# Homework Number 3

**Optional**

1. A system exists which calculates student GPAs. Each quarter when grades are filed, the student grades will be entered by the grade entry process. This information is then passed on to a process "updateGPA" which will read the existing grades from the database and combine that data with the new grades to calculate an updated GPA. The updated GPA will then be sent to the process Determine honor roll, which is responsible for writing out the names of students who are on the honor roll to a file.
    a. Based on this description, draw a data flow diagram for the system.

    b. Write a data dictionary entry for the field holding a student's GPA.

    c. What risks does this entry imply for the software?

2. A web system deployed on campus has a failure rate of $10^{-3}$ failures per hour. What is the likelihood that the system will continue operating without failure throughout the duration of exam week (Monday through Friday)? (Show all work, both assuming failure in each hour is independent and that it is dependent.)

3. Two processes are each normally distributed and have the same mean, which is unknown. Process A takes 15ms/operation, while process B takes 17 ms/operation. The sample standard deviation, averaged over both processes, is 2.5 ms.
    a. If these means come from 5 samples for each process, can we say that the processes are significantly different? (At roughly what p-value?)

    b. If these means come from 10 samples for each process, can we say that the processes are significantly different? (At roughly what p-value?)

    c. If these means come from 100 samples for each process, can we say that the processes are significantly different? (At roughly what p-value?)

    d. If we KNOW that the true standard deviation is 2 ms, what can we say about this process?

(Remaining problems **were not covered**.)

4. A web server receives a request every 50ms and processes web requests every 8 ms. Using queuing theory,
    a. What is the average response time for this system?
    b. How large should the queue be if there is to be a less than .5% chance of the queue overflowing?
    c. What would the waiting time be if the web server is modified to have 2 threads serving processes but nothing else changes?

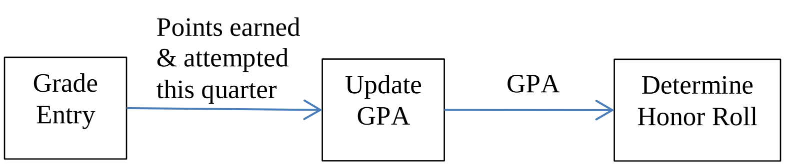
5. The following shows an output from gprof. Which functions should be improved first if you are trying to improve the performance of the system?

```
  %   cumulative   self              self     total
 time   seconds   seconds    calls  ms/call  ms/call  name
 50.4       0.59     0.59     1024     0.58     0.70  .fft [3]
 37.6       1.03     0.44        1   440.00  1160.00  .main [1]
 11.1       1.16     0.13     1024     0.13     0.13  .bit_reverse [4]
  0.9       1.17     0.01                             .__mcount [5]
  0.0       1.17     0.00      256     0.00     0.00  .cos [6]
  0.0       1.17     0.00      256     0.00     0.00  .sin [7]
  0.0       1.17     0.00       19     0.00     0.00  .fwrite [8]
  0.0       1.17     0.00       19     0.00     0.00  .memchr [9]
  0.0       1.17     0.00       11     0.00     0.00  ._flsbuf [10]
  0.0       1.17     0.00        8     0.00     0.00  .gettimeofday [11]
  0.0       1.17     0.00        8     0.00     0.00  .pwr10 [12]
  0.0       1.17     0.00        7     0.00     0.00  ._doprnt [13]
  0.0       1.17     0.00        7     0.00     0.00  ._xflsbuf [14]
  0.0       1.17     0.00        7     0.00     0.00  ._xwrite [15]
  0.0       1.17     0.00        7     0.00     0.00  .printf [16]
  0.0       1.17     0.00        7     0.00     0.00  .write [17]
  0.0       1.17     0.00        4     0.00     0.00  .__ld [18]
  0.0       1.17     0.00        4     0.00     0.00  .__nl_langinfo_std [19]
  0.0       1.17     0.00        4     0.00     0.00  .cvt [20]
  0.0       1.17     0.00        4     0.00     0.00  .cvtloop [21]
  0.0       1.17     0.00        4     0.00     0.00  .ecvt [22]
  0.0       1.17     0.00        4     0.00     0.00  .mf2x1 [23]
  0.0       1.17     0.00        4     0.00     0.00  .nl_langinfo [24]
  0.0       1.17     0.00        3     0.00     0.00  .splay [25]
  0.0       1.17     0.00        2     0.00     0.00  .free [26]
  0.0       1.17     0.00        2     0.00     0.00  .free_y [27]
  0.0       1.17     0.00        1     0.00     0.00  .__ioctl [28]
  0.0       1.17     0.00        1     0.00     0.00  ._findbuf [29]
  0.0       1.17     0.00        1     0.00     0.00  ._wrtchk [30]
  0.0       1.17     0.00        1     0.00     0.00  .atoi [31]
  0.0       1.17     0.00        1     0.00     0.00  .catopen [32]
  0.0       1.17     0.00        1     0.00     0.00  .exit [33]
  0.0       1.17     0.00        1     0.00     0.00  .ioctl [34]
  0.0       1.17     0.00        1     0.00     0.00  .isatty [35]
  0.0       1.17     0.00        1     0.00     0.00  .moncontrol [36]
  0.0       1.17     0.00        1     0.00     0.00  .monitor [37]
  0.0       1.17     0.00        1     0.00     0.00  .saved_category_name [38]
  0.0       1.17     0.00        1     0.00     0.00  .setlocale [39]
```

Answers:

1. A system exists which calculates student GPAs. Each quarter when grades are filed, the student grades will be entered by the grade entry process. This information is then passed on to a process "updateGPA" which will read the existing grades from the database and combine that data with the new grades to calculate an updated GPA. The updated GPA will then be sent to the process Determine honor roll, which is responsible for writing out the names of students who are on the honor roll to a file.

a. Based on this description, draw a data flow diagram for the system.




b. Write a data dictionary entry for the field holding a student's GPA.

*Name: GPA*
*Data type: 64 bit floating point number (Java double)*
*Valid range: 0.0-4.0, NaN*
*Units: unitless – simple ratio*

c. What risks does this entry imply for the software?

*Students who have not completed any classes may not have a defined GPA. Storing it as a NaN (as above) may result in errors when doing calculations of average GPA for all students, etc.*

*In fact, more than GPA needs to be stored. The total grade-points earned and attempted must also be stored, as indicated on the diagram above.*

2. A web system deployed on campus has a failure rate of $10^{-3}$ failures per hour. What is the likelihood that the system will continue operating without failure throughout the duration of exam week (Monday through Friday)? (Show all work, both assuming failure in each hour is independent and that it is dependent.)

*Rate: $10^{-3}$failures/hr*
*Time: 5 days = 24 hr/day * 5 days = 120 hours*

*Assuming failure in any hour is independent:*
*P(success in all hours) = $(1-10^{-3})$*
*P(failure in ANY hour) = 1 – P(success in all hours) = 1-$(1-10^{-3})$ = 11.3%*

3. Two processes are each normally distributed and have the same mean, which is unknown. Process A takes 15ms/operation, while process B takes 17 ms/operation. The sample standard deviation, averaged over both processes, is 2.5 ms.

d. If these means come from 5 samples for each process, can we say that the processes are significantly different? (At roughly what p-value?)

*Formulas in bold are given.*

*D = (17-15) ms = 2 ms*
*dof = **2N-2** = (2)(5)-2 = 10-2 = 8*
$\sigma_{measured}$ *= 2.5 ms (given)*

$$t = \sqrt{\frac{N}{2}}\boldsymbol{D/\sigma_{measured}} = \sqrt{\frac{5}{2}}2ms/2.5ms = 1.26$$

*Now that we have the t value, we need to determine if it is significant. Below is the line on the t-table with the degrees of freedom determined above in a red rectangle, and the largest t value that is less than 1.26 circled in black (1.108). This corresponds to a two-tailed p-value of 0.30 … way too large to be significant.*

## *t* Table

| cum. prob | $t_{.50}$ | $t_{.75}$ | $t_{.80}$ | $t_{.85}$ | $t_{.90}$ | $t_{.95}$ |
|---|---|---|---|---|---|---|
| one-tail | **0.50** | **0.25** | **0.20** | **0.15** | **0.10** | **0.05** |
| two-tails | **1.00** | **0.50** | **0.40** | **0.30** | **0.20** | **0.10** |
| df | | | | | | |
| 1 | 0.000 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 |
| 2 | 0.000 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 |
| 3 | 0.000 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 |
| 4 | 0.000 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 |
| 5 | 0.000 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 |
| 6 | 0.000 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 |
| 7 | 0.000 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 |
| 8 | 0.000 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 |
| 9 | 0.000 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 |
| 10 | 0.000 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 |

e. If these means come from 10 samples for each process, can we say that the processes are significantly different? (At roughly what p-value?)

*Formulas in bold are given.*

*D = (17-15) ms = 2 ms*
*dof =* ***2N-2*** *= (2)(10)-2 = 20-2 = 18*
$\sigma_{measured}$ *= 2.5 ms (given)*

$$t = \boldsymbol{\sqrt{\frac{N}{2}} D/\sigma_{measured}} = \sqrt{\frac{10}{2}} 2ms/2.5ms = 1.79$$

*Now that we have the t value, we need to determine if it is significant. Below is the line on the t-table with the degrees of freedom determined above in a red rectangle, and the largest t value that is less than 1.79 circled in black (1.734). This corresponds to a two-tailed p-value of 0.10 …* ***could be significant.***

## *t* Table

| cum. prob | $t_{.50}$ | $t_{.95}$ | $t_{.975}$ | $t_{.99}$ | $t_{.995}$ |
|---|---|---|---|---|---|
| one-tail | 0.50 | 0.05 | 0.025 | 0.01 | 0.005 |
| two-tails | 1.00 | 0.10 | 0.05 | 0.02 | 0.01 |
| df | | | | | |
| 1 | 0.000 | 6.314 | 12.71 | 31.82 | 63.66 |
| 16 | 0.000 | 1.746 | 2.120 | 2.583 | 2.921 |
| 17 | 0.000 | 1.740 | 2.110 | 2.567 | 2.898 |
| 18 | 0.000 | 1.734 | 2.101 | 2.552 | 2.878 |
| 19 | 0.000 | 1.729 | 2.093 | 2.539 | 2.861 |
| 20 | 0.000 | 1.725 | 2.086 | 2.528 | 2.845 |
| 21 | 0.000 | 1.721 | 2.080 | 2.518 | 2.831 |

f. If these means come from 100 samples for each process, can we say that the processes are significantly different? (At roughly what p-value?)

*Formulas in bold are given.*

*D = (17-15) ms = 2 ms*
*dof =* ***2N-2*** *= (2)(100)-2 = 200-2 = 198*
$\sigma_{measured}$ *= 2.5 ms (given)*

$$t = \boldsymbol{\sqrt{\frac{N}{2}} D/\sigma_{measured}} = \sqrt{\frac{100}{2}} 2ms/2.5ms = 5.67$$

*Now that we have the t value, we need to determine if it is significant. Below is the line on the t-table with the degrees of freedom determined above in a red rectangle (because 198 is not on the chart, we are using 100 (which assumes less data) to be conservative), and the largest t value that is less than 5.67 circled in black (2.626).* ***Because 5.67 is off the chart, we can say that the p value is significantly less than 0.01 (****the smallest two-tailed p-value on the chart)* ***It is certainly significant.***

## *t* Table

| cum. prob | $t_{.50}$ | $t_{.95}$ | $t_{.975}$ | $t_{.99}$ | $t_{.995}$ |
|---|---|---|---|---|---|
| one-tail | 0.50 | 0.05 | 0.025 | 0.01 | 0.005 |
| two-tails | 1.00 | 0.10 | 0.05 | 0.02 | 0.01 |
| df | | | | | |
| 1 | 0.000 | 6.314 | 12.71 | 31.82 | 63.66 |
| 100 | 0.000 | 1.660 | 1.984 | 2.364 | 2.626 |
| 1000 | 0.000 | 1.646 | 1.962 | 2.330 | 2.581 |
| *z* | 0.000 | 1.645 | 1.960 | 2.326 | 2.576 |
| | 0% | 90% | 95% | 98% | 99% |
| | dence Level | | | | |

g. If we KNOW that the true standard deviation is 2 ms, what can we say about this process?

***We need more information. Must know how many samples used to determine the mean. Suppose 10 samples for each mean.***

*Formulas in bold are given.*

*D = (17-15) ms = 2 ms*
*dof =* ***"infinite" – use "z" instead of "t"***
*$\sigma_{measured}$ = 2.5 ms (given)*
$z = \boldsymbol{\sqrt{\frac{N}{2}} D/\sigma_{measured}} = \sqrt{\frac{10}{2}} 2ms/2ms$ *=2.23* ***(Note that actual sigma is different from the sample one we used above.)***

*Now that we have the t value, we need to determine if it is significant. Below is the line on the t-table with the degrees of freedom determined above in a red rectangle, and the largest t value that is less than 2.23 circled in black (1.960).* ***This corresponds to a two-tailed p-value of 0.05, which is signifant. (Only 5% chance that this happened by chance if the true means aren't different.)***

## *t* Table

| cum. prob | $t_{.50}$ | $t_{.95}$ | $t_{.975}$ | $t_{.99}$ | $t_{.995}$ |
|---|---|---|---|---|---|
| one-tail | 0.50 | 0.05 | 0.025 | 0.01 | 0.005 |
| two-tails | 1.00 | 0.10 | 0.05 | 0.02 | 0.01 |
| df | | | | | |
| 1 | 0.000 | 6.314 | 12.71 | 31.82 | 63.66 |
| 100 | 0.000 | 1.660 | 1.984 | 2.364 | 2.626 |
| 1000 | 0.000 | 1.646 | 1.962 | 2.330 | 2.581 |
| z | 0.000 | 1.645 | 1.960 | 2.326 | 2.576 |
| | 0% | 90% | 95% | 98% | 99% |
| | dence Level | | | | |