UCLA STAT 251 / OBEE 216

Statistical Methods for the Life and Health Sciences

- **Instructor**: **Ivo Dinov**,

  Asst. Prof. In Statistics and Neurology

University of California, Los Angeles, Winter 2002

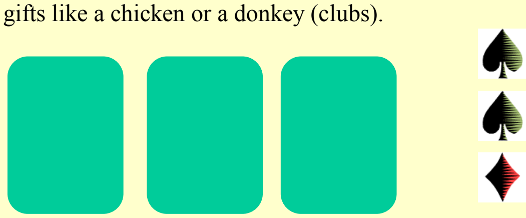
http://www.stat.ucla.edu/~dinov/

Chapter 2: Probabilities, Binomial, Poisson, Normal distributions

- Variables are classified in various ways, such as either **qualitative** (categorical) and **quantitative** (numerical) or as continuous vs. discrete and so on.
- The **mean** and the **standard deviation**, which are used summaries for distributions.
- Two special discrete distributions: the **Binomial** and **Poisson**.
- One special continuous distribution called the **Normal distribution**, both in *standard* and *nonstandard* forms. The **z score** is introduced.
- We describe how to ***approximate*** *the binomial by the normal* using the continuity correction.

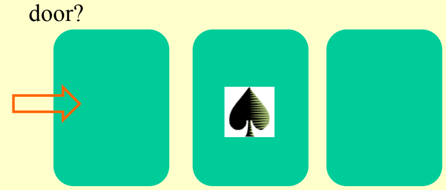
Let's Make a Deal Paradox – aka, Monty Hall 3-door problem

- This paradox is related to a popular television show in the 1970's. In the show, a contestant was given a choice of three doors/cards of which one contained a prize (diamond). The other two doors contained gag gifts like a chicken or a donkey (clubs).



Let's Make a Deal Paradox.

- After the contestant chose an initial door, the host of the show then revealed an empty door among the two unchosen doors, and asks the contestant if he or she would like to switch to the other unchosen door. The question is should the contestant switch. Do the odds of winning increase by switching to the remaining door?



1.Pick One card

2.Show one Club Card

3. Change 1st pick?

Let's Make a Deal Paradox.

- The *intuition* of most people tells them that each of the doors, the chosen door and the unchosen door, are equally likely to contain the prize so that there is a 50-50 chance of winning with either selection? This, however, is **not the case**.
- The probability of winning by using the switching technique is 2/3, while the odds of winning by not switching is 1/3. The easiest way to explain this is as follows:

Let's Make a Deal Paradox.

- The probability of picking the wrong door in the initial stage of the game is 2/3.
- If the contestant picks the wrong door initially, the host must reveal the remaining empty door in the second stage of the game. Thus, if the contestant switches after picking the wrong door initially, the contestant will win the prize.
- The probability of winning by switching then reduces to the probability of picking the wrong door in the initial stage which is clearly 2/3.

Let's Make a Deal Paradox.

- Demo: AdditionalAids.dir/StatGames.exe
  - Uncertainty→Pick a door

Long run behavior of coin tossing

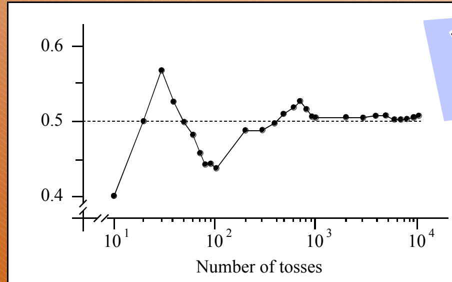


Figure 4.1.1 Proportion of heads versus number of tosses for John Kerrich's coin tossing experiment.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Definitions …

- The law of averages about the behavior of coin tosses – the relative proportion (relative frequency) of heads-to-tails in a coin toss experiment becomes more and more stable as the number of tosses increases. The law of averages applies to relative frequencies not absolute counts of #H and #T.
- Two widely held misconceptions about what the law of averages about coin tosses:
  - Differences between the actual numbers of heads & tails becomes more and more variable with increase of the number of tosses – a seq. of 10 heads doesn't increase the chance of a tail on the next trial.
  - Coin toss results are fair, but behavior is still unpredictable.

Types of Probability

- Probability models have two essential components (*sample space*, the space of all possible outcomes from an experiment; and a list of *probabilities* for each event in the sample space). Where do the outcomes and the probabilities come from?
- Probabilities from models – say mathematical/physical description of the sample space and the chance of each event. Construct a fair die tossing game.
- Probabilities from data – data observations determine our probability distribution. Say we toss a coin 100 times and the observed Head/Tail counts are used as probabilities.
- Subjective Probabilities – combining data and psychological factors to design a reasonable probability table (e.g., gambling, stock market).

Sample Spaces and Probabilities

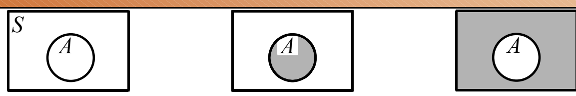
- When the relative frequency of an event in the past is used to estimate the probability that it will occur in the future, what assumption is being made?
  - The underlying process is stable over time;
  - Our relative frequencies must be taken from large numbers for us to have confidence in them as probabilities.
- All statisticians agree about how probabilities are to be combined and manipulated (in math terms), however, not all agree what probabilities should be associated with a particular real-world event.
- When a weather forecaster says that there is a 70% chance of rain tomorrow, what do you think this statement means? (Based on our past knowledge, according to the barometric pressure, temperature, etc. of the conditions we expect tomorrow, 70% of the time it did rain under such conditions.)

Sample spaces and events

- A ***sample space***, *S*, for a random experiment is the set of all possible outcomes of the experiment.
- An ***event*** is a *collection* of outcomes.
- An event ***occurs*** if any outcome making up that event occurs.

The complement of an event

- The complement of an event A, denoted $\bar{A}$, occurs *if and only if* A does not occur.

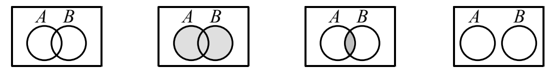


(a) Sample space containing event A (b) Event A shaded (c) $\bar{A}$ shaded

Figure 4.4.1 An event A in the sample space S.

Combining events – *all statisticians agree on*

- "***A* or *B***" contains all outcomes in A **or** B (or both).
- "***A* and *B***" contains all outcomes which are **in both** A and B.



(a) Events A and B (b) "A or B" shaded (c) "A and B" shaded (d) Mutually exclusive events

Figure 4.4.2 Two events.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Mutually exclusive events cannot occur at the same time.

Probability distributions

- Probabilities always lie between 0 and 1 and they sum up to 1 (across all simple events) .
- ***pr(A)*** can be obtained by adding up the probabilities of all the outcomes in A.

$$pr(A) = \sum_{\substack{E \text{ outcome} \\ \text{in event } A}} pr(E)$$

Review

- Law of averages for the coin-toss example.
- Sample spaces, outcomes, events, complements.
- Probabilities are always in the range [0 : 1]
- ***pr(A)*** can be obtained by adding up the probabilities of all the outcomes in A.

$$pr(A) = \sum_{\substack{E \text{ outcome} \\ \text{in event } A}} pr(E)$$

Properties of probability distributions

- A sequence of number $\{p_1, p_2, p_3, \ldots, p_n\}$ is a probability distribution for a sample space $S = \{s_1, s_2, s_3, \ldots, s_n\}$, if $pr(s_k) = p_k$, for each $1<=k<=n$. The two essential properties of a probability distribution $p_1, p_2, \ldots, p_n$?

$$p_k \geq 0; \quad \sum_k p_k = 1$$

- How do we get the probability of an event from the probabilities of outcomes that make up that event?
- If all outcomes are <u>distinct</u> & <u>equally likely</u>, how do we calculate *pr(A)* ? If $A = \{a_1, a_2, a_3, \ldots, a_9\}$ and $pr(a_1)=pr(a_2)=\ldots=pr(a_9)=p$; then

$pr(A) = 9 \times pr(a_1) = 9p$.

Example of probability distributions

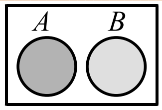
- Tossing a coin twice. *Sample space* S={HH, HT, TH, TT}, for a fair coin each outcome is equally likely, so the probabilities of the 4 possible outcomes should be identical, p. Since, p(HH)=p(HT)=p(TH)=p(TT)=p and

$$p_k \geq 0; \quad \sum_k p_k = 1$$

- p = ¼ = 0.25.

Rules for manipulating Probability Distributions

For mutually exclusive events,
pr(*A* or *B*) = pr(*A*) + pr(*B*)



Review

- If *A* and *B* are mutually exclusive, what is the probability that both occur? (0) What is the probability that at least one occurs? (sum of probabilities)
- If we have two or more mutually exclusive events, how do we find the probability that at least one of them occurs? (sum of probabilities)
- Why is it sometimes easier to compute *pr(A)* from $pr(A) = 1 - pr(\bar{A})$? (The complement of the even may be easer to find or may have a known probability. E.g., a random number between 1 and 10 is drawn. Let A ={a number less than or equal to 9 appears}. Find pr(A) = 1 – pr($\bar{A}$)). probability of $\bar{A}$ is pr({10 appears}) = 1/10 = 0.1. Also Monty Hall 3 door example!

Melanoma – type of skin cancer – an example of laws of conditional probabilities

TABLE 4.6.1: 400 Melanoma Patients by Type and Site

Type	Site: Head and Neck	Site: Trunk	Site: Extremities	Row Totals
Hutchinson's melanomic freckle	22	2	10	34
Superficial	16	54	115	185
Nodular	19	33	73	125
Indeterminant	11	17	28	56
Column Totals	68	106	226	400

Contingency table based on Melanoma histological type and its location

Conditional Probability

The ***conditional probability*** of ***A*** occurring ***given*** that *B* occurs is given by

$$\mathrm{pr}(A \mid B) = \frac{\mathrm{pr}(A \text{ and } B)}{\mathrm{pr}(B)}$$

Suppose we select one out of the 400 patients in the study and we want to find the probability that the cancer is on the *extremities* *given that* it is of type *nodular*: P = 73/125 = P(C. on Extremities | Nodular)

$$\frac{\#\text{nodular patients with cancer on extremities}}{\#\text{nodular patients}}$$

Review

$$\mathrm{pr}(A \text{ and } B) = \mathrm{pr}(A \mid B)\mathrm{pr}(B) = \mathrm{pr}(B \mid A)\mathrm{pr}(A)$$

$$\mathrm{pr}(A) = 1 - \mathrm{pr}(\bar{A})$$

1. Proportions (partial description of a real population) and probabilities (giving the chance of something happening in a random experiment) may be identical – under the experiment *choose-a-unit-at-random*
2. Properties of probabilities.

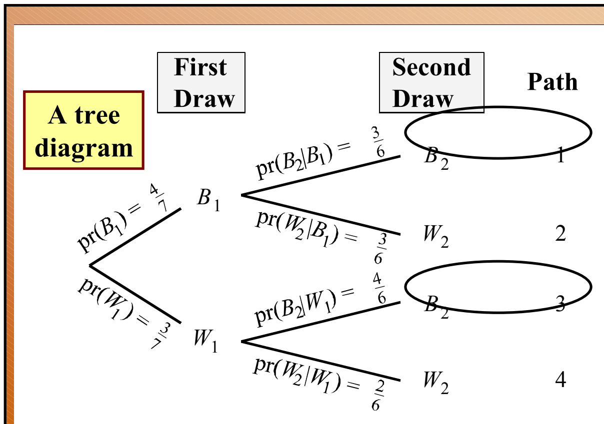
$$\{p_k\}_{k=1}^{N} \text{ define probabilities} \Leftrightarrow p_k \geq 0; \quad \sum_k p_k = 1$$

A tree diagram for computing conditional probabilities

Suppose we draw 2 balls at random one at a time *without replacement* from an urn containing **4 black** and **3 white** balls, otherwise identical. What is the probability that the *second ball is black*? Sample Spc?

Mutually exclusive

P({*2-nd ball is black*}) =
P({*2-nd is black*} &{*1-st is black*}) +
P({2-nd is black} &{*1-st is white*}) =
4/7 x 3/6 + 4/6 x 3/7 = 4/7.



Conditional probabilities and 2-way tables

- Many problems involving conditional probabilities can be solved by constructing two-way tables
- This includes *reversing the order of conditioning*

P(A & B) = P(A | B) x P(B) = P(B | A) x P(A)

Proportional usage of *oral contraceptives* and their rates of failure

We need to complete the two-way contingency table of proportions

pr(*Failed* and *Oral*) = pr(*Failed* | *Oral*) × pr(*Oral*) [= 5% of 32%]

pr(*Failed* and *IUD*) = pr(*Failed* | *IUD*) × pr(*IUD*) [= 6% of 3%]

		Method					
		Steril.	Oral	Barrier	IUD	Sperm.	Total
Outcome	Failed	0 × .38	.05 × .32	.14 × .24	.06 × .03	.26 × .03	?
	Didn't	?	?	?	?	?	?
	Total	.38	.32	.24	.03	.03	1.00

pr(*Steril.*) = .38 pr(*Barrier*) = .24 pr(*IUD*) = .03

Oral contraceptives cont.

pr(*Failed* and *Oral*) = pr(*Failed* | *Oral*) × pr(*Oral*) [= 5% of 32%]

pr(*Failed* and *IUD*) = pr(*Failed* | *IUD*) × pr(*IUD*) [= 6% of 3%]

		Method					
		Steril.	Oral	Barrier	IUD	Sperm.	Total
Outcome	Failed	0 × .38	.05 × .32	.14 × .24	.06 × .03	.26 × .03	?
	Didn't	?	?	?	?	?	?
	Total	.38	.32	.24	.03	.03	1.00

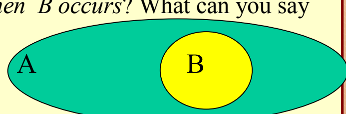
pr(*Steril.*) = .38 pr(*Barrier*) = .24 pr(*IUD*) = .03

TABLE 4.6.4 Table Constructed from the Data in Example 4.6.8

		Method					
		Steril.	**Oral**	**Barrier**	**IUD**	**Sperm.**	**Total**
Outcome	Failed	0	.0160	.0336	.0018	.0078	.0592
	Didn't	.3800	.3040	.2064	.0282	.0222	.9408
	Total	.3800	.3200	.2400	.0300	.0300	1.0000

Remarks …

- In *pr(A | B),* how should the symbol " | " is read *given that*.
- How do we interpret the fact that: *The event A always occurs when B occurs*? What can you say about pr(*A* | *B*)?



- When drawing a probability tree for a particular problem, how do you know *what events* to use for the first fan of branches and which events to use for the subsequent branching? (at each branching stage condition on all the info available up to here. E.g., at first branching use all simple events, no prior is available. At 3-rd branching condition of the previous 2 events, etc.).

TABLE 4.6.5 Number of Individuals Having a Given Mean Absorbance Ratio (MAR) in the ELISA for HIV Antibodies

MAR	Healthy Donor		HIV patients	
<2	202	} 275	0	} 2 False-Negatives (FNE)
2 - 2.99	73	Test cut-off	2	
3 - 3.99	15	False-positives	7	Power of a test is:
4 - 4.99	3		7	
5 - 5.99	2		15	1-P(FNE)=
6 -11.99	2		36	1-P(Neg|HIV)
12+	0		21	~ 0.976
Total	297		88	

Adapted from Weiss et al.[1985]

HIV cont.

pr(*HIV* and *Positive*) = pr(*Positive*|*HIV*) × pr(*HIV*) [= 98% of 1%]

pr(*Not HIV* and *Negative*) = pr(*Negative*|*Not HIV*) × pr(*Not HIV*) [= 93% of 99%]

		Test result		
		Positive	Negative	Total
Disease status	HIV	.98 × .01	?	.01 ⟵ pr(*HIV*) = .01
	Not HIV	?	.93 × .99	.99 ⟵ pr(*Not HIV*) = .99
	Total	?	?	1.00

Figure 4.6.6 Putting HIV information into the table.

HIV – reconstructing the contingency table

pr(*HIV* and *Positive*) = pr(*Positive*|*HIV*) × pr(*HIV*) [= 98% of 1%]

pr(*Not HIV* and *Negative*) = pr(*Negative*|*Not HIV*) × pr(*Not HIV*) [= 93% of 99%]

		Test result		
		Positive	Negative	Total
Disease status	HIV	.98 × .01	?	.01 ⟵ pr(*HIV*) = .01
	Not HIV	?	.93 × .99	.99 ⟵ pr(*Not HIV*) = .99
	Total	?	?	1.00

TABLE 4.6.6 Proportions by Disease Status and Test Result

		Test Result		
		Positive	**Negative**	**Total**
Disease	HIV	.0098	.0002	.01
Status	Not HIV	.0693	.9207	.99
	Total	.0791	.9209	1.00

Proportions of HIV infections by country

TABLE 4.6.7 Proportions Infected with HIV

Country	No. AIDS Cases	Population (millions)	pr(HIV)	Having | Test pr(HIV | Positive)
United States	218,301	252.7	0.00864	0.109
Canada	6,116	26.7	0.00229	0.031
Australia	3,238	16.8	0.00193	0.026
New Zealand	323	3.4	0.00095	0.013
United Kingdom	5,451	57.3	0.00095	0.013
Ireland	142	3.6	0.00039	0.005

Statistical independence

- Events *A* and *B* are *statistically independent* if knowing whether *B* has occurred gives no new information about the chances of *A* occurring,

  i.e. if **pr(*A* | *B*) = pr(*A*)**

- Similarly, P(*B* | *A*) = P(*B*), since

  P(B|A)=P(B & A)/P(A) = P(~~A~~|B)P(B)/~~P(A)~~ = P(B)

- If *A* and *B* are *statistically independent*, then

$$\text{pr}(A \text{ and } B) = \text{pr}(A) \times \text{pr}(B)$$

Example using independence

There are many genetically based blood group systems. Two of these are: *Rh* blood type system (Rh+ and Rh-) and the *Kell* system (K+ and K-). For Europeans the following proportions are experimentally obtained.

Table 4.7.1 Blood Type Data

.08×.81 .92×.81 pr(*RH+*) = .81 .08×.19

	K+	*K-*	Total
Rh+	?	?	.81
Rh-	?	?	.19
Total	.08	.92	1.00

pr(*K+*) = .08 .92 × .19

⟶

	K+	*K-*	Total
Rh+	.0648	.7452	.81
Rh-	.0152	.1748	.19
Total	.08	.92	1.00

How can we fill in the inside of the two-way contingency table?
It is known that anyone's blood type in one system is *independent* of their type in another system.

P(Rh+ and K+) = P(Rh+) x P(K+)= 0.81 x 0.08 = 0.0648

People vs. Collins

TABLE 4.7.2 Frequencies Assumed by the Prosecution

Yellow car	$\frac{1}{10}$	Girl with blond hair	$\frac{1}{3}$
Man with mustache	$\frac{1}{4}$	Black man with beard	$\frac{1}{10}$
Girl with ponytail	$\frac{1}{10}$	Interracial couple in car	$\frac{1}{1000}$

- The first occasion where a conviction was made in an American court of law, largely on statistical evidence, 1964. A woman was mugged and the offender was described as a wearing dark cloths, with blond hair in a pony tail who got into a yellow car driven by a black male accomplice with mustache and beard. The suspect brought to trial were picked out in a line-up and fit all of the descriptions. Using the *product rule for probabilities* an expert witness computed the chance that a random couple meets these characteristics, as 1:12,000,000.

Summary

- What does it mean for two events *A* and *B* to be *statistically independent*?
- Why is the working rule under independence, *P(A and B) = P(A) P(B)*, just a special case of the multiplication rule *P(A & B) = P(A | B) P(B)* ?
- *Mutual independence* of events $A_1, A_2, A_3, \ldots, A_n$ *if and only* if $P(A_1 \& A_2 \& \ldots \& A_n) = P(A_1)P(A_2)\ldots P(A_n)$
- What do we mean when we say two human characteristics are *positively associated*? *negatively associated*? (blond hair – blue eyes, pos.; black hair – blue eyes, neg.assoc.)

Review

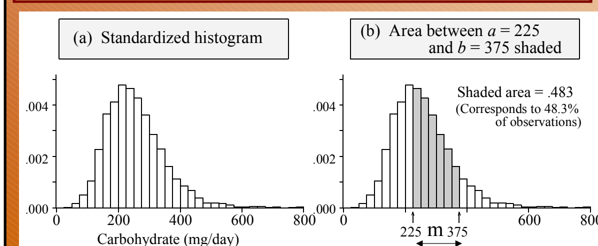
- What happens to the calculated P(*A* and *B*) if we treat positively associated events as independent? if we treat negatively associated events as independent?

  (Example, let B={A + {b}}, A & B are pos-assoc'd, P(A&B)=P(A)[P(A)+P({b})], under indep. assump's. However, P(A&B)=P(B|A)P(A)=1 x P(A) **>** P(A)[P(A)+P({b})], underestimating the real chance of events. If A & B are neg-assoc'd → A & comp(B) are pos-assoc'd. In general, this may lead to answers that are grossly too small or too large …)
- Why do people often treat events as independent? When can we trust their answers?(Easy computations! Not always!)

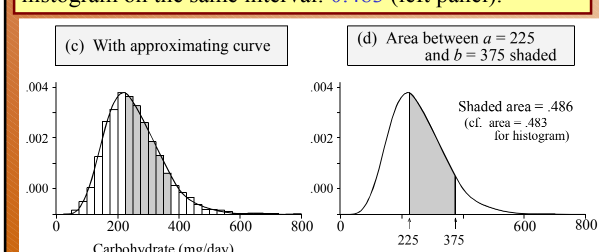
Supplement on Probabilities

- (ProbabilitiesPDF_Poisson_Ch2AppendB.pdf)
- Random variables, p. 4
- PDF, Expectation, Variance
- Binomial distribution
- Poisson distribution
- Normal Distribution – follows here!

Dietary intake of carbohydrate (mg/day) for 5929 people from a variety of work environments. Standardized histogram plot is unimodal but skewed to the right (high values). Vertical scale is (relative freq.)/(interval width) = f_j/mn. The proportion of the data in [a : b] is the area under the standardized histogram on the range [a: b].

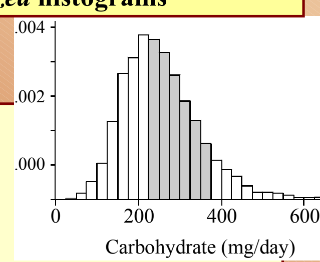


Superposition of a smooth curve (density function) on the standardized histogram (left panel). Area under the density curve on [a: b] = [225: 375] is analytically computed to be: 0.486 (right panel), which is close to the empirically obtained estimate of the area under the histogram on the same interval: 0.483 (left panel).



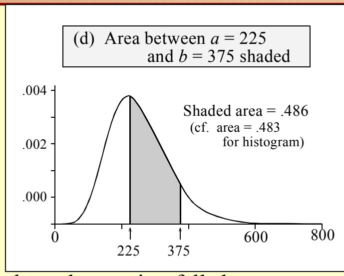
Standardized histograms

For a standardized histogram:



- The vertical scale is *Relative_frequency / Interval_width*
- ***Total area*** under histogram = 1
- ***Proportion*** of the data between *a* and *b* is the ***area*** under histogram between *a* and *b*

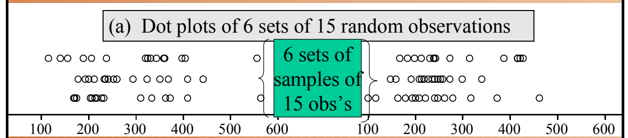
Probability and areas



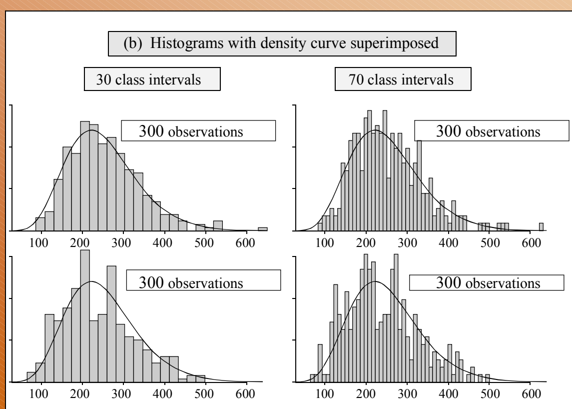
For a continuous X

- the probability a random observation falls between a and b = area under the density curve between a and b.

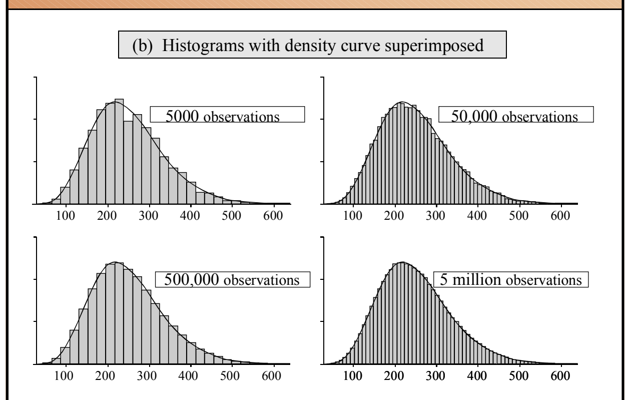
Sampling from the distribution



Note the fair amount of intra- and inter-group variability. What does that mean? Is that normal or expected?

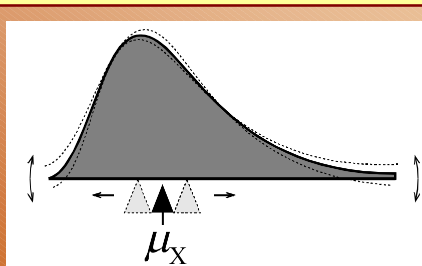


From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.



From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

Visualizing the population mean



The population mean is the imaginary value of X where the density curve balances

Interval endpoints and continuous variables

Recall a continuous variable is one where the domain has no gaps in between the values the variable can take.

In calculations involving a continuous random variable we *do not have to worry* about whether interval endpoints are included or excluded.

Review

- How does a standardized histogram differ from a relative-frequency histogram? raw histogram? (f_j/mn)
- What graphic feature conveys the proportion of the data falling into a class interval for a standardized histogram? for a relative-frequency histogram?
  (area=width . height = m f_j/mn= f_j/n)
- What are the two fundamental ways in which random observations arise? (Natural phenomena, sampling experiments – choose a student at random and use the lottery method to record characteristics, scientific experiments - blood pressure measure)
- How does a density curve describe probabilities?
  (The probability that a random obs. falls in [a:b] is the area under the PDF on the same interval.)

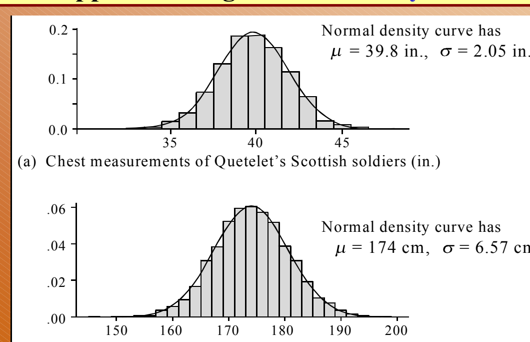
Review

- What is the total area under both a standardized histogram and a probability density curve? (1)
- When can histograms of data from a random process be relied on to closely resemble the density curve for that process? (large sample size, small histogram bin-size)
- What characteristic of the density curve does the mean correspond to? (imaginary value of X, where the density curve balances)

Review

- Does it matter whether interval endpoints are included or excluded when we calculate probabilities for a continuous random variable from the area? (No)
- Why? (Area[a:b] == Area(a:b))
- Are discrete variables the same or different in this regard, interval endpoint not effecting the area? (Different)

Two standardized histograms with approximating Normal density curve

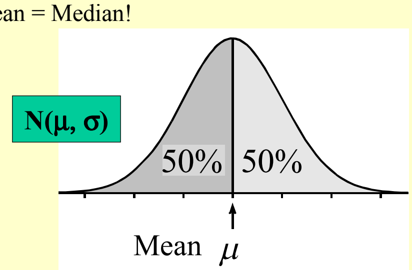


(a) Chest measurements of Quetelet's Scottish soldiers (in.)

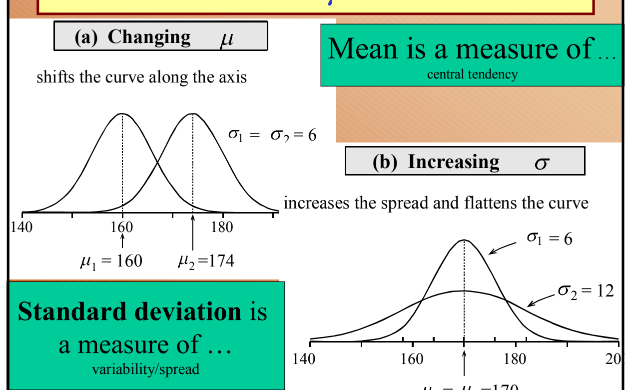
(b) Heights of the 4294 men in the workforce database (cm)

The Normal distribution density curve

- Is symmetric about the mean! Bell-shaped and unimodal.
- Mean = Median!



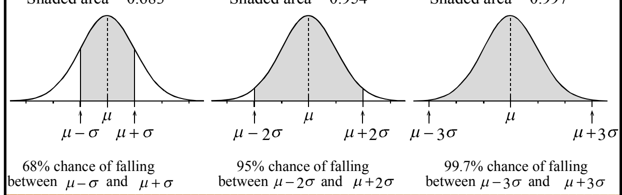
Effects of μ and σ



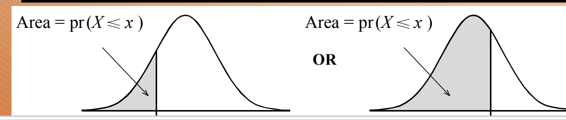
Understanding the standard deviation: σ

Probabilities/areas and numbers of standard deviations

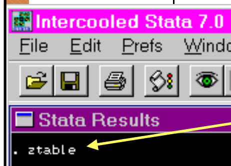
for the Normal distribution



Probabilities supplied by computer programs – Cumulative (lower-tail) probabilities



Areas in [0:Z] of the Std.Normal Distribution Are obtained by STATA command ztable



Areas between 0 & Z of the Standard Normal Distribution

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.40	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.50	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.60	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549

Probabilities supplied by computer programs – Cumulative (lower-tail) probabilities

Problem: To find P(X<=180), when μ=174 and σ=6.57

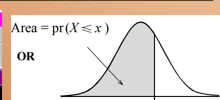
Convert to Standard units: Y = (X- μ)/ σ = 6/6.57 = 0.91

Look-up the Normal Distribution Table: 0.3186

Final cumulative (lower-tail) result: 0.5+ 0.3186= 0.819

Areas between 0 & Z of the Standard Normal Distribution

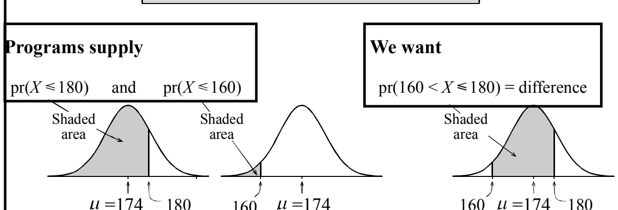
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.40	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.50	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.60	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.70	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.80	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.90	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.00	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621



Basic method for obtaining probabilities

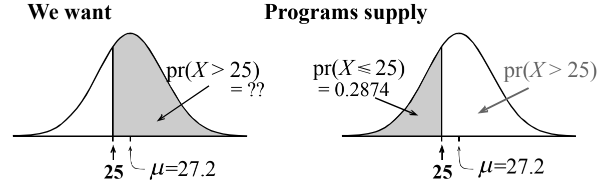
- Sketch a Normal curve, marking the mean and other values of interest.
- Shade the area under the curve that gives the desired probability.
- Devise a way of getting the desired area from lower-tail areas.
- Obtain component lower-tail probabilities from a computer program

(a) Computing pr($160 < X \leqslant 180$)



pr($160 < X \leqslant 180$) = pr($X \leqslant 180$) − pr($X \leqslant 160$)

Obtaining an upper-tail probability



Since total area under curve = 1, pr($X > 25$) = 1 - pr($X \leqslant 25$)

Generally, **pr($X > x$) = 1 − pr($X \leqslant x$)**

Review

- What features of the Normal curve do μ and σ visually correspond to? (point-of-balance; width/spread)
- What is the probability that a random observation from a normal distribution is smaller than the mean? (0.5) larger than the mean? (0.5) exactly equal to the mean? (0.0) Why?

Review

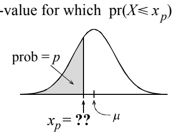
- Approximately, what is the probability that a random observation from a normal distribution falls within 1 standard deviation (SD) of the mean? (0.68) 2 SD's? (0.95) 3 SD's? (0.997)
- Computer programs may provide cumulative or partial probabilities for the normal distribution. What is a the difference between these? Can we get one from the other?

The inverse problem – Percentiles/quantiles

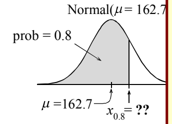
(a) *p*-Quantile

Programs supply x_p

x-value for which $\text{pr}(X \le x_p) = p$



(b) 80th percentile (0.8-quantile) **of women's heights**



Program returns
Thus 80% lie belo

80% of people have height below the 80th percentile. This is EQ to saying there's 80% chance that a random observation from the distribution will fall below the 80th percentile.

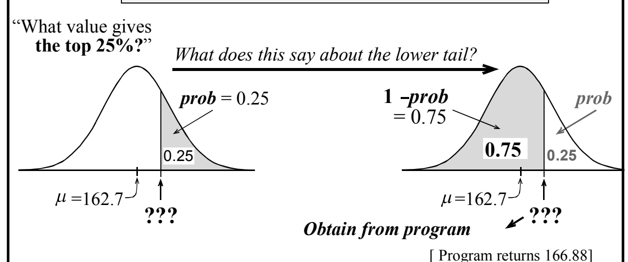
(c) Further percentiles of women's heights

Percent	1%	5%	10%	20%	30%	70%	80%	90%	95%
Propn	0.01	0.05	0.1	0.2	0.3	0.7	0.8	0.9	0.95
Percentile	*(or quantile)*								
(cm)	148.3	152.5	154.8	157.5	159.4	166.0	167.9	170.6	172.9
(ft'in")	4'10"	5'0"	5'0"	5'2"	5'2"	5'5"	5'6"	5'7"	5'8"

The inverse problem is what is the height for the 80th percentile/quantile? So far we studied given the height value what's the corresponding percentile?

The inverse problem – *upper-tail* percentiles/quantiles

Obtaining an inverse upper-tail probability



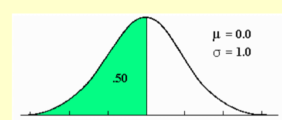
Review

- What is meant by the 60th percentile of heights?
- What is the difference between a percentile and a quantile? (percentile used in expressing results in %, whereas quantiles used to express results in term of probabilities)
- The lower quartile, median and upper quartile of a distribution correspond to special percentiles. What are they? express in terms of quantiles. (25%, 50%, 75%)
- Quantiles are sometimes called inverse cumulative probabilities. Why?

Standard Normal Curve

- The standard normal curve is described by the equation:

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

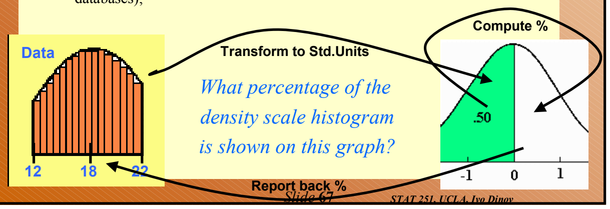


Where remember, the natural number e ~ 2.7182…

We say: X~Normal(μ, σ), or simply X~N(μ, σ)

Standard Normal Approximation

- The standard normal curve can be used to estimate the percentage of entries in an interval for any process. Here is the protocol for this approximation:
  - Convert the interval (we need the assess the percentage of entries in) to standard units. We saw the algorithm already.
  - Find the corresponding area under the normal curve (from tables or online databases);

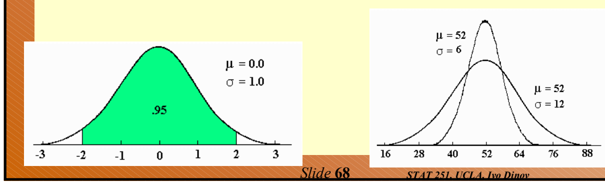


General Normal Curve

- The general normal curve is defined by:
  - Where μ is the average of (the symmetric) normal curve, and σ is the standard deviation (spread of the distribution).

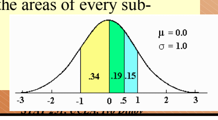
$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

  - Why worry about a standard and general normal curves?
  - How to convert between the two curves?



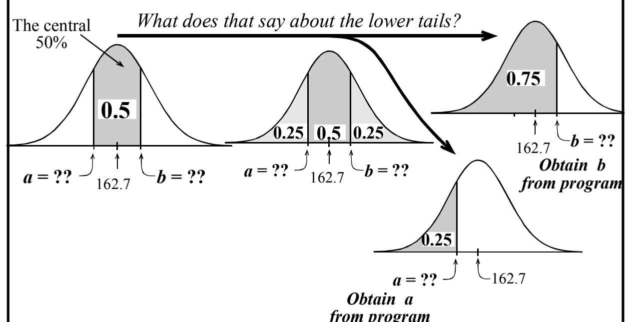
Areas under Standard Normal Curve – Normal Approximation

- Protocol:
  - Convert the interval (we need to assess the percentage of entries in) to Standard units. Actually convert the end points in Standard units.
    - In general, the transformation X → (X-µ)/σ, standardizes the observed value X, where µ and σ are the average and the standard deviation of the distribution X is drawn from.
  - Find the corresponding area under the normal curve (from tables or online databases);
    - Sketch the normal curve and shade the area of interest
    - Separate your area into individually computable sections
    - Check the Normal Table and extract the areas of every sub-section
    - Add/compute the areas of all sub-sections to get the total area.



Obtaining central range for symmetric distributions

What values contain the **central** 50%?



The z-score

- The ***z*-score** of *x* is the number of standard deviations *x* is from the mean. (Body-Mass-Index, BMI)

TABLE 6.3.1 Examples of *z* -Scores

X	z-score = $(x - \mu)/\sigma$		Interpretation
Male BMI values (kg/m^2)			
25	(25-27.3)/4.1 =	-0.56	25 kg/m^2 is 0.56 sd's below the mean
35	(35-27.3)/4.1 =	1.88	35 kg/m^2 is 1.88 sd's above the mean
Female heights (cm)			
155	(155-162.7)/6.2 =	-1.24	155cm is 1.24 sd's below the mean
180	(180-162.7)/6.2 =	2.79	180cm is 2.79 sd's above the mean

Male BMI-values: μ=27.3, σ=4.1 Females heights: μ=162.7, σ=6.2

- Which ones of these are unusually large/small/away from the mean?

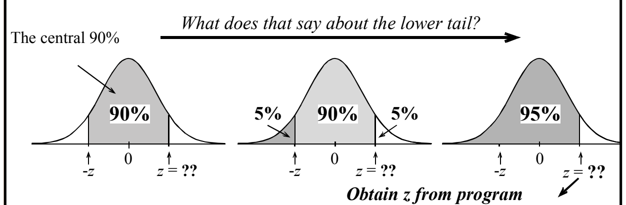
The standard Normal distribution

Standard Normal distribution:

mean(µ) = 0, SD(σ)= 1

Working in standard units

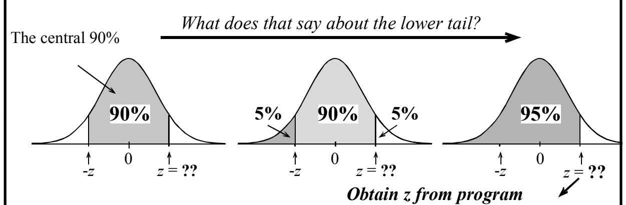
What values contain the **central** 90%?



[Program returns 1.6449]

Review, Mon., Oct. 22, 2001

What values contain the **central** 90%?



[Program returns 1.6449]

Working in standard units (# of SD's)

TABLE 6.3.2 Central Ranges

		Male BMI values		Female heights	
Percentage	z	$\mu - z\sigma$	$\mu + z\sigma$	$\mu - z\sigma$	$\mu + z\sigma$
80%	1.2816	22.05	32.55	154.8	170.6
90%	1.6449	20.56	34.04	152.5	172.9
95%	1.9600	19.26	35.34	150.5	174.9
99%	2.5758	16.74	37.86	146.7	178.7
99.9%	3.2905	13.81	40.79	142.3	183.1

Male BMI-values: μ=27.3, σ=4.1
Females heights: μ=162.7, σ=6.2

Standardizing $Z = (X - \mu)/\sigma$

Inverting $X = Z\sigma + \mu$

TABLE 6.3.3 Using z-score tables

As an example, we shall find pr($Z \le$ 1.1357) using part of the table given in Appendix A4 (reproduced below).

Step 1: Correct the z-value to two decimal places, that is, use z = 1.14.
Step 2: Look down the z column until you find 1.1. This tells you which row to look in.
Step 3: The second decimal place, here 4, tells you which column to look in.
Step 4: The entry in the table corresponding to that row and column is pr($Z \le$ 1.14) = 0.873

z	0	1	2	3	4	5	6	7	8	9
1.0	.841	.844	.846	.848	.851	.853	.855	.858	.860	.862
1.1	.864	.867	.869	.871	$\boxed{.873}$	.875	.877	.879	.881	.883
1.2	.885	.887	.889	.891	.893	.894	.896	.898	.900	.901
1.3	.903	.905	.907	.908	.910	.911	.913	.915	.916	.918
1.4	.919	.921	.922	.924	.925	.926	.928	.929	.931	.932

Quincunx (see QuincunxApplet.html)

Continuous Variables and Density Curves

- There are no gaps between the values a continuous random variable can take.
- Random observations arise in two main ways: (i) by sampling populations; and (ii) by observing processes.

The density curve

- The probability distribution of a continuous variable is represented by a density curve.
  - ***Probabilities*** are represented by ***areas under the curve***,
    - the probability that a random observation falls between a and b equal to the area under the density curve between a and b.
  - The total area under the curve equals 1.
  - The population (or distribution) mean $\mu_X = E(X)$, is where the density curve balances.
  - When we calculate probabilities for a continuous random variable, it does not matter whether interval endpoints are included or excluded.

For any random variable X

- $E(aX+b) = a\,E(X)+b$ and $SD(aX+b) = |a|\,SD(X)$

The Normal distribution

$$X \sim \text{Normal}(\mu_x = \mu,\ \sigma_x = \sigma)$$

Features of the Normal density curve:

- The curve is a symmetric bell-shape centered at μ.
- The standard deviation σ governs the spread.
  - 68.3% of the probability lies within 1 standard deviation of the mean
  - 95.4% within 2 standard deviations
  - 99.7% within 3 standard deviations

Probabilities

- Computer programs provide lower-tail (or cumulative) probabilities of the form $pr(X \le x)$
  - We give the program the x-value; it gives us the probability.
- Computer programs also provide inverse lower-tail probabilities (or quantiles)
  - We give the program the probability; it gives us the x-value.
- When calculating probabilities, we shade the desired area under the curve and then devise a way of obtaining it via lower-tail probabilities.

Standard Units

The *z-score* of a value *a* is ….

- the number of standard deviations a is away from the mean
- positive if a is above the mean and negative if a is below the mean.

The ***standard Normal*** distribution has $\mu = 0$ and $\sigma = 0$.

- We usually use Z to represent a random variable with a standard Normal distribution.

Ranges, extremes and z-scores

Central ranges:

- $P(-z \le Z \le z)$ is the same as the probability that a random observation from an arbitrary Normal distribution falls within z SD's either side of the mean.

Extremes:

- $P(Z \ge z)$ is the same as the probability that a random observation from an arbitrary Normal distribution falls more than z standard deviations above the mean.
- $P(Z \le -z)$ is the same as the probability that a random observation from an arbitrary Normal distribution falls more than z standard deviations below the mean.

Combining Random Quantities

Variation and independence:

- No two animals, organisms, natural or man-made objects are ever identical.
- There is always variation. The only question is whether it is large enough to have a practical impact on what you are trying to achieve.
- Variation in component parts leads to even greater variation in the whole.

Independence

We model variables as being independent ….

- if we think they relate to physically independent processes
- and if we have no data that suggests they are related.

Both sums and differences of independent random variables are more variable than any of the component random variables

Formulas

- For a constant number a, $E(aX) = aE(X)$ and $SD(aX) = |a|\ SD(X)$.
- Means of sums and differences of random variables act in an obvious way
  - the mean of the sum is the sum of the means
  - the mean of the difference is the difference in the means
- For independent random variables, (cf. Pythagorean theorem),

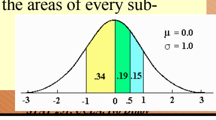
$$SD(X_1 + X_2) = SD(X_1 - X_2) = \sqrt{SD(X_1)^2 + SD(X_2)^2}$$

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

[ASIDE: Sums and differences of independent Normally distributed random variables are also Normally distributed]

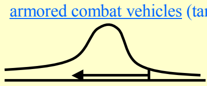
Areas under Standard Normal Curve – Normal Approximation

- Protocol:
  - Convert the interval (we need to assess the percentage of entries in) to Standard units. Actually convert the end points in Standard units.
    - In general, the transformation $X \to (X-\mu)/\sigma$, standardizes the observed value X, where μ and σ are the average and the standard deviation of the distribution X is drawn from.
  - Find the corresponding area under the normal curve (from tables or online databases);
    - Sketch the normal curve and shade the area of interest
    - Separate your area into individually computable sections
    - Check the Normal Table and extract the areas of every sub-section
    - Add/compute the areas of all sub-sections to get the total area.



Areas under Standard Normal Curve – Normal Approximation, Scottish Army Recruits

- The mean height is 64 *in* and the standard deviation is 2 *in*.
  - Only recruits shorter than 65.5 in will be trained for tank operation. What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?

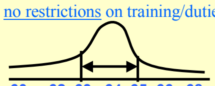


$X \to (X-64)/2$

$65.5 \to (65.5-64)/2 = 3/4$

Percentage is 77.34%

  - Recruits within ½ standard deviations of the mean will have no restrictions on duties. About what percentage of the recruits will have no restrictions on training/duties?



$X \to (X-64)/2$

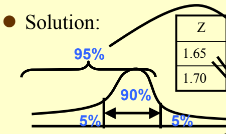
$65 \to (65-64)/2 = 1/2$

$63 \to (63-64)/2 = -1/2$

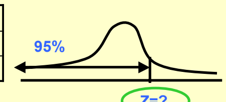
Percentage is 38.30%

Percentiles for Standard Normal Curve

- When the histogram of the observed process follows the normal curve Normal Tables (of any type, as described before) may be used to estimate percentiles. The N-th percentile of a distribution is P is N% of the population observations are less than or equal to P.
- Example, suppose the Math-part SAT scores of newly admitted freshmen at UCLA averaged 535 (out of [200:800]) and the SD was 100. Estimate the 95 percentile for the score distribution.
- Solution:



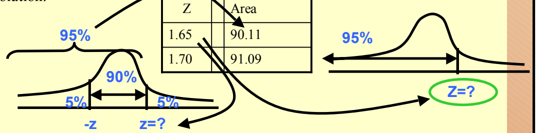
Z	Area
1.65	90.11
1.70	91.09



Percentiles for Standard Normal Curve

- Example, suppose the Math-part SAT scores of newly admitted freshmen at UCLA averaged 535 (out of [200:800]) and the SD was 100. Estimate the 95 percentile for the score distribution.
- Solution:

Z	Area
1.65	90.11
1.70	91.09



- Z=1.65 (std. Units) → 700 (data units), since

  X → (X – μ)/σ, converts data to standard units and

  X → σ X + μ, converts standard to data units!

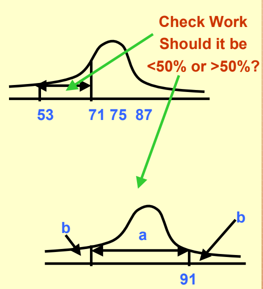
  σ = 100; μ =535, 100 x 1.65 + 535 = 700.

Summary

1. The Standard Normal curve is symmetric w.r.t. the origin (0,0) and the total area under the curve is 100% (1 unit)
2. Std units indicate how many SD's is a value below (-)/above (+) the mean
3. Many histograms have roughly the shape of the normal curve (bell-shape)
4. If a list of numbers follows the normal curve the percentage of entries falling within each interval is estimated by: 1. Converting the interval to StdUnits and, 2. Computing the corresponding area under the normal curve (Normal approximation)
5. A histogram which follows the normal curve may be reconstructed just from (μ,σ^2), mean and variance=std_dev^2
6. Any histogram can be summarized using percentiles
7. E(aX+b)=aE(X)+b, Var(aX+b)=a^{2}Var(X), where E(Y) the the mean of Y and Var(Y) is the square of the StdDev(Y),

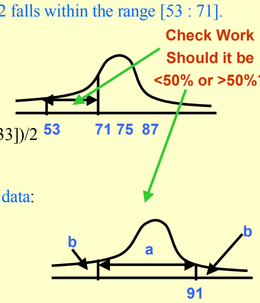
Example – work out in your notebooks

1. Compute the chance a random observation from a distribution (symmetric, bell-shaped, unimodal) with m=75 and SD=12 falls within the range [53 : 71].



Example – work out in your notebooks

1. Compute the chance a random observation from a distribution (symmetric, bell-shaped, unimodal) with m=75 and SD=12 falls within the range [53 : 71].
2. (53-75)/12 = -11/6=-1.83 Std unit
3. (71-75)/12=-0.333(3) Std units
4. Area[53:71] =
5. (SN_area[-1.83:1.83] –SN_area[-0.33:0.33])/2
6. = (93% - 25%)/2 = 34%
7. Compute the 90th percentile for the same data:
8. b+a+b=100% a=80% → A=0.8
9. a+b=90% b=10% Z=1.3 SU
10. 90% P = σ1.3 + μ =12x1.3+75=90.6

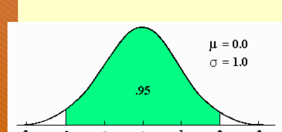


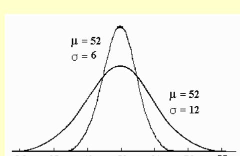
General Normal Curve

- The general normal curve is defined by:
  - Where μ is the average of (the symmetric) normal curve, and σ is the standard deviation (spread of the distribution).

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

  - Why worry about a standard and general normal curves?
  - How to convert between the two curves?





Areas under Standard Normal Curve

- Many histograms are similar in shape to the standard normal curve. For example, persons height. The height of all incoming female army recruits is measured for custom training and assignment purposes (e.g., very tall people are inappropriate for constricted space positions, and very short people may be disadvantages in certain other situations). The mean height is computed to be 64 *in* and the standard deviation is 2 *in*. Only recruits shorter than 65.5 in will be trained for tank operation and recruits within ½ standard deviations of the mean will have no restrictions on duties.

  - What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?

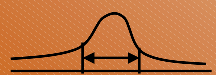
  - About what percentage of the recruits will have no restrictions on training/duties?

Standard Normal Curve – Table differences

- There are different tables and computer packages for representing the area under the standard normal curve. But the results are always interchangeable.

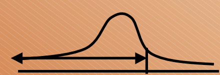
Area under Normal curve on [-z : z]

Z	Area
0.50	38.29
1.0	68.27



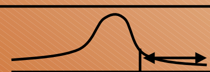
Area under Normal curve on [-infinity : z]

Z	Area
0.50	69.15
1.0	84.13



Area under Normal curve on [z: infinity]

Z	Area
0.50	30.85
1.0	15.87

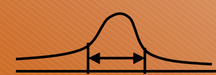


Standard Normal Curve – Table differences

- There are different tables and computer packages for representing the area under the standard normal curve. But the results are always interchangeable.

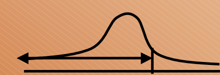
Area under Normal curve on [-z : z]

Z	Area
0.50	38.29
1.0	68.27



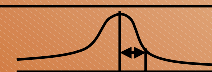
Area under Normal curve on [-infinity : z]

Z	Area
0.50	69.15
1.0	84.13



Area under Normal curve on [0 : z]

Z	Area
0.50	30.85
1.0	15.87



Summary of ideas

- The ***probabilities*** people quote come from 3 main sources:
  - (i) ***Models*** (idealizations such as the notion of equally likely outcomes which suggest probabilities by symmetry).
  - (ii) ***Data*** (e.g.relative frequencies with which the event has occurred in the past).
  - (iii) ***subjective feelings*** representing a degree of belief
- A simple probability model consists of a sample space and a probability distribution.
- A sample space, S, for a random experiment is the set of all possible outcomes of the experiment.

Summary of ideas cont.

- A list of numbers $p_1, p_2, \ldots$ is a ***probability distribution*** for $S = \{s_1, s_2, s_3, \ldots\}$, provided
  - all of the p_i's lie between 0 and 1, and
  - they add to 1.
- According to the probability model, p_i is the probability that outcome s_i occurs.
- We write $p_i = P(s_i)$.

Summary of ideas cont.

- An ***event*** is a collection of outcomes
- An event ***occurs*** if any outcome making up that event occurs
- The probability of event A can be obtained by adding up the probabilities of all the outcomes in A
- If all outcomes are equally likely,

$$\text{pr}(A) = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}}$$

Summary of ideas cont.

- The ***complement*** of an event A, denoted $\overline{A}$, occurs if A does not occur
- It is useful to represent events diagrammatically using ***Venn diagrams***
- ***A union*** of events, ***A* or *B*** contains all outcomes in A or B (including those in both). It occurs if at least one of A or B occurs
- ***An intersection*** of events, ***A and B*** contains all outcomes which are in ***both*** A and B. It occurs only if both A and B occur
- ***Mutually exclusive*** events cannot occur at the same time

Summary of ideas cont.

- The ***conditional probability*** of A occurring ***given*** that B occurs is given by $$\text{pr}(A \mid B) = \frac{\text{pr}(A \text{ and } B)}{\text{pr}(B)}$$
- Events A and B are ***statistically independent*** if knowing whether B has occurred gives no new information about the chances of A occurring, i.e. if $P(A \mid B) = P(A)$ → $P(B|A)=P(B)$.
- If events are physically independent, then, under any sensible probability model, they are also statistically independent
- Assuming that events are independent when in reality they are not can often lead to answers that are grossly too big or grossly too small

Formula Summary

- For discrete sample spaces, pr(A) can be obtained by adding the probabilities of all outcomes in A
- For equally likely outcomes in a finite sample space $$\text{pr}(A) = \frac{\text{number of outcomes in } A}{\text{total number of outcomes}}$$

Formula summary cont.

- $\text{pr}(S) = 1$
- $\text{pr}(\overline{A}) = 1 - \text{pr}(A)$
- If A and B are mutually exclusive events, then $\text{pr}(A \text{ or } B) = \text{pr}(A) + \text{pr}(B)$
  (here "or" is used in the inclusive sense)
- If $A_1, A_2, ..., A_k$ are mutually exclusive events, then
  $\text{pr}(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_k) = \text{pr}(A_1)+\text{pr}(A_2)+\dots+\text{pr}(A_k)$

Formula summary cont.

Conditional probability

- Definition: $$\text{pr}(A \mid B) = \frac{\text{pr}(A \text{ and } B)}{\text{pr}(B)}$$
- Multiplication formula: $$\text{pr}(A \text{ and } B) = \text{pr}(B|A)\text{pr}(A) = \text{pr}(A|B)\text{pr}(B)$$

Formula summary cont.

Multiplication Rule under independence:

- If A and B are independent events, then $$\text{pr}(A \text{ and } B) = \text{pr}(A)\,\text{pr}(B)$$
- If $A_1, A_2, \dots, A_n$ are mutually independent,

$\text{pr}(A_1 \text{ and } A_2 \text{ and ... and } A_n) = \text{pr}(A_1)\,\text{pr}(A_2) \dots \text{pr}(A_n)$