

Tips for Students

First of all, welcome to my research group, DatalinkX : Link data to Science. The following is some tips for you, which may be helpful to your research or life. I hope you enjoy your student life in my group.

Basic requirement

- **Master Student:** You should be professional to solve problems in your field.
- **Ph.D. Student:** You should propose new ideas on some problems in your field.

Requirement from advisor

- You should be responsible for yourself (Stay safe at any time) and your work (要有始有终).
- Finish your work as soon as possible.
- Writing your research dairy/report in detail.

About Research

First of all, **Conceptual novelty is all your need** ! that is, the idea is the most important ! **Your authorship, usually, is generated by the importance of your role, but not your time costed in this work.**

Directions of DatalinkX

- Developing statistical and machine-learning methods to have new scientific findings from massive of data
- Collaborating with other groups for exploring the mechanism of biological phenotypes

Directions of your research

- Biostatistics: You are trained to propose new data analysis problem from data and justify your method (algorithm) by real data examples.
- Statistics: You are trained to propose new statistical problem from real applications and justify your method by theoretical analysis.

Your contribution to DatalinkX (Requirement!)

Everyone in my group is required to involve in collaborative works. But, in such a collaborative work, your authorship is determined by **your advisor and the collaborator**.

Some Useful Tips

- Read more papers, as many as you can, to find a field that you are interested in.
- Be critical on results presented in papers.
- Work hard to give your understanding/perspective on the problem.
- Some research resources are available at <https://datalinkx.github.io/>.

Other suggestions

- Work hard and play hard.
- Do some exercises
- Talk with your advisor if you have any problems or confusions
- Be open minded to learn from other students, professors etc.

Recommendations on Courses

- Statistics/Probability: Bayesian Data analysis, Statistical Inference, Probability: A graduate course, Multivariate Statistical Analysis, Asymptotic Statistics, Nonparametric Statistics, Empirical Process, MCMC, Hidden Markov Model, Gaussian Process etc.
- Learning Theory: Machine/Statistical Learning, Information Theory, Convex optimization
- Biology: Genes XII, Molecular Cell Biology, Genetics: From Genes to Genomes etc.
- Programming: R, C/C++, Python.

如何高效地工作且拥有幸福生活

- **做好科研笔记**：严格按照导师的要求写研究报告，养成思考的好习惯，有序推进科研进展。
- **多总结反思**：时常确认自己的科研方向和人生方向，若发现有偏离要及时纠正。
- **多问为什么**：注意新旧知识之间的联系，把知识点串成线，组成面。弄清楚知识点或者问题的来龙去脉，把原理吃透。同时培养自己的正向思维和逆向思维。
- **多和他人交流学习**：充分学习各个组员和相关合作课题组的长处，博众家之长。
- **不断地扩展知识面**：全面扩展自己的眼界、提升自己的能力，做到厚积薄发。
 - 多读国外名校的名教授的专著和 TOP 期刊上的非华人作者的论文。这些材料开始要精度，后面可以泛读。
 - 关注有内容的公众号：比如，学术公众号会推荐一些最近的研究成果，有的知乎专栏也会专门推荐某些领域内近期的优秀成果；有的公众号会推送近期举办的学术会议或者学术报告等等。这些都是很好的学习资源。
 - 不断积累和提高自己的技能：比如编程能力，搜索能力，作图能力等等。
 - 收集和利用高效的小工具或者收集有用的案例：比如：**Mathpix Snipping** 可以准确的识别图片中的数学公式并转化为 latex 代码，**Notion** 是很灵活方便

的笔记工具；收集一些常用的技巧，比如 R 语言下的作图，拼图技巧；到知乎上学习其他人的经验等等

- **精通一种计算机语言：**利用良好的习惯和编程能力来减少重复性工作。对于数据分析而言，最好精通 R 或者 Python，另外要会 C/C++ (因为 R 和 Python 都不适合用来做循环运算)。
 - 用你精通的语言来处理重复性劳动：例子：可以用 R 语言实现重命名 1000 个文件、在服务器上批量提交任务、生成在 latex 中需要的 5 个 5*5 矩阵的 latex 代码 等等
 - 养成良好的编程习惯：代码结构清晰规范并配有注释。比如：一个分析对应一个 R 文件，一个图或者表对应一个 R 文件。每个 R 文件都包含数据读取、分析流程和结果输出三个部分。例子：当要修改某个图时，你可直接打开对应的 R 文件，对相应的代码做调整，然后重新跑一下代码即可，很方便。
 - 修改是常事，要做长远的考虑。不要觉得写程序来完成任务很麻烦而不做。通常，一份报告、一篇文章或者一个模拟需要多次修改，所以作图、做表都会是重复性劳动，一开始就用规范清晰的代码来实现会给以后带来很多便利。