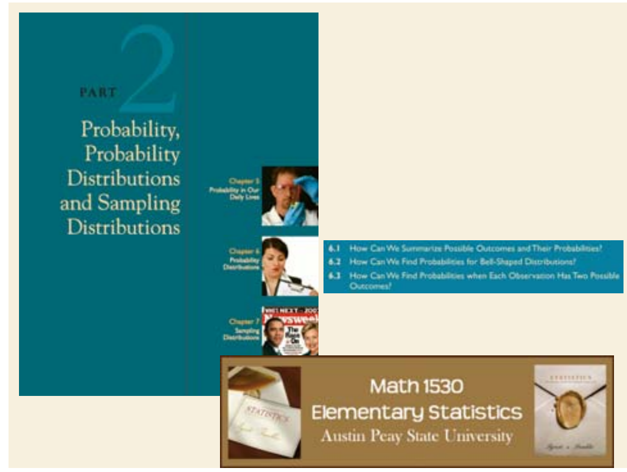
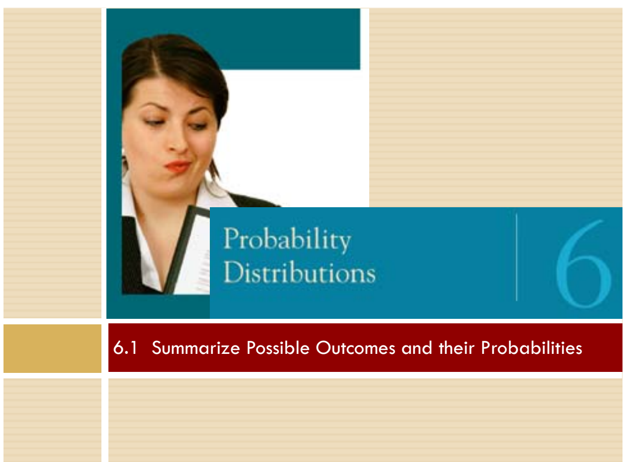
PROBABILITY DISTRIBUTIONS

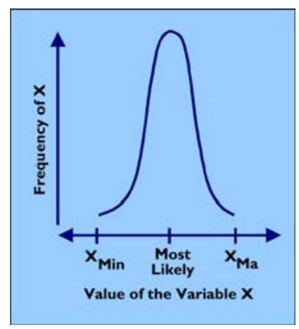
Chapter 6





Learning Objectives

1. Random variable
2. Probability distributions for discrete random variables
3. Mean of a probability distribution
4. Summarizing spread of probability distribution
5. Probability distribution for continuous random variables



cdn.information-management.com

Randomness

www.physics.umd.edu

Numerical values variable assumes result from random phenomenon:

- Selecting random sample for population or
- Performing randomized experiment

Random Variable

- Random variable is numerical outcome of random phenomenon

hoppingintopuddles.files.wordpress.com

Random Variable


www.economicpopulist.org

- Letters near end of alphabet, *x*, symbolize
  1. Variables
  2. Particular value of random variable
- Capital letter, *X*, refer to random variable itself

Ex: Flip a coin three times
- X = # of heads in 3 flips; defines random variable
- x = 2; represents value of random variable

Probability Distribution

The probability distribution of random variable specifies its possible values and their probabilities

Note: Randomness of variable allows us to give probabilities for outcomes


www.anticsland.com

Probability Distribution of Discrete Random Variable


psy2.ucsd.edu

- A discrete random variable X has separate value (0,1,2,…) outcomes
- For each *x*, probability distribution assigns P(*x*):
  - Between 0 and 1
  - Sum of all is 1

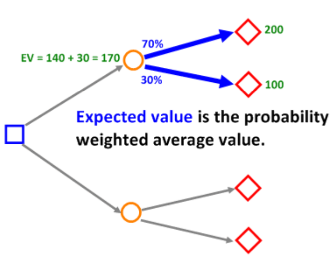
The Mean of a Discrete Probability Distribution

- The mean of a probability distribution, μ, for a discrete random variable is

$$\boxed{\mu = \sum x \cdot P(x)}$$

- Mean is a *weighted* average; more likely values receive greater weight, P(x)

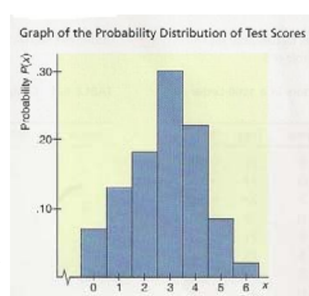
Expected Value of X



- Mean of probability distribution is also expected value of X
- Reflects not *single* observation but what we expect for *average* of lots of observations
- Often NOT a possible outcome

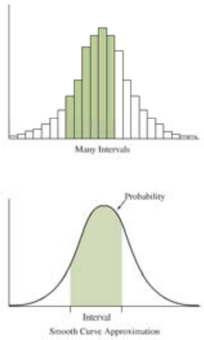
The Standard Deviation of a Probability Distribution

The standard deviation of a probability distribution, σ, measures spread
- Larger σ corresponds to greater spread
- σ describes how far random variable falls, on average, from mean



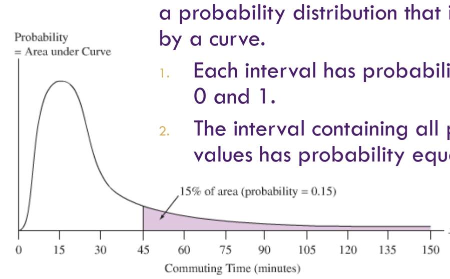
rchsbowman.files.wordpress.com

Continuous Random Variable



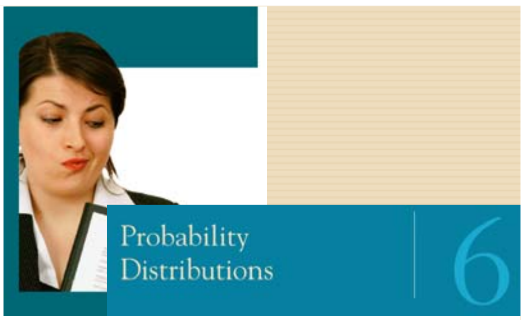
- Has infinite continuum of possible values in an interval
  - Measures: time, age, size, height, weight, …
- Continuous variables are rounded to discrete values

Probability Distribution of a Continuous Random Variable



A continuous random variable has possible values that form an interval and a probability distribution that is specified by a curve.

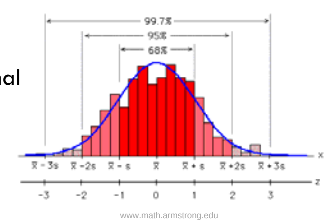
1. Each interval has probability between 0 and 1.
2. The interval containing all possible values has probability equal to 1.



6.2 Finding Probabilities for Bell-Shaped Distributions

Learning Objectives

1. Normal Distribution
2. 68-95-99.7 Rule
3. Z-Scores and Standard Normal Distribution
4. The Standard Normal Table: Finding Probabilities
5. Using the TI-calculator: find probabilities

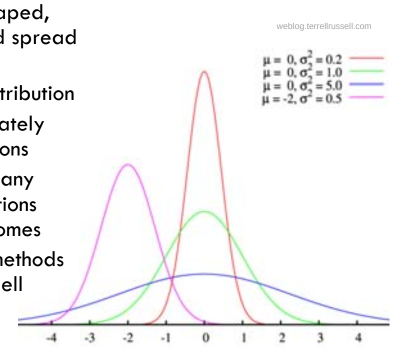


Learning Objectives

6. Standard Normal Table in Reverse
7. TI-calculator: find z-scores
8. Probabilities and Percentiles for Normally Distributed Random Variables
9. Using Z-scores to Compare Distributions

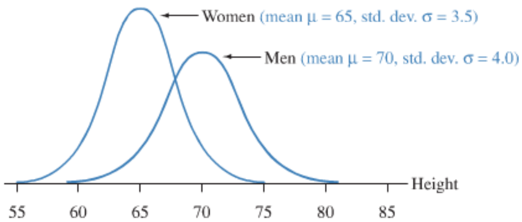
Normal Distribution

- Symmetric, bell-shaped, centered on μ, and spread determined by σ
- Most important distribution
  - Many approximately normal distributions
  - Approximates many discrete distributions with lots of outcomes
  - Used by many methods even when not bell shaped



Normal Distribution

Within what interval do almost all of the men's heights fall? Women's height?

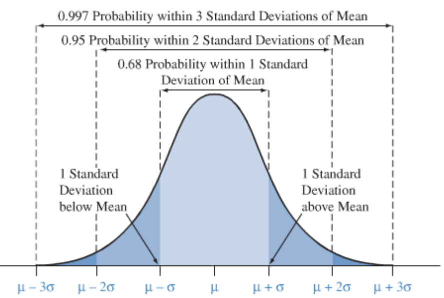


68-95-99.7 Rule for Normal Curve

68% fall within one standard deviation of the mean

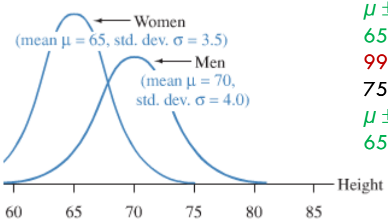
95% fall within two

99.7% fall within three



Example: 68-95-99.7 Rule

What proportion of women are less than 69 inches tall?



68% between 61.5 and 68.5"

$\mu \pm \sigma = 65 \pm 3.5$

95% between 58 and 72"

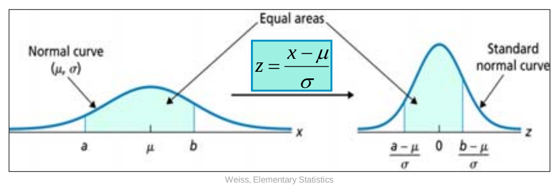
$\mu \pm 2\sigma = 65 \pm 2(3.5) = 65 \pm 7$

99.7% between 54.5 and 75.5"

$\mu \pm 3\sigma = 65 \pm 3(3.5) = 65 \pm 10.5$

z-Scores and Standard Normal Distribution

- z-score is number of standard deviations that x falls from the mean
- Negative z-score below mean; positive is above
- z-scores calculate probabilities of a normal random variable



z-Scores and Standard Normal Distribution

- μ=0 and σ=1
- Random variables with normal distribution can be converted to z-scores with the standard normal distribution

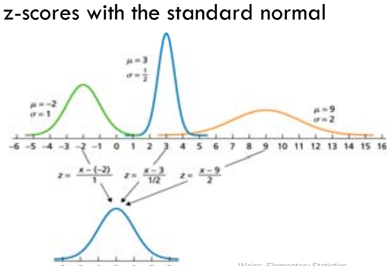


Table A: Standard Normal Probabilities

Tabulates normal cumulative probabilities below μ+zσ

- Compute z
- Look up z in table
- Body gives probability below z-score

- P(z<1.43)=
- P(z>1.43)=

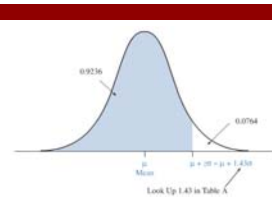


TABLE 6.3: Part of Table A for Normal Cumulative (Left-Tail) Probabilities

The top of the table gives the second digit for z. The table entry is the probability falling below μ + zσ, for instance, 0.9236 below μ + 1.43σ for z = 1.43.

Second Decimal Place of z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...										
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9139	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

Table A: Standard Normal Probabilities

- P(-1.43<z<1.43)=

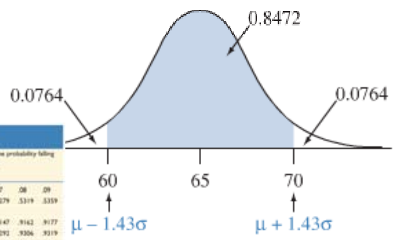


TABLE 6.3: Part of Table A for Normal Cumulative (Left-Tail) Probabilities

The top of the table gives the second digit for z. The table entry is the probability falling below μ + zσ, for instance, 0.9236 below μ + 1.43σ for z = 1.43.

Second Decimal Place of z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...										
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

Find z Given Probability?

- Use Table A in reverse
  - Find probability in body of table
  - The z-score is given by first column and row
- Find z for a probability of 0.025

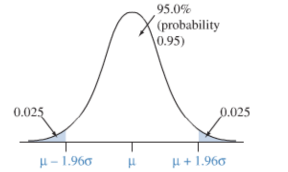


TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Find z Given Probability?

- Find z for a probability of 0.025 to the right

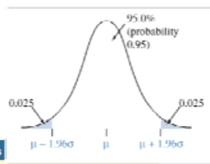


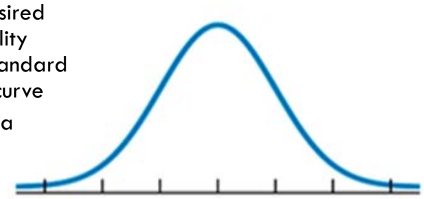
TABLE A Standard Normal Cumulative Probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Finding Probabilities for Normally Distributed Random Variables

1. State problem in terms of random variable: P(X<x)
2. Draw picture to show desired probability under standard normal curve
3. Find area

$$\boxed{P(X<x)=P\left(Z<z=\frac{x-\mu}{\sigma}\right)}$$



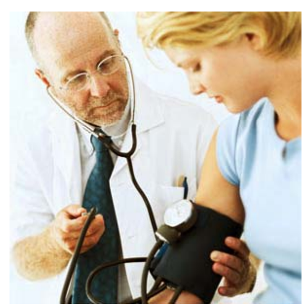
P(X<x)

www.besthealthcarerates.com

- Adult systolic blood pressure is normally distributed with μ = 120 and σ = 20. What percentage of adults have systolic blood pressure less than 100?
- P(X<100) = Normalcdf(-1E99,100,120,20)=.1587
- 15.9% of adults have systolic blood pressure less than 100

P(X>x)

- μ = 120 and σ = 20. What percentage of adults have systolic blood pressure greater than 133?
- P(X>133) = Normalcdf(133,1E99, 120,20)=.2578
- 25.8% of adults have systolic blood pressure greater than 133



www.views.pk

P(a<X<b)

- μ = 120 and σ = 20. What percentage of adults have systolic blood pressure between 100 and 133?
- P(100<X<133) = Normcdf(100,133,120,20)=.5835
- 58% of adults have systolic blood pressure between 100 and 133

Find X Value Given Area to Left

- μ = 120 and σ = 20.

  What is the 1st quartile?
- P(X<x)=.25, find x:

x = Invnorm(.25,120,20) = 106.6

Find X Value Given Area to Right

- μ = 120 and σ = 20. 10% of adults have systolic blood pressure above what level?
- P(X>x)=.10, find x.

x = Invnorm(.9,120,20) =145.6

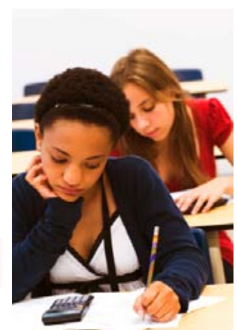
Using Z-scores to Compare Distributions

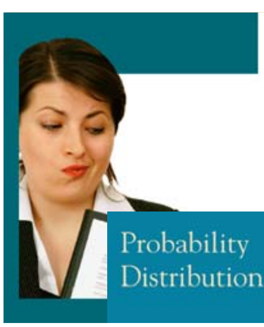
You score 650 on the SAT which has μ=500 and σ=100 and 30 on the ACT which has μ=21.0 and σ=4.7. On which test did you perform better?

- Compare z-scores

SAT: $z = \frac{650-500}{100} = 1.5$

ACT: $z = \frac{30-21}{4.7} = 1.91$





Probability Distributions 6

6.3 Probabilities When Each Observation Has Two Possible Outcomes

Learning Objectives

1. The Binomial Distribution
2. Conditions for Binomial Distribution
3. Probabilities for Binomial Distribution
4. Factorials
5. Examples using Binomial Distribution
6. Do Binomial Conditions Apply?
7. Mean and Standard Deviation of Binomial Distribution
8. Normal Approximation to Binomial

The Binomial Distribution


www.mschip.com

- Each observation is binary: has one of two possible outcomes
- Examples:
  1. Accept or decline an offer from a bank for a credit card
  2. Have or do not have health insurance
  3. Vote yes or no on a referendum

Factorials

Factorial – n factorial is n! = n*(n-1)*(n-2)…2*1

0! = 1
1! = 1
2! = 2*1
3! = 3*2*1
4! = 4*3*2*1
5! = 5*4*3*2*1
…

Conditions for Binomial Distribution

1. Each of n trials must have two possible outcomes: "success" or "failure"
2. Each must have same probability of success, p
3. Trials must be independent

The binomial random variable, X, is the number of successes in the n trials

Probabilities for a Binomial Distribution

Denote the probability of success on a trial by p. For n independent trials, the probability of x successes equals

$$P(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

Finding Binomial Probabilities: ESP


www.phantasmpsiresearch.com

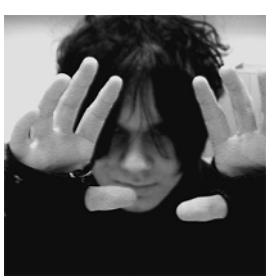
John Doe claims ESP.

- A person in one room picks an integer 1 to 5 at random
- In another room, John Doe identifies the number
- Three trials performed
- Doe got correct answer twice

Finding Binomial Probabilities: ESP

What is the probability of guessing correctly on two of the three trials?

1. SSF, SFS, and FSS
2. Each has probability: (0.2)(0.2)(0.8)=0.032
3. The total probability of two correct guesses is 3(0.032)=0.096


www.psychicsavers.com

Finding Binomial Probabilities: ESP

Outcome	Probability	Outcome	Probability
SSS	$0.2 \times 0.2 \times 0.2 = (0.2)^3$	SFF	$0.2 \times 0.8 \times 0.8 = (0.2)^1(0.8)^2$
SSF	$0.2 \times 0.2 \times 0.8 = (0.2)^2(0.8)^1$	FSF	$0.8 \times 0.2 \times 0.8 = (0.2)^1(0.8)^2$
SFS	$0.2 \times 0.8 \times 0.2 = (0.2)^2(0.8)^1$	FFS	$0.8 \times 0.8 \times 0.2 = (0.2)^1(0.8)^2$
FSS	$0.8 \times 0.2 \times 0.2 = (0.2)^2(0.8)^1$	FFF	$0.8 \times 0.8 \times 0.8 = (0.8)^3$

The probability of exactly 2 correct guesses is the binomial probability with n = 3 trials, x = 2 correct guesses and p = 0.2 probability of a correct guess.

$$P(2) = \frac{3!}{2!1!}(0.2)^2(0.8)^1 = 3(0.04)(0.8) = 0.096$$

2nd DISTR
0:binompdf(n,p,x)
Binompdf(3,.2,2)=0.096

Binomial Mean and Standard Deviation

The binomial probability distribution for n trials with probability p of success on each trial has:

$$\boxed{\mu = np, \quad \sigma = \sqrt{np(1 - p)}}$$

Racial Profiling?

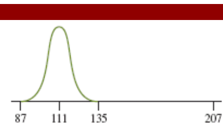
3.bp.blogspot.com

207 of 262 police car stops in Philadelphia in 1997 were African-American, which comprised 42.2% of the population at that time. Does the number of African-Americans stopped suggest possible bias?

Racial Profiling?

x = 207; n = 262; p = 0.422

$$\boxed{\mu = np, \quad \sigma = \sqrt{np(1 - p)}}$$




www.zercustoms.com

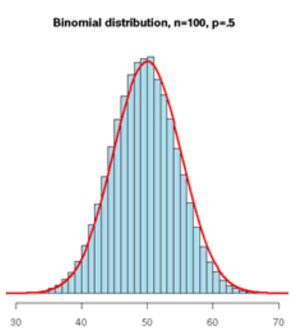
Racial Profiling?


3.bp.blogspot.com

- If no racial profiling, would we be surprised if between 87 and 135 stopped were African-American?
- What about 207?
- Different people do different amounts of driving, so we don't know that 42.2% of the potential stops were African-American.

Approximating Binomial with Normal Distribution

The binomial distribution can be well approximated by the normal distribution when the expected number of successes, *np*, and the expected number of failures, *n(1-p)* are both at least 15.



zoonek2.free.fr

Image Sources

Statistics: The Art and Science of Learning from Data, 2nd Edition, Agresti and Franklin
The Drunkard's Walk, Mlodinow
Elementary Statistics, 7th Edition, Neil Weiss
http://cdn.information-management.com/media/editorial/dmdirect/0603/062403_dutta_2.gif
http://weblog.terrellrussell.com/wp-content/uploads/2007/10/800px-normal_distribution_pdf.png
http://www.besthealthcarerates.com/Portals/41042/images//j0423046.jpg
http://www.views.pk/wp-content/uploads/2007/06/blood-pressure.jpg http://media.rd.com/rd/images/rdc/books/change-one/blood-pressure-tests-af.jpg
http://www.healthinformer.net/wp-content/uploads/2008/07/blood-presure.jpg
https://ssl.msstate.edu/virtual/hrm/DFC/images/MPj04003600000%5B1%5D.jpg
http://www.betterreportcard.com/img/SAT-ACT.jpg
http://www.mschip.com/wp-content/uploads/2009/12/affordable-health-insurance2.jpg
http://3.bp.blogspot.com/_CjXtoDbakPE/R9B0SB7M4KI/AAAAAAAAAYM/udwsbdWYLew/s400/binary+nails.bmp
http://sheerchaos.files.wordpress.com/2010/03/1-bp-blogspot-com.jpg
http://www.fallingfifth.com/files/comics/factorial.png
http://www.phantasmpsiresearch.com/JPGS/Rhine%20Institute%20ESP%20testing.jpg
http://www.psychicsavers.com/wp-content/uploads/2009/02/devin-1-main_full.jpg
http://www.columbusyouthexpressions.org/teamwork/images/stories/failure-success.jpg
http://3.bp.blogspot.com/_M8chV5JZM8c/RzxjjG6QrLI/AAAAAAAABQg/NRdAb72JU08/s320/profilingart2_large.jpg
http://www.zercustoms.com/news/images/Dodge/2009-Dodge-Charger-police-car-1.jpg
http://3.bp.blogspot.com/_30PRmkOI4ro/SfGXTA4HePI/AAAAAAAAOdk/E6r4MjvTmUc/s400/african-american-hairstyles1.jpg
http://zoonek2.free.fr/UNIX/48_R/g564.png
http://www.physics.umd.edu/grt/taj/104a/lightning2.jpg
http://hoppingintopuddles.files.wordpress.com/2008/03/random-numbers.jpg
http://www.economicpopulist.org/files/images/roulette_wheel.jpg
http://www.anticsland.com/images/games/large/billiards.jpg
http://psy2.ucsd.edu/~kang/child%20pictures/children-jump.jpg
http://www.vat19.com/webimages/covers/large/slot-machine-paperweight.jpg
http://rchsbowman.files.wordpress.com/2008/11/graph-of-discrete-probability-distribution1.jpg
http://cnfolio.com/images/img536expectedvalue.gif
http://www.math.armstrong.edu/statsonline/3/emprule.gif
https://www.glastonburyus.org/staff/MolnarL/PublishingImages/TI-84-graphing-calculator.jpg