Panel 1

I. Basic analysis of 1 variable

II. 2 variables + associations

III. Hypothesis Testing and Estimations

Need a little Probability Theory

Ex: Flip coin once, what is the probability that H is up? { 1 in 2 / 50% / 0.5

fair

1

Panel 2

Ex: Flip one coin twice. Prob of at least one Tail?

List all possibilities: TT, TH, HT, HH

P(at least one tail) = 3/4 or 75%

P(...) means: "probability of ..."

Ex: Roll a die. P(even) = 3/6 = 1/2 = 0.5

Ex: Roll two dice, record sum.

P(sum is 7 or more) = ?

2

Panel 3

Get organized:

| sum | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

$P(\text{sum} = 10) = 3/36 = 1/12$

$P(\text{sum} = 5) = 4/36 = 1/9$

$P(\text{sum at least } 5) = \frac{30}{36} = \frac{36-6}{36} = 1 - 6/36$

3

Panel 4

General Principles of Probabilities

(1) $P(\text{event})$ is always between 0 and 1

(2) $P(\text{event}) = 0$ means event is impossible
$P(\text{event}) = 1$ means event will happen for sure.

(3) $P(\text{everything}) = 1$

(4) $P(\text{event}) = 1 - P(\text{opposite event})$

4

Panel 5

Probabilities can be found by counting and getting organized.
Also possible: by experimentation.

Ex: $E$ is the event of a random person on the street making between \$30K and \$50K. Find $P(E)$

| Salary | relative % | probabilities |
|---|---|---|
| 10-20K | 6.9 | 0.069 |
| 20-30K | 49.5 | 0.495 |
| 30-40K | 22.2 | 0.222 |
| 40-50K | 6.9 | 0.069 |
| 50-60K | 5.5 | 0.055 |
| >60 | 9% | 0.09 |

Survey.

$P(E) = 0.222 + 0.069$

5

Panel 6

Using Histograms: Study to record how tall people are

| | frequencies | probabilities |
|---|---|---|
| 57 or less | 1 | 1/40 = 0.025 |
| 57 - 58.5 | 1 | 0.025 |
| 58.6 - 60.1 | 3 | 0.075 |
| 60.2 - 61.7 | 6 | 0.15 |
| 61.7 - 63.3 | 8 | 0.2 |
| 63.4 - 64.9 | 11 | 0.275 |
| 64.9 - 66.4 | 3 | 0.075 |
| 66.5 - 67.00 | 7 | 0.175 |
| | 40 | 1 |

A = 60 inches or smaller

B = 65 or taller

C = between 60 and 65

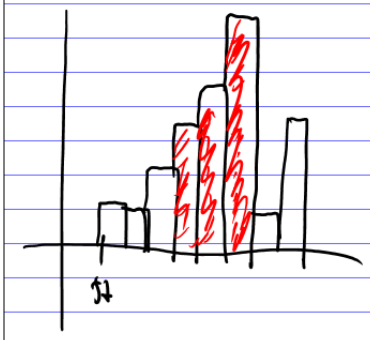
$P(A) = 5/40 = 0.125$ (best guess)

$P(B) = 10/40 = 0.4$ (best guess)

$P(C) = 1 - \frac{5}{40} - \frac{10}{40} = \frac{25}{40} = 0.625$

6

Panel 7



$P(C) = \frac{\text{red area}}{\text{total area}}$

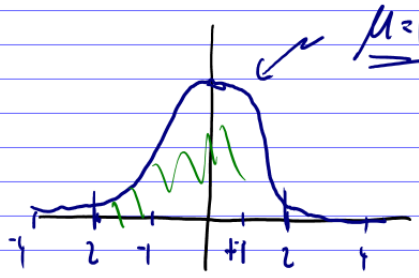
Trick: Most histograms are bell-shaped or normal

One particular distribution will be called.

7

Panel 8

Standard Normal Distribution

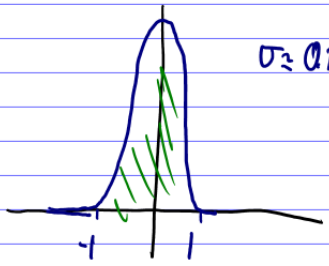


$\mu = 0$ (mean) and $\sigma = 1$

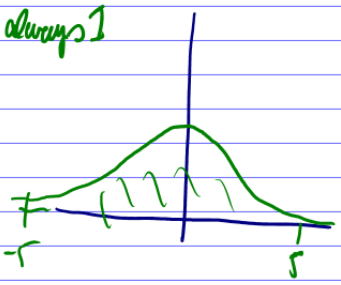
remember

$\sigma = \frac{\text{range}}{4}$

total area under curve is always 1



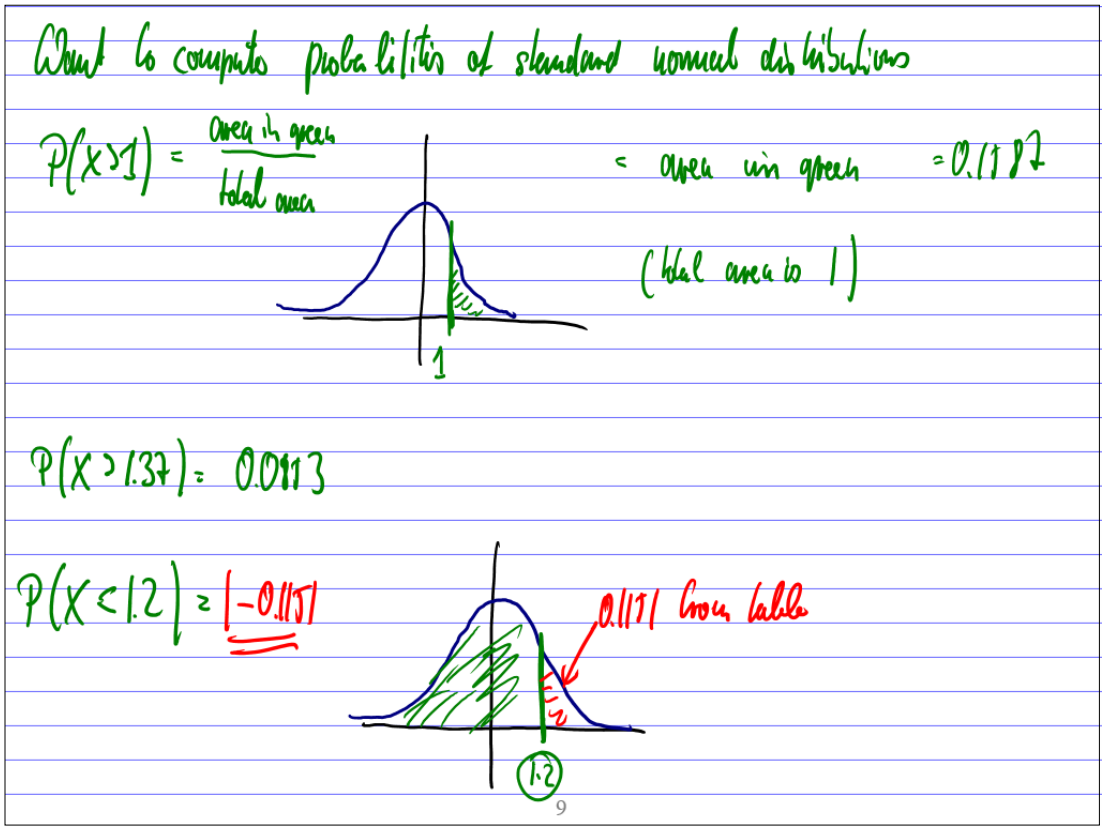
$\sigma = ?$



$\sigma = \frac{10}{4} = 2.5$

8

Panel 9


What to compute probabilities of standard normal distributions
P(X>1) = area in green / total area
= area in green = 0.1587
(total area is 1)
1
P(X>1.37) = 0.0853
P(X<1.2) = |-0.1151|
0.1151 from table
1.2
9