

China Family Panel Studies

CFPS

中国家庭动态跟踪调查

技术报告系列: CFPS-11

系列编辑: 谢宇 责任编辑: 胡婧炜

中国家庭动态跟踪调查
2010 年综合变量 (1): 字词与数学测试

徐宏伟 骆为祥

2012.12.20

1. wordtest: 字词测试原始得分

CFPS 根据受访者在字词测试部分的回答，为用户生成了字词测试得分的综合变量，命名为 wordtest。CFPS 2010 的字词测试部分为识字类题目，在设计的时候按照由易到难的顺序排列，变量 wordtest 的取值为答对的最难的一道题的题号。受访者在 T1 表中的最后学历（B4 题）不同，提问起点也会不同：如果 B4 题的答案为 1-2（小学或以下），从第一个字开始，顺序提问；如果为 3（初中），从第 9 个字开始，顺序提问；如果为 4-8（高中或以上），从第 21 个字开始，顺序提问。因此，在数据处理时，我们假定受过初中教育的受访者至少认识 8 个字，受过高中或以上教育的受访者至少认识 20 个字。如果受访者一道题也没有答对，则以其起点题的前一题的题号作为得分。如，一个受过初中教育的受访者从第 9 题开始答题，但是连续答错了第 9、10、11 题，测试终止，¹ 其得分为 8（即第 8 题）。

需要注意的是，按照以上方法计算所得的字词测试得分并未考虑随年龄增长或受教育水平提高而逐渐增强的字词认识水平。因此，该方法所得的字词测试得分是未经任何标准化处理的，在具体数据分析时需要进行额外的加工。

表 1 给出了少儿与成人样本中分性别的字词测试得分的描述性统计。

表 1. 字词测试得分（未标准化）

	样本量	均值	标准差	中位数	最小值	最大值
成人（16 岁及以上）						
男	16,270	18.9	9.9	21	0	34
女	17,293	15.2	11.5	18	0	34
总计	33,563	17.0	10.9	20	0	34
少儿（10-15 岁）						
男	1,702	21.1	7.3	22	0	34
女	1,658	22.4	7.0	23	0	34
总计	3,360	21.7	7.2	23	0	34

¹ 根据 CFPS 的设计，如果受访者有连续 3 道字词题回答错误，提问终止。

图 1 展示了少儿（15 岁及以下）样本中，字词测试得分相对于年龄的 Lowess 曲线²。由图 1 可见，随着年龄的增长，受访者的平均测试得分也随之增长。

图 2 展示了成人（16 岁及以上）样本中，字词测试得分相对于年龄的 Lowess 曲线。有趣的是，随着年龄的增长，受访者的平均测试得分反而随之降低。一个可能的解释是在成人样本中，年龄与受教育水平成负相关，因而年龄愈长者，识字（及数学）水平愈低。这个解释与图 3 中所显示的字词测试得分相对于受教育年限的 Lowess 曲线的模式相吻合——随着受教育年限的提高，受访者的平均字词测试得分亦随之提高。

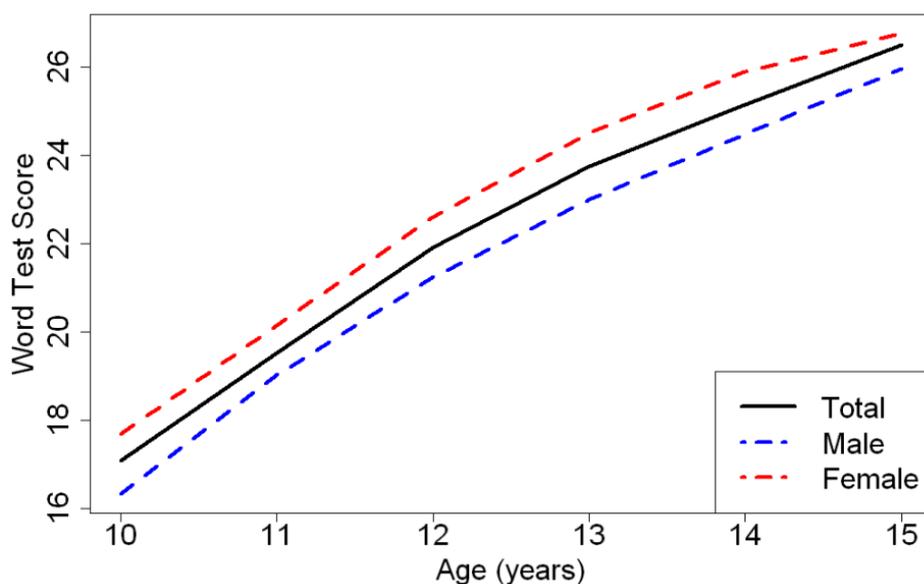


图 1. 字词测试得分对于年龄的 Lowess 曲线: CFPS 2010 少儿（15 岁及以下）样本

² Lowess 曲线是通过局部加权多项式回归分析（locally weighted polynomial regression）的技术从原始数据中绘制出的平滑曲线。对于数据中每个点附近的局部数据，拟合多项式回归曲线，并在拟合时给予与选定点较近的数据更多的权重。详情可参考 Cleveland, William S. 1979. “Robust Locally Weighted Regression and Smoothing Scatterplots.” *Journal of the American Statistical Association* 74 (368): 829-836.

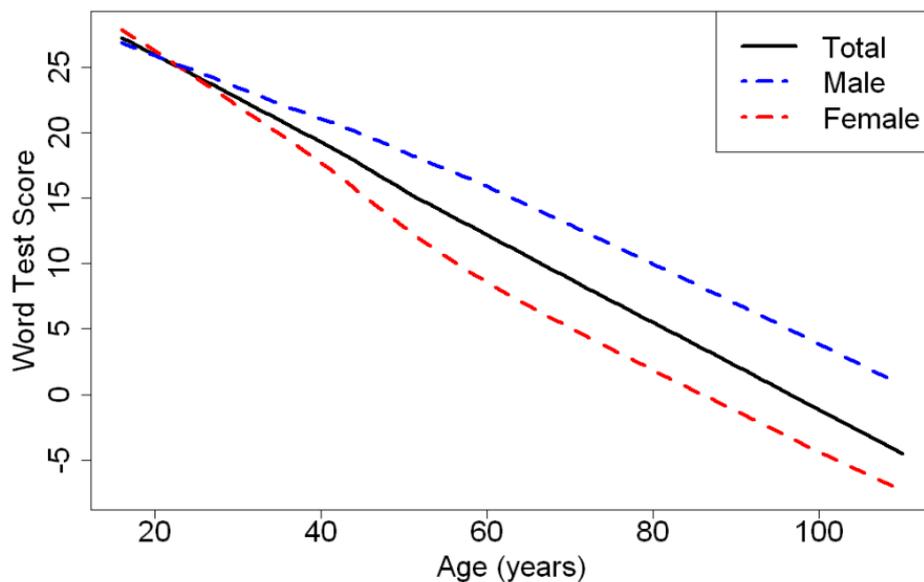


图 2. 字词测试得分对于年龄的 Lowess 曲线: CFPS 2010 成人 (16 岁及以上) 样本

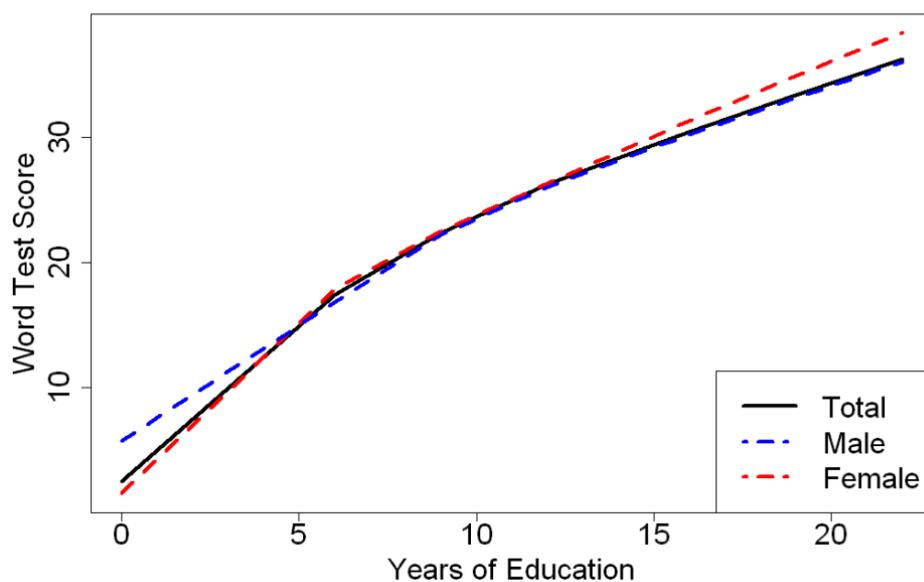


图 3. 字词测试得分对于受教育年限的 Lowess 曲线: CFPS 2010 成人 (16 岁及以上) 样本

2. mathtest: 数学测试原始得分

我们同样给出了数学测试得分的综合变量，变量名为 mathtest。数学测试题在设计的时候同样

是按照由易到难的顺序排列的，mathtest 的取值为答对的最难的一道题的题号。受访者在 T1 表中的最后学历（B4 题）决定了提问其数学测试题组的起点：如果 B4 题的答案为 1-2（小学或以下），从第 1 题开始，顺序提问；如果为 3（初中），从第 5 题开始，顺序提问；如果为 4-8（高中或以上），从第 13 题开始，顺序提问。因此，在数据处理时，我们假定受过初中教育的受访者至少能完成前 4 道题，而受过高中或以上教育的受访者至少能完成前 12 道题。类似于字词测试，如果在数学测试中，受访者一道题也没有答对，则以其起点题的前一题的题号作为得分。如，一个受过初中教育的受访者从第 5 题开始答题，但是连续答错了第 5、6、7 题，测试终止，³ 其得分为 4（即第 4 题）。

与字词测试相同，按照以上方法计算所得的数学测试得分并未考虑随年龄增长或受教育水平提高而逐渐增强的数学认识水平。因此，该方法所得的数学测试得分是未经任何标准化处理的，在具体数据分析时需要进行额外的加工。

表 2 给出了少儿与成人样本中分性别的数学测试得分的描述性统计。

表 2. 数学测试得分（未标准化）

	样本量	均值	标准差	中位数	最小值	最大值
成人（16 岁及以上）						
男	16,270	11.2	6.4	12	0	24
女	17,293	8.9	7.0	9	0	24
总计	33,563	10.0	6.8	12	0	24
少儿（10-15 岁）						
男	1,702	11.2	4.4	11	0	24
女	1,658	11.2	4.5	12	0	24
总计	3,360	11.2	4.5	12	0	24

类似的，图 4 显示，随着年龄的增长，少儿（15 岁及以下）受访者的平均数学测试得分也随之增长。

³ 根据 CFPS 的设计，如果受访者有连续 3 道数学题回答错误，提问终止。

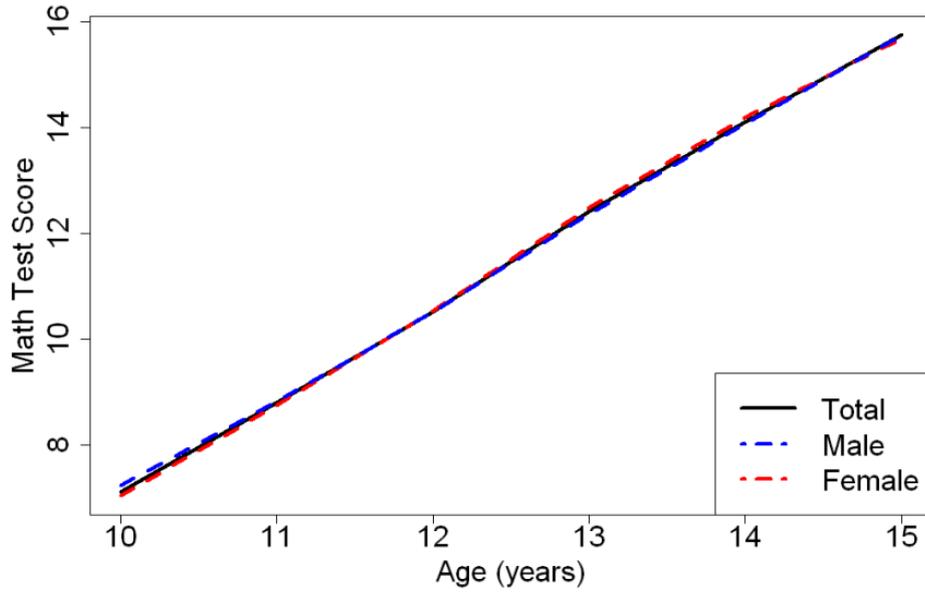


图 4. 数学测试得分对于年龄的 Lowess 曲线: CFPS 2010 少儿 (15 岁及以下) 样本

图 5 显示, 成人 (16 岁及以上) 受访者的平均数学测试得分与年龄成负相关; 而图 6 再次显示, 这很可能源于成人样本中年龄与受教育水平成负相关, 因为随着受教育水平的提高, 平均测试得分也相应提高 (见图 6)。

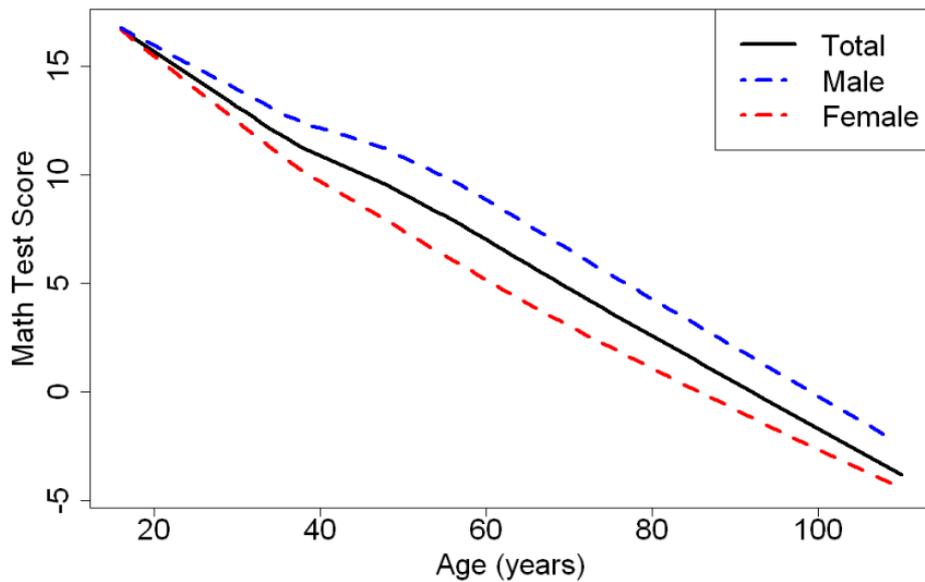


图 5. 数学测试得分对于年龄的 Lowess 曲线: CFPS 2010 成人 (16 岁及以上) 样本

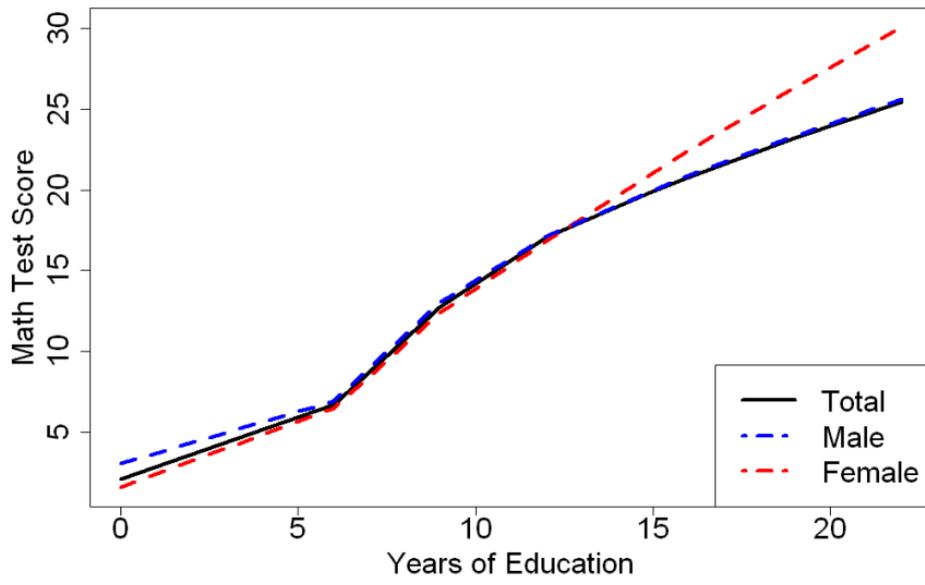


图 6. 数学测试得分对于受教育年限的 Lowess 曲线: CFPS 2010 成人 (16 岁及以上) 样本