In [2]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

1) Exercise 5.1

2) Lab: Build a linear model

a. Collect some data. Each sample should consist of an independent variable x and a dependent variable (y_i). You should collect at least 10 samples, so $N \geq 10$.
b. Write out the linear model in terms of y_i, x_i, e_i, A, and B. Also write the error function (aka., objective function, loss function) you need to minimize in order to find A and B.
c. Derive the equations that minimize the error function (i.e., the normal equations).
d. Calculate the numerical values of A and B for your data.
e. Calculate SSE, $RegSS$, TSS, r^2, and the correlation between x and y.
f. Calculate $\Sigma_i \hat{y}_i e_i$ and $\Sigma_i x_y e_i$. Verify that they are equal to zero (see the equation on pg. 86 just below figure 5.3 and exercise 5.1).
g. Describe your model in words. Does x cause y? How could you make use of a prediction given by your model?

1) Exercise 5.1

From text $\underbrace{\sum E_i = 0}_{(i)}$; $\underbrace{\sum x_i E_i = 0}_{(ii)}$

a) $\sum \hat{Y}_i E_i = \sum (A + Bx_i) E_i = A \sum E_i + B \sum x_i E_i = 0 + 0 = 0$

b) $\sum (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})$
$= \sum E_i (\hat{Y}_i - \bar{Y}) = \sum E_i \hat{Y}_i - \bar{Y} \sum E_i = 0 + 0 = 0$

from (a)

2) Lab

How does my living room's temperature depend on my bedroom's temperature?

2a) I collected temperature readings from both rooms:

In [4]:
```
data = np.array([
 [67, 69],
 [65, 69],
 [67, 71],
 [67, 71],
 [69, 71],
 [71, 73],
 [69, 73],
 [68, 69],
 [68, 69],
 [69, 71],
])

x = data[:,0]
y = data[:,1]
N = len(x)
N
```

Out[4]:
```
10
```

The first columns is from the bedroom. The second is from the living room.

2b) Linear model

$$y_i = A + Bx_i + E_i$$

The error function is the sum of squared errors:

$$SSE = \Sigma_i E_i^2$$

2c

$$SSE(A,B) = \sum E_i^2$$
$$= \sum (Y_i - A - Bx_i)^2$$

Normal Eqns

$$\frac{\partial SSE(A,B)}{\partial A} = -2\sum (Y_i - A - BX_i) = 0 \quad (1)$$

$$\frac{\partial SSE(A,B)}{\partial B} = -2\sum x_i (Y_i - A - Bx_i) = 0 \quad (2)$$

$$(1) \Rightarrow \boxed{A = \bar{Y} - B\bar{x}} \quad (3)$$

sub (3) into (2)

$$\sum x_i (Y_i - \bar{Y} + B\bar{x} - Bx_i) = 0$$
$$= \sum X_i (Y_i - \bar{Y} - B(x_i - \bar{x})) = 0 \quad (4)$$

Recall $\sum E_i = 0$, so $\bar{x} \sum E_i = 0 = \bar{x} \sum [Y_i - \bar{Y} - B(x_i - \bar{x})]$ (5)

Subtract (5) from (4)

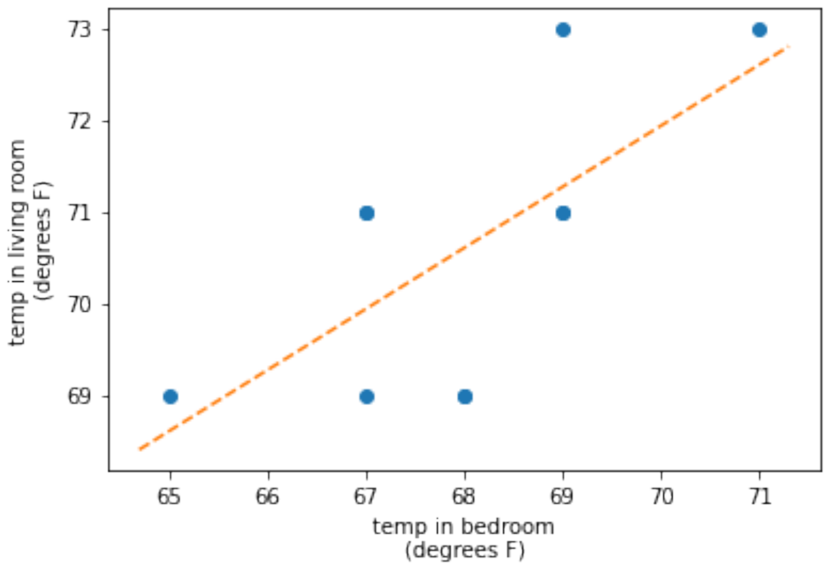
$$\sum (x_i - \bar{x}) (Y_i - \bar{Y} - B(x_i - \bar{x})) = 0$$
$$\sum (Y_i - \bar{Y})(x_i - \bar{x}) - B\sum (x_i - \bar{x})^2 = 0$$
$$\boxed{B = \frac{\sum (Y_i - \bar{Y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}}$$

2d

In [5]:
```
ym = y-y.mean()
xm = x-x.mean()
B = (xm*ym).sum() / (xm**2).sum()
A = y.mean() - B * x.mean()
print (A, B)
```

```
25.266666666666666 0.6666666666666666
```

In [15]:
```
plt.plot(x,y,'o')
c = plt.axis()
plt.plot([c[0],c[1]], [A+B*c[0], A+B*c[1]], '--');
plt.xlabel('temp in bedroom\n(degrees F)')
plt.ylabel('temp in living room\n(degrees F)')
plt.show();
```



2e

In [7]:
```
yhat = A + B*x
eps = y - yhat
TSS = ( (y-y.mean())**2 ).sum()
RSS = ( (y-yhat)**2 ).sum()
RegSS = TSS - RSS
R2 = RegSS / TSS
ym = y-y.mean()
xm = x-x.mean()
corr = (xm*ym).sum() / np.sqrt( (xm**2).sum() * (ym**2).sum() )
```

In [10]:
```
print (f"TSS = {TSS:.2f} RSS = {RSS:.2f} RegSS = {RegSS:.2f}")
print (f"R2 = {R2:.2f} corr**2 = {corr**2:.2f} corr = {corr:.2f}")
```

```
TSS = 22.40 RSS = 11.73 RegSS = 10.67
R2 = 0.48 corr**2 = 0.48 corr = 0.69
```

2f

In [13]:
```
E = (y-yhat)
print (f"(yhat * E).sum() = {(yhat * E).sum():.4f}")
print (f"(x * E).sum() = {(x * E).sum():.4f}")
```

```
(yhat * E).sum() = 0.0000
(x * E).sum() = 0.0000
```

2g

The model estimates the temperature of my living room as a function of the temperature of my bedroom.

My data is observational, so there's no evidence of causation.

If I'm feeling a little cold in my bedroom, I might consider moving to the living room to warm up, since the model always predicts warm temperatures in the living room.

In [ ]:
```

```