**Problem Set 3: RVs and MATLAB exercises**

**Worth ~7.5 points (18/2.4) + Bonus 1.67 points (4/2.4)**

**Administrative comments:**

1. Due date: Feb 10 (Mon), 11.59 pm. Remember: 5% off for each late hour!

*(The remaining administrative and general comments are all the same as for PS2.)*

1. This PS is meant to get you into the very trenches of data visualization in MATLAB. Rather than trying to set aside a large chunk of time on a single day to work on this problem set, splitting up work over a couple of days can help. Also, consider grappling with problem #2 last.
2. Upload your answer sheet and code to Gradescope. Note that there will be two assignment pages for each problem set, one for the PDF answer sheet and one for MATLAB code (single .m file, appropriately commented)
   1. Name your MATLAB code file as: ps#\_name.m (e.g., ps1\_shreesh.m)
   2. Within the code.m file, separate the code for each problem with a “section break”, which, in MATLAB, is obtained by inserting a line with %% followed by the problem number. (e.g., %% Problem 2)
   3. Name all your figures as: ps#\_name\_figX.pdf (ps1\_shreesh\_fig3.pdf, for problem 3).
3. In general, efficiency of the code will be valued, as will be validity in the choice of variable names (per Tutorial under MATLAB-REVIEW module) and the clarity of the comments. Note: using the % symbol before a line of text comments out that line.
4. Collaboration/discussion are permitted (encouraged, in fact). However, make sure that the solutions you turn in are your own and that you understand everything that you put in your solutions. The goal, here, is to have a solid foundation upon which we can build in the coming weeks.
5. Always abide by JHU honor code (see course syllabus).

**General tips and comments: For all the problems,**

1. Label the x,y (& z) axes appropriately. [Hint: “xlabel” (and ylabel and zlabel) ]
2. Make sure that the axis labels are of fontsize 15 [Hint: h=xlabel(“blood pressure”); set(h,”fontsize”,…)]
3. Set the font size of the x and y tick labels to be 10. [“help set”; the syntax that you will ultimately use will look like: set(gca,”fontsize”,…) ].
4. Provide a title for all plots [hint: “help title”]. Set title’s fontsize to be 18.
5. Use the command “print –dpdf figName” to save the current figure as a pdf file. The “current” figure is the one that you most recently clicked on with your mouse.

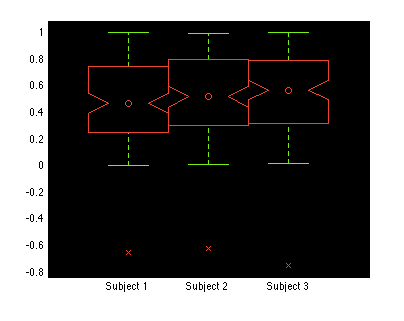
|  |  |  |  |
| --- | --- | --- | --- |
| **Qn** | **Points** | **Should take you (in min)** | **How long did it take? (approximately, min)** |
| **1** | **2** | **15’** |  |
| **2** | **2** | **15’** |  |
| **3** | **5** | **30’** |  |
| **4** | **5** | **15’** |  |
| **5** | **-** | **60’** |  |
| **6 (BONUS)** | **4** | **60’** |  |

\*Copy and paste this table into your answer sheet,

(or) just use this word document to type out your written anwers and then rename it as described in “c” above

**1. Random numbers (2 point).** Generate a vector with 30 random numbers drawn from a normal distribution with a mean of 3.2 and a variance of 16. (Hint: help randn)

**2. Variance property (2 points).** Var(X) = E[ (X-μx)2 ]. Derive what Var(aX) is. (Hint: Replace X with aX on the both sides of the formula and expand things out)

**3. Boxplot (5 points)**. Load the data set and produce a boxplot of the data to have it match (mostly) the shown figure. 

1. (1) Set color of boxes to red (don’t worry about turning some of the lines green, as in the adjacent plot).
2. (1) Plot the median as a dot instead of as a line.
3. (1) Change the outlier symbol to ‘x’
4. (1) Set the xticklabels to be ‘Subject 1’, ‘Subject 2’, and ‘Subject 3’.
5. (1) What does the interquartile range of a distribution mean (briefly and intuitively)? Show how you would calculate it for a dataset (say a vector x, containing 10 reaction times)?

For question e., write your answer down as comments below the code for a.-f (properly sectioned). Answer in equation(s) and sentence(s).

Hint: Everything you need for this, you should find under “help boxplot” + Slides\_MATLAB\_basics\_part2.ppt.

Dataset**: ps3\_3\_boxplot.mat**

**4 (5 points)** Given two random variables X and Y, what are the mathematical/statistical definitions for

* 1. Uncorrelatedness
  2. Independence
  3. Orthogonality

(Write out the ‘equation’ defining each)

* 1. Uncorrelated and zero mean RVs => independent. T/F? Explain briefly
  2. Uncorrelated => orthogonal. T/F? Explain briefly.

**5.** BONUS **(4 points)** Why is the sample variance calculated by dividing by (n-1) rather than n? Please show this explicitly with a derivation. (Hint: The answer to this was fully worked out at the end of Lecture 03A\_slides. I would like you to not only write it out, but also explain how / why you are able to go from one step to another. Either hand-written and attached as a picture in the same doc, or typed are fine.)