correlation.

ρ : greek letter "rho".

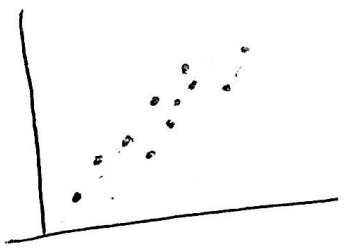
ρ: discribes the linear correlation between two variables.

for example. income and expenses.

(x_1, y_1), (x_2, y_2) ... (x_n, y_n)

x: income
y: expenses

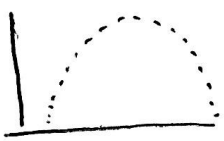
scatter plot



ρ close 1 : there is a positive linear association

ρ close -1 : there is a negative linear correlation

ρ close 0 : there is not a linear correlation.



test of hypothesis for ρ.

step 1-3 : $H_0: \rho = 0$ $H_1: \rho \neq 0$

step 4 : $\alpha = P(\text{type I error})$

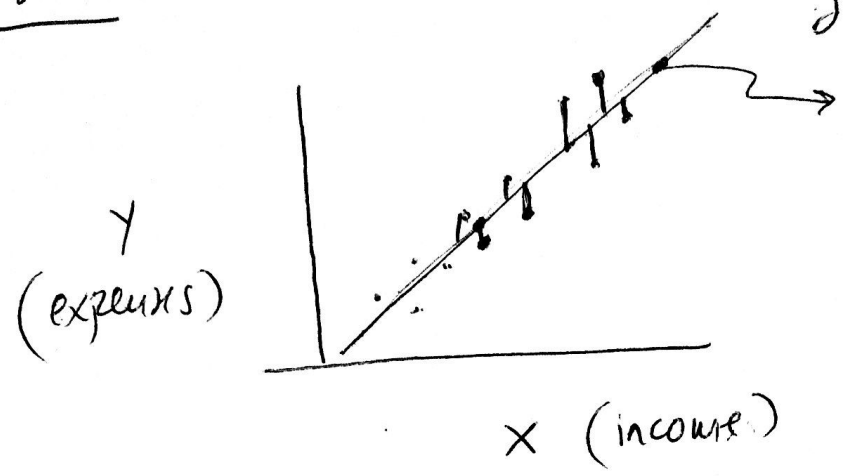
step 5 : $t^{stat} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$ follows a student t distribution with $n-2$ degrees of freedom.

step 6 : we reject H_0 if $t^{stat} > t_{\frac{\alpha}{2}}$, $t^{stat} > 0$

$t^{stat} < -t_{\frac{\alpha}{2}}$, $t^{stat} < 0$.

step 7-8 : make a decision to reject or fail to reject provide a simple non technical interpretation of decision.

regression.



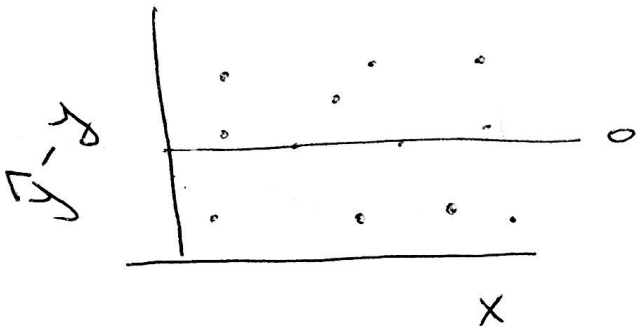
$$y = \beta_0 + \beta_1 X$$

$$\hat{y} = b_0 + b_1 X$$

$$b_0 = \bar{y} - b_1 \bar{X}$$

$$b_1 = r \frac{S_y}{S_x}$$

residuals plot



your regression equation satisfies the requirements needed to find $\hat{y}$.

then you can predict values for y.

$$\hat{y} = b_0 + b_1 X$$

practice exercises.

Exercise 1.

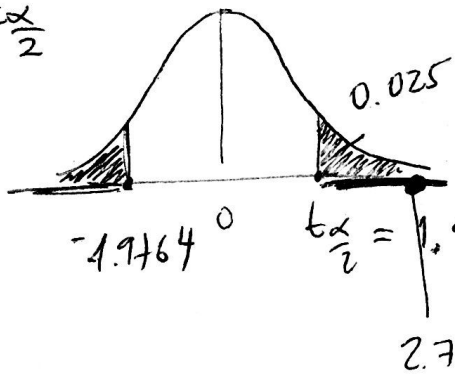
$n = 147$. $r = 0.221$ $b_0 = 4.06$ $b_1 = 0.0345$.

a) step 1-3: $H_0: \rho = 0$ $H_1: \rho \neq 0$.

step 2: $\alpha = 0.05$. $(0.01, 0.1)$

step 5: $t^{stat} = \dfrac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \dfrac{0.221}{\sqrt{\frac{1-0.221^2}{147-2}}} = 2.7286$.

step 6: $t_{\frac{\alpha}{2}}$



student t distribution with $n-2 = 145$ degrees of freedom

step 7: because $t^{stat} = 2.7286 > t_{\frac{\alpha}{2}} = 1.9764$ we decide to reject the null hypothesis.

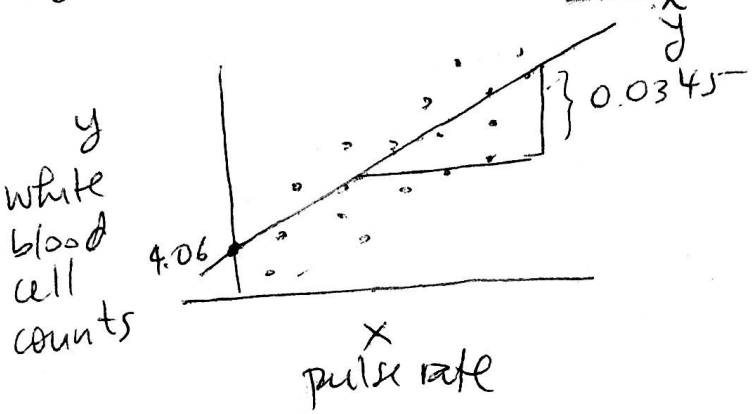
step 8: there is enough evidence to conclude that there is a linear association between pulse rate and white blood cell counts.

⊛ in this test we could be making a type I error, which is to reject the null hypothesis when the null hypothesis is actually true.

0.05 is the possibility of this error.

b) x: pulse rate
y: white blood cell count
repression equation is

$$\hat{y} = b_0 + b_1 x = 4.06 + 0.0345 x$$



- for a female with pulse rate equal to zero. the white blood cell counts is 4.06. estimates.
- if the pulse rate increases in one unit. the white blood cell counts estimate increases 0.0345 units

c)

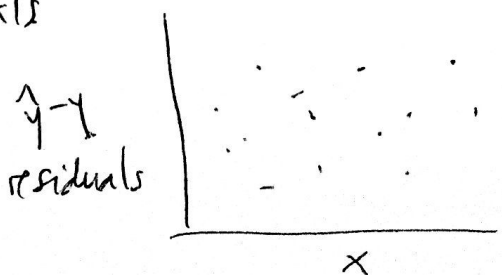
d) prediction for $x = 80$.

i) if the repression equation is a good model then the prediction is

$$\hat{y} = 4.06 + 0.0345 \cdot 80 = 6.82$$

ii) if the repression equation is not a good model. then the prediction is $\hat{y} = \bar{y}$.

e) we decide if the repression model is good or not based on the plot of residuals



2] $n = 436 \quad \bar{x} = 3.97 \quad s = 0.55$.

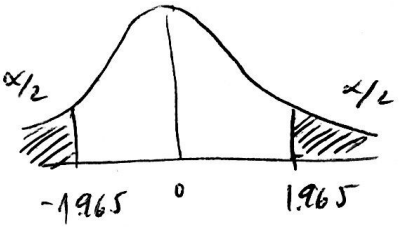
a) $H_0: \mu = 4.0 \quad H_1: \mu \neq 4.0$.

$\alpha = 0.05$

$t^{stat} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3.97 - 4.0}{0.55/\sqrt{436}} = -1.1389$, follows

Student t distribution with $n-1 = 435$ degrees of freedom.

$t_{\frac{\alpha}{2}} = 1.965$. rejection regions are



because t^{stat} is between -1.965 and 1.965 we fail to reject the null hypothesis.

so, there is not enough evidence to warrant rejection of the claim that the population of student course evaluations has a mean of 4.0.

b) the type of error is the type II error, this is, to fail to reject the null hypothesis when the null hypothesis is false.

c) no, it applies to students from the University of Texas at Austin, where the sample was obtained from.

d) a 95% confidence interval is

$\bar{x} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} = 3.97 - 1.966\frac{0.55}{\sqrt{436}} = 3.7182 < \mu < 3.97 + 1.966 \cdot \frac{0.55}{\sqrt{436}} = 4.0217$

note that $\mu = 4.0$ is contained in the confidence interval, so which agrees with the conclusion from the test of hypotheses