- last time - S_{xx}, S_{yy}, S_{xy}, $b = \frac{S_{xy}}{S_{xx}}$, $a = \bar{y} - b\bar{x}$, $\hat{y} = a + bx$

- exponential : $y = \alpha \beta^{y} + \varepsilon$
- power : $y = \alpha x^{\beta} + \varepsilon$
- reciprocal : $y = \frac{1}{\alpha + \beta x} + \varepsilon$
- Polynomial repression : $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_p x^p + \varepsilon$

APRIL 9TH
APPLIED FINAL

---

we have two random variables :, x, y

we take a random sample $(x_1, y_1) \ldots (x_n, y_n)$

The <u>sample correlation coefficient</u> r, is a number, $-1 \le r \le 1$

Describing the direction and strength of the linear relationship between the variables.

$r > 1$: the points are all on a line with a positive slope.

$r = 1$

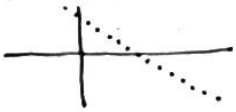
$r = 1$

If r is near 1, there is a strong positive linear relationship

As x increases, y tends to increase

Strong, positive linear relationship

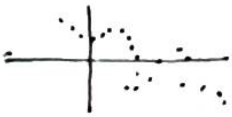
If $r = -1$, the points are all on a line with negative slope



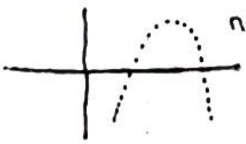
$r = -1$

If r is near -1, there is a strong negative linear relationship

As x increases, y tends to decrease



If r is near $\emptyset$, there is a nonlinear or rectilinear or no relationship between the variables.



nonlinear relationship

$\bar{x}$ = mean of x_i, $\bar{y}$ = mean of y_i

S_x = standard deviation of x_i

S_y = standard deviation of y_i

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S_x} \right) \left(\frac{y_i - \bar{y}}{S_y} \right)$$

- e.g.

x_i	1	2	3
y_i	1	3	8

$\bar{x} = \frac{1+2+3}{3} = 2$; $\bar{y} = \frac{1+3+8}{3} = 4$

$S_x^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3-1} = 1$; $S_x = 1$

$S_y^2 = \frac{(1-4)^2 + (3-4)^2 + (8-4)^2}{3-1} = 13$, $S_y = \sqrt{13}$

$\sigma = \left(\frac{1}{3-1}\right)\left(\frac{1}{1\sqrt{13}}\right)\left((1-2)(1-4) + (2-2)(3-4) + (3-2)(8-4)\right) = \frac{1}{2\sqrt{13}}(3+0+4) = \frac{7}{2\sqrt{13}}$

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

$S_x = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{S_{xx}}{n-1}}$

$S_y = \sqrt{\frac{\Sigma(y_i-\bar{y})^2}{n-1}} = \sqrt{\frac{S_{yy}}{n-1}}$

$r = \left(\frac{1}{n-1}\right)\left(\sqrt{\frac{n-1}{S_{xx}}}\right)\left(\sqrt{\frac{n-1}{S_{yy}}}\right)(S_{xy}) \Rightarrow \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

- e.g.

x_i	1	2	3
y_i	8	6	1

$\bar{x} = \frac{1+2+3}{3} = 2$; $\bar{y} = \frac{8+6+1}{3} = 5$

$S_{xx} = (1-2)^2 + (2-2)^2 + (3-2)^2 = 2$

$S_{yy} = (8-5)^2 + (6-5)^2 + (1-5)^2 = 26$

$S_{xy} = (1-2)(8-5) + (2-2)(6-5) + (3-2)(1-5) = -7$

$r = \frac{(-7)}{\sqrt{2 \cdot 26}} \cong -0.97$

Let ρ be the <u>population correlation coefficient.</u>

we will test $H_0 : \rho \leq 0$, $H_1 : \rho > 0$; $H_0\ \rho \geq 0$, $H_1\ \rho < 0$; $H_0 : \rho = 0$, $H_1 : \rho \neq 0$

Let $\gamma = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$ Then $Z = \sqrt{n-3}\,\gamma$ is standard normal

Perform Z tests as normal ($Z > Z_\alpha$, $Z < -Z_\alpha$, $|Z| > Z_{\alpha/2}$)

- e.g. Suppose we have a sample of size 10 and got a sample correlation coefficient of $r = .732$. Test the hypothesis at a .01 level of significance.

$H_0 : \rho = 0$
$H_1 : \rho \neq 0$

$\gamma = \frac{1}{2}\ln\left(\frac{1+.732}{1-.732}\right) = .933$

$Z = \sqrt{10-3}\,\gamma = 2.47$

We will reject the null hypothesis if $|Z| > Z_{.005}$

But $Z_{.005} = 2.575$

As $|Z| \ngtr Z_{.005}$, we cannot reject the null hypothesis at a .01 level of significance.