باسمه تعالی

# بررسی قانون بنفورد در نتایج آرای دهمین دوره انتخابات ریاست جمهوری ایران

**سیدمحمدصادق موحد[۱ و A]، سیدمهدی فاضلی[B] و سیدحامد سیدعلایی[۲ و A]**

[A] استادیار گروه فیزیک دانشگاه شهید بهشتی

[B] استادیار گروه فیزیک دانشگاه قم

**۞وَالَّذِينَ اجْتَنَبُوا الطَّاغُوتَ أَن يَعْبُدُوهَا وَأَنَابُوا إِلَي اللَّهِ لَهُمُ الْبُشْرَي فَبَشِّرْ عِبَادِ۞ الَّذِينَ يَسْتَمِعُونَ الْقَوْلَ فَيَتَّبِعُونَ أَحْسَنَهُ أُوْلَئِکَ الَّذِينَ هَدَاهُمُ اللَّهُ وَأُوْلَئِکَ هُمْ أُوْلُوا الْأَلْبَابِ۞**

**سوره زمر آیات ۱۷ و ۱۸**

## چکیده

در این تحقیق نتایج حاصل از شمارش آرای ۳۶۶ شهرستان که توسط وزارت کشور در تاریخ ۱۳۸۸/۳/۲۴ منتشر شده مورد بررسی قرار می‌گیرد. به طور جداگانه سازگاری نتایج هر یک نامزدها با قانون بنفورد (Benford's law) بررسی می‌شود. تا کنون در حدود ۴ مقاله در مورد این شیوه محاسباتی برای نتایج آرای انتخابات ریاست جمهوری ایران منتشر شده است که در این مقاله همچنین نتایج حاصل از مقالات قبلی را مورد کنکاش قرار داده و نقدهایی بر روی برخی از آنها ارایه می‌دهد.

## مقدمه

پس از حماسه بزرگ ۲۲ خرداد سال ۱۳۸۸، متاسفانه به دنبال بسیاری از مسایل متعدد که در این مقاله قصد به چالش کشیدن آنها را نداریم شاهد ادعاهایی نه چندان مستدل پیرامون انتخابات بودیم. تناقض‌هایی که بعضاً ناشی از بی‌اطلاعی‌ها و کج فهمی‌ها بود، جامعه ایران را دچار تنش کرد. در این میان متاسفانه برخی افراد نیز با نام علم و استفاده ابزاری از آن به همراه قولهایی مبتنی بر احساسات و تقلیدهای کورکورانه و نه مبتنی بر عقلانیت و خِردورزی سعی در تشدید ناآرامی‌ها کردند. با توجه به این مسایل نویسندگان بر خود لازم دیدند تا بار دیگر رسالت سنگین خویش را در حوزه تحقیق و پژوهش و تنها متکی بر ابزار علمی و به دور از هر گونه پیش‌داوری‌ها انجام داده و به بررسی این موضوع از نقطه نظر محاسبات آماری بپردازند تا بلکه قسمتی از حق بزرگی که کشور عزیزمان بر گردن این فرزندان خُرد دارد، در این دوران سراسر فتنه ادا گردد انشاءالله.

[1] http://faculties.sbu.ac.ir/~movahed
[2] http://faculties.sbu.ac.ir/~allaei

در این تحقیق پس از بررسی روش آماری مبتنی بر قانون تجربی که در سال ۱۹۳۸ منتشر شده و اصطلاحاً قانون بنفورد نامیده می‌شود به بررسی میزان سازگاری نتایج مربوط به شمارش آرای هر نامزد می‌پردازیم. میزان تطابق مشاهدات با مدل معرفی شده را با کمیتی به نام P-value کمّی می‌کنیم و ضمن مقایسه با نتایج منتشر شده در مقالات مرتبط اخیر که احتمالاً بسیاری از خوانندگان این مقاله پیش از این، آنها را مطالعه کرده‌اند، به تفسیر نتایج حاصل از تحلیل آماری خود می‌پردازیم. لازم است اشاره کنیم که به اقتضای چنین مطالعات آماری، نتایج خود را با احتمال و نه به عنوان نظر قطعی گزارش می‌کنیم.

## قانون بنفورد

قانون بنفورد در حقیقت در ابتدا توسط آقای Newcomb در سال ۱۸۸۱ و پس از آن توسط آقای Benford در سال ۱۹۳۸ تعمیم داده شد[۱-۴]. این قانون بیان می‌دارد که در طبیعت اگر با وقایعی روبرو شویم که مقدار آنها از شمارش رخدادهایی حاصل شوند، فراوانی عدد اول سمت چپ آن مقدار به دست آمده از معادله (۱) تبعیت می‌کند. به بیانی دیگر احتمال اینکه عدد اول سمت چپ در یک واقعه (در اینجا تعداد آرای در هر شهرستان برای هر نامزد) مقدار $i$ باشد برابر است با:

$$p_1(i) = \log_{10}\left(1+\frac{1}{i}\right) \qquad (۱)$$

اندیس $1$ در معادله (۱) شماره رقم را نشان می‌دهد. پس احتمال رخ دادن عدد ۱ به عنوان عدد اول سمت چپ رخداد (در اینجا میزان آرای مربوط به هر نامزد در هر شهرستان) برابر است با:

$$p_1(i=1) = \frac{\ln\left(1+\frac{1}{1}\right)}{\ln(10)} = 0.301029996 \qquad (۲)$$

احتمال اینکه عدد $i$ به عنوان رقم دوم مشاهده گردد توسط معادله (۳) داده‌می‌شود:

$$p_2(i) = \sum_{j=1}^{9} \frac{\ln\left(1+\frac{1}{j\times 10+i}\right)}{\ln(10)} \qquad (۳)$$

احتمال رخدادن عدد $i$ به عنوان رقم سوم از سمت چپ برابر است با:

$$p_3(i) = \sum_{k=1}^{9}\sum_{j=0}^{9} \frac{\ln\left(1+\frac{1}{k\times 100+j\times 10+i}\right)}{\ln(10)} \qquad (۴)$$

همانگونه که بیان شد اندیس در معادلات فوق نماینده چندمین رقم بودن می‌باشد. و به همین ترتیب می‌توان این احتمال را برای رخ دادن بقیه اعداد در رقمهای مورد نظر از سمت چپ محاسبه کرد. در جدول (۱) این احتمال‌ها برای رقم اول، دوم و سوم گزارش شده‌است.

**جدول (۱): احتمال رخدادن بر طبق نظریه بنفورد**

| عدد | احتمال رخدادن به عنوان رقم اول سمت چپ | احتمال رخ دادن به عنوان رقم دوم سمت چپ | احتمال رخ دادن به عنوان رقم سوم سمت چپ |
|---|---|---|---|
| 0 | 0 | 0.119679267 | 0.101784363 |
| 1 | 0.301029996 | 0.113890102 | 0.101375976 |
| 2 | 0.176091259 | 0.108821497 | 0.100972197 |
| 3 | 0.124938737 | 0.104329559 | 0.100572931 |
| 4 | 0.096910013 | 0.100308201 | 0.100178086 |
| 5 | 0.079181246 | 0.09668 | 0.099787574 |
| 6 | 0.06694679 | 0.093374734 | 0.099401309 |
| 7 | 0.057991947 | 0.090351988 | 0.099019205 |
| 8 | 0.051152522 | 0.087570052 | 0.098641183 |
| 9 | 0.045757491 | 0.084997351 | 0.098267162 |

## تحلیل داده‌ها

داده‌های مورد استفاده در این تحقیق، نتایج حاصل از شمارش آرا در ۳۶۶ شهرستان می‌باشد که در سامانه اطلاع رسانی رسمی وزارت کشور در تاریخ ۱۳۸۸/۳/۲۴ منتشر شده است[۵]. فراوانی رقم اول سمت چپ برای چهار نامزد انتخاباتی به همراه مقدار پیش‌بینی شده توسط معادله (۱) در جدول (۲) آمده است. در جدول (۳) میزان فراوانی رقم دوم به همراه میزان پیش‌بینی شده توسط قانون بنفورد آمده است.

**جدول(۲): فراوانی رقم اول سمت چپ**

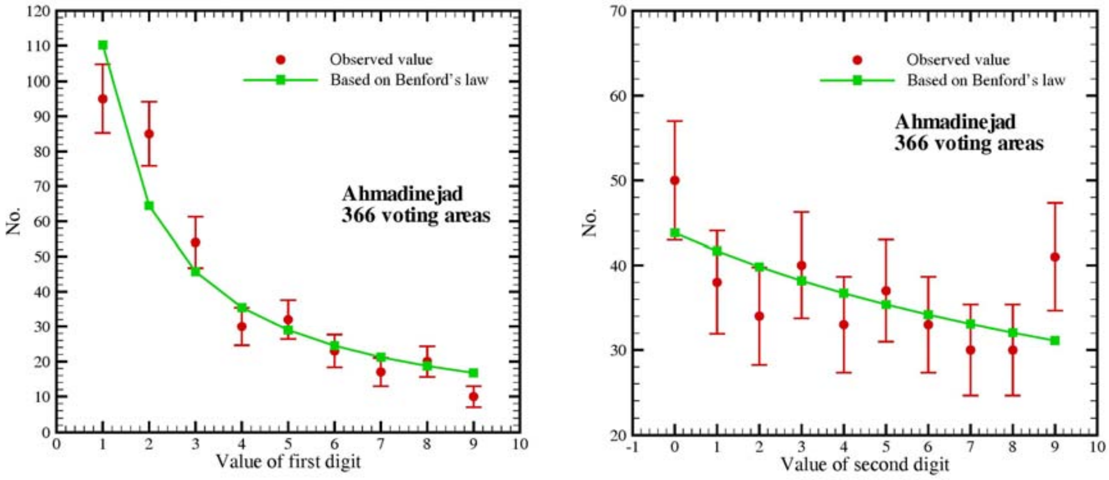
| عدد در رقم اول | احمدی نژاد | موسوی | رضایی | کروبی | پیش بینی تئوری |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 95 | 130 | 96 | 125 | 110.1769784 |
| 2 | 85 | 60 | 67 | 57 | 64.44940081 |
| 3 | 54 | 47 | 51 | 44 | 45.7275776 |
| 4 | 30 | 39 | 37 | 29 | 35.46906476 |
| 5 | 32 | 23 | 30 | 24 | 28.98033605 |
| 6 | 23 | 18 | 25 | 16 | 24.502525 |
| 7 | 17 | 17 | 24 | 41 | 21.22505259 |
| 8 | 20 | 13 | 18 | 13 | 18.72182322 |
| 9 | 10 | 19 | 18 | 17 | 16.74724155 |

**جدول(۳) میزان فراوانی رقم دوم و میزان پیش‌بینی شده توسط قانون بنفورد**

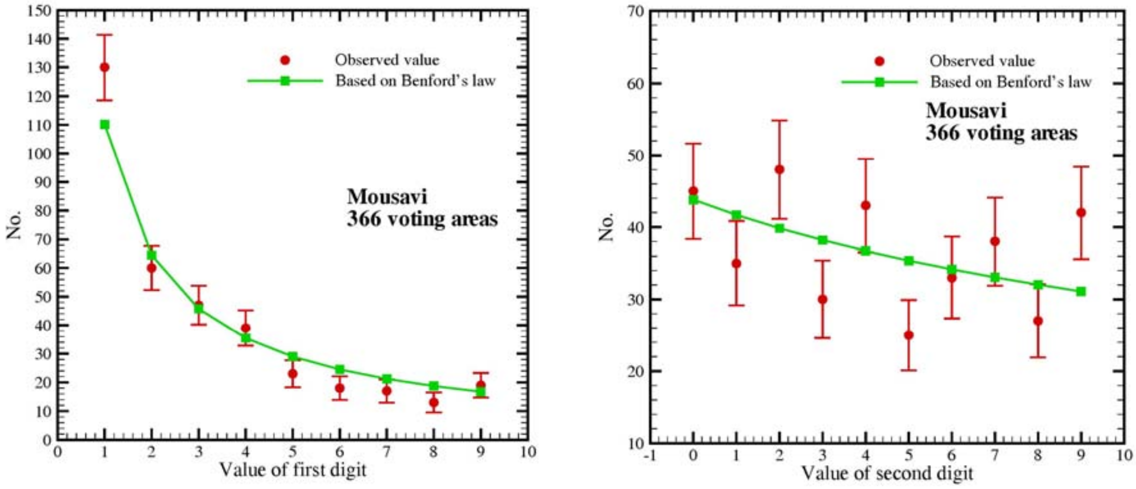
| عدد در رقم دوم | احمدی نژاد | موسوی | رضایی | کروبی | پیش بینی تئوری |
|---|---|---|---|---|---|
| 0 | 50 | 45 | 37 | 41 | 43.80261 |
| 1 | 38 | 35 | 47 | 33 | 41.68378 |
| 2 | 34 | 48 | 23 | 43 | 39.82867 |
| 3 | 40 | 30 | 38 | 48 | 38.18462 |
| 4 | 33 | 43 | 34 | 39 | 36.7128 |
| 5 | 37 | 25 | 51 | 40 | 35.38488 |
| 6 | 33 | 33 | 37 | 35 | 34.17515 |
| 7 | 30 | 38 | 35 | 32 | 33.06883 |
| 8 | 30 | 27 | 36 | 28 | 32.05064 |
| 9 | 41 | 42 | 28 | 27 | 31.10903 |

برای نمایش بهتر از نمودارهای زیر استفاده می‌کنیم. شکل (۱) نتایج مربوط به آقای احمدی‌نژاد به همراه مقدار پیش‌بینی شده بر طبق قانون بنفورد را نشان می‌دهد. شکل (۲)، (۳) و (۴) به ترتیب نتایج مربوط به آقایان

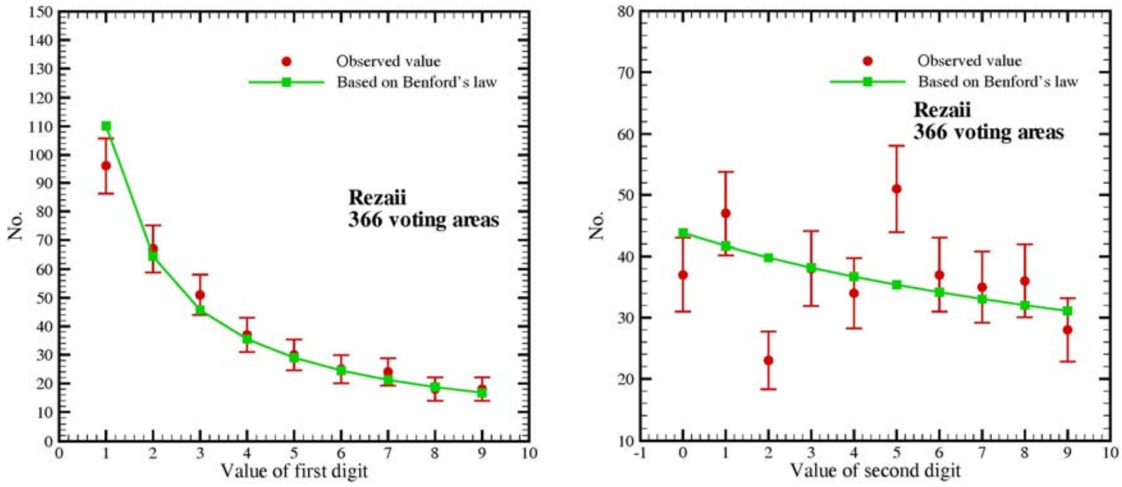
موسوی، رضایی و کروبی را نشان می‌دهد. شکل (۵) نتایج مربوط به کل آرای اخذ شده را برای رقم اول و دوم سمت



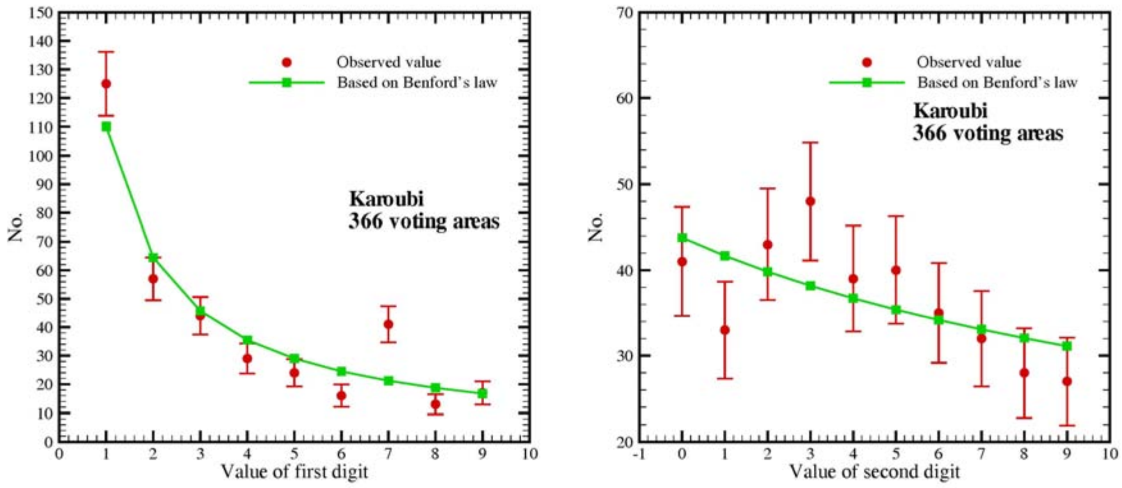

**شکل (۱): سمت چپ فراوانی رقم اول و سمت راست فراوانی رقم دوم برای نتایج مربوط به آقای احمدی‌نژاد. نقاط دایره‌ای نتایج مشاهده شده و خط پیوسته بر اساس مدل است**



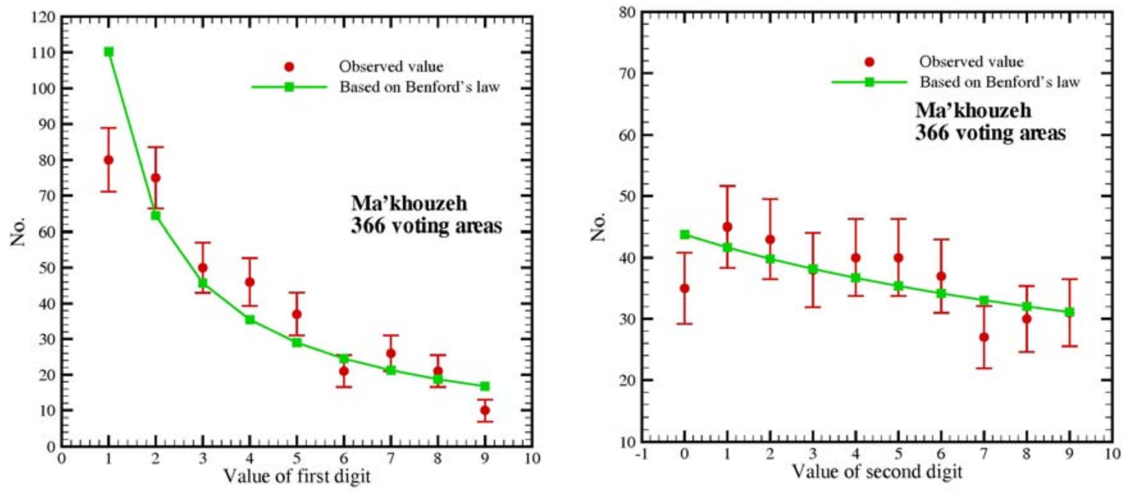

**شکل (۲): سمت چپ فراوانی رقم اول و سمت راست فراوانی رقم دوم برای نتایج مربوط به آقای موسوی. نقاط دایره‌ای نتایج مشاهده شده و خط پیوسته بر اساس مدل است**



**شکل (۳): سمت چپ فراوانی رقم اول و سمت راست فراوانی رقم برای نتایج مربوط به آقای رضایی. نقاط دایره‌ای نتایج**




**شکل (۴): سمت چپ فراوانی رقم اول و سمت راست فراوانی رقم دوم برای نتایج مربوط به آقای کروبی. نقاط دایره‌ای نتایج مشاهده شده و خط پیوسته بر اساس مدل است**




**شکل (۵): سمت چپ فراوانی رقم اول و سمت راست فراوانی رقم برای نتایج مربوط به کل آرای اخذشده. نقاط دایره‌ای نتایج مشاهده شده و خط پیوسته بر اساس مدل است**

اکنون لازم است که میزان تطابق نتایج فوق را با آنچه که از معادلات (۱) و (۳) به دست می‌آید، بررسی کنیم. برای این منظور کمیت $\chi^2$ [۱] را به صورت زیر تعریف می‌کنیم[۶]:

$$\chi^2 = \sum_{i=1}^{9} \frac{\left[n_{obs.}(i) - n_{The.}(i)\right]^2}{n_{The.}(i)} \qquad (۵)$$

[1] Pearson Chi-Square

در معادله فوق $n_{obs.}(i)$ میزان فراوانی عدد $i$ به عنوان اولین رقم سمت چپ از آرای مربوط به یک نامزد می‌باشد. همچنین $n_{The.}(i)$ میزان فراوانی تعیین شده توسط قانون بنفورد است. همین رابطه برای بررسی میزان تطابق نتایج با مدل برای دومین رقم سمت چپ به صورت زیر نوشته می‌شود:

$$\chi^2 = \sum_{i=0}^{9} \frac{\left[n_{obs.}(i) - n_{The.}(i)\right]^2}{n_{The.}(i)} \tag{۶}$$

. در جدول (۴) و (۵) به ترتیب مقادیر محاسبه شده برای هرکدام از جملات موجود در معادله (۵) و (۶) گزارش شده‌است. در جدول (۶) و (۷) به ترتیب نتایج مربوط به کل آرای اخذ شده از ۳۶۶ شهرستان که جمع آرای صحیح و باطله می‌باشد، گزارش شده‌است.

**جدول (۴): نتایج مربوط به مقدار $\chi^2$ برای رقم اول سمت چپ به همراه مقدار P-value**

| عدد در رقم اول | احمدی نژاد | موسوی | رضایی | کروبی |
|---|---|---|---|---|
| 1 | 2.090642501 | 3.566554379 | 1.824216999 | 1.994263885 |
| 2 | 6.552847995 | 0.307173804 | 0.100940523 | 0.861040937 |
| 3 | 1.496536138 | 0.035406616 | 0.607914074 | 0.065267493 |
| 4 | 0.843288921 | 0.351503592 | 0.066079067 | 1.17986756 |
| 5 | 0.314639911 | 1.234092636 | 0.035876553 | 0.855881973 |
| 6 | 0.092136683 | 1.725652006 | 0.010100239 | 2.950427821 |
| 7 | 0.8410377 | 0.8410377 | 0.362794536 | 18.42391406 |
| 8 | 0.087263717 | 1.748721828 | 0.027830022 | 1.748721828 |
| 9 | 2.718374148 | 0.303030242 | 0.093711179 | 0.003814768 |
| chi-square | 15.03676772 | 10.1131728 | 3.129463192 | 28.08320032 |
| P-value | 0.0584 | 0.2571 | 0.9259 | 0.0004 |

**جدول (۵): نتایج مربوط به مقدار $\chi^2$ برای رقم دوم سمت چپ به همراه مقدار P-value**

| عدد در رقم دوم | احمدی نژاد | موسوی | رضایی | کروبی |
|---|---|---|---|---|
| 0 | 0.876834058 | 0.03273181 | 1.056455862 | 0.179318813 |
| 1 | 0.32555147 | 1.071708984 | 0.678014949 | 1.80904881 |
| 2 | 0.852987892 | 1.676447356 | 7.110558376 | 0.252515251 |
| 3 | 0.086307264 | 1.754318436 | 0.000892611 | 2.523050313 |
| 4 | 0.375479249 | 1.076705212 | 0.200455748 | 0.142491901 |
| 5 | 0.07372111 | 3.04779139 | 6.890852042 | 0.601933159 |
| 6 | 0.040409012 | 0.040409012 | 0.233496003 | 0.019908407 |
| 7 | 0.284790953 | 0.735328793 | 0.112777714 | 0.034545902 |
| 8 | 0.131202406 | 0.79589539 | 0.486650233 | 0.511929811 |
| 9 | 3.144787153 | 3.812822732 | 0.310715893 | 0.542740509 |
| chi-square | 6.192070565 | 14.04415911 | 17.08086943 | 6.617482877 |
| P-value | 0.7205 | 0.1207 | 0.0474 | 0.6768 |

**جدول (۶): نتایج مربوط به مقدار $\chi^2$ برای رقم اول سمت چپ به همراه مقدار P-value برای کل آرای اخذ شده**

| عدد در رقم اول | کل آرای اخذ شده |
|---|---|
| 1 | 8.265338542 |
| 2 | 1.72717111 |
| 3 | 0.399181285 |
| 4 | 3.12668512 |
| 5 | 2.219263769 |
| 6 | 0.50067009 |
| 7 | 1.074208068 |
| 8 | 0.277221369 |
| 9 | 2.718374148 |
| chi-square | 20.3081135 |
| P-value | 0.0092 |

**جدول (۷): نتایج مربوط به مقدار $\chi^2$ برای رقم دوم سمت چپ به همراه مقدار P-value برای کل آرای اخذ شده**

| عدد در رقم دوم | کل آرای اخذ شده |
|---|---|
| 0 | 1.768980654 |
| 1 | 0.263827655 |
| 2 | 0.252515251 |
| 3 | 0.000892611 |
| 4 | 0.294329863 |
| 5 | 0.601933159 |
| 6 | 0.233496003 |
| 7 | 1.113757918 |
| 8 | 0.131202406 |
| 9 | 0.000382128 |
| chi-square | 4.661317649 |
| P-value | 0.8627 |

برای بررسی میزان تطابق مشاهده با مدل بایستی مقدار P-value تعیین گردد. این عبارت احتمال پذیرش مدل توسط مشاهدات انجام شده را تعیین می‌کند. به زبانی دیگر احتمال اینکه مشاهده‌ای دیگر انجام شود و مقدار $\chi^2$ آن بزرگتر از آنچه که به دست آورده‌ایم، گردد، را به دست می‌دهد. در حقیقت مقدار P-value احتمال اینکه در یک آزمایش دیگر همین اعداد را به دست آوریم را به دست می‌دهد[۶]. بر اساس قانونی که اصطلاحاً قانون سر انگشتی[1] نامیده می‌شود اگر مقدار $\chi^2 \approx \nu \pm \sqrt{2\nu}$ باشد، در آن صورت مدل توسط مشاهدات تایید می‌شود و به بیانی دیگر مشاهدات با مدل سازگاری خواهند داشت. در اینجا $\nu$ تعداد درجات آزادی[2] می‌باشد که برای رقم اول برابر با ۸ و برای رقم دوم برابر با ۹ است. این قانون از آنجا به دست می‌آید که برای تعداد درجات آزادی بزرگ یعنی $\nu \gg 1$، تابع توزیع مربوط به کمیت $\chi^2$ به تابع توزیع گوسی با انحراف معیار $\sqrt{2\nu}$ تبدیل می‌شود. از طرفی در بسیاری از موارد چنین بیان می‌شود که اگر مقدار P-value کوچکتر از ۰/۱ و بزرگتر از ۰/۹ گزارش شود مدل ارایه شده با توجه به مقادیر به دست آمده از مشاهده، مورد قبول نمی‌باشد. البته خاطر نشان می‌کنیم که P-value نزدیک به عدد ۱ یعنی تطابق فوق‌العاده، که این در مشاهدات واقعی بیشتر به داده‌سازی تعبیر می‌شود. مقدار P-value توسط رابطه زیر محاسبه می‌شود:

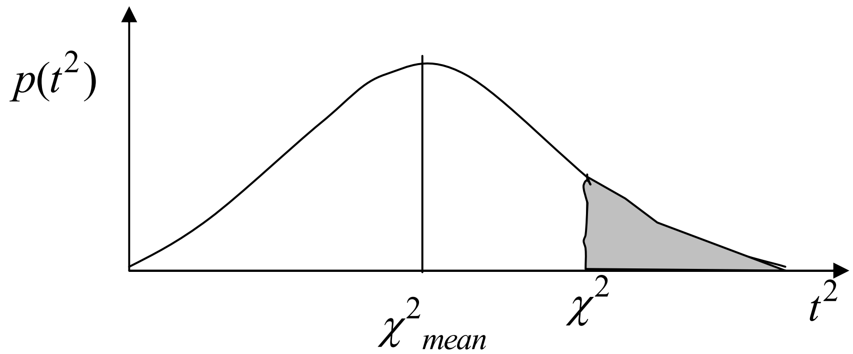
---

[1] Rule of thumb

[2] Degrees of freedom

$$P(>\chi^2;\nu)=\frac{1}{2^{\nu/2}\Gamma(\nu/2)}\int_{\chi^2}^{+\infty}\left(t^2\right)^{(\nu-2)/2}e^{-t^2/2}dt^2 \qquad (۷)$$

در معادله (۷) کمیت $\nu$ نماینده تعداد درجه آزادی می‌باشد [۶و۷]. برای توصیف بهتر می‌توان از شکل (۶) بهره گرفت. در این شکل نشان داده‌شده که منظور از مقدار P-value چیزی جز سطح زیر منحنی تابع توزیع از $\chi^2$ به دست آمده از مشاهده تا بی‌نهایت نمی‌باشد. واضح است که برای $\chi^2$ دور از مقدار متوسط مقدار P-value از مقدار بهینه خودش یعنی عدد ۵/. دور می‌شود. در جدول (۸) مقدار P-value برای تعداد درجات آزادی ۸ و ۹ برای مقادیر مختلف $\chi^2$ گزارش شده‌است. با توجه به اینکه تقریب $\chi^2\approx\nu\pm\sqrt{2\nu}$ برای تعداد درجات آزادی ۸ و ۹ دقیق نمی‌باشد بنابراین از رابطه اصلی یعنی معادله (۷) بهره می‌گیریم، بنابراین مقدار مورد انتظار برای $\chi^2$ برای ۸ و ۹ درجه آزادی برای اینکه P-value مقدار ۰/۵ را اختیار کند ترتیب برابر با ۷/۳۴۴ و ۸/۳۴۳ می‌باشد که واضحاً نسبت به قانون سرانگشتی یاد شده کمی انحراف دارد.




**شکل(۶): ناحیه سایه خورده معادل مقداری است که توسط معادله (۷) برای مقدار بدست آمده $\chi^2$ در مشاهدات، داده‌می‌شود.**

در شکل (۷) مشاهده می‌شود که هرچه مقدار $\chi^2$ به دست آمده از مشاهده به مقدار متوسط آن یعنی $\chi^2{}_{mean}$ نزدیکتر باشد در آن صورت مقدار P-value به عدد نیم نزدیکتر می‌باشد و این اتفاق برای تعداد درجات آزادی بزرگ در حدود $\nu\geq 34$ با دقت ۲ درصد منجر به مقدار $\chi^2\approx\nu\pm\sqrt{2\nu}$ می‌شود. با عنایت به شکل (۶) و از آنجا که عوامل تصادفی در مشاهدات دخیل می‌باشند بنابراین می‌توان انتظار داشت که در آزمایشی مقادیر دور از مقدار متوسط برای $\chi^2$ به دست آید. بنابراین توجه به این نکته ضروری است که مقدار P-value فقط احتمال رخ دادن آزمایشی که منجر به مقدار بیشتری برای $\chi^2$ شود، به دست می‌دهد و اصطلاحاً کیفیت همخوانی مشاهدات کنونی را تعیین می‌کند.

**جدول(۸): مقدار مربوط به $\chi^2$ مترادف با مقادیر مختلف P-value**

| تعداد درجه آزادی | 0.99 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 | 0.001 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8 | 1.648 | 3.488 | 4.592 | 5.528 | 6.424 | 7.344 | 8.352 | 9.528 | 11.032 | 13.36 | 15.504 | 18.168 | 20.088 | 26.128 |
| 9 | 2.088 | 4.167 | 5.382 | 6.39 | 7.353 | 8.343 | 9.414 | 10.656 | 12.24 | 14.688 | 16.92 | 19.683 | 21.663 | 27.873 |

یک نکته دیگری که به نظر می‌رسد مناسب است به آن پرداخت باطلنمای Lindley [1] است. در برخی موارد مقدار P-value منجر به عدم همخوانی مشاهدات با مدل می‌شود در حالی که اگر آنالیز درستنمایی[2] به همراه یک تابع توزیع اولیه که حاوی اطلاعات اولیه پیرامون پارامترهای آزاد مدل می‌باشد واصطلاحاً تابع توزیع اولیه[3] نامیده می‌شود، به کار رود نتیجه عکس به دست می‌دهد و بالعکس[۸]. قبل از پرداختن به این موضوع بیان می‌کنیم که این شیوه آنالیز اصطلاحاً تحلیل بایاسین[4] نام دارد و در مقابل آنالیز تکراری[5] مطرح می‌شود[۷و۹]. به طور خلاصه این تحلیل به صورت زیر می‌باشد:

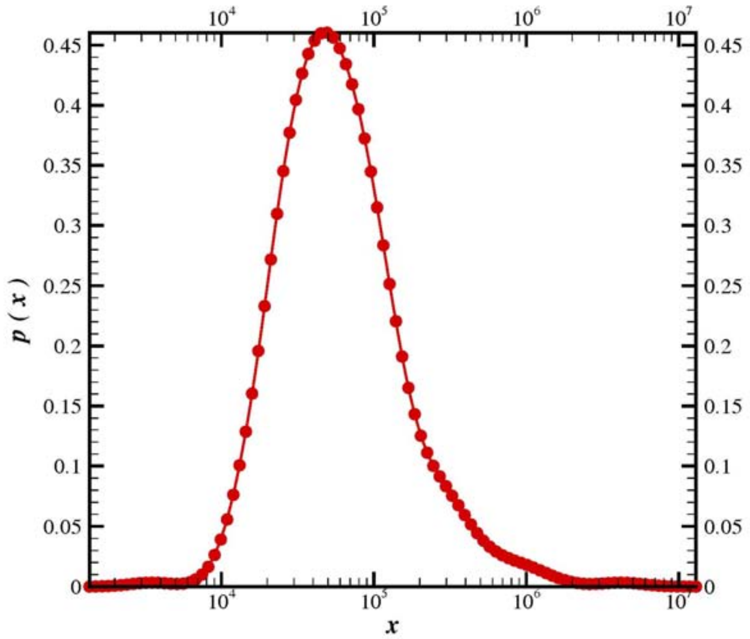
$$p(\Theta \mid x) = \frac{p(x \mid \Theta)\, p(\Theta)}{\int p(x \mid \Theta)\, p(\Theta)\, d\Theta} \qquad (۸)$$

در معادله بالا $p(\Theta \mid x)$ احتمالی که مشاهدات به مقادیر مورد نظر برای کمیت‌های مدل منجر شوند، به دست می‌دهد و اصطلاحاً Posterior نامیده می‌شود . کمیت $p(x \mid \Theta)$ احتمالی که مقادیر مشخص برای پارامترهای مدل منجر به ایجاد مشاهدات گردد، تعیین می‌کند و اصطلاحاً Likelihood نامیده می‌شود. کمیت $p(\Theta)$ اطلاعات اولیه پیرامون مدل را مشخص می‌کند. ضمناً $\{\Theta\}$ نشاندهنده مجموعه پارامترهای آزاد مدل و $\{x\}$ مجموعه مشاهدات را نمایش می‌دهند. در برخی مسایل ما اطلاعی از $p(\Theta)$ نداریم بنابراین بیشینه کردن posterior با بیشینه کردن Likelihood یکسان بوده و در صورتی که مشاهدات از یکدیگر مستقل باشند با مفهوم $\chi^2$ مواجه خواهیم بود که بیشینه شدن درستنمایی با کمینه شدن $\chi^2$ یکسان می‌باشد. در چنین وضعیتی اصلاً باطلنمای Lindley موضوعیت ندارد. در صورتی که احتمال اولیه وجود داشته باشد در آن صورت ممکن است به ازای مقدار مشخصی از $\chi^2$ به این نتیجه برسیم که مدل قابل قبول نیست این درحالی است که در حضور تابع اولیه ما اصلاً با درستنمایی روبرو نخواهیم بود و بایستی در حضور prior درستنمایی را بکار بریم که اصطلاحاً با یک تحلیل Bayesian روبرو هستیم و احتمال بایاسین میزان تطابق را تعیین می‌کند. با توجه به اینکه در مورد تحلیل آرا ما تابع احتمال اولیه‌ای نداریم بنابراین به نظر می‌رسد که باطلنمای Lindley موضوعیت قابل توجهی ندارد.

برای بررسی دقیق تر دلیل انحراف برخی از اعداد به عنوان رقم اول و دوم سمت چپ آرا، مناسب است تابع توزیع تعداد کل آرای اخذ شده در هر شهرستان را مورد بررسی قرار دهیم. در شکل (۷) این تابع توزیع رسم شده‌است. با توجه به وجود تعداد زیادی از حوزه‌ها که بین ۴۰۰۰۰ تا ۶۰۰۰۰ آرای ماخوذه داشته‌اند، که این موضوع ناشی از تقسیمات کشوری (در سالهای قبل) است، همین موضوع می‌تواند باعث ایجاد انحرافی برای رخ دادن تعداد همین اعداد (۴ و ۵) به عنوان رقم اول سمت چپ گردد که مطابق جدول (۶) برابر با ۳/۱۳ و ۲/۲۲ (عدد موجود در ردیف پنجم و ششم ستون دوم) برای کل آرای ماخوذه و ۶/۵۵ (عدد موجود در ردیف سوم و ستون دوم در جدول (۴)) برای آرای نامزدی که حدود ۶۰ درصد آرا را کسب کرده، و همچنین عدد ۳/۵۷ برای نامزدی که حدود ۳۵ درصد آرا را کسب کرده (عدد موجود در ردیف دوم و ستون سوم در جدول (۴)) می‌شود.

---

[1] Lindley's paradox
[2] Likelihood analysis
[3] Prior probability
[4] Bayesian analysis
[5] Frequentist




**شکل (۷): تابع توزیع تعداد آرای اخذ شده در هر شهرستان**

## نتیجه‌گیری

یکی از مهمترین آزمونهایی که به صورت احتمالی انحرافات در رای‌گیری را مورد بررسی قرار می‌دهد، آزمون مبتنی بر قانون بنفورد می‌باشد. بر اساس تحلیلی که انجام شد به ترتیب نتایج زیر به دست آمد:

۱- بر اساس آزمون رقم اول و با توجه به مقادیر P-value گزارش شده در جدول (۴)، مدل تئوری با احتمال ۵/۸۴ درصد و ۰/۰۴ درصد به ترتیب با نتایج آقایان احمدی‌نژاد و کروبی همخوانی دارد. به بیانی دیگر اگر مشاهده دیگری رخ دهد با احتمال ۹۴/۱۶ درصد برای نتایج آقای احمدی‌نژاد و با احتمال ۹۹/۹۶ درصد برای نتایج آقای کروبی، منجر به مقداری کمتری برای $\chi^2$ خواهد شد. از سویی دیگر انحراف قابل توجهی در نتایج آقای رضایی مشاهده نمی‌شود. با توجه به ماهیت P-value نتایج آقای موسوی با مدل سازگاری قابل قبولی دارد.

۲- برای رقم دوم نیز با توجه به مقادیر P-value گزارش شده در جدول (۵)، مدل با احتمال ۴/۷۴ درصد در نتایج آرای آقای رضایی تایید می‌گردد و نتایج آقای موسوی با احتمال ۱۲/۰۷ درصد با مدل سازگاری دارد. در بقیه موارد انحراف قابل ملاحظه‌ای دیده نمی‌شود.

۳- برای کل آرای اخذ شده نیز با توجه به مقادیر گزارش شده در جداول (۶) و (۷)، مشاهده می‌شود که آمار رقم اول با مدل سازگاری ندارد در حالی که برای رقم دوم انحراف عمده‌ای مشاهده نمی‌شود. لازم است اشاره شود که این عدم سازگاری ممکن است مربوط به قوانین تقسیمات کشوری ایران باشد که تقسیمات بین بخش و شهرستان بر اساس جمعیت صورت گرفته است.

## مقایسه با نتایج حاصل از مطالعات منتشر شده قبلی

مقاله‌های دیگری نیز با استفاده از قانون بنفورد، نتایج دهمین دوره انتخابات ریاست جمهوری ایران را تحلیل کرده‌اند. یکی از کارهای شاخص، مقاله Walter R. Mebane استاد علوم سیاسی و آمار دانشگاه میشگان است[۱۰]. او در این مقاله علاوه بر سایر تحلیل‌ها، به بررسی فراوانی دومین رقم از سمت چپ پرداخته و نتیجه

گرفته است که رقم دوم از سمت چپ برای آقای دکتر رضائی از قانون بنفورد پیروی نمی‌کند و پیشنهاد می‌دهد که در مورد این نامزد نیاز به بررسی‌های بیشتری وجود دارد.

کار شاخص بعدی، مقاله Boudewijn F. Roukema از دانشگاه نیکلاس کوپرنیک است[۱۱]. او نیز در این مقاله از چند زاویه به نتایج انتخابات نگاه کرده است که یکی از آنها قانون بنفورد برای رقم اول از سمت چپ است. یافته جالب ایشان فراوانی بیش از انتظار عدد هفت در نتایج آقای کروبی است. بر اساس قانون بنفورد، باید در حدود ۲۱ عدد ۷ در نتایج داشته باشیم در حالیکه در نتایج گزارش شده، ۴۱ بار عدد ۷ در سمت چپ مشاهده می‌شود. ایشان با اشاره به اینکه در فراوانی رقم اول از سمت چپ آقای دکتر احمدی‌نژاد عدد یک کمتر از مقدار مورد انتظار ظاهر شده و عدد دو بیشتر از مقدار مورد انتظار ظاهر شده، این فرض را مطرح می‌کنند که در تعدادی از حوزه‌ها رقم سمت چپ آرای ایشان از یک به دو تغییر یافته‌است. تفسیر دیگری که می‌توانیم مطرح کنیم این است که در جدول (۴) مقدار تفاوت رخدادن عدد ۱ از مقدار تئوری برای آرای آقای احمدی‌نژاد در حدود ۲/۰۹ می‌باشد که در همان جدول جملاتی بزرگتر از این نیز وجود دارد بنابراین استدلالی مبتنی بر شائبه‌ی تحلیل سیاسی را قوت خواهد بخشید .در ویرایش دوم این مقاله که در تاریخ ۱۳۸۸/۳/۳۱ در سامانه arXiv قرار گرفت با کمک پراکندگی در نسبت آرای هر نامزد بر حسب کل آرا در هر شهرستان و با تکیه بر اینکه در سه شهر بزرگ شیراز، اصفهان و مشهد که آرای آقای احمدی‌نژاد نسبت به بقیه نامزدها قابل مقایسه می‌باشد و درست در همین حوزه‌ها آرای آقای کروبی با عدد ۷ شروع می‌گردد، تخمینی برای میزان تفاوت آرای آقای احمدی‌نژاد و آقای موسوی با درنظر گرفتن اینکه عدد مربوط به کل آرا ثابت بماند، ارایه می‌کند.

علاوه بر این دو کار، دو مقاله به زبان فارسی نیز منتشر شده‌اند که به نوعی سعی کرده‌اند که دو کار فوق را تکرار کنند. اما نتایج متفاوتی گرفته‌اند. در مقاله آقای رسول رستگاری منتشر شده در سایت خبرگزاری انتخاب[۱۲]، به نظر می‌رسد که با استفاده از قانون بنفورد برای دومین رقم از سمت چپ به این نتیجه رسیده که نتایج تمام نامزدها با قانون بنفورد هم‌خوانی ندارد. نتایج ایشان شامل مقادیر $\chi^2$ و P-value می‌باشد که با نتایج ما و Mebane هم‌خوانی ندارد. ممکن است ایشان مرتکب اشتباه محاسباتی در محاسبه مقادیر $\chi^2$ شده باشند.

مقاله دیگری که به بررسی این موضوع پرداخته، مقاله آقای سیدمهدی سیدنصرالله است[۱۳] . ایشان نیز با استفاده از قانون بنفورد به بررسی فراوانی ارقام اول، دوم و سوم از سمت چپ در نتایج اعلامی پرداخته‌اند و نتیجه بر صحت انتخابات گرفته‌اند. مقادیر $\chi^2$ محاسبه شده توسط ایشان، با نتایج ما و کارهای Mebane و Roukema هم‌خوانی دارد. اما ایشان به جای محاسبه P-value از روی $\chi^2$، بدون اشاره در مقاله به نظر می‌رسد کمیت زیر را به عنوان خطای نسبی محاسبه نموده‌اند:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\frac{\left[n_{obs.}(i) - n_{The.}(i)\right]^2}{n_{The.}(i)^2} \qquad (۸)$$

و آنرا ملاکی برای هم‌خوانی نتایج با قانون بنفورد در نظر گرفته‌اند. لازم است اشاره کنیم که برای بررسی میزان تطابق[1] مشاهدات و مدل بایستی در هر تراز تطابق، احتمال یافتن مقداری برای $\chi^2$ به صورت $\chi^2 > \chi^2_\alpha$ را

[1] Goodness of fit

محاسبه نمود که این کار توسط معادله (۷) انجام می‌شود[۷و۱۴]. در اینجا کمیت $\alpha$ تراز مورد نظر را نشان می‌دهد. ضمن اینکه ایشان همان کمیت خطای نسبی را به صورت توان دوم گزارش کرده‌اند که حتی در صورتی که این کمیت در اینجا موضوعیتی داشت بایستی از آن جذر می‌گرفتند زیرا در گزارش خطا بایستی خود انحراف معیار[۱] یا انحراف معیار میانگین[۲] را گزارش کرد نه توان دوم آن. بنابراین چنین استدلالی اشتباه می‌باشد. نکته دیگر اینکه در نمودارها بایستی انحراف معیار را نیز در مقادیر مشاهده شده بر طبق آمار مربوطه نشان داده‌شود که در نمودارهای ایشان وجود ندارد و حتی به صورت کیفی نیز نمی‌توان برداشت درستی از نمودارها داشت. از طرفی دیگر لزوماً همخوانی نتایج با قانون بنفورد شرط لازم و کافی برای صحت انتخابات از این منظر نمی‌باشد همچنین اینکه عدم همخوانی نیز دلیل بر تقلب نیست. که در همین جا توجه همه خوانندگان را به آن جلب می‌کنیم.

از بین این چهار مقاله که به بررسی قانون بن‌فورد در نتایج انتخابات پرداخته‌اند، دو مقاله Mebane و Roukema با دقت بیشتری نوشته شده‌اند و توصیه می‌شود مراجع تصمیم گیرنده توجهی به نکات مطرح شده در این دو مقاله داشته باشند. این دو مقاله نکات دیگری را هم علاوه بر قانون بنفورد مطرح کردند که در اینجا به آنها نپرداخته‌ایم.

## مراجع


[1] Benford, F. (1938). *The law of anomalous numbers*, in *Proceedings of the American Philosophical Society*, 78, 551–572.

[2] George Judge, *"Detecting Problems in Survey Data using Benford's Law"*, University of California at Berkeley, Laura Schechter, University of Wisconsin at Madison, November 1, 2007.

[3] Cindy Durtschi,William Hillison, Carl Pacini, *"The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data"*, , Journal of Forensic Accounting Vol. V 2004, pp 17-34.

[4] Walter R. Mebane, *"Election Forensics: Vote Counts and Benford's Law"*, Jr., July 17, 2006.

[5] Ministry of the Interior, Islamic Republic of Iran (2009a). xls file http://www.moi.ir/Portal/File/ShowFile.aspx?ID=0793459f-18c3-4077-81ef-b6ead48a5065, 2009-06-14.

[6] Philip R. Bevington , D. Keith Robinson," Data Reduction and Error Analysis for The Physical Sciences", McGraw-Hill Inc., US, 2002.

[7] M. Sadegh Movahed,Lecture notes for Advance Topics in Data Analysis, http://faculties.sbu.ac.ir/~movahed/attachments/modeling.pdf

[8] Lindley, Dennis V. (1957). "A Statistical Paradox". *Biometrika* **44**: 187–192

[9] Andrew Gelman, et. Al., "Gayesian Data Analysis", CHAPMAN & HALL/CRC, 2004.

[10] Mebane, W.R., Jr. (2009). Note on the presidential election in Iran, June 2009, http://www.umich.edu/~wmebane/note20jun2009.pdf

[11] Boudewijn F. Roukema, arXiv:0906.2789

[۱۲] رسول رستگاری، خبرگزاری انتخاب، http://www.entekhabnews.us/portal/index.php?news=6435&print،۱۳۸۸/۳/۲۷

[۱۳] سیدمهدی سیدنصرالله، http://alef.ir/1388/content/view/48019، ۱۳۸۸/۴/۱

[14] I. Narsky, *PHYSTAT2003, SLAC, Stanford, California, September 8-11, 2003*


[1] Standard Deviation
[2] Mean Standard deviation