves
- Assessments for reports made by experts due to lack of data
  Next set of slides show the results of a project to determine fragilities for power line outages which are considered "failures"

9/26/2011 week1.ppt 17

Nisqually Earthquake Analysis

- February 28, 2001
- Magnitude 6.8
- Epicenter 58 km southwest of Seattle, Washington, USA
- Focal depth of 60 km
- 11.4% outages with flickers; 8.7% without

9/26/2011 week1.ppt 18

Modified Mercalli Intensity MMI

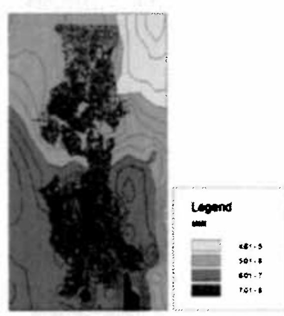
- Definition:
  - I. Not felt except by a very few under especially favorable circumstances…
  - VI. Felt by all, many frightened… Damage slight…
  - VIII. Damage slight to partial collapse, depending on design…
  - IX. Damage considerable…buildings shifted off foundations…
- *Relationships between Peak Ground Acceleration, Peak Ground Velocity and Modified Mercalli Intensity in California*
  - **David J. Wald, Vincent Quitoriano, Thomas H. Heaton, Hiroo Kanamori**
  - ***Earthquake Spectra, Vol. 15, No. 3, 557-564, 1999***

$$I_{mm} = ModifiedMercalliIntensity$$
$$I_{mm} = 2.20\log(PGA) + 1.00$$
$$and$$
$$I_{mm} = 2.10\log(PGV) + 3.40$$

9/26/2011 week1.ppt 19

Nisqually Earthquake Analysis

- Modified Mercalli Damage contours [Dept. of Earth & Space Sciences, UW,USA] with entire distribution system overlay
- Primary damage at levels 5 to 8



9/26/2011 week1.ppt 20

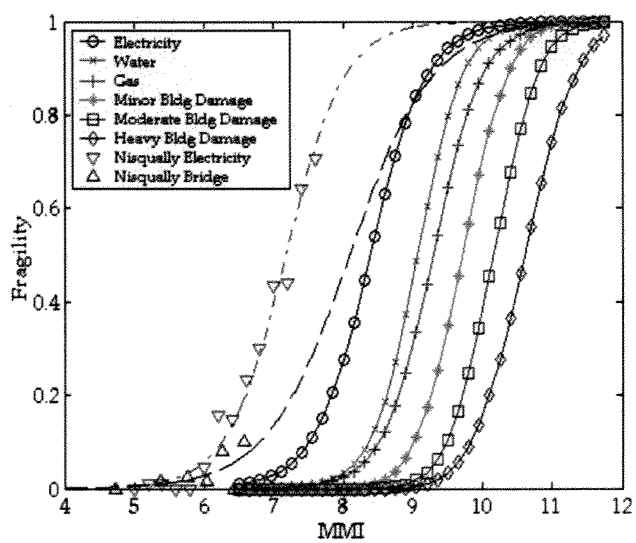
Ranges of Ground Motion

MMI	I	II-III	IV	V	VI	VII	VIII	IX	X+
Peak Acc. (%g)	<0.17	0.17-1.4	1.4-3.9	3.9-9.2	9.2-18	18-34	34-65	65-124	>124
Peak Vel. (cm/s)	<0.1	0.1-1.11	1.1-3.4	3.4-8.1	8.1-16	16-31	31-60	60-116	>116

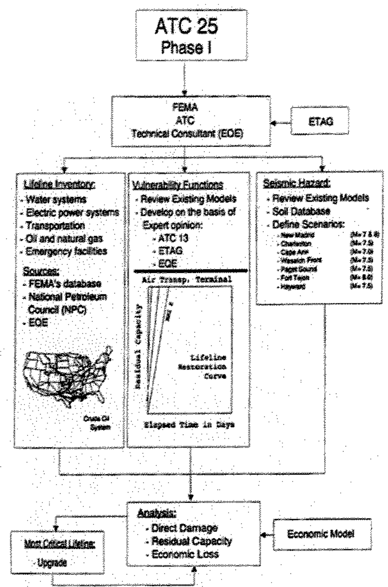
Instrumental MMI=I_{mm}	Fragility=Probability of damage per unit length [%]
5-6	1.0
6-7	16.7
7-8	41.2

- ATC25: For a Puget Sound M7, weeks estimated to restore system to full capacity.
- Why are the values so different?

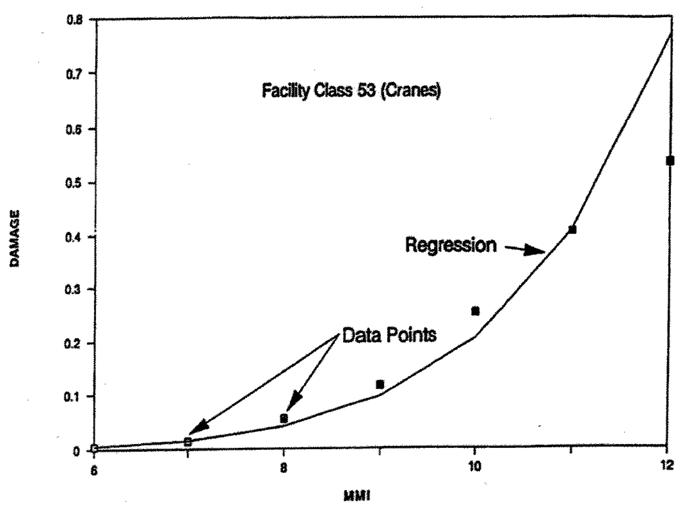
9/26/2011 week1.ppt 21



9/26/2011 week1.ppt 22

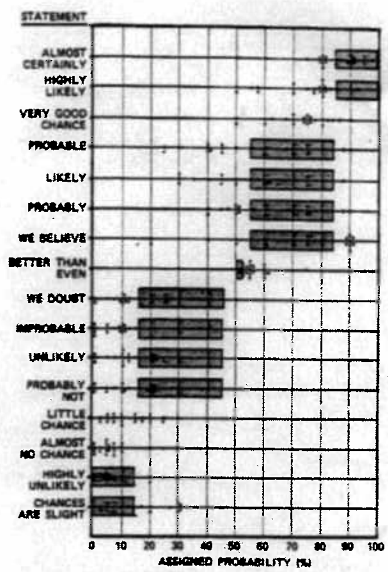


Notation: ATC-13: ATC-13 Report, *Earthquake Damage Evaluation Data for California* (ATC, 1985)
ETAG: Expert Technical Advisory Group (project advisory panel)
EQE: EQE Engineering, Inc. (project subcontractor)



Comparison of ATC-13 Appendix C data (Statistics of Expert Responses for Motion-Damage Relationships) versus regression curve.

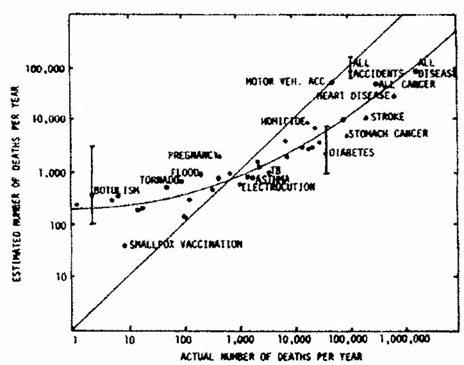
Study of probability and descriptions



VonWinterfeldt & Edwards

9/26/2011 week1.ppt 25

Judgments under Uncertainty:Heuristics and Biases
Kahneman, Slovic and Tversky



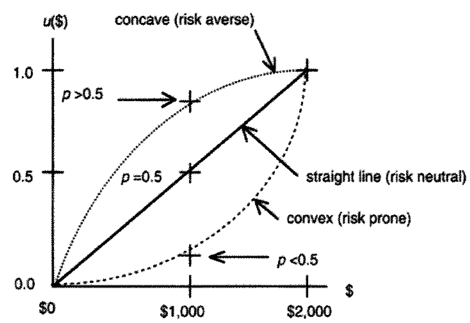
9/26/2011 week1.ppt 26

- Representativeness
  - Information relating to base rates is ignored when probability judgments are made upon the basis of similarity. Judgments are made on the basis of stereotypes: "All engineers are nerds."
- Availability
  - We judge the probability that an event will occur according to the ease with which we can retrieve similar events from memory. For example, we will overestimate the probability of being in a car accident if we are involved in one, or if we have seen reports in the news about accidents recently.
- Anchoring and adjusting
  - We often choose an initial anchor and make forecasts based on that anchor. This frequently occurs in sales forecasting. In engineering, experts frequently overestimate damage predictions because they are adjusting anchored values based on their experience.
- Motivational bias
  - Incentives often lead people to over- or under-estimate probabilities to make themselves look good. For example, a sales rep may forecast low sales to make herself look good when she does exceed her forecasted sales for a time period. Weather reports may be gloomier than anticipated so that people don't get disappointed.

9/26/2011 week1.ppt 27

What about perceived risk as viewed by the decision maker?

- Attitudes towards risk are incorporated through utility functions in decision making.



Hyman 1998

Figure 9-26. General Form of Risk Averse, Risk Neutral, and Risk Prone Utility Functions

9/26/2011 week1.ppt 28

What about the frequentist theory?

- Data analysis typically exists of identifying descriptive statistics and comparing with previous results.
- Excel, Matlab, Mathematica, etc.,are useful in this analysis.
- Definitions
  - Capital letters denote uncertain quantities:
  - P(X=3) is the probability that the uncertain quantity X equals 3.
  - P(Y>0) is the probability that the uncertain quantity Y is greater than 0.
  - So P(X=x) refers to the uncertain quantity X where x is the actual outcome.
  - Probability Mass Function

$$\text{probability mass function } p_X(x) = P[X = x]$$
$$0 \le p_X(x) \le 1 \text{ for all possible } x$$
$$\sum p_X(x) = 1$$

9/26/2011 week1.ppt 29

Measures of randomness

Mean or expected value

E(X) or μ_X

Measures central tendency of the data, so also known as the first central moment

By definition :

$$E(X) = \mu_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

9/26/2011 week1.ppt 30

Variance, Var(X)
Measures central dispersion about the mean
"Second central moment"

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2$$

Standard devation, σ_X

$$\sigma_X = \sqrt{Var(X)}$$

Coefficient of variation, COV(X) or δ_X

$$COV(X) = \delta_X = \frac{\sigma_X}{\mu_X}$$

In wind engineering, X = wind velocity V, and COV(V) is called the "intensity of turbulence".

9/26/2011 week1.ppt 31

Skewness is the third central moment.

$$Skewness = \frac{1}{n}\sum_{i-1}^{n}(x_i - \mu_X)^3$$

Nondimensional format θ_X

$$\theta_X = \frac{skewness}{\sigma_X^3}$$

0 : symmetric randomness
(+) : more dispersion above mean
(-) : more dispersion below mean

9/26/2011 week1.ppt 32

Histogram

- Arrange data in increasing order.
- Subdivide data into equal intervals ["bins"] and count the number of points per interval.
- Plot the number of points or observations in each interval.
- Bin size may be difficult to estimate
  - Use estimate number of bins k
    - $k = 1+3.3\log_{10}n$
  - Many programs do this for you

9/26/2011 week1.ppt 33

Example: Problem 3.1, H&M, pp. 58-59

3.1 In an examination for a class of 30 students, the following scores were obtained: 99, 45, 60, 80, 95, 100, 95, 91, 85, 87, 77, 75, 61, 71, 85, 88, 83, 85, 79, 81, 82, 55, 63, 75, 82, 88, 77, 78, 41, and 70.

(a) Draw the histogram for the data.

(b) Draw the frequency diagram for the data.

(c) Calculate the mean, variance, standard deviation, coefficient of variation, skewness, and skewness coefficient for the test scores.

(d) Assume that a student must score at least 85 to receive an A grade. What is the probability that any student in the class will receive an A, using the actual scores only? What will be the corresponding probability if the frequency diagram is used instead?

9/26/2011 week1.ppt 34

Sorting by hand

Bin size $k = 1+3.3\log_{10}n = 1+3.3\log_{10}(30) \sim 6$

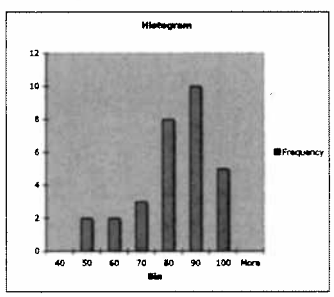
Score	Number of Observations	Fraction of Observations
40-50	2	2/30 = 0.0667
50-60	2	0.0667
60-70	3	0.10
70-80	8	0.2667
80-90	10	0.3333
90-100	5	0.1666
Sum =	*30*	*1.0*

9/26/2011 week1.ppt 35

Setting bin size

Scores: Data_>Sort	Input Bin Range
41	40
45	50
55	60
60	70
61	80
63	90
70	100
71	
75	
75	
77	
77	
78	
79	
80	
81	
82	
82	
83	
85	
85	
85	
87	
88	
88	
91	
95	
95	
99	
100	

Bin	*Frequency*
40	0
50	2
60	2
70	3
80	8
90	10
100	5
More	0

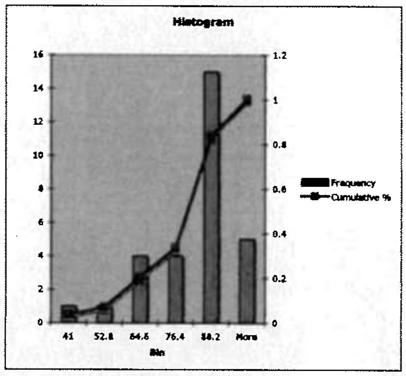


9/26/2011 week1.ppt 36

Default bin size

Scores
99
45
60
80
95
100
95
91
85
87
77
75
61
71
85
88
83
85
79
81
82
55
63
75
82
88
77
78
41
70

Tools-> Data Analysis -> Histogram



Bin	Frequency	Cumulative %
41	1	3.33%
52.8	1	6.67%
64.6	4	20.00%
76.4	4	33.33%
88.2	15	83.33%
More	5	100.00%

9/26/2011 week1.ppt 37

let S = RV representing "Score"

$$E(S) = \mu_S = \frac{1}{30}\sum_{i=1}^{30} s_i = \frac{23.3}{30} = 77.8$$

$$Var(S) = \frac{1}{29}\sum_{1}^{30}(s_i - 77.8)^2 = 211.3$$

$$\sigma_S = \sqrt{211.3} = 14.5$$

$$COV(S) = \frac{14.5}{77.8} = 0.19$$

$$skewness = \frac{1}{30}\sum_{1}^{30}(s_i - 77.8)^3 = -2450.6$$

$$\theta_S = \frac{-2450.6}{14.5^3} = -0.80$$

9/26/2011 week1.ppt 38

part (d)

From the actual data,

$$P(\text{a student will receive an A}) = \frac{11}{30} = 0.3667$$

Using the frequency diagram, we can show that

$$P(\text{student will receive an A}) = \frac{5*10+10*5}{10*30} = 0.333$$

Tools->Data Analysis->Descriptive Statistics

	Scores, S
Mean	77.7666667
Standard Error	2.65385572
Median	80.5
Mode	85
Standard Deviation	14.5357664
Sample Variance	211.288506
Kurtosis	0.62920417
Skewness	-0.88441019
Range	59
Minimum	41
Maximum	100
Sum	2333
Count	30

Skewness [Excel]

SKEW

Returns the skewness of a distribution. Skewness characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values.

Syntax

SKEW(number1,number2,...)

Number1,number2... are 1 to 30 arguments for which you want to calculate skewness. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If there are fewer than three data points, or the sample standard deviation is zero, SKEW returns the #DIV/0! error value.
- The equation for skewness is defined as:

$$\frac{n}{(n-1)(n-2)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^3$$

9/26/2011 week1.ppt 41

Kurtosis

KURT

Returns the kurtosis of a data set. Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution.

Syntax

KURT(number1,number2, ...)

Number1,number2,... are 1 to 30 arguments for which you want to calculate kurtosis. You can also use a single array or a reference to an array instead of arguments separated by commas.

Remarks

- The arguments must be either numbers or names, arrays, or references that contain numbers.
- If an array or reference argument contains text, logical values, or empty cells, those values are ignored; however, cells with the value zero are included.
- If there are fewer than four data points, or if the standard deviation of the sample equals zero, KURT returns the #DIV/0! error value.
- Kurtosis is defined as:

$$\left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i-\bar{x}}{s}\right)^4\right\}-\frac{3(n-1)^2}{(n-2)(n-3)}$$

where:
s is the sample standard deviation.

9/26/2011 week1.ppt 42

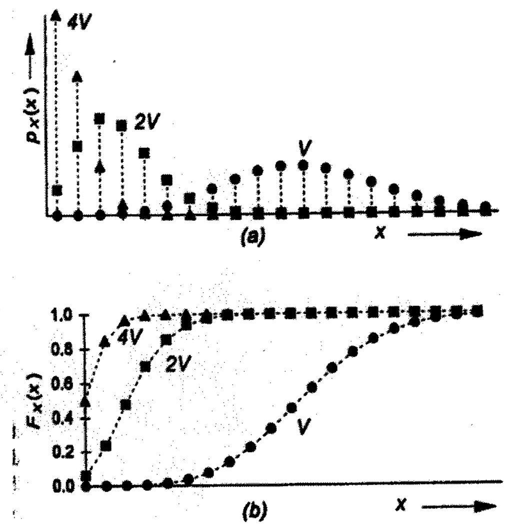


FIGURE 3.2.1
Schematic diagrams of the (*a*) probability mass functions and (*b*) the cumulative distribution functions of three discrete variables *X* with different coefficients of variation *V*.

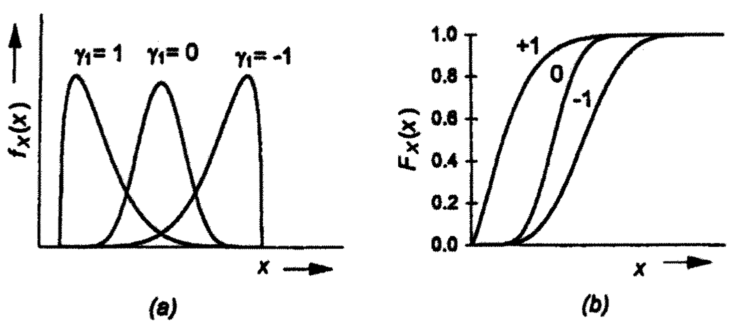


FIGURE 3.2.3
Schematic diagrams of (*a*) the probability density function and (*b*) the cumulative distribution functions of three continuous variables *X* with coeffcients of skewness $\gamma_1 = 1, 0,$ and -1.

Discrete RV

$$E(X) = \mu_X = \sum_{all\ x_i} x_i p_X(x_i)$$

$$Var(X) = \sum_{all\ x_i} (x_i - \mu_X)^2 p_X(x_i)$$

$$skewness = \sum_{all\ x_i} (x_i - \mu_X)^3 p_X(x_i)$$

9/26/2011 week1.ppt 45

Continuous RV

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx$$

$f_X(x) \equiv pdf \equiv$ probability density function

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

Expected value of a function $g(X)$

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

$$Skewness = \int_{-\infty}^{\infty} (x - \mu_X)^3 f_X(x) dx$$

9/26/2011 week1.ppt 46

Continuous RV

$$P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x)dx$$

$$P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Discrete RV

$$F_X(x) = P(X \le x) = \sum_{x_i \le x} p_X(x_i)$$

9/26/2011 week1.ppt 47

Multiple RV: Joint Distributions for RV X and Y:

$$F_{XY}(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)dvdu$$

$$F_{XY}(x,y) = \sum_{x_i \le x}\sum_{y_j \le y} p_{XY}(x_i,y_j)$$

Conditionals

$$f_{X|Y}(x\,|\,y) = \frac{f_{XY}(x,y)}{f_Y(y)}; \quad f_{Y|X}(y\,|\,x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

$$p_{X|Y}(x\,|\,y) = \frac{p_{XY}(x,y)}{p_Y(y)}; \quad p_{Y|X}(y\,|\,x) = \frac{p_{XY}(x,y)}{p_X(x)}$$

Marginals

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx$$

$$p_X(x) = \sum_{all\,y_j} p_X(x,y_j); \quad p_Y(y) = \sum_{all\,x_i} p_X(x,y_j)$$

9/26/2011 week1.ppt 48

Covariance and Correlation

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$
$$= E[XY - \mu_X Y - X\mu_Y + \mu_X \mu_Y]$$
$$= E(XY) - \mu_X \mu_Y$$
$$= E(XY) - E(X)E(Y)$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)\,dx\,dy$$

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

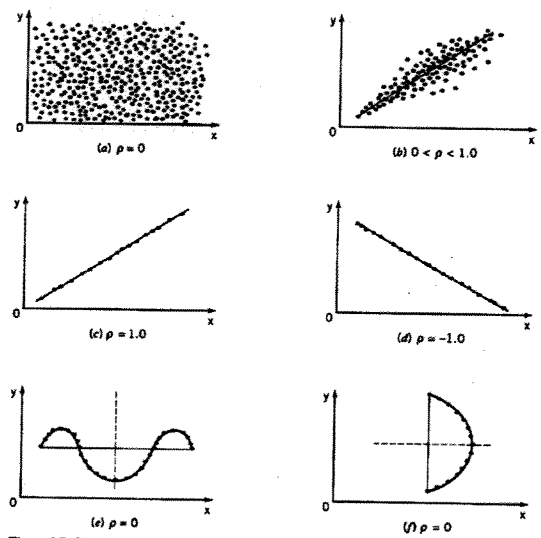


Figure 3.7 Correlation of Two Random Variables

Example problem from H&M: Multiple random variables

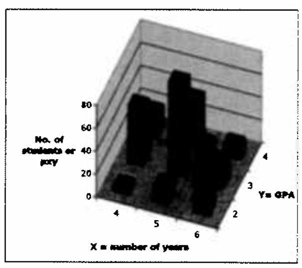
3.9 The study duration and grade point average (GPA) of students graduating with a B.S. degree from an engineering school were studied. With X defined as the number of years it takes to graduate and Y as the GPA, it was observed that X could be 4, 5, or 6 years and Y could be 2, 3, or 4. The following table shows the number of students for each combination of X and Y.

Y \ X	4	5	6
2	5	15	60
3	50	80	20
4	20	40	10

(a) Plot the joint PMF of X and Y.
(b) Plot the marginal PMF of X and Y.
(c) If only a GPA of 3 is under consideration (i.e., $Y = 3$) plot the conditional PMF of X.
(d) Determine Cov(X, Y) and the corresponding correlation coefficient between X and Y.

X	Y	No. of students
4	2	5
5	2	15
6	2	60
4	3	50
5	3	80
6	3	20
4	4	20
5	4	40
6	4	10

Y=GPA/ X=No. of years	2	3	4	sum across
4	5	50	20	75
5	15	80	40	135
6	60	20	10	90
			sum down	300

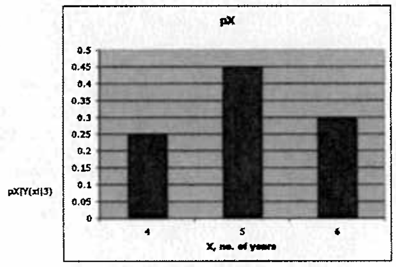


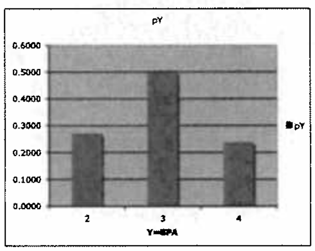
	2	3	4	sum across
4	5	50	20	75
5	15	80	40	135
6	60	20	10	90
			sum down	300

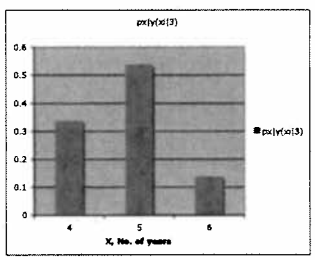
	2	3	4	sums	E(X)=x*sum
4	0.0167	0.1667	0.0667	0.25	5.05
5	0.05	0.2667	0.1333	0.45	Var(x)=
6	0.2	0.0667	0.0333	[illegible]	0.55
sums	0.2667	0.5	0.2333	1	sDEV=
E(Y)	Var(Y)=	sdev=			0.740
2.97	0.499	0.706			

E(XY)->

	2	3	4		
				14.77	
4	0.0167	0.1667	0.0667		
5	0.05	0.2667	0.1333	top=	rho XY=
6	0.2	0.0667	0.0333	-0.215	-0.411







Recall $Cov(X,Y) = E(XY) - E(X)E(Y)$

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Find using table:

$E(X) = 4(p_X(4)) + 5(p_X(5)) + 6(p_X(6)) = 4*0.25 + ...$

$Var(X) = (4 - E(X))^2(0.25) + ...$

$\sigma_X = \sqrt{Var(X)}...$

Same for Y

Then

$E(XY) = \sum(\text{product of x, y, relative frequency})$

9/26/2011 week1.ppt 55

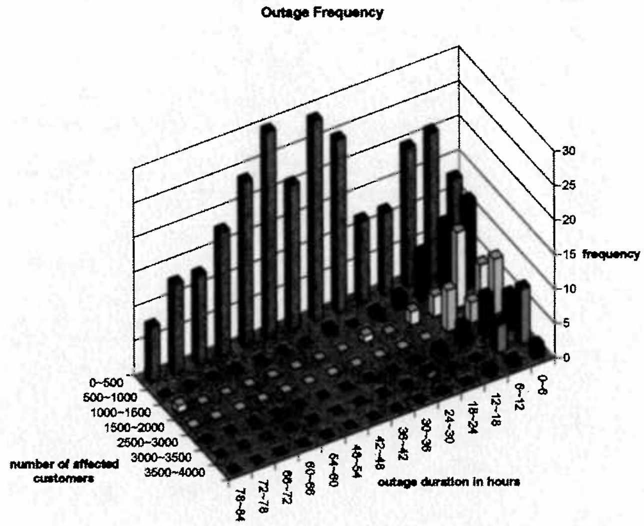
	2	3	4	sums	E(X)=x*sum
4	0.0167	0.1667	0.0667	*0.25*	5.05
5	0.05	0.2667	0.1333	*0.45*	Var(x)=
6	0.2	0.0667	0.0333	*0.3*	0.55
sums	*0.2667*	*0.5*	*0.2333*	*1*	sDEV=
E(Y)	Var(Y)=	sdev=			0.740
2.97	0.499	0.706			

E(XY)->

	2	3	4		
				14.77	
4	0.0167	0.1667	0.0667		
5	0.05	0.2667	0.1333	top=	rho XY=
6	0.2	0.0667	0.0333	-0.215	-0.411

9/26/2011 week1.ppt 56

Research Example



9/26/2011 week1.ppt 57

3.10 A person's commuting time from home to the workplace (X) and from the workplace to home (Y) is studied for 100 days. Assume that the commuting time each way can be approximated as 30, 40, or 50 minutes. The following table shows the number of days for each combination of X and Y.

Y \ X	30	40	50
30	10	20	25
40	5	30	4
50	3	2	1

(a) Plot the joint PMF of X and Y.
(b) Plot the marginal PMF of X and Y.
(c) Considering 30 minutes commuting time from home to workplace (i.e., $X = 30$), calculate the conditional PMF of Y.
(d) What is the probability that the commuting time in each direction on a particular day will be at least 40 minutes?
(e) Determine Cov(X, Y) and the corresponding correlation coefficient.

9/26/2011 week1.ppt 58

References

- Behn and Vaupel, (1982), *Quick Analysis for Busy Decision Makers.*
- Clemen and Reilly, (2001), *Making Hard Decisions.*
- Dai and Wang, (1992), *Reliability Analysis in Engineering Applications.*
- Haldar and Mahadevan, (2000), *Probability, Reliability and Statistical Methods in Engineering Design.*
- Hyman, (1998), *Engineering Design.*
- Kahneman, Slovic and Tversky, (1982), *Judgments Under Uncertainty.*
- Kapur and Lamberson,(1977), *Reliability in Engineering Design.*
- Kottegoda and Rosso, (1997), *Probability, Statistics and Reliability for Civil and Environmental Engineers.*
- Reiter, (1990), *Earthquake Hazard Analysis.*
- Von Winterfeldt and Edwards, (1986), *Decision Analysis and Behavioral Research.*
- Washington, Karlaftis and Mannering, (2004), *Statistical and Econometric Methods for Transportation Data Analysis.*