

中山大学本科生数据分析报告

(2024 学年秋季学期)

课程名称: 概率论与数理统计

研究对象: 诺贝尔物理学奖得主年龄

学号: xxxxxxxx

姓名: XXX

专业(方向): 计算机科学与技术 (系统结构方向)

Email: longyh27@mail2.sysu.edu.cn

诺贝尔物理学奖得主年龄调研分析报告

一、研究背景和目标

诺贝尔奖自 1901 年首次颁发以来，到 2024 年已历经 124 届（不包括 1940 至 1942 年未颁发的年份）。其中物理学奖、化学奖、生理学和医学奖等领域的获奖者对世界科学的进步作出了巨大贡献。本研究旨在分析诺贝尔物理学奖得主在其重要贡献发现时的年龄规律、我们假设样本来自正态总体，并通过抽样样本进行参数估计和假设检验。

本研究的目标包括：

- 1、研究诺贝尔物理学奖得主贡献发现时的年龄分布；
- 2、在一定置信水平下，求解诺贝尔物理学奖得主贡献发现年龄的置信区间；
- 3、在一定置信水平下，进行假设检验，检验诺贝尔物理学奖得主贡献发现年龄的样本均值是否符合某一假设。

二、数据收集与准备

数据来源：本研究所使用的诺贝尔奖数据来源于公开的维基百科数据库，涵盖 1901 年到 2024 年诺贝尔物理学奖得主的出生年份、获奖年份以及所获奖项的科学贡献。

每位获奖者的年龄计算公式为：

$$\text{年龄} = \text{获奖年份} - \text{出生年份}$$

在 1901 年到 2024 年间，在剔除了缺失的数据之后，我们选取了 224 条诺贝尔物理学奖获奖数据作为总体样本空间，并从中随机抽取 30 个作为样本进行分析。

三、数据初步分析与展示

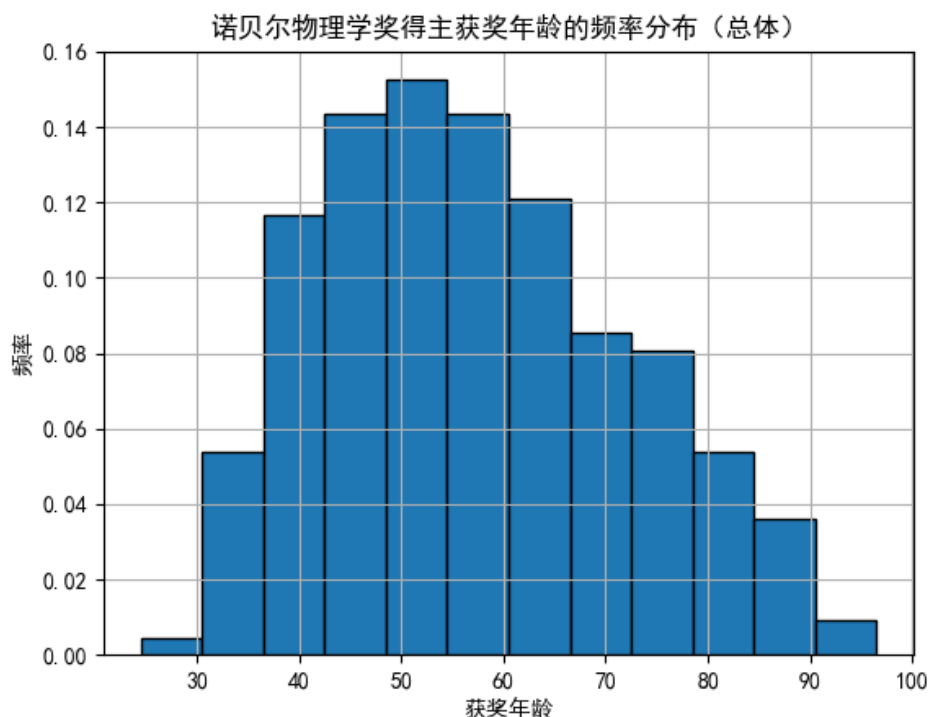
通过对 224 名获奖者获奖时的年龄统计（总体数据见附录），可以得到获奖年龄最小值为 25，最大值为 96，即所有数据落在区间 $[25, 96]$ ，现取区间 $[24.5, 96.5]$ (比数据精度高

一位，避免数据落在分点上)，覆盖区间 $[25, 96]$ ，将 $[24.5, 96.5]$ 等分为 12 个小区间，小区间的长度记为 Δ ， $\Delta = \frac{96.5-24.5}{12} = 6$ 。 Δ 称为组距。小区间的端点称为组限。数出落在每个

小区间内的数据的频数 $\frac{f_i}{n}$ ($n = 224, i = 1, 2, \dots, 12$)如下表：

组限	频数 f_i	频率 f_i/n	累积频率
24.5 ~ 30.5	1	0.004	0.004
30.5 ~ 36.5	12	0.053	0.058
36.5 ~ 42.5	26	0.116	0.174
42.5 ~ 48.5	32	0.143	0.318
48.5 ~ 54.5	34	0.152	0.470
54.5 ~ 60.5	32	0.143	0.614
60.5 ~ 66.5	27	0.121	0.735
66.5 ~ 72.5	19	0.085	0.820
72.5 ~ 78.5	18	0.080	0.901
78.5 ~ 84.5	12	0.053	0.955
84.5 ~ 90.5	8	0.035	0.991
90.5 ~ 96.5	2	0.008	1

频率直方图如下：



通过计算得到总体均值 $\mu = 57.02$, 方差 $\sigma^2 = 225.52$

假设总体服从 $X \sim N(\mu, \sigma^2)$ 正态分布

从中随机抽取 30 个样本数据(X_1, X_2, \dots, X_{30})如下:

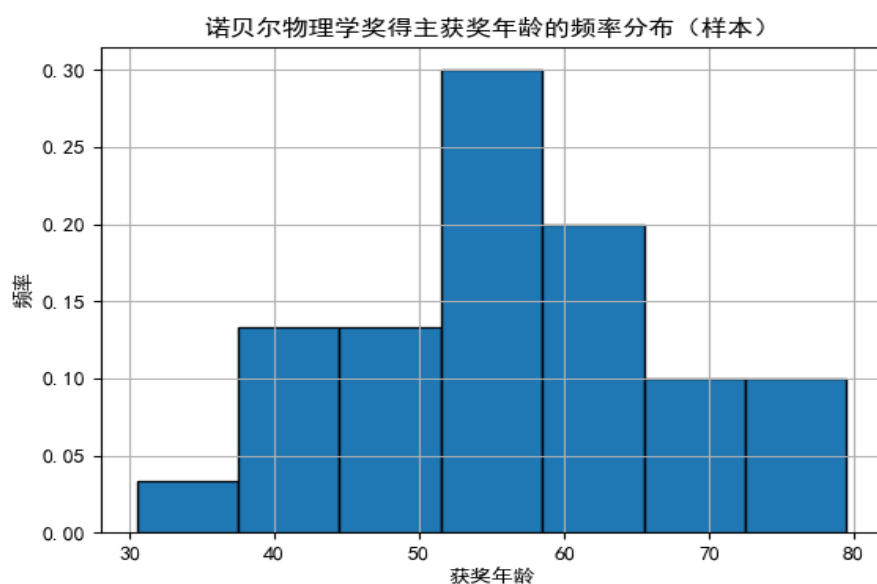
55	47	63	76	65	53
54	45	61	79	48	43
72	38	56	38	55	48
59	44	64	53	56	68
77	52	68	31	63	53

样本最小值为 31, 最大值为 79, 即所有数据落在区间 $[31, 79]$, 现取区间 $[30.5, 79.5]$ (比数据精度高一位, 避免数据落在分点上), 覆盖区间 $[31, 79]$, 将 $[30.5, 79.5]$ 等分为 7 个小区间, 小区间的长度记为 Δ , $\Delta = \frac{79.5-30.5}{7} = 7$. Δ 称为组距. 小区间的端点称为组限. 数出

落在每个小区间内的数据的频数 $\frac{f_i}{n}$ ($n = 30, i = 1, 2, \dots, 7$)如下表:

组限	频数 f_i	频率 f_i/n	累积频率
30.5 ~ 37.5	1	0.033	0.033
37.4 ~ 44.5	4	0.133	0.166
44.5 ~ 51.5	4	0.133	0.300
51.5 ~ 58.5	9	0.300	0.600
58.5 ~ 65.5	6	0.200	0.800
65.5 ~ 72.5	3	0.100	0.900
72.5 ~ 79.5	3	0.100	1

频率直方图如下：



四、数据分析与结果

- 1) 总体 $X \sim N(\mu, \sigma^2)$, $\sigma^2 = 225.52$, 假设均值 μ 未知, 下面求 μ 的置信水平为 $1 - \alpha$ 的置信区间

我们知道 \bar{X} 是 μ 的无偏估计, 且有

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

即

$$P\left\{\bar{X} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right\} = 1 - \alpha$$

$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 所服从的分布 $N(0, 1)$ 不依赖于任何未知参数, 按标准正态分布的上 α 分

位数的定义, 有:

$$P\left\{\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right\} = 1 - \alpha$$

化简可得到 μ 的一个置信水平为 $1 - \alpha$ 的置信区间

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}}z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$$

取 $1 - \alpha = 0.95$, 即 $\alpha = 0.05$, $n = 30$, 查表得 $z_{\alpha/2} = z_{0.025} = 1.96$, 通过计算得到样本均值的观察值 $\bar{x} = 56.13$, 数据代入上述公式计算可以得到 μ 的一个置信水平为 0.95 的置信区间 (56.13 ± 0.98) , 即 $(55.15, 57.11)$ 。

故在置信水平为 0.95 的诺贝尔物理学奖获得者重要贡献发现的平均年龄的置信区间为 $(55.15, 57.11)$

- 2) 总体 $X \sim N(\mu, \sigma^2)$, 总体均值 $\mu_0 = 57.02$, $\sigma^2 = 225.52$, 我们需要根据样本值来判断样本均值 $\mu = 57.02$ 还是 $\mu \neq 57.02$ 。为此, 我们提出两个相互对立的假设:

$$H_0: \mu = \mu_0 = 57.02$$

和

$$H_1: \mu \neq \mu_0$$

我们知道, \bar{X} 是 μ 的无偏估计, \bar{X} 的观察值 \bar{x} 的大小在一定程度上反映 μ 的大小。

因此, 如果假设 H_0 为真, 则观察值 \bar{x} 与 μ_0 的偏差 $|\bar{x} - \mu_0|$ 一般不应太大, 我们就

怀疑假设 H_0 的正确性而拒绝 H_0 , 并考虑到当 H_0 为真时 $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$, 而衡

量 $|\bar{x} - \mu_0|$ 的大小可归结为衡量 $\left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right|$ 的大小。基于上面的想法, 我们可适当选定一

正数 k ，使当观察值 \bar{x} 满足 $\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| \geq k$ 时就拒绝假设 H_0 ，反之，若 $\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| < k$ ，则接受假设 H_0 。然而，由于我们以样本为依据进行决策，当实际上 H_0 为真时仍可能做出拒绝 H_0 的决策（这种错误的可能性是无法消除的），因此，我们考虑统计量

$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ ，只允许犯这类错误的概率最大为 α ，即令

$$P\{\text{当} H_0 \text{ 为真时拒绝} H_0\} = P_{\mu_0} \left\{ \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \geq k \right\} = \alpha$$

由于当 H_0 为真时， $Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ ，由标准正态分布分位数的定义得 $k = z_{\alpha/2}$ ，因此，若 Z 的观察值满足

$$|z| = \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| \geq k = z_{\alpha/2}$$

则拒绝 H_0 ，而若

$$|z| = \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| < k = z_{\alpha/2}$$

则接受 H_0 。

取 $\alpha = 0.05$ ，则有 $z_{\alpha/2} = z_{0.025} = 1.96$ ，又已知 $n = 30$ ， $\sigma^2 = 225.52$ ， $\bar{x} = 56.13$ ，通过计算可得到

$$\left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| = 0.32 < 1.96$$

于是接受假设 H_0 ，即在显著性水平 $\alpha = 0.05$ 的条件下，我们接受诺贝尔物理学奖获得者重要贡献发现的样本平均年龄 $\mu = 57.02$ 的假设。

五、结论

通过对诺贝尔物理学奖得主贡献发现年龄的分析，可以得出以下结论：

1) 年龄分布特征：

诺贝尔物理学奖得主在其重大科学贡献时的年龄呈现出一定的分布规律。根据对 224 名获奖者的年龄数据分析，发现年龄分布大致符合正态分布，均值约为 57.02 岁。大多数

获奖者的贡献发生在 40 岁到 70 岁之间，这表明科学家的重要发现往往是在其职业生涯的成熟期。

2) 置信区间:

通过样本数据的分析，计算得出诺贝尔物理学奖得主贡献发现的平均年龄的 95% 置信区间为 $(55.15, 57.11)$ 。这意味着在 95% 的置信度下，诺贝尔物理学奖得主的贡献发现时的平均年龄位于这个区间内。

3) 假设检验:

在假设检验中，针对诺贝尔物理学奖得主贡献发现的平均年龄是否为 57.02 岁，我们通过样本均值和标准误差进行计算。结果表明，样本数据未能显著偏离假设均值 57.02，因此我们接受原假设，认为诺贝尔物理学奖得主贡献发现的平均年龄为 57.02 岁。

4) 学术价值与启示:

本研究通过概率论与数理统计方法对诺贝尔物理学奖得主贡献发现的平均年龄进行了深入分析，为理解科学家的创新与贡献提供了数据支持。研究结果表明，诺贝尔奖得主通常在职业生涯的中后期取得重要突破，强调了经验积累和科研环境对科学创新的重要性。此外，研究为未来对其他领域诺贝尔奖得主年龄分布的探讨提供了借鉴。

综上所述，诺贝尔物理学奖得主的贡献发现年龄呈现出特定的年龄规律，在统计上符合正态分布，且其平均贡献发现年龄为 57.02 岁，具有一定的科学和实践意义。

附录:

1901~2024 诺贝尔物理学奖得主年龄（来自维基百科）

颁奖年 份	人名	出生年 份	获奖年 龄
1901	威廉·伦琴	1845	56
1902	亨得里克·洛伦兹	1853	49
1902	彼得·塞曼	1865	37
1903	亨利·贝克勒	1852	51
1903	皮埃尔·居里	1859	44
1903	玛丽·居里	1867	36
1904	约翰·斯特拉特	1842	62
1905	菲利普·莱纳德	1862	43
1906	约瑟夫·汤姆孙	1856	50
1907	阿尔伯特·迈克耳孙	1852	55
1908	加布里埃尔·李普曼	1845	63
1909	古列尔莫·马可尼	1874	35
1909	卡尔·费迪南德·布劳恩	1850	59
1910	约翰内斯·范德瓦耳斯	1837	73
1911	威廉·维恩	1864	47
1912	古斯塔夫·达伦	1869	43
1913	海克·卡末林·昂内斯	1853	60
1914	马克斯·冯·劳厄	1879	35
1915	威廉·亨利·布拉格	1862	53
1915	威廉·劳伦斯·布拉格	1890	25
1917	查尔斯·巴克拉	1877	40
1918	马克斯·普朗克	1858	60
1919	约翰尼斯·斯塔克	1874	45
1920	夏尔·纪尧姆	1861	59
1921	阿尔伯特·爱因斯坦	1879	42
1922	尼尔斯·波尔	1885	38
1923	罗伯特·密立根	1868	55
1924	曼内·西格巴恩	1886	38
1925	詹姆斯·弗兰克	1882	43
1925	古斯塔夫·赫兹	1887	38
1926	让·佩兰	1870	56
1927	阿瑟·康普顿	1892	35

1927	查尔斯·威耳逊	1869	58
1928	欧文·理查森	1879	49
1929	路易·德布羅意公爵	1892	37
1930	钱德拉塞卡拉·拉曼	1888	42
1932	维尔纳·海森堡	1901	31
1933	薛定谔	1887	46
1933	保罗·狄拉克	1902	31
1935	詹姆斯·查德威克	1891	44
1936	维克托·赫斯	1883	53
1936	卡尔·戴维·安德森	1905	31
1937	克林顿·戴维孙	1881	56
1937	乔治·汤姆孙	1892	45
1938	恩里科·费米	1901	37
1939	欧内斯特·劳伦斯	1901	38
1943	奥托·施特恩	1888	55
1944	伊西多·拉比	1898	46
1945	沃尔夫冈·泡利	1900	45
1946	珀西·布里奇曼	1882	64
1947	爱德华·阿普尔顿	1892	55
1948	帕特里克·布莱克特	1897	51
1949	汤川秀树	1907	42
1950	塞西尔·鲍威尔	1903	47
1951	约翰·考克饶夫	1897	54
1951	欧内斯特·沃吞	1903	48
1952	费利克斯·布洛赫	1905	47
1952	爱德华·珀塞尔	1912	40
1953	弗里茨·塞尔尼克	1888	65
1954	马克斯·玻恩	1882	72
1954	瓦尔特·博特	1891	63
1955	威利斯·兰姆	1913	42
1955	波利卡普·库施	1911	44
1956	威廉·肖克利	1910	46
1956	约翰·巴丁	1908	48
1956	沃尔特·布喇顿	1902	54
1957	杨振宁	1994	37
1957	李政道	1926	31
1958	帕维尔·切连科夫	1904	54

1958	伊利亚·弗兰克	1908	50
1958	伊戈尔·塔姆	1895	63
1959	埃米利奥·塞格雷	1905	54
1959	欧文·张伯伦	1920	39
1960	唐纳德·格拉泽	1926	34
1961	罗伯特·霍夫施塔特	1915	46
1961	鲁道夫·穆斯堡尔	1929	32
1962	列夫·朗道	1908	54
1963	尤金·维格纳	1902	61
1963	玛丽亚·格佩特-梅耶	1906	57
1963	約翰內斯·延森	1907	56
1964	查尔斯·汤斯	1915	49
1964	尼古拉·巴索夫	1922	42
1964	亚历山大·普罗霍罗夫	1916	48
1965	朝永振一郎	1906	59
1965	朱利安·施温格	1918	47
1965	理查德·费曼	1918	47
1966	阿尔弗雷德·卡斯特勒	1902	64
1967	汉斯·贝特	1906	61
1968	路易斯·阿尔瓦雷茨	1911	57
1969	默里·盖尔曼	1929	40
1970	汉尼斯·阿尔文	1908	62
1970	路易·奈爾	1904	66
1971	伽博·丹尼斯	1900	71
1972	约翰·巴丁	1908	64
1972	利昂·库珀	1930	42
1972	约翰·施里弗	1931	41
1973	江崎玲於奈	1925	48
1973	伊瓦尔·贾埃弗	1929	44
1973	布赖恩·约瑟夫森	1940	33
1974	马丁·赖尔	1918	56
1974	安东尼·休伊什	1924	50
1975	奥格·玻尔	1922	53
1975	本·莫特森	1926	49
1975	利奥·雷恩沃特	1917	58
1976	伯顿·里克特	1931	45
1976	丁肇中	1936	40

1977	菲利普·安德森	1923	54
1977	内维尔·莫特	1905	72
1977	約翰·凡扶累克	1899	78
1978	彼得·卡皮查	1894	84
1978	阿诺·彭齐亚斯	1933	45
1978	羅伯特·伍德羅·威爾遜	1936	42
1979	谢尔登·格拉肖	1932	47
1979	阿卜杜勒·萨拉姆	1926	53
1979	史蒂文·温伯格	1933	46
1980	詹姆斯·克罗宁	1931	49
1980	瓦尔·菲奇	1923	57
1981	凯·西格巴恩	1918	63
1981	尼古拉斯·布隆伯根	1920	61
1981	阿瑟·肖洛	1921	60
1982	肯尼斯·威耳孙	1936	46
	苏布拉马尼扬·钱德拉塞		
1983	卡	1910	73
1983	威廉·福勒	1911	72
1984	卡洛·鲁比亚	1934	50
1984	西蒙·范德梅尔	1925	59
1985	克劳斯·冯·克利青	1943	42
1986	恩斯特·鲁斯卡	1906	80
1986	格尔德·宾宁	1947	39
1986	海因里希·罗雷尔	1933	53
1987	约翰内斯·贝德诺尔茨	1950	37
1987	卡尔·米勒	1927	60
1988	利昂·莱德曼	1922	66
1988	梅尔文·施瓦茨	1932	56
1988	杰克·施泰因贝格尔	1921	67
1989	诺曼·拉姆齐	1915	74
1989	汉斯·德莫尔特	1922	67
1989	沃尔夫冈·保罗	1913	76
1990	杰尔姆·弗里德曼	1930	60
1990	亨利·肯德尔	1926	64
1990	理查·爱德华·泰勒	1929	61
1991	皮埃尔-吉勒·德热纳	1932	59
1992	乔治·夏帕克	1924	68

1993	拉塞尔·赫尔斯	1950	43
1993	约瑟夫·泰勒	1941	52
1994	伯特伦·布罗克豪斯	1918	76
1994	克利福德·沙尔	1915	79
1995	马丁·佩尔	1927	68
1995	弗雷德里克·莱因斯	1918	77
1996	道格拉斯·奥谢罗夫	1945	51
1996	罗伯特·科尔曼·理查森	1937	59
1997	朱棣文	1948	49
1997	克洛德·科昂-唐努德日	1933	64
1997	威廉·丹尼尔·菲利普斯	1948	49
1998	羅伯特·勞夫林	1950	48
1998	霍斯特·施特默	1949	49
1998	崔琦	1939	59
1999	杰拉德·特·胡夫特	1946	53
1999	马丁纽斯·韦尔特曼	1931	68
2000	若雷斯·阿尔费罗夫	1930	70
2000	赫伯特·克勒默	1928	72
2000	傑克·基爾比	1923	77
2001	埃里克·康奈尔	1961	40
2001	卡尔·威曼	1951	50
2001	沃尔夫冈·克特勒	1957	44
2002	雷蒙德·戴维斯	1914	88
2002	小柴昌俊	1926	76
2002	里卡尔多·贾科尼	1931	71
2003	阿列克谢·阿布里科索夫	1928	75
2003	维塔利·金兹堡	1916	87
2003	安東尼·萊格特	1938	65
2004	戴维·格娄斯	1941	63
2004	休·波利策	1949	55
2004	弗兰克·维尔切克	1951	54
2005	罗伊·格劳伯	1925	80
2005	约翰·霍尔	1934	71
2005	特奥多尔·亨施	1941	64
2006	约翰·马瑟	1945	61
2006	乔治·斯穆特	1945	61
2007	艾尔伯·费尔	1994	13

2007	彼得·格林贝格	1939	68
2008	小林诚 (物理学家)	1944	64
2008	益川敏英	1940	68
2008	南部阳一郎	1921	87
2009	高錕	1933	76
2009	威拉德·博伊尔	1924	85
2009	乔治·埃尔伍德·史密斯	1930	79
2010	安德烈·海姆	1958	52
2010	康斯坦丁·诺沃肖洛夫	1974	36
2011	索尔·珀尔马特	1959	52
2011	布莱恩·施密特	1967	44
2011	亚当·里斯	1969	42
2012	塞尔日·阿罗什	1944	68
2012	戴维·瓦恩兰	1944	68
2013	弗朗索瓦·恩格勒	1932	81
2013	彼得·希格斯	1929	84
2014	赤崎勇	1929	85
2014	天野浩	1960	54
2014	中村修二	1954	60
2015	梶田隆章	1959	56
2015	阿瑟·麦克唐纳	1943	72
2016	戴维·索利斯	1934	82
2016	邓肯·霍尔丹	1951	65
2016	约翰·科斯特利茨	1943	73
2017	莱纳·魏斯	1932	85
2017	巴里·巴里什	1936	81
2017	基普·索恩	1940	77
2018	阿瑟·阿什金	1922	96
2018	热拉尔·穆鲁	1944	74
2018	唐娜·斯特里克兰	1959	59
2019	吉姆·皮布尔斯	1935	84
2019	米歇尔·麦耶	1942	77
2019	迪迪埃·奎洛兹	1966	53
2020	羅傑·潘洛斯	1931	89
2020	赖因哈德·根策尔	1952	68
2020	安德烈娅·盖兹	1965	55
2021	真鍋淑郎	1931	90

2022	阿兰·阿斯佩	1947	75
2022	约翰·克劳泽	1942	80
2022	安东·蔡林格	1945	77
2023	皮埃尔·阿戈斯蒂尼	1941	82
2023	费伦茨·克劳斯	1962	61
2023	安妮·吕利耶	1958	65
2024	约翰·霍普菲尔德	1933	91
2024	杰弗里·辛顿	1947	77

维基百科数据爬取脚本：

```
import requests
from bs4 import BeautifulSoup

from openpyxl import load_workbook

import re

file_path = 'excel/诺贝尔物理学奖得主获奖年龄.xlsx'
sheet_name = 'Sheet1'
def append_row(file_path,sheet_name,data):
    workbook = load_workbook(file_path)
    sheet = workbook[sheet_name]

    # 整行追加写入 Excel
    sheet.append(data)

    workbook.save(file_path)

headers = {
    "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/131.0.0.0 Safari/537.36
Edg/131.0.0.0"
}

year = []
response =
requests.get("https://zh.wikipedia.org/wiki/%E8%AF%BA%E8%B4%9D%E5%B
```

```

0%94%E7%89%A9%E7%90%86%E5%AD%A6%E5%A5%96%E5%BE%97%E4%B8%BB%E5%88%97%E8%A1%A8")
html = response.text
soup = BeautifulSoup(html, "lxml")
table = soup.find_all('table', class_='wikitable')
for tbody in table:
    trs = tbody.find_all('tr')
    for tr in trs:
        tds = tr.find_all('td')
        for td in tds:
            if (td.string):
                year = re.findall(r'\d{4}', td.string)
                if (year):
                    print(year[0])
            else:
                if (td.a and td.span == None and td.a.has_attr('title')):
                    href = td.a['href']
                    title = td.a['title']
                    response2 = requests.get(f'https://zh.wikipedia.org{href}')
                    html2 = response2.text
                    soup2 = BeautifulSoup(html2, "lxml")
                    p = soup2.p
                    # print(p.prettify())
                    # print(href)
                    year2 = re.findall(r'\d{4}', p.text)
                    if year and year2: # 检查 year 和 year2 是否为空
                        age = int(year[0]) - int(year2[0])
                        data = [year[0], title, year2[0], age]
                        print(data)
                        append_row(file_path, sheet_name, data)
                    else:
                        print("year 或 year2 列表为空")

```

数据处理以及绘图脚本:

```

import matplotlib.pyplot as plt
import numpy as np
from openpyxl import load_workbook

```



```
# 设置中文字体
plt.rcParams['font.sans-serif'] = ['SimHei'] # 指定默认字体为黑体
plt.rcParams['axes.unicode_minus'] = False # 解决保存图像时负号 '-' 显示为方块的问题

# 加载工作簿
workbook = load_workbook(filename='D:\\code\\python\\excel\\诺贝尔物理学奖得主获奖年龄.xlsx')

# 选择工作表
sheet = workbook['Sheet1']

# 读取单元格内容
cell_value = sheet['A1'].value
print(f'A1 单元格的值是: {cell_value}')

# 读取整列数据
column_values = [cell.value for cell in sheet['D'] if isinstance(cell.value, int)]
print(len(column_values))
print(f'第 D 列的值是: {column_values}')

# 手动指定直方图的开始和结束区间
start = 24.5 # 开始区间
end = 96.5 # 结束区间
bins = 12 # 区间的数量

# 绘制频率分布直方图
hist, bin_edges = np.histogram(column_values, bins=bins, range=(start, end))
bin_centers = (bin_edges[:-1] + bin_edges[1:]) / 2

# 计算频率和累积频率
frequency = hist / np.sum(hist)
cumulative_frequency = np.cumsum(frequency)

# 输出每个区间的频数、频率和累积频率
for i in range(len(bin_centers)):
```

```
print(f'区间 {bin_edges[i]} - {bin_edges[i+1]} 的频数: {hist[i]}, 频率: {frequency[i]}, 累积频率: {cumulative_frequency[i]}')

# 计算均值和方差
mean = np.mean(column_values)
variance = np.var(column_values)

# 输出均值和方差
print(f'均值: {mean}, 方差: {variance}')

# 绘制直方图
plt.bar(bin_centers, frequency, width=np.diff(bin_edges), edgecolor='black') # 使用自定义的区间
plt.title('诺贝尔物理学奖得主获奖年龄的频率分布（总体）')
plt.xlabel('获奖年龄')
plt.ylabel('频率')
plt.grid(True)
plt.show()

# 设置随机种子
np.random.seed(42)

random_samples = np.random.choice(column_values, size=30, replace=False)
print(random_samples)

# 手动指定直方图的开始和结束区间
start = 30.5 # 开始区间
end = 79.5 # 结束区间
bins = 7 # 区间的数量

# 绘制频率分布直方图
hist, bin_edges = np.histogram(random_samples, bins=bins, range=(start, end))
bin_centers = (bin_edges[:-1] + bin_edges[1:]) / 2

# 计算频率和累积频率
frequency = hist / np.sum(hist)
cumulative_frequency = np.cumsum(frequency)
```

```
# 输出每个区间的频数、频率和累积频率
for i in range(len(bin_centers)):
    print(f'区间 {bin_edges[i]} - {bin_edges[i+1]} 的频数: {hist[i]}, 频率: {frequency[i]}, 累积频率: {cumulative_frequency[i]}')

# 计算均值和方差
mean = np.mean(random_samples)
variance = np.var(random_samples)

# 输出均值和方差
print(f'均值: {mean}, 方差: {variance}')

# 绘制直方图
plt.bar(bin_centers, frequency, width=np.diff(bin_edges), edgecolor='black') # 使用自定义的区间
plt.title('诺贝尔物理学奖得主获奖年龄的频率分布（样本）')
plt.xlabel('获奖年龄')
plt.ylabel('频率')
plt.grid(True)
plt.show()
```