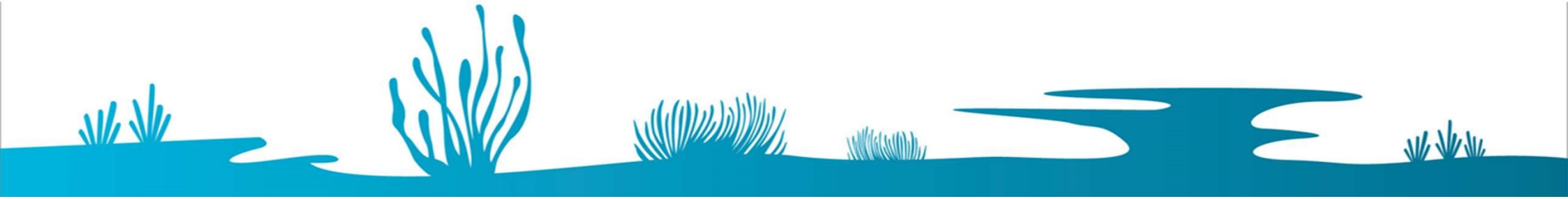
Climate and Oceans Support
Program in the Pacific

# ACCESS-S Workshop

## MODULE: Statistics – mean, anomaly, median



# Topics in this module

- Mean / average
- Anomalies
- Median

**Expected learning outcomes**

- Understanding the mean, median and anomalies and how to calculate them

Climate and Oceans Support Program in the Pacific

# Operations with the mean

$$mean = \frac{Sum\ of\ all\ the\ observations}{Number\ of\ all\ observations}$$

**Means can be added if variable is cumulative (like rainfall)**

$$Mean\ Jan\ to\ March\ rainfall = Jan\ mean + Feb\ mean + March\ mean$$

**Means can be averaged if variable is not cumulative (like temperature)**

$$Mean\ Jan\ to\ March\ maximum\ temperature = \frac{Jan\ mean + Feb\ mean + March\ mean}{3}$$

# Operations with the mean example

## Adding mean rainfall for Port Vila

| Year | Jan | Feb | Mar | Jan-Mar |
|---|---|---|---|---|
| 1998 | 141.2 | 204.9 | 531.1 | 877.2 |
| 1999 | 568.1 | 377.6 | 196.3 | 1142.0 |
| 2000 | 416.6 | 600.7 | 256.5 | 1273.8 |
| 2001 | 304.0 | 573.7 | 286.9 | 1164.6 |
| 2002 | 206.5 | 454.9 | 289.8 | 951.2 |
| 2003 | 130.7 | 098.6 | 144.0 | 373.3 |
| 2004 | 023.4 | 480.1 | 238.9 | 742.4 |
| 2005 | 200.4 | 133.0 | 195.4 | 528.8 |
| 2006 | 192.9 | 141.6 | 123.8 | 458.3 |
| 2007 | 211.1 | 368.4 | 406.5 | 986.0 |
| 2008 | 649.1 | 282.4 | 284.8 | 1216.3 |
| 2009 | 420.3 | 344.9 | 351.2 | 1116.4 |
| 2010 | 085.4 | 375.0 | 428.2 | 888.6 |
| 2011 | 325.4 | 345.8 | 182.0 | 853.2 |
| 2012 | 243.0 | 316.5 | 069.5 | 629.0 |
| 2013 | 102.1 | 211.9 | 290.1 | 604.1 |
| 2014 | 394.1 | 102.7 | 339.0 | 835.8 |
| 2015 | 365.0 | 421.4 | 792.8 | 1579.2 |
| 2016 | 024.8 | 053.1 | 119.5 | 197.4 |
| 2017 | 107.5 | 248.0 | 286.8 | 642.3 |
| 2018 | 181.5 | 194.0 | 156.5 | 532.0 |
| 2019 | 218.0 | 244.0 | 407.0 | 869.0 |
| 2020 | 112.2 | 204.3 | 320.7 | 637.2 |
| Mean | 244.5 | 294.7 | 291.2 | 830.4 |

244.5 + 294.7 + 291.2

= 830.4

## Averaging mean minimum temperature for Melbourne Airport

| Statistics | Jan | Feb | Mar |
|---|---|---|---|
| Temperature | | | |
| Mean maximum temperature (°C) | 26.6 | 26.6 | 24.2 |
| Mean minimum temperature (°C) | 13.8 | 14.1 | 12.7 |

(13.8 + 14.1 + 12.7) / 3

= 13.5

Climate and Oceans Support Program in the Pacific

# Anomalies: Differences from the mean

$$anomaly = observation - mean$$

- The word "**anomaly**" is used to describe the difference between an observation and a measure of "average" of those observations, most commonly the mean

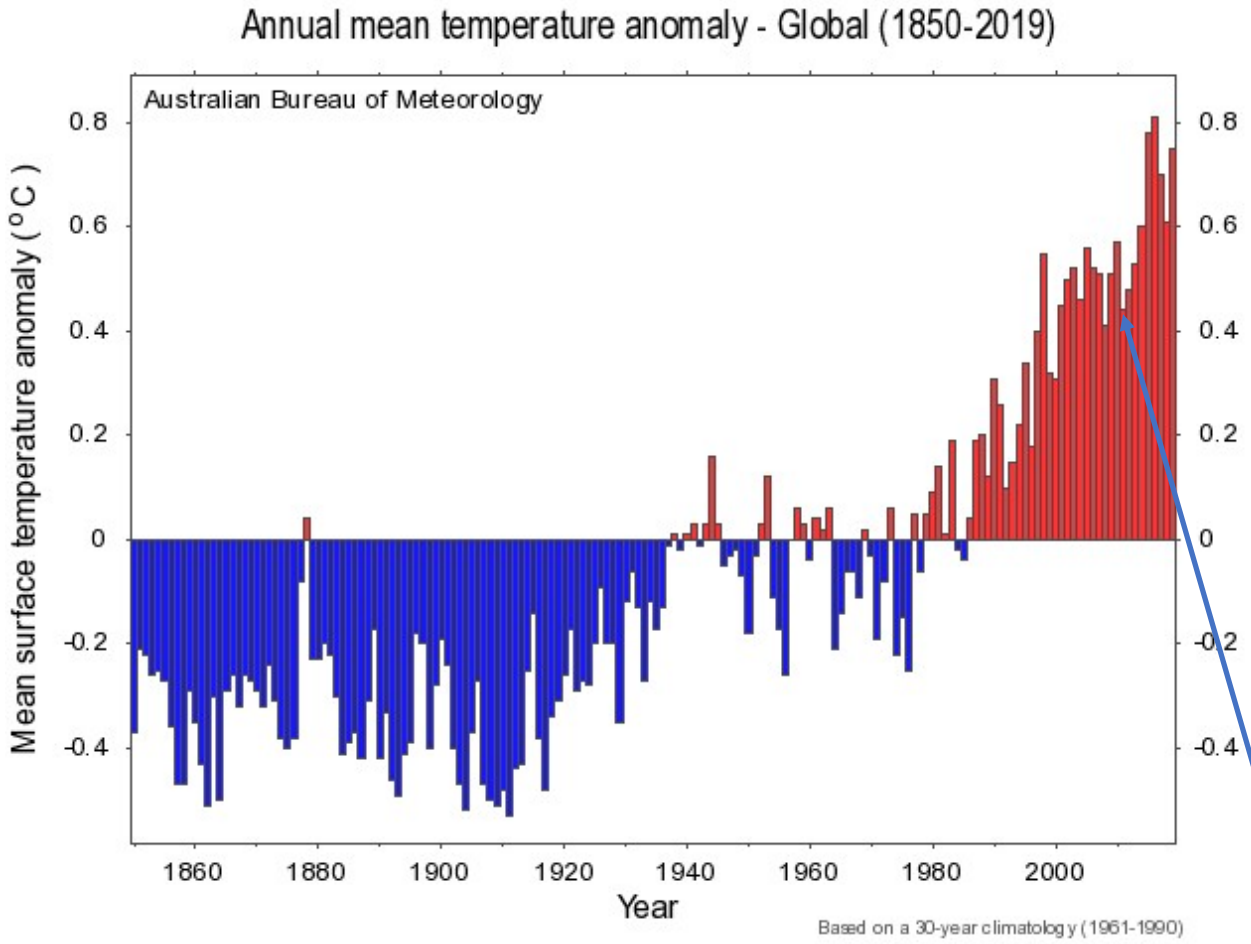
- It can be a **scaler** quantity (i.e. simply a number, positive or negative) or a vector (e.g. wind anomalies)

- **Anomalies are used to determine how different or unusual a climate pattern or event may be**

- **E.g. the 2020 temperature anomaly is how much above or below the climatological average the temperature was**

Climate and Oceans Support Program in the Pacific

# Anomalies: Differences from the mean (scalar)




$$anomaly = observation - mean$$

**Global Mean Temperature (any year)**

Reference period mean
$= T_m$
Observed Annual Mean for any year
$= T_y$
Observed Annual Anomaly for any year
$A_y = T_y - T_m$

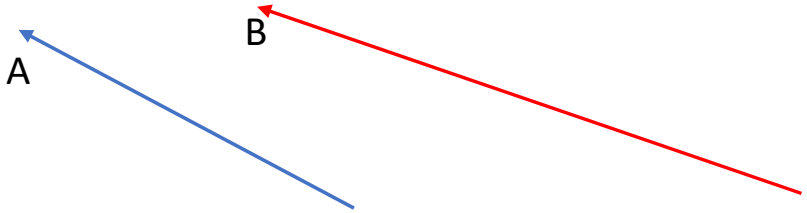
**2011 example**
1961-1990 reference period mean = 13.97 °C

2011 annual temperature = 14.41 °C

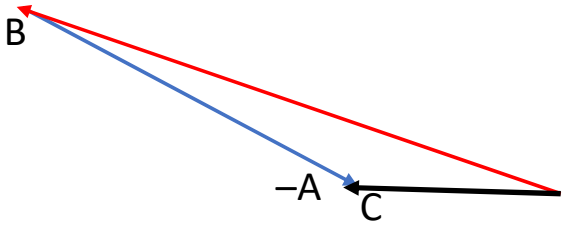
2011 annual temperature anomaly = +0.44 °C

Climate and Oceans Support Program in the Pacific

# Anomalies: Differences from the mean (vector)

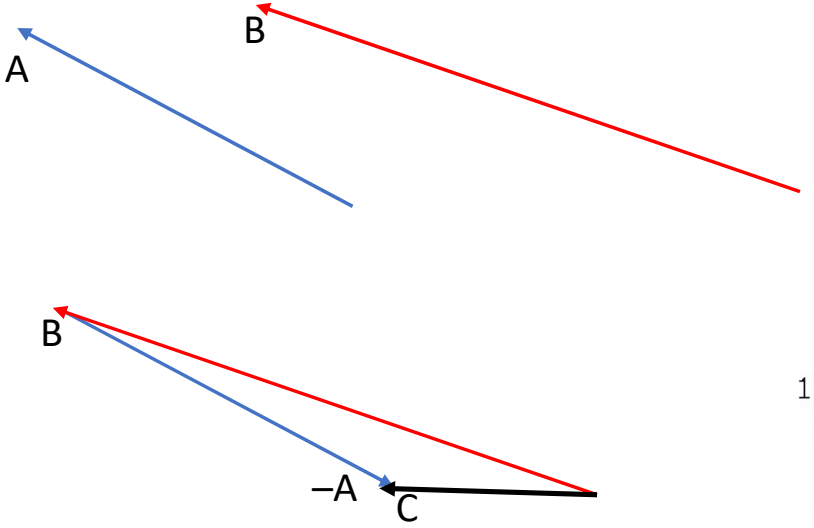



**Blue arrow A** represents the long-term **average** wind at a location
**Red arrow B** is the **observed** wind at the location

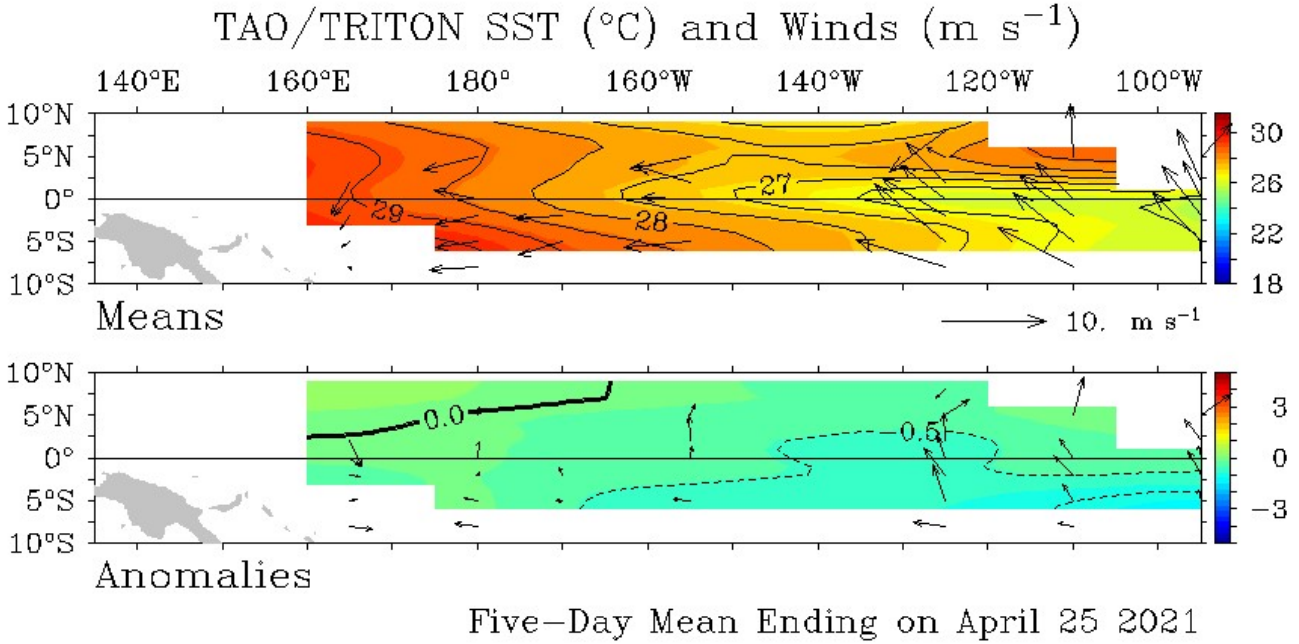



The **anomaly** = B – A is shown by **black arrow C**

# Anomalies: Differences from the mean (vector)




**Wind vector anomalies are used for tracking the trade winds**



# The Median

**ACCESS-S predictions often use the Median because the mean does not always represent the population.**

**The median is the "middle" value in a data set.**

### Median Example 1:

Consider the following **nine** numbers:
3, 13, 7, 5, 21, 23, 39, 300, 500

Put numbers in numerical order:
3, 5, 7, 13, 21, 23, 39, 300, 500

Median is the middle number. Our middle number will be the fifth number:
3, 5, 7, 13, **21**, 23, 39, 300, 500

So the **Median is 21**, and **the Mean is 101.2**.

**Extreme values can skew the mean** but have little effect on the median.

### Median Example 2:

Consider the following **eight** numbers:
3, 13, 7, 5, 21, 23, 23, 40

Put numbers in numerical order :
3, 5, 7, 13, 21, 23, 23, 40

Median is the middle number. We have two middle numbers:
3, 5, 7, **13, 21**, 23, 23, 40

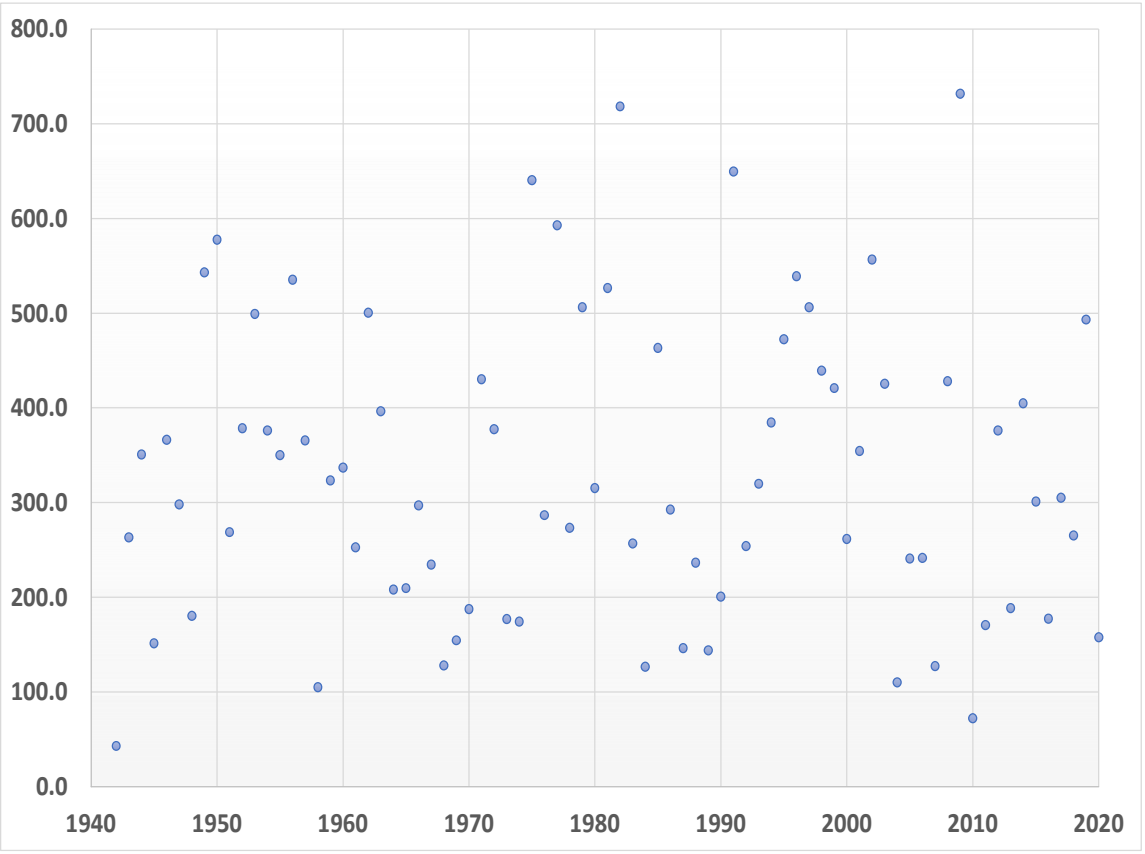
**Median = (13 +21) / 2**
**= 17**

17 was not in the list of numbers…but that is OK because half the numbers in the list are less, and half the numbers are greater.)

Climate and Oceans Support Program in the Pacific

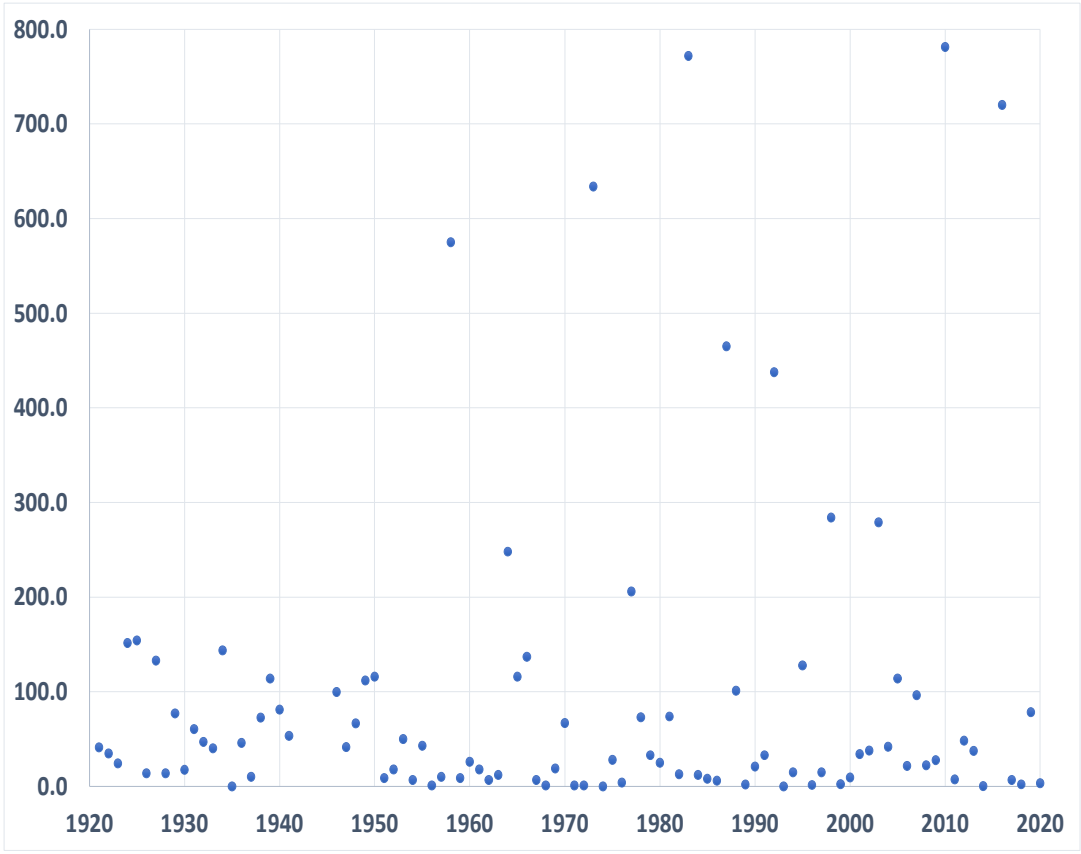
# Means and Medians

## Suva, Fiji. January Rainfall
## Mean and median are similar




Mean 331mm, Median 305mm

## Kiritimati, Kiribati. January Rainfall
## Extreme values show up in the mean, not the median




Mean 95mm, Median 34mm

Climate and Oceans Support Program in the Pacific

# Operations with the Median

**Medians must not be added if variable is cumulative (like rainfall)**

$$Median\ June\ to\ August\ rainfall \neq June\ median + July\ median + August\ median$$

**Medians must not be averaged if variable is not cumulative (like temperature)**

$$Median\ June\ to\ August\ maximum\ temperature \neq \frac{June\ median + July\ median + August\ median}{3}$$

Climate and Oceans Support Program in the Pacific

# Operations with the Median examples

**Do not add medians**

**Rainfall for Nadi, Fiji**

| Year | Jun | Jul | Aug | Jun-Aug |
|---|---|---|---|---|
| **2001** | 17.4 | 96.6 | 146.6 | 250.6 |
| **2002** | 53.2 | 102.8 | 48.5 | 204.5 |
| **2003** | 29.6 | 28.8 | 97.6 | 156.0 |
| **2004** | 69.9 | 59.2 | 256.2 | 385.3 |
| **2005** | 80.0 | 61.0 | 52.0 | 193.0 |
| **2006** | 39.9 | 35.6 | 104.4 | 179.9 |
| **2007** | 1.5 | 69.3 | 1.1 | 71.9 |
| **2008** | 50.1 | 73.2 | 2.9 | 126.2 |
| **2009** | 113.5 | 73.9 | 55.9 | 243.3 |
| **2010** | 4.6 | 33.5 | 22.0 | 60.1 |
| **2011** | 76.4 | 101.4 | 80.4 | 258.2 |
| **2012** | 311.9 | 32.7 | 66.3 | 410.9 |
| **2013** | 124.9 | 6.3 | 10.7 | 141.9 |
| **2014** | 1.3 | 13.3 | 0.0 | 14.6 |
| **2015** | 14.8 | 15.6 | 49.9 | 80.3 |
| **2016** | 28.1 | 13.4 | 197.0 | 238.5 |
| **2017** | 2.4 | 20.5 | 60.6 | 83.5 |
| **2018** | 53.3 | 0.5 | 5.5 | 59.3 |
| **2019** | 58.1 | 65.0 | 20.3 | 143.4 |
| **2020** | 32.5 | 59.7 | 17.3 | 109.5 |
| **Median** | 45.0 | 47.4 | 51.0 | 149.7 |

If we **sum the individual medians** for June, July and August, i.e. 45.0 + 47.4 + 51.9, we get **143.4**

This is different to the actual value of 149.7

If we **average the individual medians** for October, November and August, i.e. (24.0 + 24.3 + 26.0)/3, we get **24.8**

This is close but **not equal** to the actual value of **24.6**

**Do not average medians**

**Mean maximum temperature for Sydney, Australia**

| Year | Oct | Nov | Dec | Oct-Dec |
|---|---|---|---|---|
| **2001** | 24.0 | 23.5 | 26.2 | 24.6 |
| **2002** | 24.5 | 25.3 | 25.4 | 25.1 |
| **2003** | 21.6 | 22.7 | 26.0 | 23.4 |
| **2004** | 23.3 | 25.4 | 25.6 | 24.8 |
| **2005** | 24.3 | 23.9 | 28.6 | 25.6 |
| **2006** | 24.2 | 24.3 | 24.2 | 24.2 |
| **2007** | 25.0 | 23.5 | 24.7 | 24.4 |
| **2008** | 23.2 | 23.7 | 25.9 | 24.3 |
| **2009** | 21.4 | 26.1 | 25.5 | 24.3 |
| **2010** | 22.0 | 23.5 | 25.8 | 23.8 |
| **2011** | 22.2 | 25.2 | 23.1 | 23.5 |
| **2012** | 23.3 | 24.1 | 26.0 | 24.5 |
| **2013** | 25.7 | 24.0 | 26.4 | 25.4 |
| **2014** | 24.7 | 25.7 | 25.8 | 25.4 |
| **2015** | 24.9 | 26.2 | 26.1 | 25.7 |
| **2016** | 23.9 | 26.1 | 28.1 | 26.0 |
| **2017** | 24.0 | 23.8 | 27.8 | 25.2 |
| **2018** | 21.5 | 24.8 | 26.8 | 24.4 |
| **2019** | 24.4 | 26.6 | 26.8 | 25.9 |
| **Median** | 24.0 | 24.3 | 26.0 | 24.6 |

# Mean, median and anomaly summary

$$\boldsymbol{mean} = \frac{Sum\ of\ all\ the\ observations}{Number\ of\ all\ observations}$$

Means **can be added** if variable is cumulative (like rainfall)

Means **can be averaged** if variable is not cumulative (like temperature**)**

Mean is most representative of the data.

**Discussion question: When have you used a mean?**

The **median** is the "middle value" in a dataset.

The median is not skewed by extreme values

**Discussion question: When have you seen the median being used?**

$$anomaly = observation - mean$$

Anomalies are used to determine how different or unusual a climate pattern or event may be.

Anomalies can be a scalar or vector

**Discussion question: When have you used an anomaly?**