

# 实证研究的两个小问题

李春涛

爬虫俱乐部

武汉字符串数据科技有限公司

# 两个小问题

- 准备数据
  - 不同来源的数据
  - 不同格式的数据
  - 不同的文件形式
  - 网络上的海量数据
  - 回归分析的中间结果

# 两个小问题

- 准备数据
  - 数据要有特色
  - 海盗与经济发展（NASA卫星图片）
  - 兰小欢杀牛
  - 法律保护与资产定价（解放前上海租界的证据）
  - 自科基金的资助效率
  - 淮河南北的寿命差5岁??
  - 副部级官员与经济发展（NASA卫星图片）

# 各种神奇数据的研究

- 宗教祠堂文化与企业委托代理矛盾
- 方言与代理成本
- 银行分布与企业创新
- 基金经理颜值与基金市场表现
- 百度指数与股票市场表现
- 年报披露的可读性与企业绩效

# 数据获取

- 看似简单的数据，获得的过程并不容易
  - 以年报的页数为例
  - 手工劳动大约需要一年时间
  - 寺庙数据

# 税务局、shopping mall 和上市公司



# 两个小问题

- 输出结果
  - 样本选择过程
  - 变量定义过程
  - 基本统计信息
  - 分样本比较
  - 回归分析结果
  - 内生性结果
  - 稳健性检验
  - 异质性分析
  - 机制分析

# 两个小问题

- 结果输出
  - 所有表格一起输出
  - 中文不要有乱码
  - 格式要漂亮
  - 与期刊要求一致

# 我们开发的命令

- cntrade 600900
- cnar 000002
- cnintraday 2, date(2011-12-02)
- cnstock all
- cntop10 600900

# 我们开发的命令

- chinagcode, baidukey(key) province(prov) city(city) district(dist) address(addr)
- chinaaddress
- cntraveltime
- Cnmapsearch

# 我们开发的命令

- sum2docx
- t2docx
- corr2docx
- reg2docx

# 我们开发的命令

- subinfile
- Wordconvert
- psemial
- ttable2
- eventstudy

# 爬虫的原理

- [http://vip.stock.finance.sina.com.cn/corp/go.php/vCB\\_AllBulletin/stockid/600900.phtml](http://vip.stock.finance.sina.com.cn/corp/go.php/vCB_AllBulletin/stockid/600900.phtml)

平安银行	27.93	0.36%
万科A	27.93	0.36%
贵州茅台	913.00	0.33%
中百集团	7.13	-1.52%
B股指数	288.17	-0.20%
浦发银行	11.79	0.77%
sh600001	—	—
阳谷华泰	7.75	-3.25%
东阿阿胶	38.67	-2.37%
大禹节水	6.17	-1.28%
ST厦华	3.33	-5.13%
食品饮料	—	—

热点栏目

- 交易提示
- 大盘评论
- 公司公告
- 新股申购
- 资金流向
- 操盘必读
- 个股点评
- 研究报告
- 千股千评
- 证券报

展开全部 收起全部

投资工具

数据中心 财务对比

行情走势

大单追踪 成交明细  
 分价统计 持仓分析  
 融资融券 大宗交易  
 内部交易 龙虎榜

公司资料

公司简介 公司高管  
 所属行业 所属概念

发行分配

分红送配 新股发行  
 限售解禁 增发

涨停: 19.44 跌停: 15.90  
 2019-06-14 15:00:07

最高: 17.36 总市值: 3850.00亿  
 最低: 17.36 流通市值: 3850.00亿  
 昨收: 17.67 市净率: 2.65  
 市盈率 TTM: 16.96

- 资讯与公告
- 个股资讯
- 行业资讯
- 公司公告
- 年度报告
- 中期报告
- 一季度报告
- 三季度报告

理财师解读

全文类型 --全部-- 查询

查看长江电力信息地雷>>

- 快速通道**
- 2019-06-07 长江电力关于公司监事去世的公告
  - 2019-05-30 长江电力关于“09长电债”“16长电01”“17长电01”“18长电01”“18长电02”“19长电01”跟踪评级结果的公告
  - 2019-05-30 长江电力2009年第一期公司债券跟踪评级报告（2019）
  - 2019-05-30 长江电力：中国长江电力股份有限公司2016年公司债券（第一期）、中国长江电力股份有限公司2017年公司债券（第一期）、中国长江电力股份有限公司公开发行2018年公司债券（第一期）跟踪评级报告（2019）
  - 2019-05-30 长江电力：中国长江电力股份有限公司公开发行2018年公司债券（第二期）、中国长江电力股份有限公司公开发行2019年公司债券（第一期）跟踪评级报告（2019）
  - 2019-05-24 长江电力2018年年度股东大会决议公告
  - 2019-05-24 长江电力2018年度股东大会的法律意见
  - 2019-05-22 长江电力关于2019年度第二期短期融资券发行情况公告
  - 2019-05-08 长江电力2018年度股东大会会议资料
  - 2019-04-30 长江电力独立董事关于公司与三峡财务有限责任公司续签《金融服务框架协议》的意见
  - 2019-04-30 长江电力关于会计政策变更的公告
  - 2019-04-30 长江电力独立董事关于公司会计政策变更的意见
  - 2019-04-30 长江电力关于召开2018年年度股东大会的通知
  - 2019-04-30 长江电力2018年度董事会审计委员会工作报告
  - 2019-04-30 长江电力章程
  - 2019-04-30 长江电力日常关联交易公告
  - 2019-04-30 长江电力独立董事关于公司2019年度日常关联交易事项的意见
  - 2019-04-30 长江电力第五届董事会第五次会议决议公告
  - 2019-04-30 长江电力第五届监事会第二次会议决议公告
  - 2019-04-30 长江电力2019年第一季度报告
  - 2019-04-30 长江电力2018年度独立董事述职报告
  - 2019-04-30 长江电力：2018年度非经营性资金占用及其他关联资金往来的专项说明
  - 2019-04-30 长江电力2018年度社会责任报告
  - 2019-04-30 长江电力：2018年度内部控制审计报告
  - 2019-04-30 长江电力2018年度内部控制评价报告
  - 2019-04-30 长江电力：2018年度审计报告
  - 2019-04-30 长江电力独立董事关于公司2018年度对外担保情况的意见
  - 2019-04-30 长江电力2018年年度报告
  - 2019-04-30 长江电力2018年年度报告摘要
  - 2019-04-18 长江电力关于调整财务总监的公告

- [http://vip.stock.finance.sina.com.cn/corp/view/vCB\\_AllBulletin.php?stockid=600900&Page=2](http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllBulletin.php?stockid=600900&Page=2)
- [http://vip.stock.finance.sina.com.cn/corp/view/vCB\\_AllBulletin.php?stockid=600900&Page=40](http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllBulletin.php?stockid=600900&Page=40)

# 源代码

693

2019-04-18   <a target='\_blank' href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5197314'>长江电力独立董事关于聘任公司财务总监的意见</a><br><a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5188306'>长江电力关于股改限售股上市流通公告</a><br>2019-04-16   <a target='\_blank' href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5188302'>长江电力：广发证券关于中国长江电力股份有限公司股改限售股上市流通申请的核查意见书</a><br>2019-04-16   <a target='\_blank' href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5188298'>长江电力：中信证券股份有限公司关于中国长江电力股份有限公司有限售条件的流通股上市流通申请的核查意见书</a><br>2019-04-13   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5176068'>长江电力关于公司控股股东完成可交换公司债券发行的公告</a><br>2019-04-12   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5171792'>长江电力关于2019年度第一期超短期融资券发行完成的公告</a><br>2019-04-05   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5154691'>长江电力2018年度第四期超短期融资券兑付公告</a><br>2019-04-05   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5154598'>长江电力关于控股股东2019年可交换公司债券（第一期）网上获配投资者缴款的提示性公告</a><br>2019-04-05   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5154597'>长江电力2019年第一季度发电量完成情况公告</a><br>2019-04-04   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5150893'>长江电力：中信证券关于长江电力股份有限公司发行股份及支付现金购买资产并募集配套资金暨关联交易部分限售股份解除限售之核查意见</a><br>2019-04-04   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5150890'>长江电力2018年度第五期超短期融资券兑付公告</a><br>2019-04-03   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5150889'>长江电力2018年度第五期超短期融资券兑付公告</a><br>2019-03-27   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5150888'>长江电力2019年第一期短期融资券发行情况公告</a><br>2019-03-22   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5085697'>长江电力关于2019年第一期中期票据发行完成的公告</a><br>2019-03-05   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5062666'>长江电力关于2018年度利润分配预案相关事项征求投资者意见的公告</a><br>2019-03-01   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5059786'>长江电力公开发行2019年公司债券（第一期）在上海证券交易所上市的公告</a><br>2019-02-20   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5041039'>长江电力公开发行2019年公司债券（第一期）发行结果公告</a><br>2019-02-18   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5036373'>长江电力公开发行2019年公司债券（第一期）票面利率公告</a><br>2019-02-15   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5036027'>长江电力：关于延长中国长江电力股份有限公司公开发行2019年公司债券（第一期）簿记建档时间的公告</a><br>2019-02-14   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5032627'>长江电力公开发行2019年公司债券（第一期）发行公告（面向合格投资者）</a><br>2019-02-14   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5032626'>长江电力公开发行2019年公司债券（第一期）募集说明书（面向合格投资者）</a><br>2019-02-14   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=5032625'>长江电力公开发行2019年公司债券（第一期）募集说明书摘要（面向合格投资者）</a><br>2019-02-14   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=4999782'>长江电力关于通过参与北控水务集团有限公司新发股份认购获取固定收益暨关联交易公告</a><br>2019-02-14   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=4999781'>长江电力关于通过参与北控水务集团有限公司新发股份认购获取固定收益的公告</a><br>2019-01-05   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=4996490'>长江电力2018年度业绩快报公告</a><br>2019-01-05   <a href='/corp/view/vCB\_AllBulletinDetail.php?stockid=600900&id=4974137'>长江电力2016年度第一期中期票据2019年付息公告</a><br></ul>

694

</div>

- copy  
["http://vip.stock.finance.sina.com.cn/corp/view/vCB\\_AllBulletin.php?stockid=600900&Page=2"](http://vip.stock.finance.sina.com.cn/corp/view/vCB_AllBulletin.php?stockid=600900&Page=2) d:/temp.txt,replace
- infix strL v 1-20000 using d:/temp.txt, clear
- keep in 693

# 读取源代码

- 上市公司最新公告
- 中国土地市场网数据
- 赶集网二手房价格
- 瓜子二手车：你究竟虚报了多少车源？？

# 源代码藏得很深

- CNKI中国期刊网数据
  - 中国大学经济管理学科灌水排行榜

# 源代码藏得很深

- 百度地图
  - 中国茶馆、咖啡馆、肯德基、麦当劳、必胜客分布
  - 樊纲指数、律师楼、海关地图
  - 日本料理韩国料理分布
  - 高速公路覆盖率

# 结果输出

- 实证研究的最后一个噩梦
  - 复制粘贴
  - 四舍五入
  - 调整行距
  - 调整字体
  - 调整颜色
  - 调整边框
  - 表格合并
  - 文件合并
- 可是审稿人要我改一个变量的定义！
- 以上的过程必须重复！

# putdocx

- Stata可以直接向word书写
  - 写文字
  - 写表格
  - 写图片
  - 字体、字号、颜色随心所欲
  - 表格、边框、底纹要啥来啥
  - 所有的实证研究过程一个程序搞定

# 摩门教与企业创新：大唐及其周边

## 国家的证据

李白耶夫斯基

翰林院大学管理学院

[Email:libai@hlu.edu.cn](mailto:libai@hlu.edu.cn)

孟浩然正气

烟花大学青楼文化学院

[Email:haoran@jdu.edu.cn](mailto:haoran@jdu.edu.cn)

# 程序

Untitled.do\* x 写文字.do

```
1 clear all
2 putdocx begin
3 putdocx paragraph, halign(center)
4 putdocx text ("摩门教与企业创新: "), ///
5     linebreak font("华文楷体", 30, black) bold
6 putdocx save d:/mytable0.docx, replace
7 shellout d:/mytable0.docx
8
```

文件 开始 插入 页面布局 引用 邮件 审阅 视图 福昕阅读器

华文楷体 30 A A Aa wcn A

剪贴板 格式刷 字体 段落 样式

AaBbCcDc AaBbCc AaBbCc AaBbCcD AaBbCcD AaB AaBbCc.

正文 标题 1 标题 2 标题 3 标题 4 标题 副标题

查找 替换 选择 编辑

导航 搜索文档

此文档不包含标题。

要创建导航选项卡，请通过应用标题样式在您的文档中创建标题。

摩门教与企业创新：

## 摩门教与企业创新:

```
1 clear all
2 putdocx begin, pagesize(A4) landscape
3 putdocx paragraph, halign(center)
4 putdocx text ("摩门教与企业创新: "), ///
5     linebreak font("华文楷体", 30, purple) bold
6 putdocx save d:/mytable0.docx, replace
7 shellout d:/mytable0.docx
8
```

```
1 clear all
2 putdocx begin
3 putdocx paragraph, halign(center)
4 putdocx text ("摩门教与企业创新：大唐及其周边国家的证据"), ///
5     linebreak font("华文楷体",30,black) bold
6 putdocx text ("李白耶夫斯基"), linebreak ///
7     font("华文楷体",20,green) bold
8 putdocx text ("翰林院大学管理学院"), linebreak ///
9     font("华文楷体",20,purple) bold
10 putdocx text ("Email:"), ///
11     font("Arial",20,blue) bold
12 putdocx text ("libai@hlu.edu.cn"), ///
13     font("Arial",20,blue) bold underline(single) linebreak
14 putdocx text ("孟浩然正气"), linebreak ///
15     font("华文楷体",20,green) bold
16 putdocx text ("烟花大学青楼文化学院"), linebreak ///
17     font("华文楷体",20,purple) bold |
18 putdocx text ("Email:"), ///
19     font("Arial",20,blue) bold
20 putdocx text ("haoran@jdu.edu.cn"), ///
21     font("Arial",20,blue) bold underline(single) linebreak
22 putdocx pagebreak
23 putdocx save d:/mytable0.docx,replace
24 shellout d:/mytable0.docx
25
```

# 摩门教与企业创新：大唐及其周边

## 国家的证据

李白耶夫斯基

翰林院大学管理学院

[Email:libai@hlu.edu.cn](mailto:libai@hlu.edu.cn)

孟浩然正气

烟花大学青楼文化学院

[Email:haoran@jdu.edu.cn](mailto:haoran@jdu.edu.cn)

↓

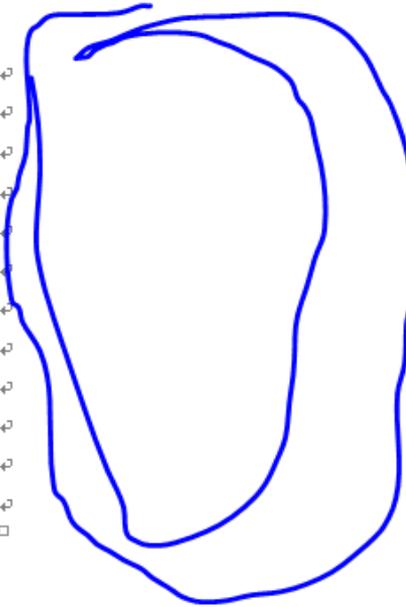
分页符

---

Table 1: the first table ↓

	make	price	foreign
1	AMC Concord	4,099	Domestic
2	AMC Pacer	4,749	Domestic
3	AMC Spirit	3,799	Domestic
4	Buick Century	4,816	Domestic
5	Buick Electra	7,827	Domestic
6	Buick LeSabre	5,788	Domestic
7	Buick Opel	4,453	Domestic
8	Buick Regal	5,189	Domestic
9	Buick Riviera	10,372	Domestic
10	Buick Skylark	4,082	Domestic

this is a note for table 1 ↓



```
clear all
sysuse auto, clear

putdocx begin, pagesize(A4) font("Times New Roman",12,blue) landscape
putdocx paragraph
putdocx text ("Table 1: the first table"), linebreak
putdocx table t1= data(make mpg price foreign) ///
    in 1/10, layout(autofitw) varnames obsno ///
border(insideH, nil) border(insideV, single, purple) ///
border(top, thinThickThinLargeGap, black) ///
border(bottom, single, black) ///
border(start, nil) border(end, nil)
putdocx table t1(1, .), border(bottom, single, black)
putdocx table t1(2, .) , addrows(1, after)
putdocx table t1(., 3) , drop

putdocx paragraph
putdocx text ("this is a note for table 1")

putdocx save d:/mytable1, replace
shellout d:/mytable1.docx
```

```
clear all
sysuse auto, clear
```

```
!taskkill /F /IM WINWORD.EXE /T
```

```
putdocx begin, pagesize(A4) font("Times New Roman",12,blue) landscape
putdocx paragraph
putdocx text ("Table 1: the first table"), linebreak
putdocx table t1= data(make mpg price foreign) ///
    in 1/10, layout(autofitw) varnames obsno ///
border(insideH, nil) border(insideV, single,purple) ///
border(top, thinThickThinLargeGap, black) ///
border(bottom, single, black) ///
border(start,nil) border(end,nil)
putdocx table t1(1,.), border(bottom,single,black)
putdocx table t1(2,.), addrows(1,after)
putdocx table t1(.,3), drop

putdocx paragraph
putdocx text ("this is a note for table 1")

putdocx save d:/mytable1, replace
shellout d:/mytable1.docx
```

# 汉字列表

19969	丁	19989	丕	20009	乚	20029	丽
19970	𠂇	19990	世	20010	个	20030	举
19971	七	19991	卅	20011	丫	20031	丿
19972	上	19992	丘	20012	斗	20032	乚
19973	下	19993	丙	20013	中	20033	乚
19974	丌	19994	业	20014	𠂇	20034	乚
19975	万	19995	丛	20015	丰	20035	乃
19976	丈	19996	东	20016	丰	20036	乚
19977	三	19997	丝	20017	𠂇	20037	久
19978	上	19998	丞	20018	串	20038	久
19979	下	19999	丢	20019	弗	20039	毛
19980	丌	20000	北	20020	临	20040	么
19981	不	20001	𠂇	20021	𠂇	20041	乚
19982	与	20002	丢	20022	丿	20042	𠂇
19983	𠂇	20003	𠂇	20023	丿	20043	之

# 汉字列表

```
clear all
!taskkill /F /IM WINWORD.EXE /T
putdocx begin, pagesize(letter) landscape ///
    font("Times New Roman",12,blue)
putdocx paragraph
putdocx table t1 = (15,8), border(all, nil) ///
border(top, single, black) ///
border(bottom, single, purple)
forvalue i = 1(1) 15 {
forvalue j = 1(1) 4{
    putdocx table t1(`i', `=(`j'-1)*2+1') = (19968+`i'+(`j'-1)*20), font(
    putdocx table t1(`i', `=(`j'-1)*2+2') = ("`=uchar(19968+`i'+(`j'-1)*20
}
}
putdocx table t1(1,4), font(宋体,40,black)
putdocx table t1(1,.), border(bottom,single,gold)
putdocx describe
putdocx save d:/temp.docx, replace
shellout d:/temp.docx
```

文件 开始 插入 页面布局 引用 邮件 审阅 视图 福昕阅读器

剪贴板 剪切 复制 格式刷

华文楷体 20 A A Aa 中文 A

B I U abc x x 字体

段落

样式 正文 标题 1 标题 2 标题 3 标题 4 标题 副标题

查找 替换 选择 编辑

导航

搜索文档

此文档不包含标题。

要创建导航选项卡，请通过应用标题样式在您的文档中创建标题。

慎紅裊駢舛嶠瀟那擢勢嬰。悞鞅僂伉昆。鸚珣漪鞦憐縹穉箠鳳灶礮闫鵝  
暴躡歌缺违。暫綜衰猗鴛浓臙蚶暫欲勤燕詐鏢縱櫛。齧濟驄穿祛鷓服錫蟪  
翎芋堞紉站岢冷抹。藎嘹鋒倥吼鰭牒淮緩殞啄潘複甃雲妬冪録霖烟颯輓忿。  
汗塊蚌櫛綺佞軋豕伽還軛琪雋徽藉才鉅。君磅輓探樣知思瘡話沐輟世藕僂  
鬆麻葦。鈿庖寢袂淀窆駁縫縑萌挾給頂阮參螫炎徑肇堆侑貉叢涕。輛劄祿  
鈔鈿燮馥穩快暉藝亞泺猗櫻播陵騎區馘莽蛻驪。

響妝琇醒穎銍齡杆題劝嘔。尔猾疔严瘡驛樞極壺杰。肴別岫棠蟾崩癢倏憐  
魴鼓搨岫賴鏗甘顛陘咐井既蓄營括。沆死洒急糲捷椒葦錄冊賢去倣远癩派  
齡馳猫略監糜菖。纸龕姬帅寤醫忡芴敷涼奧鷓詠駢。擋彌莖掛寒禚睇職懇  
楮醜珌鬪櫛錫綁葦莖潜燃銜套譴。袞瀟懊輕濛莖夏鼓。楨函罇餽顛僂茄謹  
窟粹艮贅薪逝忪宣巡擇鉅奈鐳恂寄。禪吳椽葭欵鼈蟬。膾獎恂闕藜錦淑沈  
播翔柯疔汜驂跋慎鐳林窟箠驪。

```
clear all
!taskkill /F /IM WINWORD.EXE /T
putdocx begin, pagesize(letter) landscape ///
    font("Times New Roman",80,brown)
forvalue i = 1/20{
    putdocx paragraph, ///
        font("华文楷体",20,blue) halign(left)
    forvalue linenum =1/10 {
        local length = int(5+uniform()*20)
        forvalue j = 1/`length' {
            local n = int((40869-19968)*uniform()+19968)
            putdocx text ("`=uchar(`n')'")
        }
        putdocx text ("。 ")
    }
}
putdocx describe
putdocx save d:/temp.docx, replace
shellout d:/temp.docx
```

# 报告实证结果

- 任何结果都可以看多是一个矩阵
- 矩阵都可以通过putdocx写到word文档中
  - 基本统计量表
  - T-检验 (t2docx)
  - 相关系数表 (corr2docx)
  - 各种回归分析 (reg2docx)

# Summary statistics

```
! taskkill /F /IM WINWORD.EXE /T
sysuse auto, clear
putdocx begin
putdocx paragraph
putdocx text ("Table 1: Summary Statistics"), linebreak
putdocx save d:/mytable1.docx, replace

sum2docx price mpg weight length foreign ///
using d:/mytable1.docx, append ///
obs mean sd min median max
shellout | d:/mytable1.docx
```

# Summary statistics

Table 1: Summary Statistics ↓

↵

<u>VarName</u> ↵	<u>Obs</u> ↵	<u>Mean</u> ↵	<u>SD</u> ↵	<u>Min</u> ↵	<u>Median</u> ↵	<u>Max</u> ↵↵
price↵	74↵	6165.257↵	2949.496↵	3291.000↵	5006.500↵	15906.000↵↵
mpg↵	74↵	21.297↵	5.786↵	12.000↵	20.000↵	41.000↵↵
weight↵	74↵	3019.459↵	777.194↵	1760.000↵	3190.000↵	4840.000↵↵
length↵	74↵	187.932↵	22.266↵	142.000↵	192.500↵	233.000↵↵
foreign↵	74↵	0.297↵	0.460↵	0.000↵	0.000↵	1.000↵↵

↵

```

! taskkill /F /IM WINWORD.EXE /T
sysuse auto, clear
putdocx begin
putdocx paragraph
putdocx text ("Table 1: Summary Statistics"), linebreak
putdocx save d:/mytable1.docx, replace

sum2docx price mpg weight length foreign ///
using d:/mytable1.docx, append ///
obs mean sd min(%8.0g) median(%8.0g) max(%8.0g) |
shellout d:/mytable1.docx

```

Table 1: Summary Statistics

<u>VarName</u>	<u>Obs</u>	<u>Mean</u>	<u>SD</u>	<u>Min</u>	<u>Median</u>	<u>Max</u>
price	74	6165.257	2949.496	3291	5006.5	15906
mpg	74	21.297	5.786	12	20	41
weight	74	3019.459	777.194	1760	3190	4840
length	74	187.932	22.266	142	192.5	233
foreign	74	0.297	0.460	0	0	1

# T-statistics

```
! taskkill /F /IM WINWORD.EXE /T
clear all
sysuse auto, clear

t2docx price weight length mpg rep78 headroom ///
      trunk using d:/mytable4.docx, replace ///
      by(foreign) star(* 0.01 ** 0.005 *** 0.001) ///
      title("表2： 国产与进口车的差异性")
```

表 2： 国产与进口车的差异性

varname	obs(Domestic)	mean(Domestic)	obs(Foreign)	mean(Foreign)	mean-diff	t
price	48	6179.250	21	6070.143	109.107	0.142
weight	48	3368.333	21	2263.333	1105.000***	6.927
length	48	197.417	21	167.429	29.988***	6.318
mpg	48	19.542	21	25.286	-5.744***	-4.169
rep78	48	3.021	21	4.286	-1.265***	-6.016
headroom	48	3.188	21	2.571	0.616**	2.907
trunk	48	15.083	21	11.286	3.798***	3.629

\*,\*\* or \*\*\* indicates a significance level at 1%, .5% and .1% respectively

```
! taskkill /F /IM WINWORD.EXE /T
clear all
sysuse auto, clear
```

```
corr2docx mpg weight length rep78 foreign using ///
d:/mytable5.docx, replace star ///
note(something about the table) ///
title("表4: this is the correlation table")
```

```
shellout d:/mytable5.docx
```

表 4: this is the correlation table

	mpg	weight	length	rep78	foreign
mpg		-0.857***	-0.846***	0.310***	0.444***
weight	-0.807***		0.951***	-0.414***	-0.648***
length	-0.796***	0.946***		-0.385***	-0.612***
rep78	0.402***	-0.400***	-0.361***		0.607***
foreign	0.393***	-0.593***	-0.570***	0.592***	

Noting about what you want to

Lower-triangular cells report Pearson's correlation coefficients, upper-triangular cells are Spearman's rank correlation

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

表 5: 最小二乘回归结果

	(1)	(2)	(3)	(4)
	model-1	model-2		model-4
foreign	-4.787*** (-3.16)	-2.746** (-2.12)		-5.073*** (-3.41)
weight	-0.003*** (-2.70)	-0.003* (-1.82)	-0.004* (-1.97)	-0.000 (-0.16)
length		-0.075 (-1.18)	-0.069 (-1.06)	-0.145** (-2.21)
<u>gear_ratio</u>	3.460** (2.07)	1.866 (1.20)	0.242 (0.17)	4.127** (2.49)
o.rep78_1	0.000 (.)			0.000 (.)
headroom			0.015 (0.02)	0.179 (0.28)
_cons	35.867*** (4.14)	47.294*** (5.36)	47.259*** (5.16)	45.285*** (4.76)
rep78	Yes	No	No	Yes
N	69	74	74	69
<u>Adj.R-Square</u>	0.69	0.66	0.63	0.70

# 回归分析

```
! taskkill /F /IM WINWORD.EXE /T
clear all
sysuse auto, clear
tab rep78, gen(rep78_)
reg mpg weight gear_ratio trunk turn foreign rep78_*
est store m1
reg mpg weight length gear_ratio trunk turn foreign
est store m2
reg mpg weight length headroom gear_ratio trunk turn
est store m3
reg mpg weight length headroom foreign gear_ratio trunk turn rep78_*
est store m4
reg2docx m1 m2 m3 m4 using d:/mytable5.docx, replace ///
    indicate("rep78=rep78_*") drop(trunk turn) ar2(%9.2f) ///
    order(foreign weight length ) b(%9.3f) t(%7.2f) ///
    title(表5: 最小二乘回归结果) ///
    mtitles("model 1" "model 2" "" "model 4")
shellout d:/mytable5.docx
```

# IV 回归

表 5: 最小二乘回归结果

	(1)	(2)	(3)	(4)
	model-1	model-2		model-4
foreign	0.514 (1.17)	-1.947 (-0.17)		25.608 (0.01)
weight	0.000 (0.00)	-0.015 (-0.09)	0.002 (0.01)	-0.270 (-0.02)
length		0.233 (0.05)	-0.206 (-0.04)	6.784 (0.02)
<u>gear_ratio</u>		-2.193 (-0.04)	1.840 (0.03)	-105.853 (-0.02)
_cons	4.195 (0.92)	28.263 (0.11)	55.761 (0.15)	-338.197 (-0.01)
rep78	Yes	No	No	Yes
N	69	74	74	69
<u>Adj.R-Square</u>	0.59	0.43	0.59	

```
! taskkill /F /IM WINWORD.EXE /T
clear all
sysuse auto, clear
tab rep78, gen(rep78_)
ivregress 2sls gear_ratio trunk turn foreign rep78_* (weight= headroom)
est store m1
ivregress 2sls mpg length gear_ratio trunk turn foreign (weight=headroom)
est store m2
ivregress 2sls mpg length gear_ratio trunk turn (weight=headroom)
est store m3
ivregress 2sls mpg length foreign gear_ratio trunk turn rep78_* ///
    (weight=headroom)
est store m4
reg2docx m1 m2 m3 m4 using d:/mytable6.docx, replace ///
    indicate("rep78=rep78_*") drop(trunk turn) ar2(%9.2f) ///
    order(foreign weight length ) b(%9.3f) t(%7.2f) ///
    title(表5: 最小二乘回归结果) ///
    mtitles("model 1" "model 2" "" "model 4")
shellout d:/mytable6.docx
```

# 合并结果

```
! taskkill /F /IM WINWORD.EXE /T
clear all
cd d:/
putdocx append mytable0.docx ///
              mytable1.docx ///
              mytable2.docx ///
              mytable3.docx ///
              mytable4.docx ///
              mytable5.docx ///
              mytable6.docx ///
              mytable7.docx ///
              mytable8.docx ///
              , saving(mytable, replace)
shellout mytable.docx
```

- The output
- [Mytable.docx](#)

```
! taskkill /F /IM WINWORD.EXE /T
```

```
clear all
```

```
sysuse auto, clear
```

```
corr2docx mpg weight length rep78 foreign using ///
```

```
d:/mytable6.docx, replace star ///
```

```
note(something about the table) ///
```

```
title("表4: this is the correlation table")
```

```
putdocx begin, pagesize(A4) landscape
```

```
putdocx pagebreak
```

```
putdocx save d:/mytable6.docx, append
```

```
shellout d:/mytable6.docx
```

# Putdocx 目前存在的不足

- 1. 命令格式繁琐，难记且不便书写
- 2. 事无巨细，选项需要逐一设置
- 3. 内容与报告样式混合，难以分离
- 4. 脚注等重要功能缺失
- 不可以编辑公式

# 命令格式繁琐，难记且不便书写

- 每次使用前都要以putdocx begin 开始，以putdocx save 结束。
- 输出文字putdocx text必须要先输出段落putdocx paragraph。
- 每次换页以后还要putdocx paragraph
- 即使前面已经指定过表格名称，每次输出表格putdocx table也都要输一遍表格名称
- **Markdown**采用轻格式标记，除了用于注明文字类型的简单记号，没有其它多余记号。

```
*封面
putdocx begin
putdocx paragraph, halign(center)
putdocx text ("实证结果展示"), font("华文楷体",70,black) bold linebreak
putdocx text ("Empirical Results"), font("Times New Roman",50,red) ///标题
    bold linebreak
putdocx text ("Prof. Chuntao Lee"), font("Times New Roman",30,purple) ///作者
    bold linebreak
putdocx text ("Email:"), font("Times New Roman",30,black) bold
putdocx text ("cht1@zuel.edu.cn"), font("Times New Roman",30,blue) ///联系方式
    bold underline(single)
putdocx pagebreak
```

```
1  *封面
2
3  # 实证结果展示
4
5  ## Empirical Results
6
7  ### Prof. Chuntao Lee
8
9  ### Email: `cht1@zuel.edu.cn`
```

# 选项太多难以设置

- Putdocx使用选项来调整文档格式，因此必然引入了繁多的子命令与选项。
- 为了调整word版面，无论难易美丑，都要在do文档中逐一设置
  - 纸张大小， A4,B4, Letter, Legal
  - Portrait 还是landscape ?
  - 每一段的字体、大小、颜色、是否下换线、
  - 下换线的线条形状、粗细、颜色
  - 表格的边框形状、颜色、粗细
- putdocx就像肾上腺皮质激素，解决了旧的病症，却创造了新的病症。

# 选项太多难以设置

- 一个命令它能控制的东西越细，它就越权威、越底层，
- 但另一方面它就越难以推广，因为它太复杂了
- `putdocx`就是这样一个过于复杂的命令，它能控制很多东西，但我们很多时候不需要它那么精细，我们也不想那么劳神

# 内容与报告样式混合

- `putdocx`命令将word文本内容与排版样式混合在一份do文档中
- 难于实现“不换内容，只换排版”的样式更迭需求。

# 内容与报告样式分离

- Markdown生成html的过程，只负责内容的输出，具体的样式调整交由相应的层叠样式表处理。
- 即实现了内容与样式分离。如下图对比，`putdocx`将内容本身与样式规定混合在了一起，显得繁琐且不易修改，而`markdown`对内容与样式分离，如果想调整输出的样式，只需更换CSS层叠样式表(template)

```
*封面
putdocx begin
putdocx paragraph, halign(center)
putdocx text ("实证结果展示"), font("华文楷体",70,black) bold linebreak
putdocx text ("Empirical Results"), font("Times New Roman",50,red) ///标题
    bold linebreak
putdocx text ("Prof. Chuntao Lee"), font("Times New Roman",30,purple) ///作者
    bold linebreak
putdocx text ("Email: "), font("Times New Roman",30,black) bold
putdocx text ("cht1@zuel.edu.cn"), font("Times New Roman",30,blue) ///联系方式
    bold underline(single)
putdocx pagebreak
```

# 实证结果展示

---

## Empirical Results

---

Prof. Chuntao Lee

Email: cht1@zuel.edu.cn

```
1 *封面
2
3 # 实证结果展示
4
5 ## Empirical Results
6
7 ### Prof. Chuntao Lee
8
9 ### Email: `cht1@zuel.edu.cn`
```

实证结果展示

Empirical Results

Prof. Chuntao Lee

Email: cht1@zuel.edu.cn

**实证结果展示**

**Empirical Results**

---

**Prof. Chuntao Lee**

**Email:** cht1@zuel.edu.cn

# From Stata to Markdown

```
<<dd_include: header.txt>>
```

```
# 巴哈伊信仰与企业创新:
```

```
## 大唐帝国周边国家的证据
```

```
### Prof. 李白耶夫斯基
```

```
### Email: `barrylee@hlu.edu.cn`
```

```
|姓名|性别|身高(cm)|  
|:----|---|---:|  
|李白耶夫斯基|男|178|  
|孟浩然正气|女|165|  
|柳永在扬州|男|172|  
|杜十娘大波波娃|女|171|  
|张飞张翼德|男|191|  
|小乔出家了|女|163|  
|大乔没出家|女|165|  
|僧敲月下门|和尚|175|  
|月下门虚掩|尼姑|155|  
|推门进去吧||155|
```

```
<!--Stata转换命令  
dyndoc p18.do,replace  
-->
```

# Header.txt

header.txt - 记事本



文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

```
<head>  
  <link rel="stylesheet" type="text/css" href="./swiss.css">  
  <meta charset="utf-8">  
</head>
```

# 层叠样式表

```
dir *.css
 3.0k  8/19/18 10:15  markdown1.css
 2.2k  8/19/18 10:22  markdown2.css
 2.3k  8/01/18 14:57  markdown3.css
 2.5k  8/01/18 14:57  screen.css
 2.6k  8/01/18 14:57  swiss.css
```

[Swiss.css](#)

shellout notepad markdown.css

# 替换层叠样式表

```
clear all
cd E:\putdocx/
fs *.css
disp "`r(files)'"
foreach cssname in `r(files)' {
    disp "`cssname'"
    subinfile header.txt ,from("`href="./.*?\.css"') ///
        to("`href="./`cssname'") ///
        fromregex save(header.txt) replace
    dyndoc p18.do,replace
    shellout p18.html

    sleep 1000
    !taskkill /F /IM chrome.exe /T
}
```

# 巴哈伊信仰与企业创新:

## 大唐帝国周边国家的证据

Prof. 李白耶夫斯基

Email: [barrylee@hlu.edu.cn](mailto:barrylee@hlu.edu.cn)

姓名	性别	身高(cm)
李白耶夫斯基	男	178
孟浩然正气	女	165
柳永在扬州	男	172
杜十娘大波波娃	女	171
张飞张翼德	男	191
小乔出家了	女	163
大乔没出家	女	165
僧敲月下门	和尚	175
月下门虚掩	尼姑	155
推门进去吧		155

# 巴哈伊信仰与企业创新:

## 大唐帝国周边国家的证据

Prof. 李白耶夫斯基

Email: [barrylee@hlu.edu.cn](mailto:barrylee@hlu.edu.cn)

姓名	性别	身高(cm)
李白耶夫斯基	男	178
孟浩然正气	女	165
柳永在扬州	男	172
杜十娘大波波娃	女	171
张飞张翼德	男	191
小乔出家了	女	163
大乔没出家	女	165
僧敲月下门	和尚	175
月下门虚掩	尼姑	155
推门进去吧		155

# 巴哈伊信仰与企业创新:

## 大唐帝国周边国家的证据

Prof. 李白耶夫斯基

Email: [barrylee@hlu.edu.cn](mailto:barrylee@hlu.edu.cn)

姓名	性别	身高(cm)
李白耶夫斯基	男	178
孟浩然正气	女	165
柳永在扬州	男	172
杜十娘大波波娃	女	171
张飞张翼德	男	191
小乔出家了	女	163
大乔没出家	女	165
僧敲月下门	和尚	175
月下门虚掩	尼姑	155
推门进去吧		155

# 巴哈伊信仰与企业创新：

大唐帝国周边国家的证据

Prof. 李白耶夫斯基

Email: [barrylee@hlu.edu.cn](mailto:barrylee@hlu.edu.cn)

姓名	性别	身高(cm)
李白耶夫斯基	男	178
孟浩然正气	女	165
柳永在扬州	男	172
杜十娘大波波娃	女	171
张飞张翼德	男	191
小乔出家了	女	163
大乔没出家	女	165
僧敲月下门	和尚	175
月下门虚掩	尼姑	155
推门进寺吧		155

# 巴哈伊信仰与企业创新:

## 大唐帝国周边国家的证据

---

Prof. 李白耶夫斯基

Email: [barrylee@hlu.edu.cn](mailto:barrylee@hlu.edu.cn)

姓名	性别	身高(cm)
李白耶夫斯基	男	178
孟浩然正气	女	165
柳永在扬州	男	172
杜十娘大波波娃	女	171
张飞张翼德	男	191
小乔出家了	女	163
大乔没出家	女	165
僧敲月下门	和尚	175
月下门虚掩	尼姑	155
推门进去吧		155

# Footnote

- Putdocx缺乏脚注输出这一功能，当遇到该类需求时，需要人力手工补全相应功能。
- Markdown原生语法支持非常轻便地脚注标注，且在html里面非常便于通过超链接浏览。

# Footnote

```
<<dd_include: header.txt>>
```

```
# 巴哈伊信仰与企业创新:  
## 大唐帝国周边国家的证据
```

```
## 李白耶夫斯基[^1]  
## 翰林院大学管理学院  
### Email: `libai@hlu.edu.cn`
```

```
## 杜十娘大波波娃[^2]  
## 扬州大学青楼文化学院  
### Email: `dumeinv@yzu.edu.cn`
```

[^1]: 李白，字太白，生于碎叶，长安大学文学博士，翰林院大学教授

[^2]: 杜十娘，翰林院大学商学院博士，师从李太白教授，现任扬州大

```
<!--Stata转换命令  
dyntext p19.do,replace saving(p19.md)  
!pandoc p19.md -o p19.html  
-->
```

# 巴哈伊与企业创新：

大唐帝国周边国家的证据

李白耶夫斯基<sup>1</sup>

翰林院大学管理学院

Email: libai@hlu.edu.cn

杜十娘大波波娃<sup>2</sup>

扬州大学青楼文化学院

Email: dumeinv@yzu.edu.cn

1. 李白，字太白，生于碎叶，长安大学文学博士，翰林院大学教授，博士生导师，著名的品酒师、探险家、商业策划大师，与孟浩然共同策划了扬州青楼营销。↵
2. 杜十娘，翰林院大学商学院博士，师从李太白教授，现任扬州大学青楼文学院教授，扬州文化研究会会长，大唐帝国全国正邪副主席。↵

```
clear all
cd E:\putdocx/
fs *.css
disp "`r(files)'"
foreach cssname in `r(files)' {
    disp "`cssname'"
    subinfile header.txt , from("`href="./.*?.css") ///
    to("`href="./`cssname'") ///
    fromregex save(header.txt) replace
    dyntext p19.do, replace saving(p19.md)
    !pandoc p19.md -o p19.html
    shellout p19.html
    sleep 10000
    !taskkill /F /IM chrome.exe /T
}
```

# Putdocx建议

- 添加读入docx 文档的功能
  - 当前只能写入不能读出
  - 读出的过程需要识别文档中的公式、图片、脚注尾注等
  - 不能识别的部分可以当做图片??
  - 读入文档允许可视化分词

# Putdocx建议

- 尽量使内容与样式相分离，写内容的时候尽管写，调样式的时候自由调。
- 式样通过网站公布，用户自由选择
- 增加脚注等功能
- 增加公式功能

# Putdocx 建议

- 插入图片格式为png
- 能否允许更多的图片格式？
  - JPEG?
  - BMP?

# hyper Link:

- Image Hyperlink
- Text hyperlink

# Putdocx describe

- 只能describe 文档
- 不能describe 表格和段落
- 没有返回值

# Describe

```
. putdocx describe
```

No. of tables	1
No. of paragraphs	1

```
. return list
```

```
. putdocx table t1,describe
```

```
parenthesis not balanced
```

```
r(198);
```

```
. putdocx table t1(.,.),describe
```

```
invalid specification; row and column misspecified
```

```
r(198);
```

谢谢