# Introduction to Earth System

Working with data

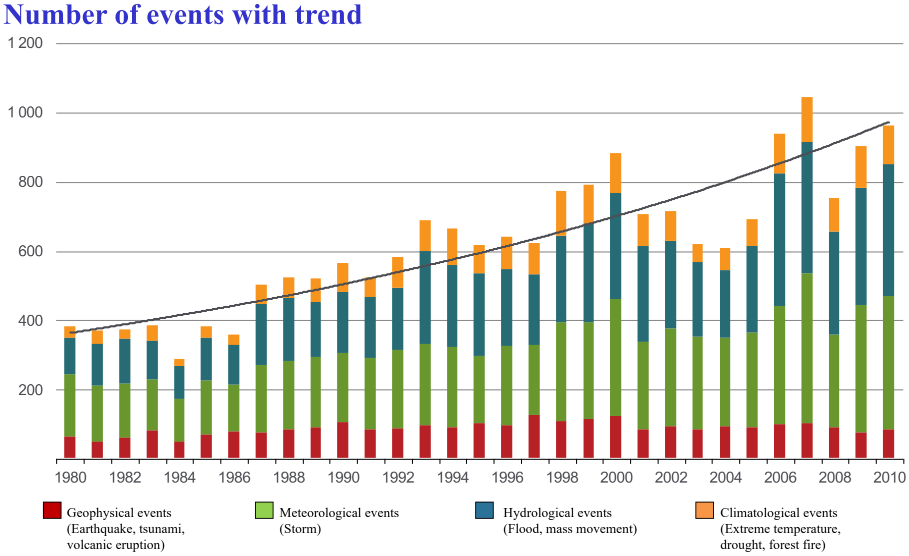
**Trend analysis**



1

## Natural Catastrophes Worldwide, 1980-2010

### Number of events with trend




Source: Geo Risks Research, NatCatSERVICE.



2

## Trend analysis

Purpose:

To determine if a series of observations of a random variable is generally increasing or decreasing with time (or to find a mathematic expression of a series in function of time)

Be careful with data with seasonal pattern



# I. Linear Trend

**Recall: Covariance**

$$cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Interpreting Covariance

| | |
|---|---|
| cov(X,Y) > 0 | X and Y are positively correlated |
| cov(X,Y) < 0 | X and Y are inversely correlated |
| cov(X,Y) = 0 | X and Y are independent |

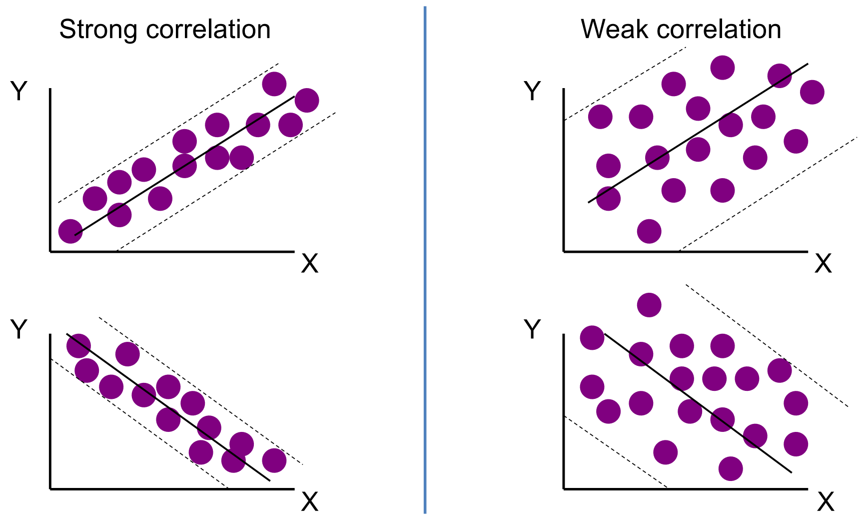
# Correlation coefficient

Pearson's Correlation Coefficient is standardized covariance:

$$r = \frac{cov(x,y)}{\sqrt{var(x) * var(y)}} = \frac{cov(x,y)}{s_x * s_y}$$

- ✓ Measures the relative strength of the *linear* relationship between two variables
- ✓ Unit-less
- ✓ between –1 and 1
- ✓ The closer to –1, the stronger the negative linear relationship
- ✓ The closer to 1, the stronger the positive linear relationship
- ✓ The closer to 0, the weaker any positive linear relationship





# Linear Regression Analysis…

- Regression analysis is used to predict the value of one variable (the **dependent variable**) on the basis of other variables (the **independent variables**).

- Dependent variable: **Y**
- Independent variables: $X_1, X_2, \ldots, X_k$

**Simple linear regression** $y = b_0 + b_1 x + \varepsilon$

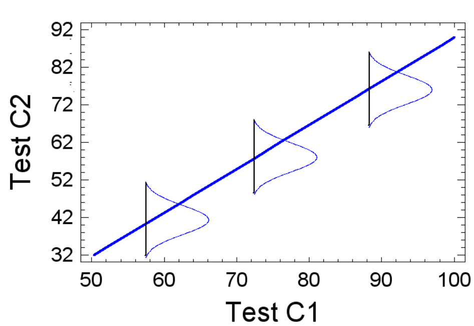


7

# Linear Regression Analysis

$y = b_0 + b_1 x + \varepsilon$

- Parameters:
- $b_0$ = Y-Intercept
- $b_1$ = Slope
  - > 0 [positive slope]
  - < 0 [negative slope]




**Assumptions:** The errors $\varepsilon$ are uncorrelated (i.e. **Independent**) in successive observations. The errors $\varepsilon$ are **Normally** distributed with mean 0 and variance $\sigma^2$. That is: $\varepsilon \sim N(0, \sigma^2)$

Theoretical Linear Model

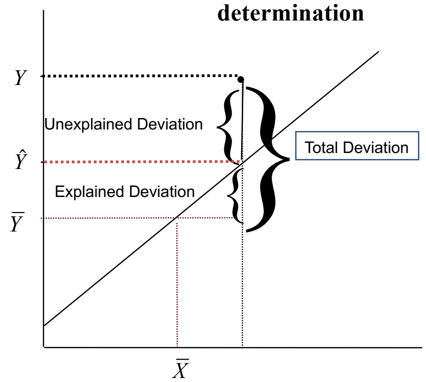


8

# How Good is the Regression?

The **coefficient of determination,** $R^2$, is a descriptive measure of the strength of the regression relationship, a measure how well the regression line fits the data.

**$R^2$: coefficient of determination**




$$(y-\bar{y}) = (y-\hat{y}) + (\hat{y}-\bar{y})$$

Total Deviation = Unexplained Deviation (Error) + Explained Deviation (Regression)

$$\sum (y-\bar{y})^2 = \sum (y-\hat{y})^2 + \sum (\hat{y}-\bar{y})^2$$

SST = SSE + SSR

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Percentage of total variation explained by the regression.



9

# some equations ...

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$$

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

$$b_1 = \frac{s_{xy}}{s_x^2}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$$

The sum of squares for error is:

$$SSE = (n-1)\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right)$$

Standard error of estimate

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}}$$



10

# Significant Testing for the Slope

If no linear relationship exists between the two variables → expect to have the regression line to be ***horizontal*** → a ***slope of zero***.

To test if there is a linear relationship, i.e. we want to see if $\beta_1 \neq 0$ → our research hypothesis becomes:

- $H_1$: $\beta_1 \neq 0$

→ the null hypothesis becomes:

- $H_0$: $\beta_1 = 0$



# Testing the Slope…

- We can implement this test statistic to try our hypotheses:
- $H_0$: $\beta_1 = 0$

$$t = \frac{b_1 - \beta_1}{s_{b_1}}$$

- Where $s_{b1}$ is the standard deviation of $b_1$, defined as:

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$$

- If the error variable is normally distributed, the test statistic has a **Student t-distribution** with n–2 degrees of freedom.

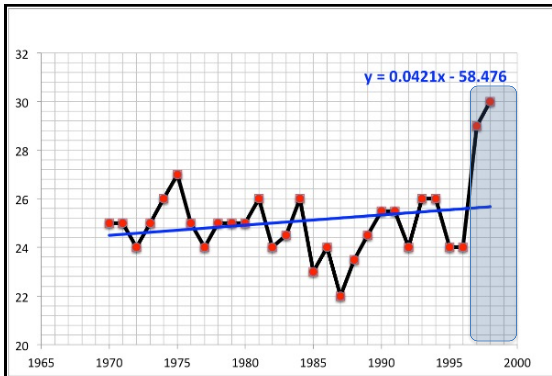
# t-table

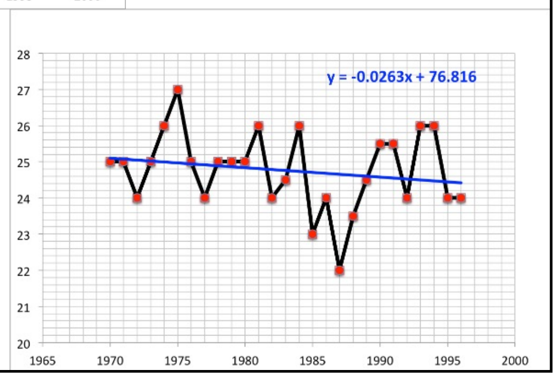
| cum. prob | $t_{.50}$ | $t_{.75}$ | $t_{.80}$ | $t_{.85}$ | $t_{.90}$ | $t_{.95}$ | $t_{.975}$ | $t_{.99}$ | $t_{.995}$ | $t_{.999}$ | $t_{.9995}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **one-tail** | **0.50** | **0.25** | **0.20** | **0.15** | **0.10** | **0.05** | **0.025** | **0.01** | **0.005** | **0.001** | **0.0005** |
| **two-tails** | **1.00** | **0.50** | **0.40** | **0.30** | **0.20** | **0.10** | **0.05** | **0.02** | **0.01** | **0.002** | **0.001** |
| **df** | | | | | | | | | | | |
| 1 | 0.000 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 31.82 | 63.66 | 318.31 | 636.62 |
| 2 | 0.000 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 | 31.599 |
| 3 | 0.000 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.215 | 12.924 |
| 4 | 0.000 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 0.000 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 0.000 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 0.000 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 0.000 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 0.000 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 0.000 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 0.000 | 0.697 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 0.000 | 0.695 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 0.000 | 0.694 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 0.000 | 0.692 | 0.868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 0.000 | 0.691 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 0.000 | 0.690 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 0.000 | 0.689 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 0.000 | 0.688 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 0.000 | 0.688 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 0.000 | 0.687 | 0.860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |
| 21 | 0.000 | 0.686 | 0.859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 0.000 | 0.686 | 0.858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 0.000 | 0.685 | 0.858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.768 |
| 24 | 0.000 | 0.685 | 0.857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 0.000 | 0.684 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725 |
| 26 | 0.000 | 0.684 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 0.000 | 0.684 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690 |
| 28 | 0.000 | 0.683 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 0.000 | 0.683 | 0.854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.659 |
| 30 | 0.000 | 0.683 | 0.854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 | 3.646 |
| 40 | 0.000 | 0.681 | 0.851 | 1.050 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551 |
| 60 | 0.000 | 0.679 | 0.848 | 1.045 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| 80 | 0.000 | 0.678 | 0.846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.374 | 2.639 | 3.195 | 3.416 |
| 100 | 0.000 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.364 | 2.626 | 3.174 | 3.390 |
| 1000 | 0.000 | 0.675 | 0.842 | 1.037 | 1.282 | 1.646 | 1.962 | 2.330 | 2.581 | 3.098 | 3.300 |
| ***z*** | 0.000 | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 |
| | 0% | 50% | 60% | 70% | 80% | 90% | 95% | 98% | 99% | 99.8% | 99.9% |
| | | | | | | **Confidence Level** | | | | | |



# Mann-Kendall test for a monotonic trend

- a non-parametric test for identifying trends in time series data.
- compare the relative magnitudes of sample data rather than the data values themselves
- One benefit of this test is that the data need not conform to any particular distribution








# Mann-Kendall test

Let $x_1$ , $x_2$ , … $x_n$ represent n data points where $x_j$ represents the data point at time j → the Mann-Kendall statistic (S) is given by

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign(x_j - x_k)$$

where :

$$sign(x_j - x_k) = 1 \text{ if } x_j - x_k > 0$$
$$= 0 \text{ if } x_j - x_k = 0$$
$$= \text{-}1 \text{ if } x_j - x_k < 0$$

- positive value of S → indicator of an increasing trend
- negative value of S → indicator of a decreasing trend

# Quick exercise #1

2m- daily temperature for the 6 consecutive days: {30, 31, 34, 31, 44, 31}

→ Estimate the Mann-Kendall statistic?



For the current data set:

`=sign($B3-$B$2)`

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1985 | 151 | | | | | | | | | | | | |
| 1986 | 151 | 0 | | | | | | | | | | | |
| 1987 | 147 | -1 | -1 | | | | | | | | | | |
| 1988 | 149 | -1 | -1 | 1 | | | | | | | | | |
| 1989 | 146 | -1 | -1 | -1 | -1 | | | | | | | | |
| 1990 | 142 | -1 | -1 | -1 | -1 | -1 | | | | | | | |
| 1991 | 143 | -1 | -1 | -1 | -1 | -1 | 1 | | | | | | |
| 1992 | 145 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | | | | | |
| 1993 | 141 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | | | | |
| 1994 | 143 | -1 | -1 | -1 | -1 | -1 | 1 | 0 | -1 | 1 | | | |
| 1995 | 145 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | 0 | 1 | 1 | | |
| 1996 | 138 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | |
| | | | | | | | | | | | | | *T* |
| | SUMS: | -10 | -10 | -7 | -8 | -7 | 2 | 0 | -3 | 1 | 0 | -1 | -43 |

Now, is $S = -43$ enough negatively large to show evidence for a trend?

**Need to statistically quantify the significance of the trend**

# Sen slope

$$Slope = median\left\{\frac{x_j - x_k}{j-k}\right\} \forall j > k$$

**Quick exercise #2:**

2m- daily temperature for the 4 consecutive days: {30, 31, 34, 31}

→ Estimate the Mann-Kendall statistic?
→ Estimate the Sen slope?



**Hypothesis testing**

→ Calculation of probability associated with the Mann-Kendall statistic S

$H_0$: No trend

$H_A$ : Negative/Positive monotonic trend

Variance of S:

$$VAR(S) = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{g} t_p(t_p - 1)(2t_p + 5)\right]$$

Where

- n is the number of data points,
- g is the number of tied groups (a tied group is a set of sample data having the same value);
- $t_p$ is the number of data points in the $p^{th}$ group

**Quick exercise #3:**

2m- daily temperature for the 5 consecutive days:

{30, 31, 34, 44, 31}

→? n, g, $t_1$, …, $t_g$

→? S, var(S)





Compute a normalized test statistic Z :

$$Z = \frac{S-1}{[VAR(S)]^{1/2}} \text{ if } S > 0$$
$$= 0 \text{ if } S = 0$$
$$= \frac{S+1}{[VAR(S)]^{1/2}} \text{ if } S < 0$$

→ $Z \in N(0,1)$

→ Decide on a probability level of significance (95% typically)

**Z-table**

This table gives a probability that a statistic is greater than *Z*.

- The label for rows contains the integer part and the first decimal place of *Z*.
- The label for columns contains the second decimal place of *Z*.
- The values within the table are the probabilities corresponding to the table type.

| *z* | +0.00 | +0.01 | +0.02 | +0.03 | +0.04 | +0.05 | +0.06 | +0.07 | +0.08 | +0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.49601 | 0.49202 | 0.48803 | 0.48405 | 0.48006 | 0.47608 | 0.47210 | 0.46812 | 0.46414 |
| 0.1 | 0.46017 | 0.45620 | 0.45224 | 0.44828 | 0.44433 | 0.44038 | 0.43640 | 0.43251 | 0.42858 | 0.42465 |
| 0.2 | 0.42074 | 0.41683 | 0.41294 | 0.40905 | 0.40517 | 0.40129 | 0.39743 | 0.39358 | 0.38974 | 0.38591 |
| 0.3 | 0.38209 | 0.37828 | 0.37448 | 0.37070 | 0.36693 | 0.36317 | 0.35942 | 0.35569 | 0.35197 | 0.34827 |
| 0.4 | 0.34458 | 0.34090 | 0.33724 | 0.33360 | 0.32997 | 0.32636 | 0.32276 | 0.31918 | 0.31561 | 0.31207 |
| 0.5 | 0.30854 | 0.30503 | 0.30153 | 0.29806 | 0.29460 | 0.29116 | 0.28774 | 0.28434 | 0.28096 | 0.27760 |
| 0.6 | 0.27425 | 0.27093 | 0.26763 | 0.26435 | 0.26109 | 0.25785 | 0.25463 | 0.25143 | 0.24825 | 0.24510 |
| 0.7 | 0.24196 | 0.23885 | 0.23576 | 0.23270 | 0.22965 | 0.22663 | 0.22363 | 0.22065 | 0.21770 | 0.21476 |
| 0.8 | 0.21186 | 0.20897 | 0.20611 | 0.20327 | 0.20045 | 0.19766 | 0.19489 | 0.19215 | 0.18943 | 0.18673 |
| 0.9 | 0.18406 | 0.18141 | 0.17879 | 0.17619 | 0.17361 | 0.17106 | 0.16853 | 0.16602 | 0.16354 | 0.16109 |
| 1.0 | 0.15866 | 0.15625 | 0.15386 | 0.15151 | 0.14917 | 0.14686 | 0.14457 | 0.14231 | 0.14007 | 0.13786 |
| 1.1 | 0.13567 | 0.13350 | 0.13136 | 0.12924 | 0.12714 | 0.12507 | 0.12302 | 0.12100 | 0.11900 | 0.11702 |
| 1.2 | 0.11507 | 0.11314 | 0.11123 | 0.10935 | 0.10749 | 0.10565 | 0.10383 | 0.10204 | 0.10027 | 0.09853 |
| 1.3 | 0.09680 | 0.09510 | 0.09342 | 0.09176 | 0.09012 | 0.08851 | 0.08692 | 0.08534 | 0.08379 | 0.08226 |
| 1.4 | 0.08076 | 0.07927 | 0.07780 | 0.07636 | 0.07493 | 0.07353 | 0.07215 | 0.07078 | 0.06944 | 0.06811 |
| 1.5 | 0.06681 | 0.06552 | 0.06426 | 0.06301 | 0.06178 | 0.06057 | 0.05938 | 0.05821 | 0.05705 | 0.05592 |
| 1.6 | 0.05480 | 0.05370 | 0.05262 | 0.05155 | 0.05050 | 0.04947 | 0.04846 | 0.04746 | 0.04648 | 0.04551 |
| 1.7 | 0.04457 | 0.04363 | 0.04272 | 0.04182 | 0.04093 | 0.04006 | 0.03920 | 0.03836 | 0.03754 | 0.03673 |
| 1.8 | 0.03593 | 0.03515 | 0.03438 | 0.03362 | 0.03288 | 0.03216 | 0.03144 | 0.03074 | 0.03005 | 0.02938 |
| 1.9 | 0.02872 | 0.02807 | 0.02743 | 0.02680 | 0.02619 | 0.02559 | 0.02500 | 0.02442 | 0.02385 | 0.02330 |
| 2.0 | 0.02275 | 0.02222 | 0.02169 | 0.02118 | 0.02068 | 0.02018 | 0.01970 | 0.01923 | 0.01876 | 0.01831 |
| 2.1 | 0.01786 | 0.01743 | 0.01700 | 0.01659 | 0.01618 | 0.01578 | 0.01539 | 0.01500 | 0.01463 | 0.01426 |
| 2.2 | 0.01390 | 0.01355 | 0.01321 | 0.01287 | 0.01255 | 0.01222 | 0.01191 | 0.01160 | 0.01130 | 0.01101 |
| 2.3 | 0.01072 | 0.01044 | 0.01017 | 0.00990 | 0.00964 | 0.00939 | 0.00914 | 0.00889 | 0.00866 | 0.00842 |
| 2.4 | 0.00820 | 0.00798 | 0.00776 | 0.00755 | 0.00734 | 0.00714 | 0.00695 | 0.00676 | 0.00657 | 0.00639 |
| 2.5 | 0.00621 | 0.00604 | 0.00587 | 0.00570 | 0.00554 | 0.00539 | 0.00523 | 0.00508 | 0.00494 | 0.00480 |

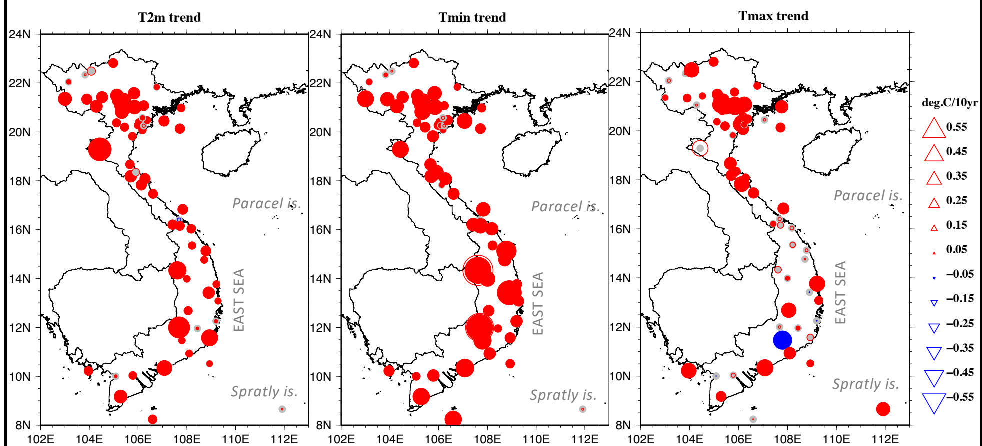


**Quick exercise #4:**

2m- daily temperature for the 5 consecutive days:
{30, 31, 34, 44, 31}

→? Z
→Is the trend statistically significance?

# 1961-2007 Temperature trend

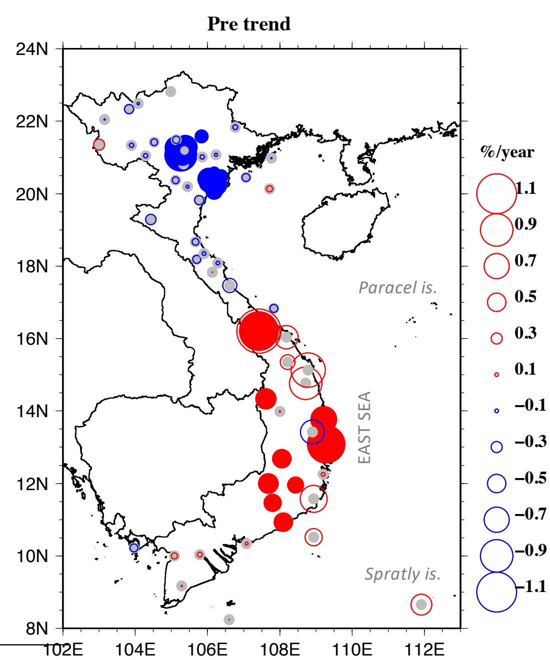



B2-SA-NDT-Earth System

25

25

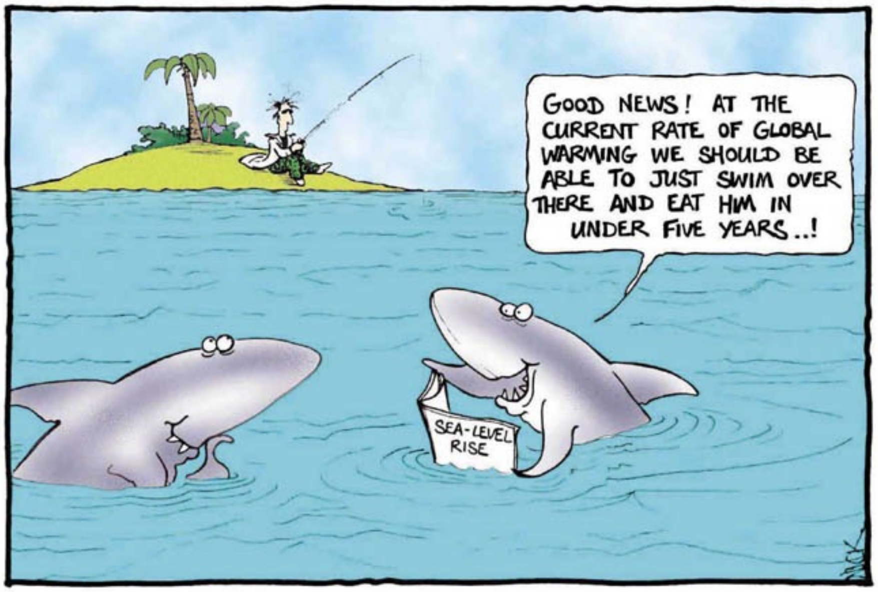
# 1961-2007 Rainfall trend




B2-SA-NDT-Earth System

26

26



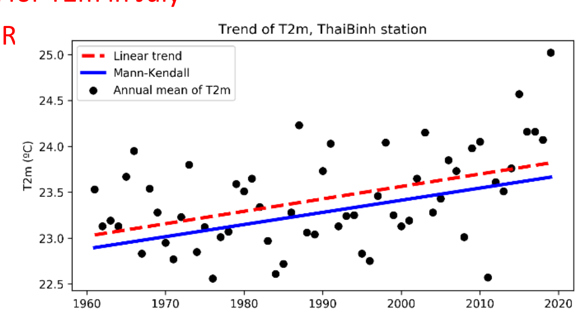

B2-SA-NDT-Earth System 27

27

# Practice Python #11

Working with the daily 2m-temperature data at ThaiBinh (or VietTri) station :
http://remosat.usth.edu.vn/~thanhnd/Download/dat_THAIBINH/

1. Plot annual mean of T2m
2. Plot the linear trend of the annual mean of T2m
3. Plot the Kendall trend of the annual mean of T2m
4. Plot a similar figure but for T2m in January
5. Plot a similar figure but for T2m in July
6. Plot a similar figure for R




B2-SA-NDT-Earth System

28