Panel 1

Goal: Investigate 2 numeric variables:

1. Graph – Scatter Plot
2. Best-fit line (least-squares regression line) $y = mx + b$
3. Estimate correlation coeff. r
4. With formulas

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$
$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$
$$S_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}, \quad m = \frac{S_{xy}}{S_{xx}}, \quad b = \bar{y} - m\bar{x}$$

Panel 2

pre-test	post-test
1	2
2	4
3	4
4	7
5	6

Sketchpad

Simple linear regression results:
Dependent Variable: post-test
Independent Variable: pre-test
post-test = 1.3 + 1.1 pre-test
Sample size: 5
R (correlation coefficient) = 0.9022
R-sq = 0.79605263
Estimate of error standard deviation: 1.01652

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	1.3	1.0801387	≠ 0	3	1.2135436	0.3093
pre-test	1.1	0.3214953	≠ 0	3	3.4214286	0.0419

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	12.1	12.1	11.769670	0.0418
Error	3	21	1.0333333		
Total	4	15.2			

$y = mx + b$

$y = 1.1x + 1.3$

$r = 0.9022$ (strong pos. relation)

Predict y if $x = 9$ (say). $y = 1.1 \cdot 9 + 1.3 = 11.2$

scatter plot ⇒

Panel 3

For quiz:

x	y	x^2
3	4	9
2	4	4
1	1	1
6		14

$\bar{x} = \frac{6}{3} = 2$

$\bar{y} = \frac{13}{3} = 4.33$

Find S_{xx} → 14

Given that $S_{xx} = 6$, $S_{yy} = 5$, $S_{xy} = 4$, find m, b, r

$r = \frac{4}{\sqrt{6 \cdot 5}} = \frac{4}{\sqrt{30}} = 0.73$

$m = \frac{4}{6} = \frac{2}{3} = 0.67$

$b = 4.33 - 0.67 \cdot 2 = \#$

$\bar{y} = m\bar{x}$

Panel 4

Suppose $y = 3x + 2$. Predict y if $x = 4$.

$r = 0.037$

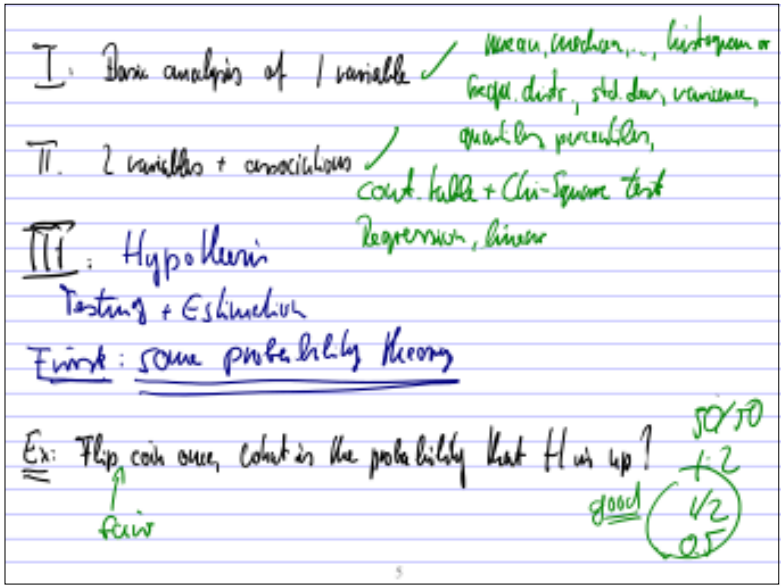
$y = 3(4) + 2 = 14$

Prediction is not believable because $r \approx 0$

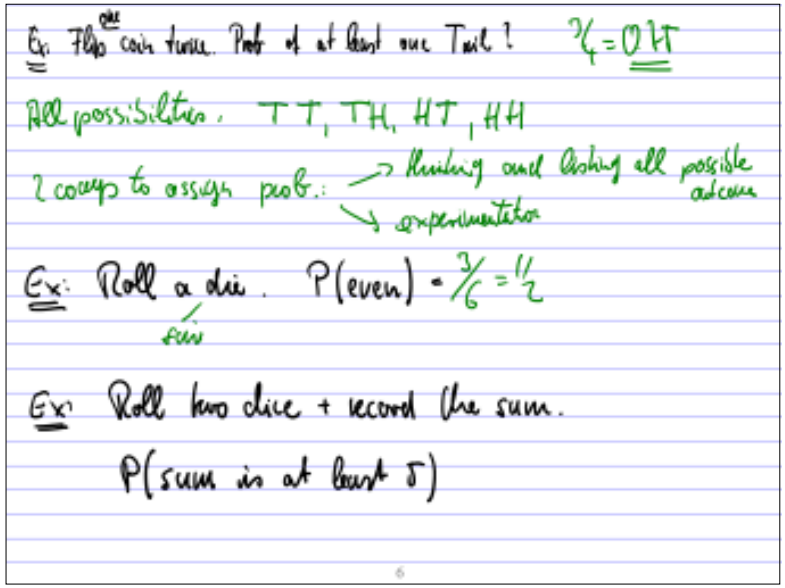
Draw best-fit line and estimate $y = mx + b$.

$m = \frac{2}{1} = 2$, $b = -2$

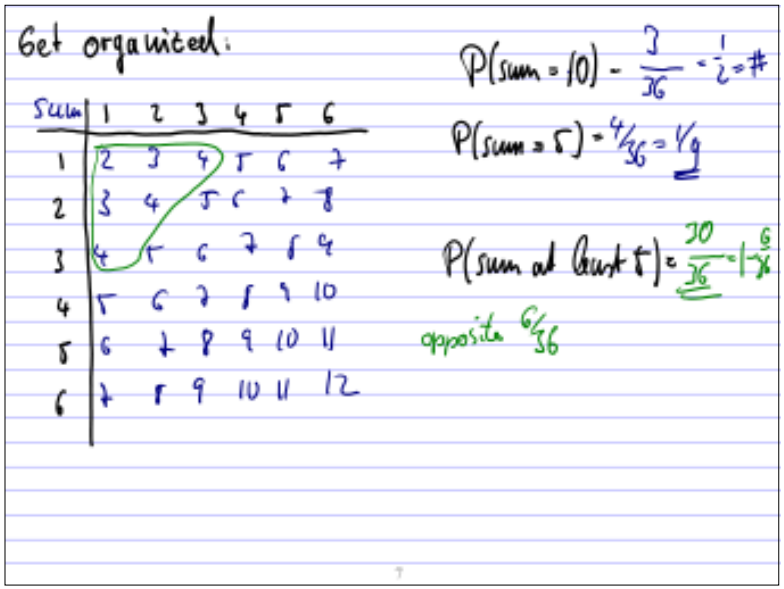
Panel 5



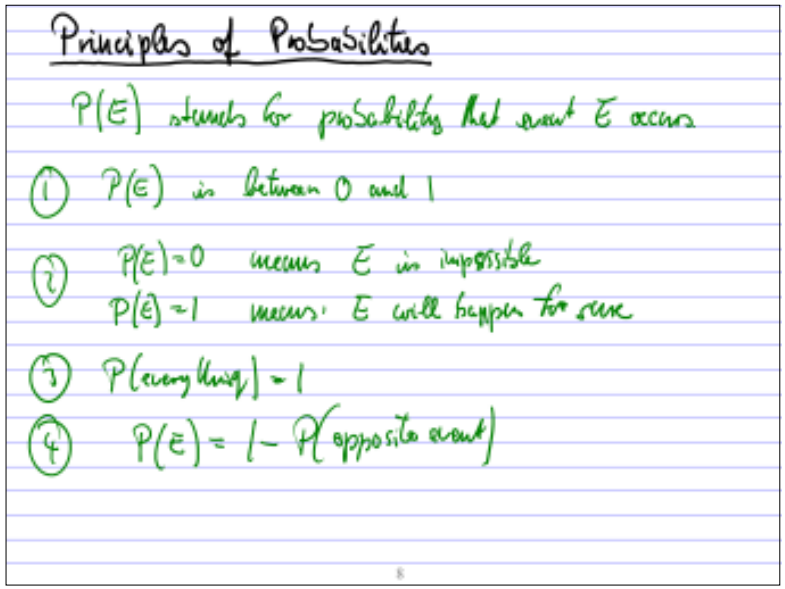
Panel 6



Panel 7



Panel 8



Panel 9

Probabilities can be found by counting + organizing, or by experimentation

Ex: E is the event of a random person on the street making between $30K and $50K. Find P(E) = 0.225 + 0.069

Salary	relat %	probabilities
10-20K	6.9	0.069
20-30K	49.5	0.495
30-40K	22.5	0.225
40-50K	6.9	0.069
50-60K	5.5	0.055
>60	9.5	0.095

9

Panel 10

Using Histograms: Study to record how tall people are

	frequency	probabilities
57 & less	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	7	0.075
66.5 - 67.00	4	0.125
	40	

A = 60 inches or smaller

B = 65 or taller

C = between 60 and 65

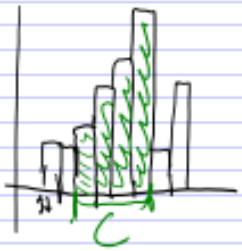
P(A) = 5/40

P(B) = 10/40

P(C) = 1 - 5/40 - 10/40 = 25/40

10

Panel 11



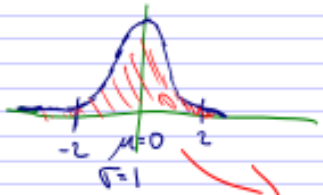
$P(C) = \frac{\text{area of } C}{\text{total area}}$

Most distributions are bell-shaped or normal. One of these bell-shaped distr. will be called Standard Normal Distr (mean=0, std. dev.=1)

11

Panel 12

Standard Normal Distribution

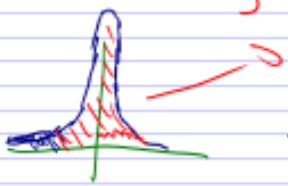


$\frac{\text{range}}{4} = \sigma \Rightarrow \text{range} = 4$

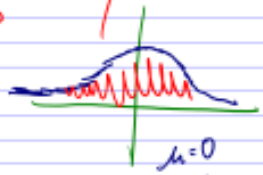
-2 $\mu=0$ 2

$\sigma=1$

red area is 1



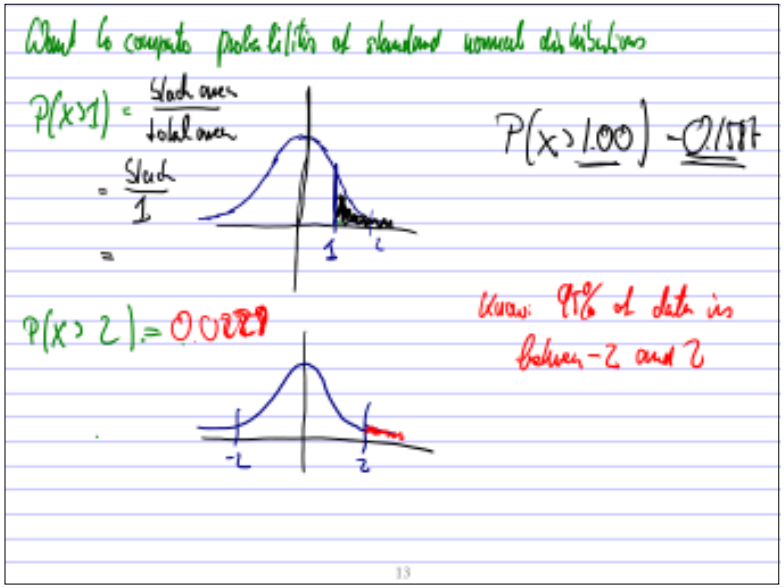
$\mu=0$ $\sigma=1/2$ skinnier + higher



$\mu=0$ $\sigma=2$

12

Panel 13



Panel 14