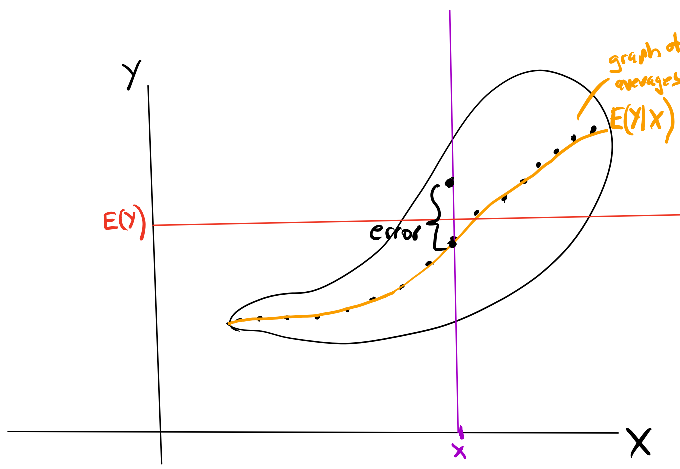
Stat 134 lec 39

Pre 6.5 Predicting Y from X.



In this picture x, y aren't in standard units

Properties (from lec 34)

① $E(Y) = E(E(Y|X))$ iterated expectation

② $E(aY+b|X) = aE(Y|X)+b$

③ $E(Y+Z|X) = E(Y|X)+E(Z|X)$

④ $E(g(X)|X) = g(X)$

⑤ $E(g(X)Y|X) = g(X)E(Y|X)$

⑥ $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$

total variance decomposition (see 6.2.18)

tot var = avg( ) + — var of $E(Y|X)$ (between group variance)

within grp variance.

In this picture x, y aren't in standard units

Note
$E(E(Y|X)) = E(Y)$

Last time:

Predictor 1: use $E(Y)$ to predict Y for every x.

$mse = E(error^2)$ is $Var(Y)$

Predictor 2: Use $E(Y|X)$ to predict Y for every x.

$$mse = E(error^2) = E\left((Y - E(Y|X))^2\right)$$

$$error|X=x = Y|X=x - E(Y|X=x)$$

$$E(error|X) = E(Y|X) - \underbrace{E(E(Y|X))}_{= E(Y)}$$

$$E(error) = E(E(error|X)) = E(E(Y|X)) - E(E(Y)) = \boxed{0}$$

$$mse = E(error^2) = var(error).$$

Compare mse predictor 2 $= var(error)$ to $Var(Y)$ ← mse predictor 1

$$Var(Y) = E\left((Y - E(Y))^2\right)$$

$$Var(Y|X=x) = E\left(\left(\underbrace{Y|X=x - E(Y|X=x)}_{error|X=x}\right)^2\right)$$

$$= E(error^2 | X = x)$$

so

$$E(Var(Y|X)) = E(E(error^2|X)) = E(error^2) = var(error)$$

Have $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$ by variance decomp formula

$Var(Y)$ = mse predictor 1, $E(Var(Y|X))$ = mse predictor 2 = mse predictor 2, $Var(E(Y|X)) \geq 0$

$\Rightarrow$ mse pred 2 $\leq$ mse pred 1.

Claim $E(Y|X)$ is the least square estimate of Y among functions of X.

Proof

Let $h(X)$ be an arbitrary function of X.

We need to compute $E((\overbrace{Y-h(X)}^{\text{error}})^2)$

and show it is greater than $E((\overbrace{Y-E(Y|X)}^{\text{error}})^2)$.

write $Y-h(X) = \overbrace{Y-E(Y|X)}^{A} + \overbrace{E(Y|X)-h(X)}^{B}$

$$E((Y-h(X))^2) = E((A+B)^2)$$

$$= E(A^2) + E(B^2) + 2E(AB)$$

$$E(A^2) = E((Y-E(Y|X))^2)$$

$$E(B^2) \geq 0$$

property of iterated expectation

$$E(AB) = E(E(AB|X))$$

Note that $E(AB|X) = BE(A|X)$ since B is a function of X not Y.

so:

$$E(AB) = E(B\underbrace{E(A|X)}_{=0}) = 0$$

since $E(A|X)$
$= E((Y-E(Y|X))|X)$
$= E(Y|X - E(Y|X)) = E(Y)-E(Y) = 0$

So

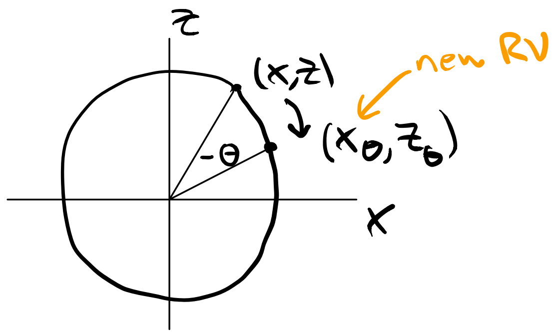
$$E((Y-h(X))^2) \geq E((Y-E(Y|X))^2)$$

$\square$

Sec 6.5 Bivariate normal

Here is a recipe how to make two correlated std normals!

let X, Z iid standard normals



We saw in lecture 30

$$X_\theta = \underbrace{\cos\theta}_{\rho} X + \underbrace{\sin\theta}_{\sqrt{1-\rho^2}} Z$$

call this Y

$$Y = \rho X + \sqrt{1-\rho^2}\, Z \sim N(0,1)$$

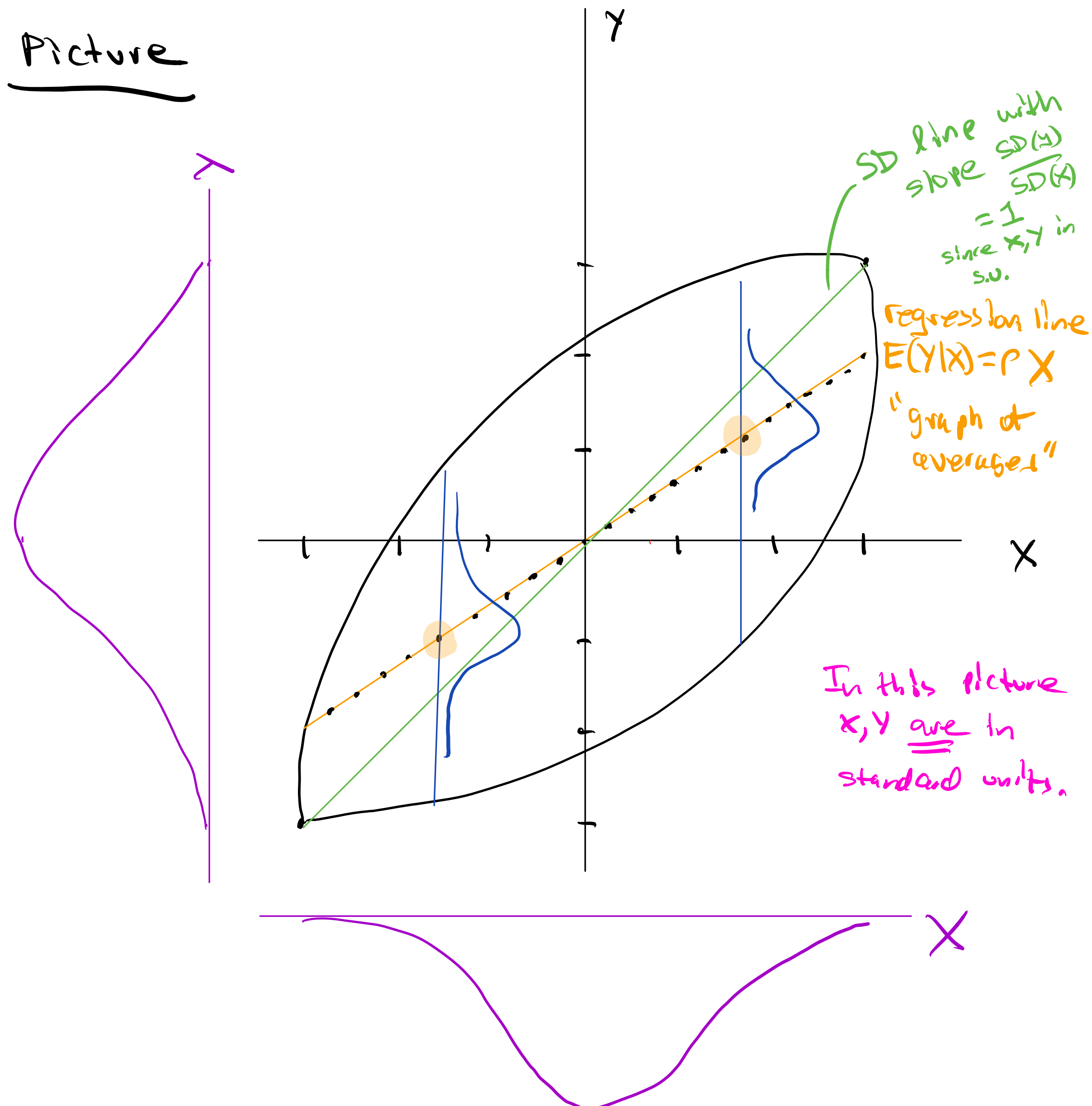
$$\begin{aligned}
Cov(X,Y) &= Cov(X, \rho X + \sqrt{1-\rho^2}\, Z) \\
&= Cov(X, \rho X) + Cov(X, \sqrt{1-\rho^2}\, Z) \\
&= \rho \underbrace{Var(X)}_{1} + \sqrt{1-\rho^2}\, \underbrace{Cov(X,Z)}_{0} \\
&= \boxed{\rho}
\end{aligned}$$

$$\Rightarrow Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = \boxed{\rho}$$ correlation coeff.

$$\begin{aligned}
E(Y|X) &= E(\rho X + \sqrt{1-\rho^2}\, Z | X) \\
&= E(\rho X | X) + E(\sqrt{1-\rho^2}\, Z | X) \\
&= \rho X + \sqrt{1-\rho^2}\, \underbrace{E(Z)}_{0} \\
&= \rho X
\end{aligned}$$

since X, Z indep.

Picture



$E(Y|X)$ is best prediction of Y among all functions $h(X)$. In the case where X, Y bivariate normal, it is called the regression line.

$$\mathrm{Var}(Y|X) = \mathrm{Var}\left(\rho X + \sqrt{1-\rho^2}\, Z \mid X\right)$$
$$= \mathrm{Var}\left(\sqrt{1-\rho^2}\, Z \mid X\right)$$
$$= (1-\rho^2)\,\underbrace{\mathrm{Var}(Z|X)}_{\mathrm{Var}(Z)=1}$$

$$\Rightarrow \boxed{SD(Y|X) = \sqrt{1-\rho^2}}$$

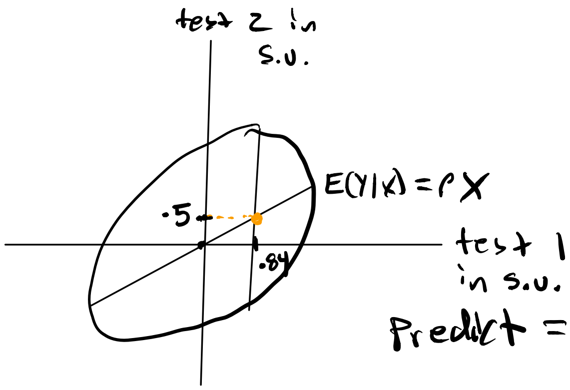
We have $Y = \rho X + \sqrt{1-\rho^2}\, Z$

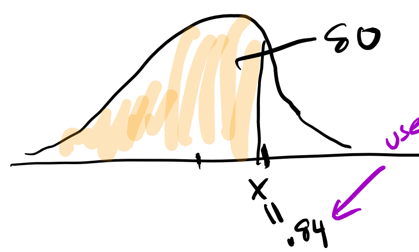
so $\underbrace{Y|X=x = \rho x + \sqrt{1-\rho^2}\, Z}$ is normal

we have

$$Y|X=x \sim N(\rho x, 1-\rho^2)$$

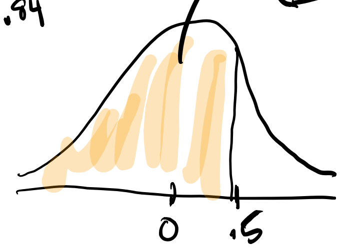
$\underline{\text{ex}}$ Two test, scores are bivariate normal $\rho = .6$. Given 80^{th} percentile on one test, estimate the percentile rank on the other.





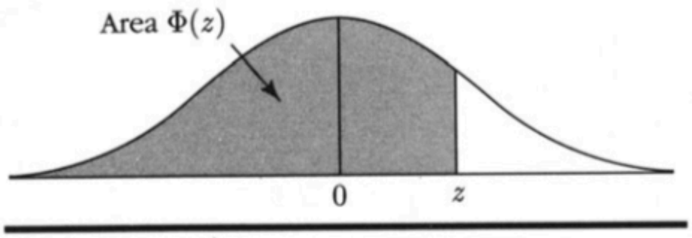
use normal table on next page

$\Phi(.5) = \boxed{.69}$



Predict $= .6(.84) = .5$

.5 for test 2 is only 69^{th} percentile!



Appendix 5
Normal Table

Table shows values of $\Phi(z)$ for z from 0 to 3.59 by steps of .01. Example: to find $\Phi(1.23)$, look in row 1.2 and column .03 to find $\Phi(1.2 + .03) = \Phi(1.23) = .8907$. Use $\Phi(z) = 1 - \Phi(-z)$ for negative z.

	.0	.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
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
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
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
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
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Regression effect

The regression line has slope $-1 \le \rho \le 1$ compared with the SD line which has slope 1. For a fixed x you predict $y = \rho x$ which will be less than x if $x > 0$ and greater than x if $x < 0$. This means that if you do really well on a midterm (at least greater than 50^{th} percentile — so $x > 0$ in s.u.) then you won't do as well on the final relative to the class (i.e you won't go all the way up to the SD line. The opposite is true however if you do poorly on the midterm.

Picture