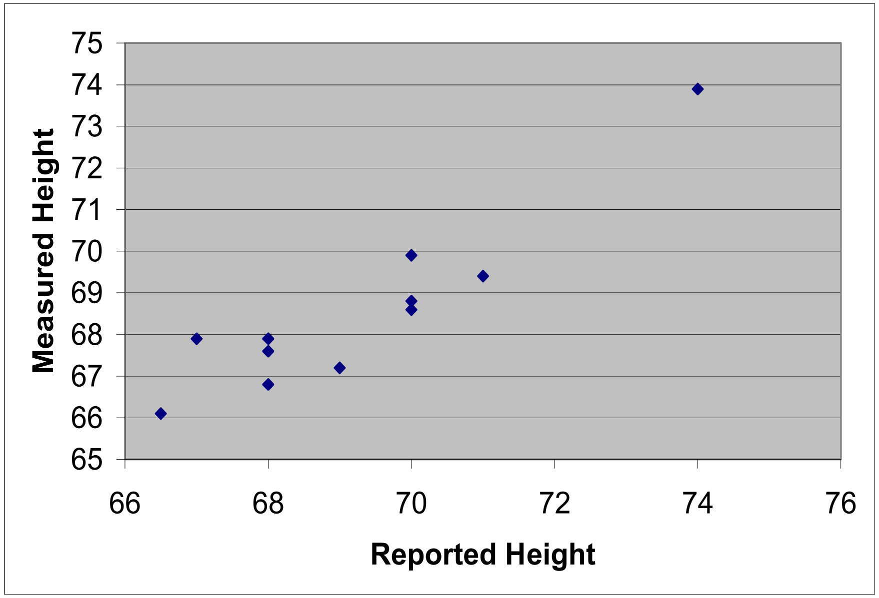

<u>Correlation—jump ahead to Chapter 10-2</u>

Correlation is a measure of the linear association between two variables, x and y.

Pearson product moment coefficient of correlation, $\boldsymbol{r}$, is a measure of the strength of the linear relationship between two variables x and y. It is computed for a ***<u>sample</u>*** of n measurements on x and y as:

$$r = \frac{1}{n-1}\sum_i \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

Can also be written as

$$r = \frac{\mathrm{SS}_{xy}}{\sqrt{\mathrm{SS}_{xx}\mathrm{SS}_{yy}}}$$

$$\mathrm{SS}_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

$$\mathrm{SS}_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

$$\mathrm{SS}_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}$$

In Excel use:
=Pearson(A1:An,B1:Bn)

1. r is positive when the slope is positive and likewise negative when the slope is negative

2. If $SS_{xy} = 0$, then $r = 0$ ➔ that there is no association between the magnitudes of the two variables; or, a change in the magnitude of one variable does not imply a change in the magnitude of the other variable

3. The correlation coefficient, r, is unitless and assumes a value between –1 and +1, regardless of the units of x and y

4. The $\text{Cor}(X,Y) = \text{Cor}(Y,X)$

For the reported/measured height data, r is strongly positive, indicating that reported height tends to increase as measured height increases---***for this sample of 11 individuals***

Interpretations of an observed association

1. Correlation does not prove causation, although it is possible that there is causation.

2. There may be causation, but confounding factors contribute as well, making this causation difficult to prove.

3. There is no causation. The association is explained by how the explanatory and response variables are both affected by other variables.

4. The response variable is causing a change in the explanatory variable.

Population Correlation Coefficient

Denoted by ρ (rho)

estimated by the sample statistic, r

Hypothesis test for ρ given on page 527—we will not specifically go over this

Reported	Measured	Reported-Measured
68	66.8	1.2
74	73.9	0.1
66.5	66.1	0.4
69	67.2	1.8
68	67.9	0.1
71	69.4	1.6
70	69.9	0.1
70	68.6	1.4
67	67.9	-0.9
68	67.6	0.4
70	68.8	1.2

Correlation	0.923
Covariance	4.086

Mean	69.227	68.555	0.673
SD	2.114	2.094	0.826
Var	4.468	4.387	**0.682**
SE	**0.897**		**0.249**

Let R = reported height; M = measured height

Var(R) + Var(M) - 2(Cov(R,M)) = **0.682**

$$\operatorname{cor}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} \Rightarrow \operatorname{cov}(X,Y) = \operatorname{cor}(X,Y)(\sigma_X \sigma_Y)$$

$H_0 : \mu_R = \mu_M$ versus $H_1 : \mu_R > \mu_M$

$H_0 : \mu_R - \mu_M = 0$ versus $H_1 : \mu_R - \mu_M > 0$

t-Test: Two-Sample Assuming Unequal Variances

	Reported	*Measured*
Mean	69.227	68.555
Variance	4.468	4.387
Observations	11.000	11.000
Hypothesized Mean Difference	0.000	
df	20.000	
t Stat	0.750	
P(T<=t) one-tail	0.231	
t Critical one-tail	1.725	
P(T<=t) two-tail	0.462	
t Critical two-tail	2.086	

$H_0 : \mu_d = 0$ versus $H_1 : \mu_d > 0$

t-Test: Paired Two Sample for Means

	Reported	*Measured*
Mean	69.227	68.555
Variance	4.468	4.387
Observations	11.000	11.000
Pearson Correlation	0.923	
Hypothesized Mean Difference	0.000	
df	10.000	
t Stat	2.701	
P(T<=t) one-tail	0.011	
t Critical one-tail	1.812	
P(T<=t) two-tail	0.022	
t Critical two-tail	2.228	