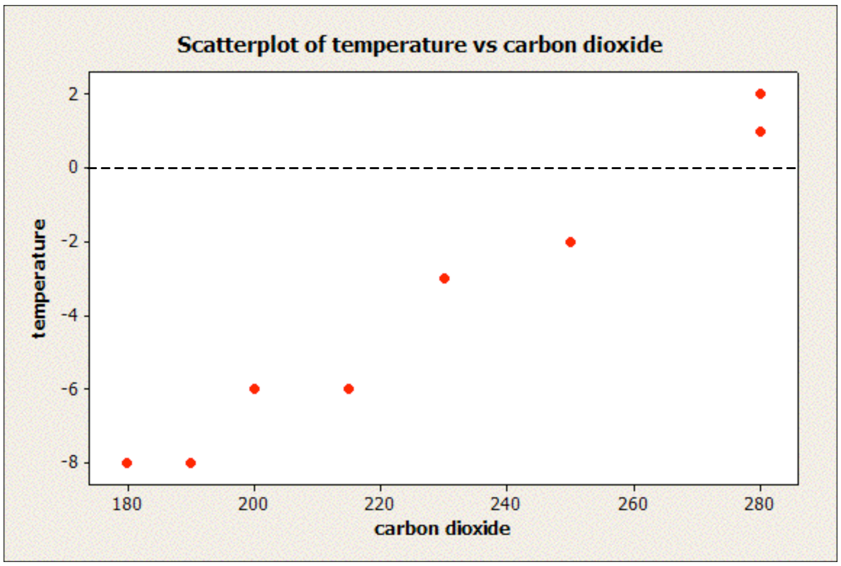
Correlation

Example: Correlation between temperature (measured as change from present) and carbon dioxide concentration in the atmosphere over the past 400,000 years.

Data: Example taken from Groom et al. Principles of Conservation Biology; they state that "there is a strong correlation between greenhouse gas abundance and mean temperatures". Using data from ice cores (bubbles of air trapped in deep ice cores preserve air from the past – can be then measured for atmospheric CO_2 levels, and use ratio of two isotopes of oxygen to calculate atmospheric temperatures at the time the ice was formed), can look at relationship between CO_2 levels in the atmosphere and the mean global temperature.

Some data from the Vostok ice core:



1. Sample correlation coefficient (r) and SE_r

Carbon dioxide	Temperature					
X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$(X-\bar{X})(Y-\bar{Y})$
280	2	51.875	5.75	2691.02	33.0625	298.281
180	-8	-48.125	-4.25	2316.02	18.0625	204.531
250	-2	21.875	1.75	478.52	3.0625	38.281
200	-6	-28.125	-2.25	791.02	5.0625	63.281
190	-8	-38.125	-4.25	1453.52	18.0625	162.031
230	-3	1.875	0.75	3.52	0.5625	1.406
215	-6	-13.125	-2.25	172.27	5.0625	29.531
280	1	51.875	4.75	2691.02	22.5625	246.406
Mean:	Mean:			Sum:	Sum:	Sum:
228.125	-3.75			1324.61	13.1875	130.469

$$r=\frac{\sum\left(X-\bar{X}\right)\left(Y-\bar{Y}\right)}{\sqrt{\sum\left(X-\bar{X}\right)^2}\sqrt{\sum\left(Y-\bar{Y}\right)^2}}=\frac{130.469}{\sqrt{1324.61}\sqrt{13.1875}}=\frac{130.469}{132.168}=0.987$$

$$SE_r=\sqrt{\frac{1-r^2}{n-2}}=\sqrt{\frac{1-\left(0.987\right)^2}{8-2}}=\sqrt{\frac{0.025831}{6}}=\sqrt{0.004305}=0.0656$$

2. Hypothesis test for the population correlation coefficient (ρ)

Step 1: Is there a correlation between carbon dioxide and temperature change?

Step 2: H_0: There is no correlation between carbon dioxide and temperature change ($\rho = 0$).
H_A: There is a correlation between carbon dioxide and temperature change ($\rho \neq 0$).

Steps 3 and 4: Collect a random sample and calculate descriptive statistics (see above).

Step 5: Determine how compatible data are with results expected under H_0 (calculate test statistic, compare to null distribution, and assign *P*-value).

test statistic for *t*-test of ρ: $t=\frac{r-0}{SE_r}$ $df = n-2$

for our example: $t=\frac{r-0}{SE_r}=\frac{0.987}{0.0656}=15.05$

With *t* distribution and df = 6, *P*-value < 0.0001

Step 6: Make a decision regarding H_0. *P*-value < α, we reject H_0.

Step 7: Conclusions.

With $\alpha = 0.05$, we have significant evidence ($t_6 = 15.05$, *P*-value < 0.0001) from 8 ice core samples that there is a positive correlation ($r = 0.987$, $SE_r = 0.0656$) between carbon dioxide concentration in the atmosphere and temperature change.