

BIOS 567 Statistical Methods for High-Throughput Genomic Data I

	QAQC1	order 1	QAQC2	order 2	QAQC33	order 3	QAQC4	order 4
AFFX-HUMGAPDH/M33197_3_at1	17216	1	20255.5	1	18504.5	1	20635.8	1
AFFX-HUMGAPDH/M33197_3_at2	15823	2	19793.8	2	18622.5	2	20943	2
AFFX-HUMGAPDH/M33197_3_at3	13721.3	3	14155	3	14772.3	3	17608.5	3
AFFX-HUMGAPDH/M33197_3_at4	11332.8	4	12888.3	4	12073	4	13596.8	4
AFFX-HUMGAPDH/M33197_3_at5	8739	5	10341.5	5	10599.3	5	11188	5
AFFX-HUMGAPDH/M33197_3_at6	3565	6	4365.3	6	4902.3	6	4748.8	6
AFFX-HUMGAPDH/M33197_3_at7	6118.3	7	6873	7	6488.8	7	7187	7
AFFX-HUMGAPDH/M33197_3_at8	6811.3	8	8704.8	8	8197	8	8906.3	8
AFFX-HUMGAPDH/M33197_3_at9	12177	9	12241.5	9	12356	9	15262.5	9
AFFX-HUMGAPDH/M33197_3_at10	9727.5	10	10249.3	10	11112.8	10	10160.3	10
AFFX-HUMGAPDH/M33197_3_at11	17385.5	11	20579.5	11	19017	11	21311.8	11
AFFX-HUMGAPDH/M33197_3_at12	7961.5	12	8773.5	12	8994	12	9931.8	12
AFFX-HUMGAPDH/M33197_3_at13	9297.5	13	10425.5	13	9768	13	11831.5	13
AFFX-HUMGAPDH/M33197_3_at14	14613.8	14	17664.8	14	16041	14	17952	14
AFFX-HUMGAPDH/M33197_3_at15	11090	15	11528.5	15	12117	15	14837.5	15
AFFX-HUMGAPDH/M33197_3_at16	17957.3	16	18999.5	16	17884.5	16	19188	16
AFFX-HUMGAPDH/M33197_3_at17	2570.8	17	2476.8	17	2646.3	17	2843	17
AFFX-HUMGAPDH/M33197_3_at18	19599.5	18	25077.5	18	17462	18	25121.8	18
AFFX-HUMGAPDH/M33197_3_at19	22619.5	19	27955.8	19	26247.5	19	28640.5	19
AFFX-HUMGAPDH/M33197_3_at20	930.5	20	1318.8	20	1719.3	20	1252.8	20

Step a. Let X represent a matrix with N columns, 1 column for each GeneChip, and G rows where each row represents a probe (usually both PM and MM). As a small example, we have the X matrix for PM probes 1 – 20 for AFFX-HUMGAPDH/M33197_3 probe set, QAQC Arrays 1 – 4 where $N = 4$, $G = 20$.

BIOS 567 Statistical Methods for High-Throughput Genomic Data I

	QAQC1	order 1	QAQC2	order 2	QAQC3	order 3	QAQC4	order 4
AFFX-HUMGAPDH/M33197_3_at1	930.5	20	1318.8	20	1719.3	20	1252.8	20
AFFX-HUMGAPDH/M33197_3_at2	2570.8	17	2476.8	17	2646.3	17	2843	17
AFFX-HUMGAPDH/M33197_3_at3	3565	6	4365.3	6	4902.3	6	4748.8	6
AFFX-HUMGAPDH/M33197_3_at4	6118.3	7	6873	7	6488.8	7	7187	7
AFFX-HUMGAPDH/M33197_3_at5	6811.3	8	8704.8	8	8197	8	8906.3	8
AFFX-HUMGAPDH/M33197_3_at6	7961.5	12	8773.5	12	8994	12	9931.8	12
AFFX-HUMGAPDH/M33197_3_at7	8739	5	10249.3	10	9768	13	10160.3	10
AFFX-HUMGAPDH/M33197_3_at8	9297.5	13	10341.5	5	10599.3	5	11188	5
AFFX-HUMGAPDH/M33197_3_at9	9727.5	10	10425.5	13	11112.8	10	11831.5	13
AFFX-HUMGAPDH/M33197_3_at10	11090	15	11528.5	15	12073	4	13596.8	4
AFFX-HUMGAPDH/M33197_3_at11	11332.8	4	12241.5	9	12117	15	14837.5	15
AFFX-HUMGAPDH/M33197_3_at12	12177	9	12888.3	4	12356	9	15262.5	9
AFFX-HUMGAPDH/M33197_3_at13	13721.3	3	14155	3	14772.3	3	17608.5	3
AFFX-HUMGAPDH/M33197_3_at14	14613.8	14	17664.8	14	16041	14	17952	14
AFFX-HUMGAPDH/M33197_3_at15	15823	2	18999.5	16	17462	18	19188	16
AFFX-HUMGAPDH/M33197_3_at16	17216	1	19793.8	2	17884.5	16	20635.8	1
AFFX-HUMGAPDH/M33197_3_at17	17385.5	11	20255.5	1	18504.5	1	20943	2
AFFX-HUMGAPDH/M33197_3_at18	17957.3	16	20579.5	11	18622.5	2	21311.8	11
AFFX-HUMGAPDH/M33197_3_at19	19599.5	18	25077.5	18	19017	11	25121.8	18
AFFX-HUMGAPDH/M33197_3_at20	22619.5	19	27955.8	19	26247.5	19	28640.5	19

Step b. Sort each column in X and define the sorted matrix as X_{sort} . In this example, each of the 4 QAQC Arrays PM probe intensities for the AFX-HUMGAPDH/M33197_3 probe set have been sorted in increasing order.

BIOS 567 Statistical Methods for High-Throughput Genomic Data I

	QAQC1	order 1	QAQC2	order 2	QAQC3	order 3	QAQC4	order 4	Row Average
AFFX-HUMGAPDH/M33197_3_at1	930.5	20	1318.8	20	1719.3	20	1252.8	20	1305.35
AFFX-HUMGAPDH/M33197_3_at2	2570.8	17	2476.8	17	2646.3	17	2843	17	2634.225
AFFX-HUMGAPDH/M33197_3_at3	3565	6	4365.3	6	4902.3	6	4748.8	6	4395.35
AFFX-HUMGAPDH/M33197_3_at4	6118.3	7	6873	7	6488.8	7	7187	7	6666.775
AFFX-HUMGAPDH/M33197_3_at5	6811.3	8	8704.8	8	8197	8	8906.3	8	8154.85
AFFX-HUMGAPDH/M33197_3_at6	7961.5	12	8773.5	12	8994	12	9931.8	12	8915.2
AFFX-HUMGAPDH/M33197_3_at7	8739	5	10249.3	10	9768	13	10160.3	10	9729.15
AFFX-HUMGAPDH/M33197_3_at8	9297.5	13	10341.5	5	10599.3	5	11188	5	10356.575
AFFX-HUMGAPDH/M33197_3_at9	9727.5	10	10425.5	13	11112.8	10	11831.5	13	10774.325
AFFX-HUMGAPDH/M33197_3_at10	11090	15	11528.5	15	12073	4	13596.8	4	12072.075
AFFX-HUMGAPDH/M33197_3_at11	11332.8	4	12241.5	9	12117	15	14837.5	15	12632.2
AFFX-HUMGAPDH/M33197_3_at12	12177	9	12888.3	4	12356	9	15262.5	9	13170.95
AFFX-HUMGAPDH/M33197_3_at13	13721.3	3	14155	3	14772.3	3	17608.5	3	15064.275
AFFX-HUMGAPDH/M33197_3_at14	14613.8	14	17664.8	14	16041	14	17952	14	16567.9
AFFX-HUMGAPDH/M33197_3_at15	15823	2	18999.5	16	17462	18	19188	16	17868.125
AFFX-HUMGAPDH/M33197_3_at16	17216	1	19793.8	2	17884.5	16	20635.8	1	18882.525
AFFX-HUMGAPDH/M33197_3_at17	17385.5	11	20255.5	1	18504.5	1	20943	2	19272.125
AFFX-HUMGAPDH/M33197_3_at18	17957.3	16	20579.5	11	18622.5	2	21311.8	11	19617.775
AFFX-HUMGAPDH/M33197_3_at19	19599.5	18	25077.5	18	19017	11	25121.8	18	22203.95
AFFX-HUMGAPDH/M33197_3_at20	22619.5	19	27955.8	19	26247.5	19	28640.5	19	26365.825

Step c1. Calculate the average of each row in X_{sort} .

BIOS 567 Statistical Methods for High-Throughput Genomic Data I

	QAQC1	order 1	QAQC2	order 2	QAQC3	order 3	QAQC4	order 4
AFFX-HUMGAPDH/M33197_3_at1	1305.350	20	1305.350	20	1305.350	20	1305.350	20
AFFX-HUMGAPDH/M33197_3_at2	2634.225	17	2634.225	17	2634.225	17	2634.225	17
AFFX-HUMGAPDH/M33197_3_at3	4395.350	6	4395.350	6	4395.350	6	4395.350	6
AFFX-HUMGAPDH/M33197_3_at4	6666.775	7	6666.775	7	6666.775	7	6666.775	7
AFFX-HUMGAPDH/M33197_3_at5	8154.850	8	8154.850	8	8154.850	8	8154.850	8
AFFX-HUMGAPDH/M33197_3_at6	8915.200	12	8915.200	12	8915.200	12	8915.200	12
AFFX-HUMGAPDH/M33197_3_at7	9729.150	5	9729.150	10	9729.150	13	9729.150	10
AFFX-HUMGAPDH/M33197_3_at8	10356.575	13	10356.575	5	10356.575	5	10356.575	5
AFFX-HUMGAPDH/M33197_3_at9	10774.325	10	10774.325	13	10774.325	10	10774.325	13
AFFX-HUMGAPDH/M33197_3_at10	12072.075	15	12072.075	15	12072.075	4	12072.075	4
AFFX-HUMGAPDH/M33197_3_at11	12632.200	4	12632.200	9	12632.200	15	12632.200	15
AFFX-HUMGAPDH/M33197_3_at12	13170.950	9	13170.950	4	13170.950	9	13170.950	9
AFFX-HUMGAPDH/M33197_3_at13	15064.275	3	15064.275	3	15064.275	3	15064.275	3
AFFX-HUMGAPDH/M33197_3_at14	16567.900	14	16567.900	14	16567.900	14	16567.900	14
AFFX-HUMGAPDH/M33197_3_at15	17868.125	2	17868.125	16	17868.125	18	17868.125	16
AFFX-HUMGAPDH/M33197_3_at16	18882.525	1	18882.525	2	18882.525	16	18882.525	1
AFFX-HUMGAPDH/M33197_3_at17	19272.125	11	19272.125	1	19272.125	1	19272.125	2
AFFX-HUMGAPDH/M33197_3_at18	19617.775	16	19617.775	11	19617.775	2	19617.775	11
AFFX-HUMGAPDH/M33197_3_at19	22203.950	18	22203.950	18	22203.950	11	22203.950	18
AFFX-HUMGAPDH/M33197_3_at20	26365.825	19	26365.825	19	26365.825	19	26365.825	19

Step c2. Substitute the average row value for each of the individual elements in that row to get X'_{sort} .

BIOS 567 Statistical Methods for High-Throughput Genomic Data I

	QAQC1	order 1	QAQC2	order 2	QAQC3	order 3	QAQC4	order 4
AFFX-HUMGAPDH/M33197_3_at1	18882.525	1	19272.125	1	19272.125	1	18882.525	1
AFFX-HUMGAPDH/M33197_3_at2	17868.125	2	18882.525	2	19617.775	2	19272.125	2
AFFX-HUMGAPDH/M33197_3_at3	15064.275	3	15064.275	3	15064.275	3	15064.275	3
AFFX-HUMGAPDH/M33197_3_at4	12632.200	4	13170.950	4	12072.075	4	12072.075	4
AFFX-HUMGAPDH/M33197_3_at5	9729.150	5	10356.575	5	10356.575	5	10356.575	5
AFFX-HUMGAPDH/M33197_3_at6	4395.350	6	4395.350	6	4395.350	6	4395.350	6
AFFX-HUMGAPDH/M33197_3_at7	6666.775	7	6666.775	7	6666.775	7	6666.775	7
AFFX-HUMGAPDH/M33197_3_at8	8154.850	8	8154.850	8	8154.850	8	8154.850	8
AFFX-HUMGAPDH/M33197_3_at9	13170.950	9	12632.200	9	13170.950	9	13170.950	9
AFFX-HUMGAPDH/M33197_3_at10	10774.325	10	9729.150	10	10774.325	10	9729.150	10
AFFX-HUMGAPDH/M33197_3_at11	19272.125	11	19617.775	11	22203.950	11	19617.775	11
AFFX-HUMGAPDH/M33197_3_at12	8915.200	12	8915.200	12	8915.200	12	8915.200	12
AFFX-HUMGAPDH/M33197_3_at13	10356.575	13	10774.325	13	9729.150	13	10774.325	13
AFFX-HUMGAPDH/M33197_3_at14	16567.900	14	16567.900	14	16567.900	14	16567.900	14
AFFX-HUMGAPDH/M33197_3_at15	12072.075	15	12072.075	15	12632.200	15	12632.200	15
AFFX-HUMGAPDH/M33197_3_at16	19617.775	16	17868.125	16	18882.525	16	17868.125	16
AFFX-HUMGAPDH/M33197_3_at17	2634.225	17	2634.225	17	2634.225	17	2634.225	17
AFFX-HUMGAPDH/M33197_3_at18	22203.950	18	22203.950	18	17868.125	18	22203.950	18
AFFX-HUMGAPDH/M33197_3_at19	26365.825	19	26365.825	19	26365.825	19	26365.825	19
AFFX-HUMGAPDH/M33197_3_at20	1305.350	20	1305.350	20	1305.350	20	1305.350	20

Step d. X_{norm} by rearranging each column of X'_{sort} to have the same ordering as original