Independent vs. dependent samples

Matched pairs

Stat 943
4-21-15

The primary sampling unit is a pair

$H_0: \mu_d = 0$	$H_0: \mu_d \leq 0$	$H_0: \mu_d \geq 0$
$H_1: \mu_d \neq 0$	$H_1: \mu_d > 0$	$H_1: \mu_d < 0$

Analysis consists of a 1-sample t-test on the differences

Reaction time example

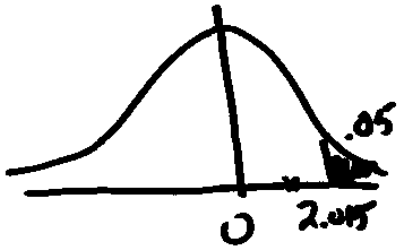
	Left	Right	d = L-R
1	.582	.408	.174
2	.481	.404	.077
3	.841	.542	.299
4	.267	.402	-.135
5	.685	.456	.229
6	.450	.533	-.083

Construct a test to show that times are faster with the right hand.

$\alpha = .05$

$H_0: \mu_d \leq 0$

$H_1: \mu_d > 0$



invT(.95,5)

Test stat = 1.32

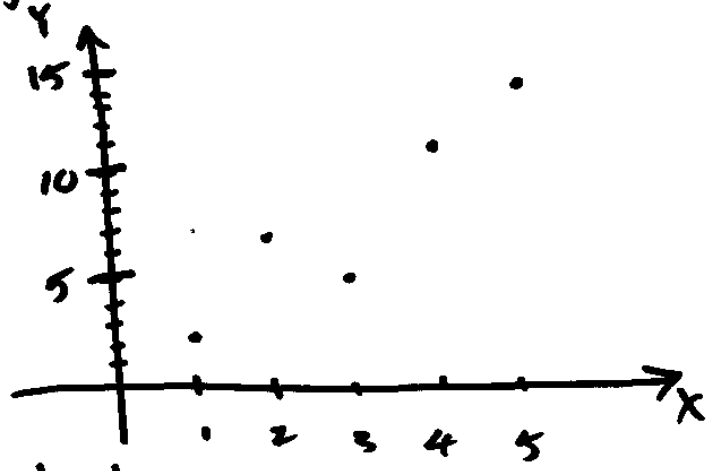
pval = .122 Fail to reject H_0

Regression

① Is there a trend?
② Is it increasing? decreasing?
③ Is it linear?

Simple linear regression

X	Y
1	3
2	7
3	5
4	11
5	14



X = # yrs Coffee shop has been open

Y = # employees

④

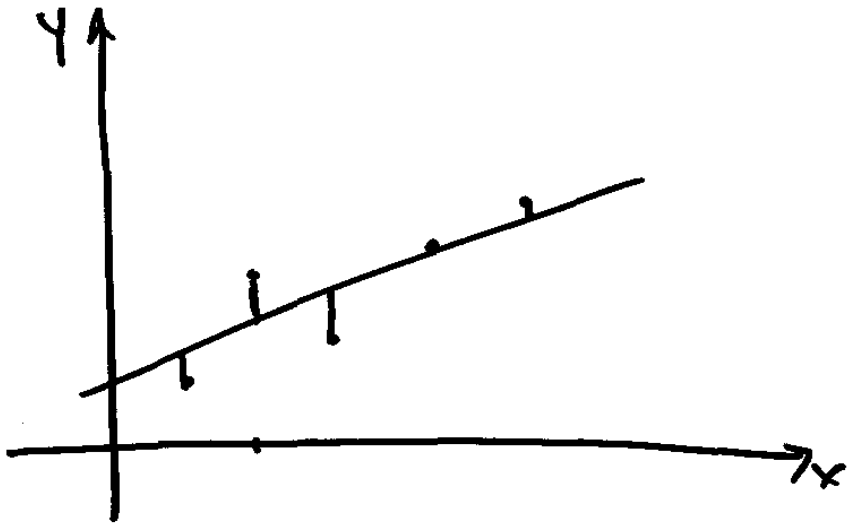
Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

- β_0: intercept
- β_1: slope
- ε_i: unexplained error

$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$ is also a linear model because it is a linear combination of the parameters

$y_i = \beta_0 e^{\beta_1 x_i} + \varepsilon_i$ is not a linear model

(5)



Least squares method: minimizes the sum of squares of the vertical errors.

(6)

From the calculator (LinRegTTest)

$a = .2 \quad b = 2.6$

$$\hat{y} = .2 + 2.6x$$

Note: this line will always pass through the center of mass of the data $(\bar{x}, \bar{y})$

Coeff of determination

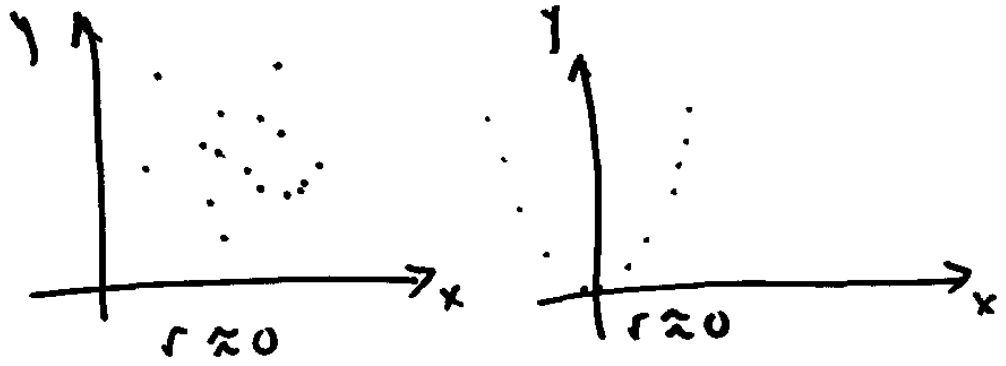
$r^2 = .845$

$r = .919$ correlation

$s = 2.033$ is an estimate of the standard deviation of the vertical errors

(7)

$-1 \leq r \leq 1$ 1 and -1 are perfect



r measures the linear relationship between y and x (straight line)

(8)

$0 \leq r^2 \leq 1$

r^2 is the proportion of the total variation in the y-values that is explained by x.