this time: covariance, correlation, next time: moments, conditional expectation

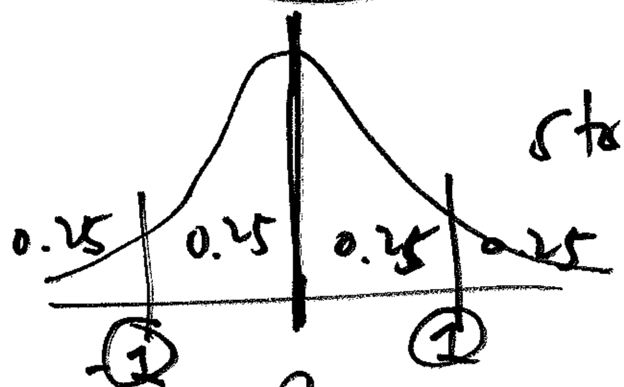
quiz 6 due tomorrow night

AMS 131
19 Aug 19

(1)

$$V(cX) = c^2 V(X)$$

$$V(X + c) = V(X)$$



standard Cauchy

median = 50th percentile = 0.5 quantile

$E(X) =$

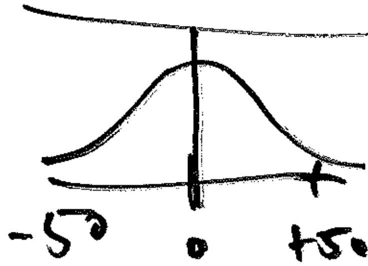
$E(X^2)$

$V(X) = E(X^2) - [E(X)]^2$

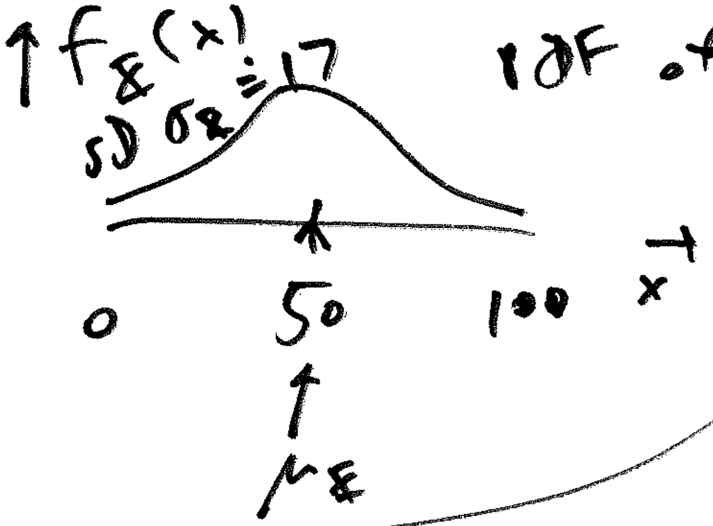
$$V(X) = E\left[(X - \mu)^2\right]$$

$E(X) = 50 = \mu$

$E(X - \mu) = 0$



conj. → skewness $E\left[(X - \mu)^3\right] = 0$



PDF of X

empirical rule (graphical interpretation of SD)

1. start at mean μ_X, go $\boxed{1}$ SD σ_X either way: $(\mu_X - \sigma_X, \mu_X + \sigma_X)$ you will usually enclose about 2/3 (68%) of the total probability (data)

2. ——— $\boxed{2}$ SDs ——— $(\mu_X - 2\sigma_X, \mu_X + 2\sigma_X)$ ——— most (95%) ———

③ ______________ $\boxed{3}$ SDs

______________ $(\mu_{\bar{X}} - 3\sigma_{\bar{X}}, \mu_{\bar{X}} + 3\sigma_{\bar{X}})$

______________ $\boxed{\text{almost all}}$

(99.7%) ______________

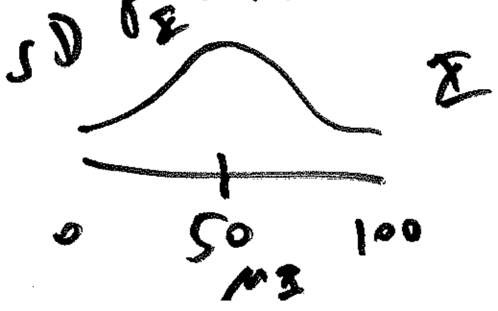
---

① $\sigma_{\bar{X}} = 2$ $(48, 52)$
way too small

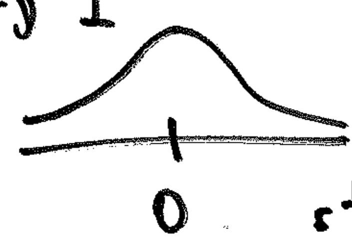
② $\sigma_{\bar{X}} = 45$ $(5, 95)$
way too big

$$50 - 3\sigma_{\bar{X}} = 0 \rightarrow \sigma_{\bar{X}} \doteq 17$$

SD $\sigma_{\bar{X}} \doteq 17$



SD 1



$$\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

standard Normal dist.

$$e^{x} = \frac{1}{0!} + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$= \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$e^{X} = 1 + X + \frac{X^2}{2} + \frac{X^3}{6} + \cdots$$

$$e^{tX} = 1 + (tX) + \frac{(tX)^2}{2} + \frac{(tX)^3}{6} + \cdots$$

$$E\left(e^{tX}\right) = 1 + E(tX) + E\left(\frac{t^2X^2}{2}\right) + E\,\frac{t^3X^3}{6} + \cdots$$

$$= 1 + t\,\boxed{E(X)} + \frac{t^2}{2} E(X^2) + \frac{t^3}{6} E(X^3) + \cdots$$

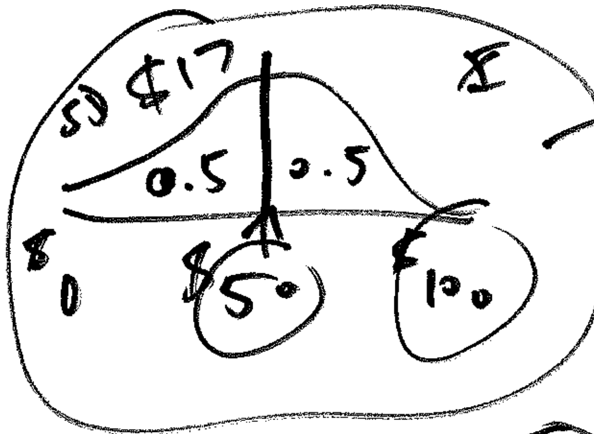
$$\frac{d}{dt} E\left(e^{tX}\right) = E(X) + \frac{2t}{2} E(X^2) + \frac{3t^2}{6} E(X^3) + \dots$$

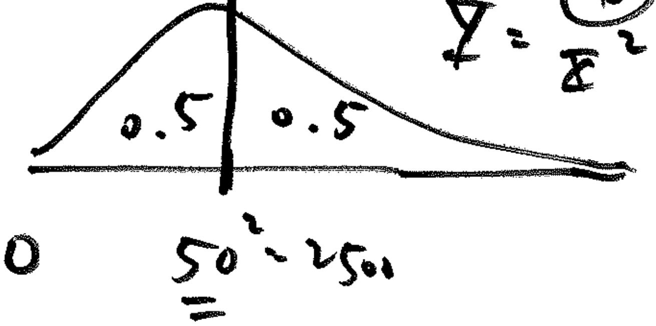
$$\left[\frac{d}{dt} E\left(e^{tX}\right)\right]_{t=0} = E(X)$$

$$\left[\frac{d^2}{dt^2} E\left(e^{tX}\right)\right]_{t=0} = E(X^2)$$

etc.

(9.54) $E\left(e^{itX}\right)$ = <u>characteristic function</u> of X





$\hat{x}$	X	$\hat{x} - X$
50	57	−7
50	39	+11
$\vdots$	$\vdots$	$\vdots$
50		

$E(\hat{x} - X)^2$ Gauss

want small

Laplace

typical prediction error $E|\hat{x} - X|$

$$RMSE(\hat{x}) = SD(X) = \sigma_X$$

↑ root

I always predict $\hat{x} = \$50 = \mu_X$

& I expect to be wrong

(10.42) each time by about

$\sigma_X = \$17$.

$$
\begin{array}{cccc}
 & & \text{height of father} & \\
g_1 & y_1 & x_1 & \underset{\sim}{x}_1 \\
g_2 & y_2 & x_2 & \underset{\sim}{x}_2 \\
\vdots & \vdots & \vdots & n = 1{,}000 \\
 & y_i & x_i & \underset{\sim}{x}_i \\
\vdots & \vdots & \vdots & \\
g_n & y_n & x_n & \underset{\sim}{x}_n
\end{array}
$$

height of son

1 row for each British family in 1885 with at least 1 son

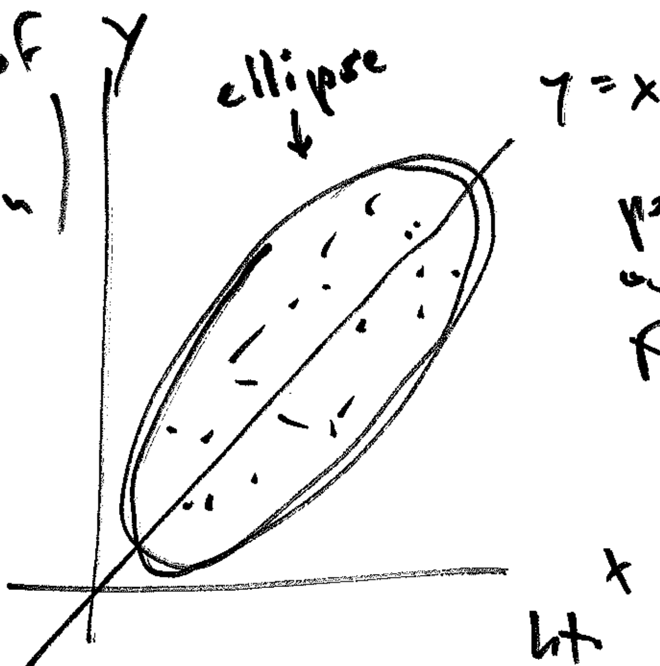
Sir Francis Galton (1890) (eugenics)

(Mendel 1860s plants)

	mean	SD
x	58 in. $= \bar{x}$	2.5 in $= s_x$
y	59 in $= \bar{y}$	2.5 in $= s_y$

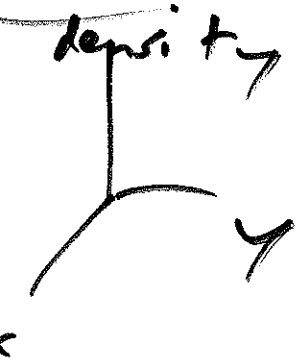
secular trend in height (better nutrition)

ht. of son (-1 in)



ht of father

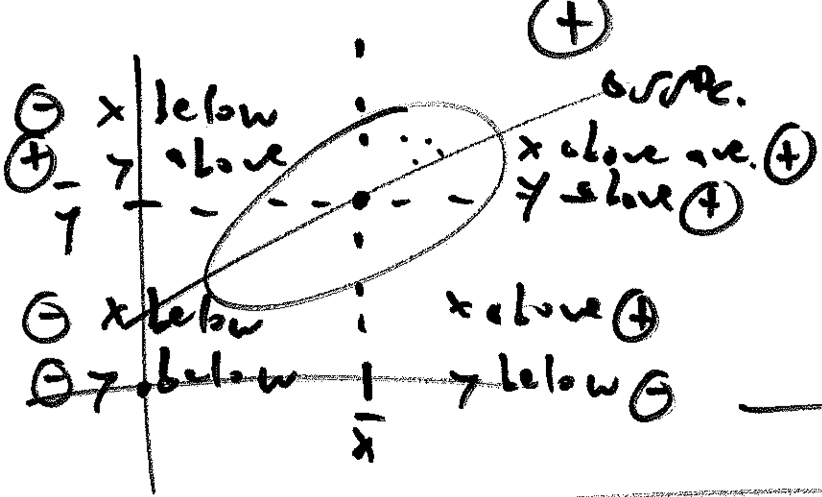
density

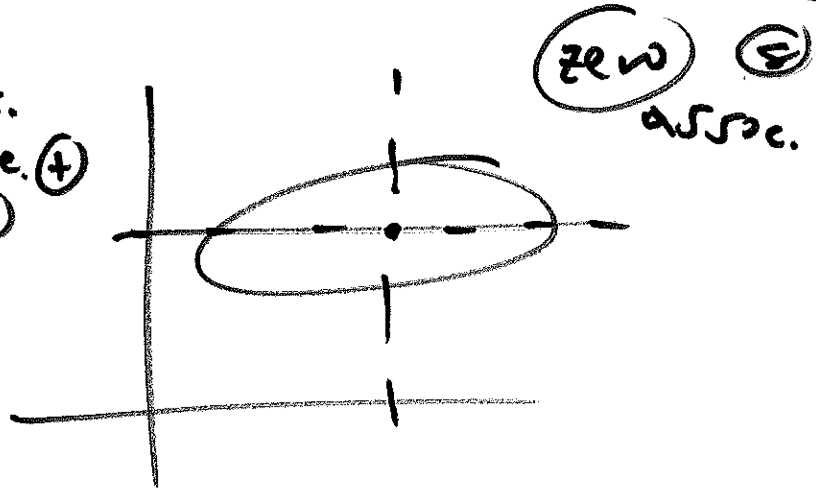


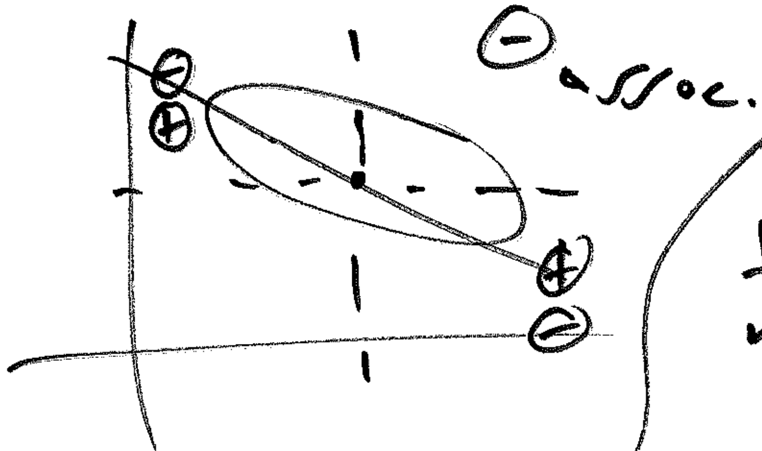
positive association: how strong?

scatterplot

Karl Pearson







$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\cdot\left(\frac{y_i-\bar{y}}{s_y}\right)$$

$= r$

correlation (coefficient)

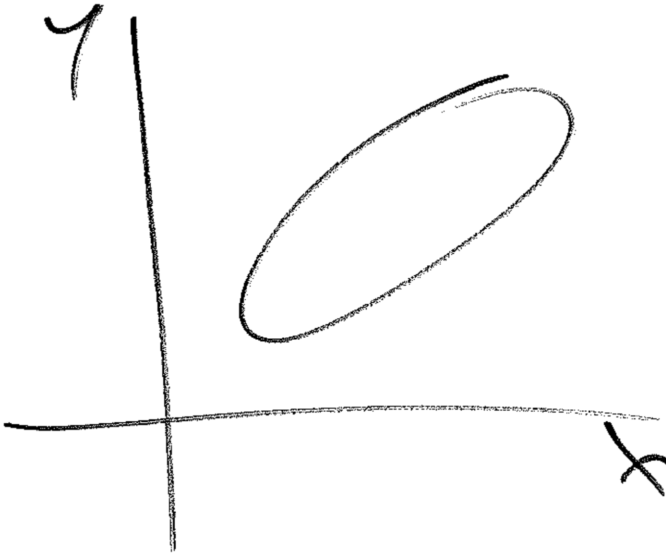
$$E\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\cdot\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right]$$

rv version

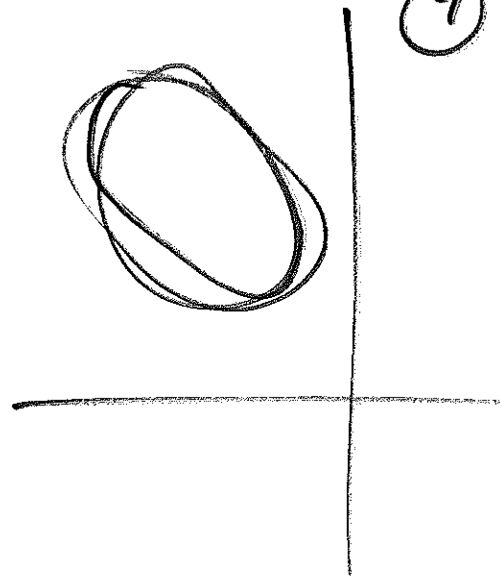
numerator $E\left[(X-\mu_X)(Y-\mu_Y)\right]$

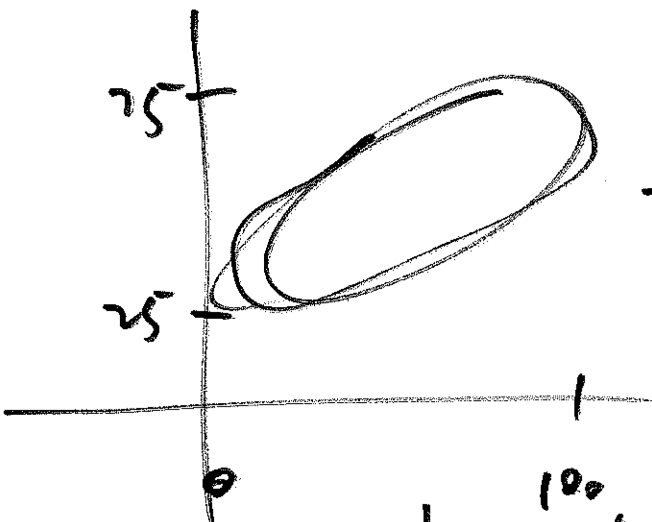
Cov

$C(X,Y)$ = covariance of X and Y

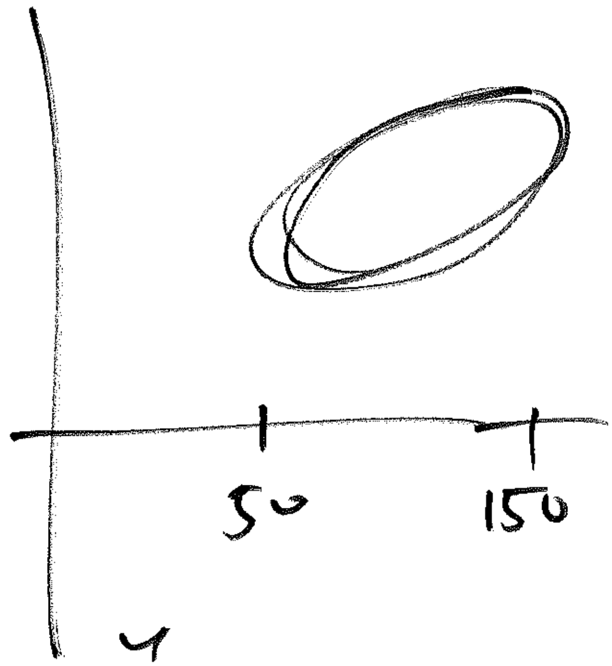


multiply x by -1

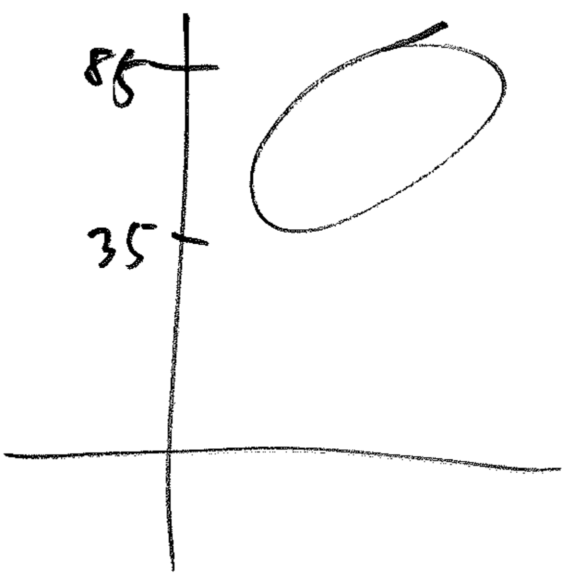


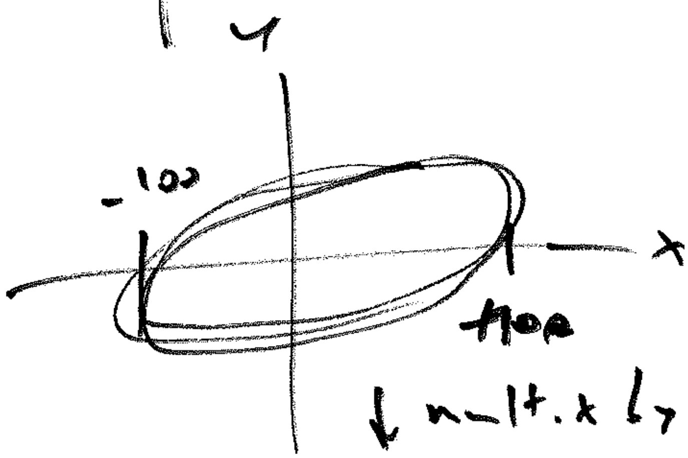


add 50 to x



add 10 to y





mult. x by 2