

## CIMT Statistics p222 Example: Worked Solution for Hypothesis test on $\rho$

Assumptions: a linear model fits the population;  
pairs of observations are independent;  
data are samples from normal distributions

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Assume  $H_0$  to be true

$\alpha = 5\%$ , two tailed test

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 1060.1 - \frac{108^2}{12} = 88.1$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 3396942 - \frac{6372^2}{12} = 13410$$

$$S_{xy} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} = 56825.4 - \frac{108 \times 6372}{12} = -522.6$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{-522.6}{\sqrt{88.1 \times 13410}} = -0.480803$$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0.480803\sqrt{12-2}}{\sqrt{1-(-0.480803)^2}} = -1.73401$$

Under  $H_0$ , the test statistic,  $t$ , has a  $t_{12-2}$  distribution

$$p\text{-value} = 2 \times P(t_{10} < -1.73401)$$

$$= 2 \times 0.056789$$

$$= 0.113577$$

$$> 0.05$$

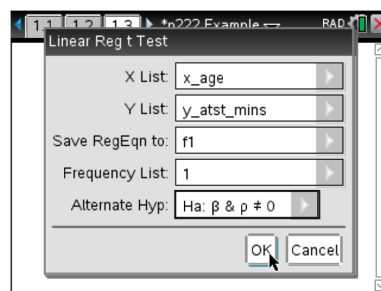
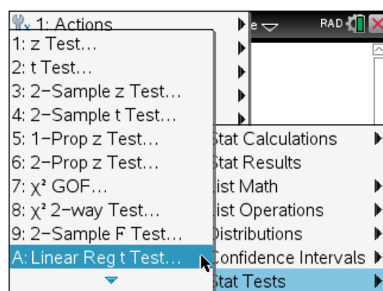
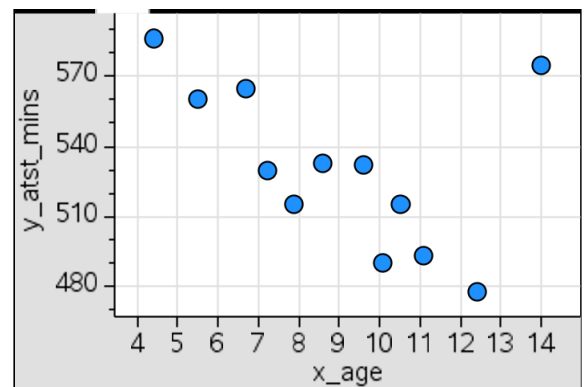
Hence we conclude that we do not have sufficient evidence to reject  $H_0$  and therefore there is no obvious linear correlation between age and average total sleep time.

We shall look at the scatterplot to see whether this seems reasonable.

We appear to have an outlier (Child I)

If we **remove Child I** from the sample, and then re-perform the test, we obtain a different result. With the raw data available, this re-calculation can be performed very quickly using the TI-Nspire:

$s_{xx} = 1060.1 - \frac{108^2}{12}$	88.1
$s_{yy} = 3396942 - \frac{6372^2}{12}$	13410
$s_{xy} = 56825.4 - \frac{108 \cdot 6372}{12}$	-522.6
$r = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}}$	-0.480803
$t = \frac{r \cdot \sqrt{12-2}}{\sqrt{1-r^2}}$	-1.73401
$t\text{Cdf}(-9.999, t, 10)$	0.056789
$p\_val = 2 \cdot t\text{Cdf}(-9.999, t, 10)$	0.113577



LinRegTTest x_age, y_atst_mins, 1, 0: CopyVar	
"Title"	"Linear Reg t Test"
"Alternate Hyp"	" $\beta \neq 0$ "
"RegEqn"	" $a + b \cdot x$ "
"t"	-7.03796
"PVal"	0.000061
"df"	9
"a"	634.136
"b"	-12.5371
"s"	13.8932
"SESlope"	1.78136
"r^2"	0.84624
"r"	-0.919913
"Resid"	"{...}"

Now, the p-value is well under 5% and we have strong evidence to suggest that there is a linear correlation between age and average total sleep time, once outlying data have been removed.