

## Reported Diversion Statistics

### Data Processing Module:

ExpectedDemand\_ExceedsFV\_UnitConversion\_UseVsDiv\_Statistics.xlsx

**Description:** Calculate various statistics from reported diversion and use to gain insight on expected diversion amounts, and identify outlier reports that may be in error

**Example:** QAQC staff quickly identify a water right with a relatively large average monthly standard deviation. Further review shows there was 1 report with substantially different values than the rest of the reports, and it seems to be reported in the wrong units. Staff correct the errant records and move on to the next record.

**What (is being flagged):** Average monthly reported values of direct diversion, and diversion to storage, monthly standard deviations (by month, across reporting years), and average monthly standard deviation.

**Why:** Pre-calculate several classic statistical metrics to assist staff in identifying reporting outliers, to prioritize and target further investigation as to whether a report is in error or not, and make changes as needed.

**How:** Reshape the RMS reported data into rows of reporting years with columns of months, for direct diversion and diversion to storage. Then calculate monthly standard deviations (by column for each water right) to provide the value of a standard deviation across various reports for a given month. Then calculate the average of all months for each water right. While the average standard deviation is not a standard statistical metric, it is useful in collapsing all the statistical calculations into 1 value for each water right, which can be sorted to assist prioritizing review of water right reports.

**Resolution:** If it is determined the outlier is due to a reporting error, the errant reports can be overridden by QAQC staff to better represent actual/expected diversion.

### Data Source(s):

ewrims\_flat\_file.csv

water\_use\_report.csv

### Existing Fields:

APPL\_ID - Data from water\_use\_report.csv

YEAR - Data from water\_use\_report.csv

MONTH - Data from water\_use\_report.csv

AMOUNT- Data from water\_use\_report.csv

DIVERSION\_TYPE - Data from water\_use\_report.csv

APPLICATION\_NUMBER – Data from ewrims\_flat\_file.csv

INI\_REPORTED\_DIV\_AMOUNT – Data from ewrims\_flat\_file.csv

INI\_REPORTED\_DIV\_UNIT – Data from ewrims\_flat\_file.csv

FACE\_VALUE\_AMOUNT – Data from ewrims\_flat\_file.csv

FACE\_VALUE\_UNITS – Data from ewrims\_flat\_file.csv

**New Fields:**

REPORTING\_YEAR

INDEX

APPLID

JAN\_DIRECT\_DIVERISION

FEB\_DIRECT\_DIVERISION

MAR\_DIRECT\_DIVERISION

APR\_DIRECT\_DIVERISION

MAY\_DIRECT\_DIVERISION

JUN\_DIRECT\_DIVERISION

JUL\_DIRECT\_DIVERISION

AUG\_DIRECT\_DIVERISION

SEP\_DIRECT\_DIVERISION

OCT\_DIRECT\_DIVERISION

NOV\_DIRECT\_DIVERISION

DEC\_DIRECT\_DIVERISION

JAN\_STORAGE\_DIVERISION

FEB\_STORAGE\_DIVERISION

MAR\_STORAGE\_DIVERISION

APR\_STORAGE\_DIVERISION

MAY\_STORAGE\_DIVERISION

JUN\_STORAGE\_DIVERISION

JUL\_STORAGE\_DIVERISION

AUG\_STORAGE\_DIVERISION

SEP\_STORAGE\_DIVERISION

OCT\_STORAGE\_DIVERISION

NOV\_STORAGE\_DIVERISION

DEC\_STORAGE\_DIVERISION

ANNUAL\_DIRECT

ANNUAL\_STORAGE

DUPLICATE\_DIRECT\_STORAGE

CALENDAR\_YEAR\_TOTAL

MAY\_TO\_SEP\_TOTAL\_DIVERISION

FaceValue

Initial\_Reported\_Diversion  
Units\_IniDiv  
IniDiv\_Converted\_to\_AF  
AnnualTotalDiversion  
Annual\_Use  
Diversion\_as\_Percent\_of\_FV  
Amount\_over\_FV  
Diversion\_as\_Percent\_of\_IniDiv  
Amount\_over\_IniDiv  
Annual\_Diversion\_if\_reported\_in\_Gallons  
Annual\_Diversion\_if\_reported\_in\_GPM  
Annual\_Diversion\_if\_reported\_in\_GPD  
Annual\_Diversion\_if\_reported\_in\_CFS  
UNIQUE\_APPL\_ID  
JAN\_AVERAGE\_DIRECT\_DIVERSION  
FEB\_AVERAGE\_DIRECT\_DIVERSION  
MAR\_AVERAGE\_DIRECT\_DIVERSION  
APR\_AVERAGE\_DIRECT\_DIVERSION  
MAY\_AVERAGE\_DIRECT\_DIVERSION  
JUN\_AVERAGE\_DIRECT\_DIVERSION  
JUL\_AVERAGE\_DIRECT\_DIVERSION  
AUG\_AVERAGE\_DIRECT\_DIVERSION  
SEP\_AVERAGE\_DIRECT\_DIVERSION  
OCT\_AVERAGE\_DIRECT\_DIVERSION  
NOV\_AVERAGE\_DIRECT\_DIVERSION  
DEC\_AVERAGE\_DIRECT\_DIVERSION  
JAN\_AVERAGE\_STORAGE\_DIVERSION  
FEB\_AVERAGE\_STORAGE\_DIVERSION  
MAR\_AVERAGE\_STORAGE\_DIVERSION  
APR\_AVERAGE\_STORAGE\_DIVERSION  
MAY\_AVERAGE\_STORAGE\_DIVERSION  
JUN\_AVERAGE\_STORAGE\_DIVERSION  
JUL\_AVERAGE\_STORAGE\_DIVERSION  
AUG\_AVERAGE\_STORAGE\_DIVERSION  
SEP\_AVERAGE\_STORAGE\_DIVERSION  
OCT\_AVERAGE\_STORAGE\_DIVERSION  
NOV\_AVERAGE\_STORAGE\_DIVERSION  
DEC\_AVERAGE\_STORAGE\_DIVERSION  
JAN\_EXPECTED\_TOTAL\_DIVERSION  
FEB\_EXPECTED\_TOTAL\_DIVERSION  
MAR\_EXPECTED\_TOTAL\_DIVERSION  
APR\_EXPECTED\_TOTAL\_DIVERSION  
MAY\_EXPECTED\_TOTAL\_DIVERSION

JUN\_EXPECTED\_TOTAL\_DIVERSION  
JUL\_EXPECTED\_TOTAL\_DIVERSION  
AUG\_EXPECTED\_TOTAL\_DIVERSION  
SEP\_EXPECTED\_TOTAL\_DIVERSION  
OCT\_EXPECTED\_TOTAL\_DIVERSION  
NOV\_EXPECTED\_TOTAL\_DIVERSION  
DEC\_EXPECTED\_TOTAL\_DIVERSION  
ANNUAL\_TOTAL\_DIVERSION  
MAY\_TO\_SEP\_TOTAL\_DIVERSION  
TOTAL\_ANNUAL\_USE  
JAN\_STDEV  
FEB\_STDEV  
MAR\_STDEV  
APR\_STDEV  
MAY\_STDEV  
JUN\_STDEV  
JUL\_STDEV  
AUG\_STDEV  
SEP\_STDEV  
OCT\_STDEV  
NOV\_STDEV  
DEC\_STDEV  
AVERAGE\_STDEV  
Total\_Cumulative\_Diverted  
Total\_Cumulative\_Use  
Total\_Use\_as\_a\_Percent\_of\_Total\_Diverted