



Course material

Course:

ENG606 / PHYS 442

Video:

DOE_lesson1_part3

Concepts (extracted from automatically generated subtitles):

Type of data. Quantitative data. Very important health problem. Only thing. Group of students. Much problem. Okay pareto principle. First cookie. Second world war. Obscure important variation. Product b. United states. Highest value. Percent rule. Children of a very tall person.



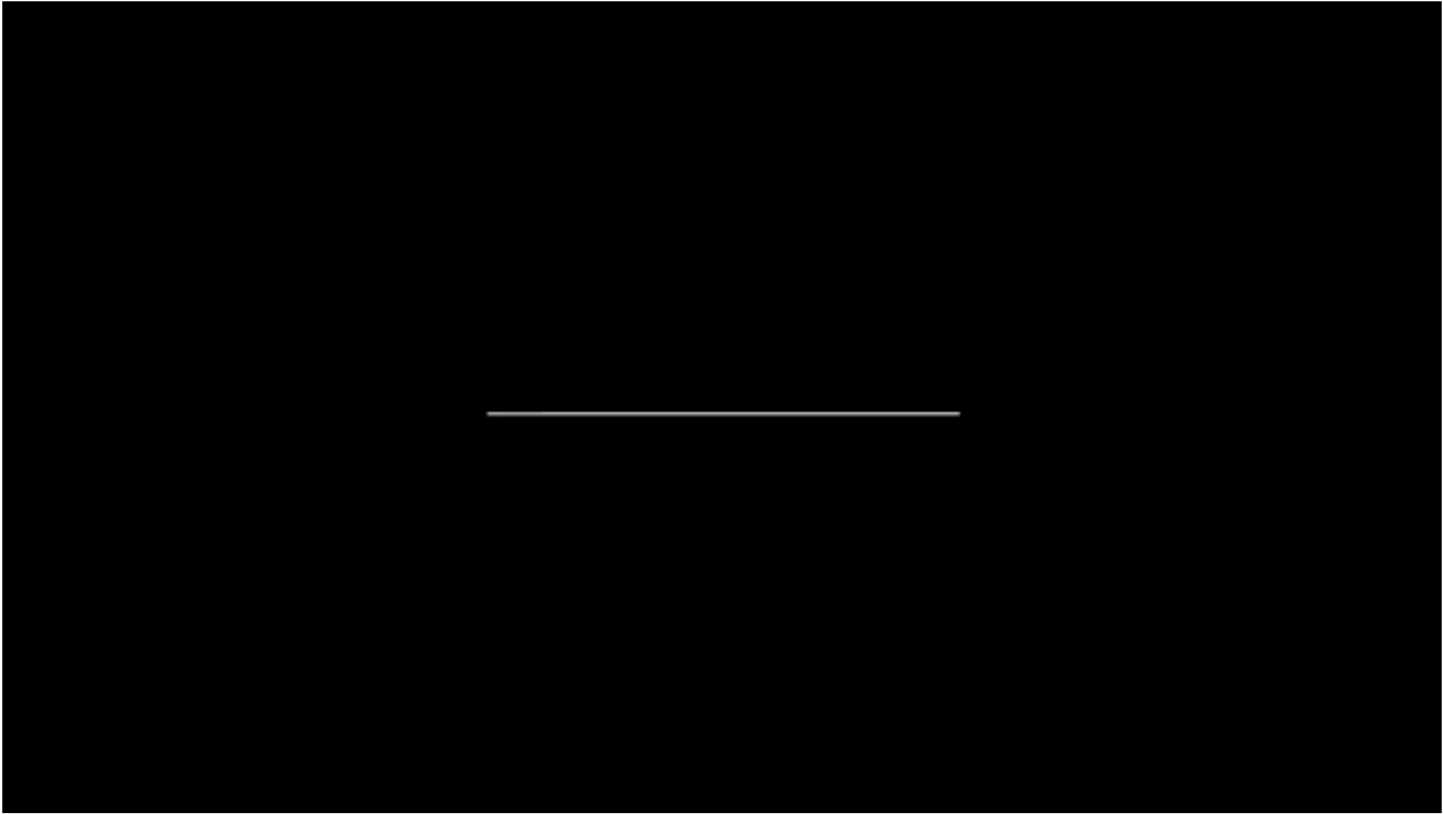
[to video sequence search](#)
(within ENG606 / PHYS 442.)



[to video](#)

Center for Digital Education. More educational support material here:

<https://www.epfl.ch/education/educational-initiatives/cede/educational-technologies-gallery/boocs-en/>
page 1/43



...

notes

.....

.....

.....

.....

.....

.....

.....

.....

.....


.....

summary

.....

.....

0m 0s



.....

.....

1.3.1 Introduction to Data

- ▶ **Data** : Raw facts and figures collected through observations or measurements.
- ▶ **Types of Data** :
 - ▶ *Quantitative Data* : Numerical, can be discrete or continuous.
 - ▶ *Qualitative Data* : Categorical, can be nominal or ordinal.
- ▶ **Examples** : Test scores, survey responses, temperature readings.

These subtitles have been generated automatically

notes

summary

0m 1s



1.3.2 Situations Based on Types of Data

Type of X	Type of Y	Situation Description
Qualitative	Qualitative	Comparing frequencies between groups
Qualitative	Quantitative	Comparing means or medians between groups
Quantitative	Qualitative	Logistic regression
Quantitative	Quantitative	Regression

I asked yesterday, chatGPT to make this slide. So I said, OK, I had one slide quite obscure. I was not happy for many years. And I said, chatGPT. So the next slide are from chatGPT. They should write chatGPT. OK, so type of data is something very, very, very important to understand. You have quantitative and qualitative data. And they are not treated the same way. We cannot do the same. So what we call quantitative data. So it's a value, 5, 3, 4, 1,000, 2 millions, et cetera. And when we do qualitative, so it's a category. So it was the product A or the product B or the product C. We first cookie. Cookie. Cookie. Or we first dry it. So it's going to be operations. So it's a do not represent by number. You represent them by your category. When it's ordinal so that you can make an order between them, it is quite a little bit the question. You can sometimes put something. You can transfer them as quite quantitative order, and became quantitative. There's a quantity, but a quantity. So type of data, desk, course, survey, response, temperature readings, et cetera. So this are the possibilities. So you have your factor and you have your answer. So we usually use X for factors and Y for the responses, or the responses, just the usual choice. So what you can do. So if you have balls that are qualitative, what's the only thing that you can try to do is to compare the frequency between groups. And after they are statistic, but I will not do that in this question. When they are qualitative and quantitative, you can compare means, millions, and both. We will do that. We'll make something called constant post-vision model. I've been showing a way of treating this. It's happened, but all happens, but they say, qualitative, qualitative in engineering. Why

notes

summary

0m 5s



1.3.2 Situations Based on Types of Data

Type of X	Type of Y	Situation Description
Qualitative	Qualitative	Comparing frequencies between groups
Qualitative	Quantitative	Comparing means or medians between groups
Quantitative	Qualitative	Logistic regression
Quantitative	Quantitative	Regression

don't I think so much? So we try to usually to avoid, even because it's not science. It's also science, but sometimes complicated science. Quantitative, quantitative. Sometimes people call it, I know that I don't appreciate calling that constant post-mission or random post-mission model. Like that, you want to understand if it's better to attend the course, make the exercise, make reading books in the library. So we can take a group of students and make tests on the results. So it's possible to work with. When it's quantitative, qualitative, typically, it makes some value for your inputs and the answer is fail or not. So we use to make a logistic regression. It's possible for the work. The regulation, the name doesn't... But there are many situations in science where we work with. And watch our course will be most of the time when we have quantitative and quantitative. So you put some value, not more, some resistance and some inductance in your electrical network and you are playing in some position, attention, and current. So you have numbers on one side and numbers on the other side. So in this course, we will work mainly on... Let's say we will work a lot with that and we will work a little bit... With that in the last chapter. I didn't resist to the word regression. I don't know if you know where it's coming. Word regression. It's Galton who has coined this term regression. And its interest was to discover what's happened. Galton is known for something we call regression to the mean. What's happened was the children of a very tall person.

notes

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

summary

.....

.....

.....

.....

1.3.3 Introduction to Metadata

- ▶ **Metadata** : Data about data ; provides context and additional information.
- ▶ **Purpose** :
 - ▶ Helps in understanding the context, quality, and limitations of the data.
 - ▶ Supports data management and interpretation.
- ▶ **Examples** : Date of data collection, methodology, units of measurement, source.

because finally children of tall people sometimes are not as tall as their parents and so the words of repression. I tell you, those three persons have really invented quite everything.

notes

summary

5m 37s



Introduction

Course organization
Why learning DOE?
A few basic statistical concepts
Weighing three objects

on to Metadata

data about data ; provides context and additional

Understanding the context, quality, and limitations of

data management and interpretation.

ate of data collection, methodology, units of

source.


an-Marie FürbringerModelling and design of experiments

After when we have data, we have metadata, it's also very important. So metadata, the data, the data, the data. They tell you when, where, exactly in which configuration, who and the measurement that we've done. It's very important to keep a straight relation between the data and the metadata. It will help you to understand and sometimes it will make a big change to analyze the wrong data. Maybe because what you have in mind, but not only the conclusion would be completely wrong if you don't take it. So I really insist on the necessity to keep a very straight link between data and the data. It's also why I recommend you to use in MATLAB more tables and arrays for your data. Or when you make Excel files, a lot of people keep the data in Excel files. It's very, very important to keep track. And another point, laboratory notebook, I see students taking notes. This is the metadata portion. But for the metadata in an experiment, use a laboratory notebook. Unfortunately, in this school, I don't understand why it's not a culture. Now, some laboratory insists obliged that in the general culture of EPFL, unfortunately, it's not. We didn't learn that students in TPP are not used to make a laboratory notebook. It's very important because it's also a good way to be sure that this file is the file you have taken. When is that, the fire clock, when we have this, I don't know, the window was open or it was sunny or the temperature of the laboratory was very low. It really lets you to understand what you are. So very, very important. The unit is also the stabilization of your device and anything which was the device that you use in reality, etc.

notes

summary

6m 2s



Graunt's book 1663

Natural and Political Observations Made upon the Bills of Mortality

The Table of CASUALTIES.

The Years of our Lord	1647	1648	1649	1650	1651	1652	1653	1654	1655	1656	1657	1658	1659	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670	1671	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688	1689	1690	1691	1692	1693	1694	1695	1696	1697	1698	1699	1700	1701	1702	1703	1704	1705	1706	1707	1708	1709	1710	1711	1712	1713	1714	1715	1716	1717	1718	1719	1720	1721	1722	1723	1724	1725	1726	1727	1728	1729	1730	1731	1732	1733	1734	1735	1736	1737	1738	1739	1740	1741	1742	1743	1744	1745	1746	1747	1748	1749	1750	1751	1752	1753	1754	1755	1756	1757	1758	1759	1760	1761	1762	1763	1764	1765	1766	1767	1768	1769	1770	1771	1772	1773	1774	1775	1776	1777	1778	1779	1780	1781	1782	1783	1784	1785	1786	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	1800	1801	1802	1803	1804	1805	1806	1807	1808	1809	1810	1811	1812	1813	1814	1815	1816	1817	1818	1819	1820	1821	1822	1823	1824	1825	1826	1827	1828	1829	1830	1831	1832	1833	1834	1835	1836	1837	1838	1839	1840	1841	1842	1843	1844	1845	1846	1847	1848	1849	1850	1851	1852	1853	1854	1855	1856	1857	1858	1859	1860	1861	1862	1863	1864	1865	1866	1867	1868	1869	1870	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2433	2434	2435	2436	2437	2438	2439	2440	2441	2442	2443	2444	2445	2446	2447	2448	2449	2450	2451	2452	2453	2454	2455	2456	2457	2458	2459	2460	2461	2462	2463	2464	2465	2466	2467	2468	2469	2470	2471	2472	2473	2474	2475	2476	2477	2478	2479	2480	2481	2482	2483	2484	2485	2486	2487	2488	2489	2490	2491	2492	2493	2494	2495	2496	2497	2498	2499	2500	2501	2502	2503	2504	2505	2506	2507	2508	2509	2510	2511	2512	2513	2514	2515	2516	2517	2518	2519	2520	2521	2522	2523	2524	2525	2526	2527	2528	2529	2530	2531	2532	2533	2534	2535	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545	2546	2547	2548	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558	2559	2560	2561	2562	2563	2564	2565	2566	2567	2568	2569	2570	2571	2572	2573	2574	2575	2576	2577	2578	2579	2580	2581	2582	2583	2584	2585	2586	2587	2588	2589	2590	2591	2592	2593	2594	2595	2596	2597	2598	2599	2600	2601	2602	2603	2604	2605	2606	2607	2608	2609	2610	2611	2612	2613	2614	2615	2616	2617	2618	2619	2620	2621	2622	2623	2624	2625	2626	2627	2628	2629	2630	2631	2632	2633	2634	2635	2636	2637	2638	2639	2640	2641	2642	2643	2644	2645	2646	2647	2648	2649	2650	2651	2652	2653	2654	2655	2656	2657	2658	2659	2660	2661	2662	2663	2664	2665	2666	2667	2668	2669	2670	2671	2672	2673	2674	2675	2676	2677	2678	2679	2680	2681	2682	2683	2684	2685	2686	2687	2688	2689	2690	2691	2692	2693	2694	2695	2696	2697	2698	2699	2700	2701	2702	2703	2704	2705	2706	2707	2708	2709	2710	2711	2712	2713	2714	2715	2716	2717	2718	2719	2720	2721	2722	2723	2724	2725	2726	2727	2728	2729	2730	2731	2732	2733	2734	2735	2736	2737	2738	2739	2740	2741	2742	2743	2744	2745	2746	2747	2748	2749	2750	2751	2752	2753	2754	2755	2756	2757	2758	2759	2760	2761	2762	2763	2764	2765	2766	2767	2768	2769	2770	2771	2772	2773	2774	2775	2776	2777	2778	2779	2780	2781	2782	2783	2784	2785	2786	2787	2788	2789	2790	2791	2792	2793	2794	2795	2796	2797	2798	2799	2800	2801	2802	2803	2804	2805	2806	2807	2808	2809	2810	2811	2812	2813	2814	2815	2816	2817	2818	2819	2820	2821	2822	2823	2824	2825	2826	2827	2828	2829	2830	2831	2832	2833	2834	2835	2836	2837	2838	2839	2840	2841	2842	2843	2844	2845	2846	2847	2848	2849	2850	2851	2852	2853	2854	2855	2856	2857	2858	2859	2860	2861	2862	2863	2864	2865	2866	2867	2868	2869	2870	2871	2872	2873	2874	2875	2876	2877	2878	2879	2880	2881	2882	2883	2884	2885	2886	2887	2888	2889	2890	2891	2892	2893	2894	2895	2896	2897	2898	2899	2900	2901	2902	2903	2904	2905	2906	2907	2908	2909	2910	2911	2912	2913	2914	2915	2916	2917	2918	2919	2920	2921	2922	2923	2924	2925	2926	2927	2928	2929	2930	2931	2932	2933	2934	2935	2936	2937	2938	2939	2940	2941	2942	2943	2944	2945	2946	2947	2948	2949	2950	2951	2952	2953	2954	2955	2956	2957	2958	2959	2960	2961	2962	2963	2964	2965	2966	2967	2968	2969	2970	2971	2972	2973	2974	2975	2976	2977	2978	2979	2980	2981	2982	2983	2984	2985	2986	2987	2988	2989	2990	2991	2992	2993	2994	2995	2996
-----------------------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

1.3.5 Categorization : Titanic passengers

Age	Class	Dead		Survivors	
		M	F	M	F
Children	1st	0	0	5	1
	2nd	0	0	11	13
	3rd	35	17	13	14
	Crew	0	0	0	0
Adults	1st	118	4	57	140
	2nd	154	13	14	80
	3rd	387	89	75	76
	Crew	670	3	192	20

"[...] Those methods deal comprehensively with entire species, and with entire groups of influences, just as if they were single entities, and express the relations between them in an equally compendious manner. They commence by marshalling the values in order of magnitude from the smallest up to the largest, thereby converting a mob into an orderly array, which like a regiment thenceforth becomes a tactical unit.
", F. Galton, Biometrika, Volume 1, Issue 1, October 1901

So data, this is an example of very, very old data, perhaps first data for making statistics. It's about the bill of mortality in London or in England. Look at the years, 1640 something, etc. trying to understand. It's data that is not to prove that is also data that have been organized. And so we start to eventually understand something because we see frequencies being higher than others and eventually I don't know. Yeah, I cannot read that but those are people that died from gangrene. So you see the problem of gangrene is a very important health problem, etc. Data organized by categories is the first part of the analysis. Here is an example of the people who died in the Titanic. And observe that when you organize your data, you start to understand something that a story behind the organization of the data. Yeah, we have young boys and young girls have died also.

notes

summary

9m 1s



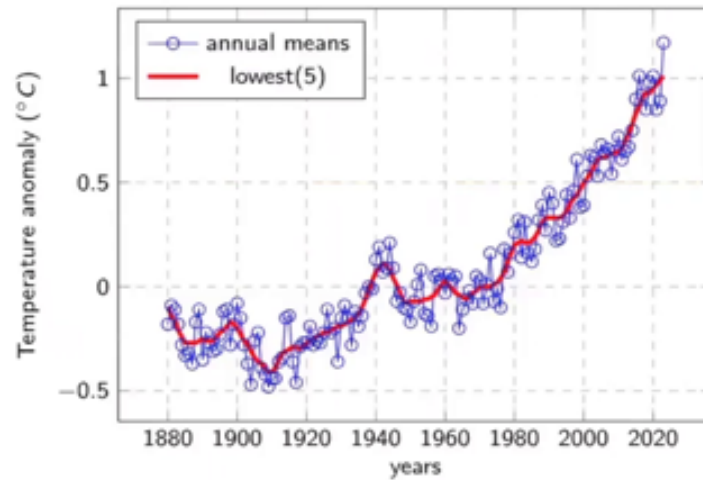
year	crude	lowest(5)	year	crude	lowest(5)	year	crude	lowest(5)
1901	-0.15	-0.23	1951	-0.07	-0.07	2001	0.53	0.52
1902	-0.28	-0.26	1952	0.01	-0.07	2002	0.63	0.55
1903	-0.37	-0.28	1953	0.08	-0.07	2003	0.62	0.58
1904	-0.47	-0.31	1954	-0.13	-0.06	2004	0.53	0.61
1905	-0.26	-0.34	1955	-0.14	-0.06	2005	0.68	0.62
1906	-0.22	-0.36	1956	-0.19	-0.05	2006	0.64	0.62
1907	-0.39	-0.37	1957	0.05	-0.04	2007	0.66	0.63
1908	-0.42	-0.39	1958	0.06	-0.01	2008	0.54	0.64
1909	-0.48	-0.41	1959	0.03	0.01	2009	0.65	0.64
1910	-0.44	-0.41	1960	-0.03	0.03	2010	0.72	0.65
1911	-0.44	-0.39	1961	0.06	0.01	2011	0.61	0.66
1912	-0.36	-0.35	1962	0.03	-0.01	2012	0.65	0.7
1913	-0.34	-0.32	1963	0.05	-0.03	2013	0.67	0.74
1914	-0.15	-0.31	1964	-0.2	-0.04	2014	0.75	0.78
1915	-0.14	-0.3	1965	-0.11	-0.05	2015	0.9	0.83
1916	-0.36	-0.29	1966	-0.06	-0.06	2016	1.01	0.87
1917	-0.46	-0.29	1967	-0.02	-0.05	2017	0.92	0.91
1918	-0.29	-0.3	1968	-0.08	-0.03	2018	0.85	0.93
1919	-0.27	-0.29	1969	0.05	-0.02	2019	0.98	0.93
1920	-0.27	-0.27	1970	0.03	0	2020	1.01	0.95
1921	-0.19	-0.26	1971	-0.08	0	2021	0.85	0.97
1922	-0.28	-0.25	1972	0.01	0	2022	0.89	0.99
1923	-0.26	-0.24	1973	0.16	0	2023	1.17	1.01
1924	-0.27	-0.23	1974	-0.07	0.01			
1925	-0.22	-0.22	1975	-0.01	0.02			

notes

11m 13s



1.3.7 A picture is worth a thousand words



Dr Jean-Marie Fürbringer

Modelling and design of experiments

So here is another type of data. Something funny about my relational strategy, I had my data when one day or during the weekend, one day it's like, what could be the data I can present to my students for understanding what is important of data? It gave me a few and it gives me temperature of the earth, of the surface of the earth, the average temperature. And okay, you see the data and you are, it's not possible to classify, we classify by the date, it's a rolling date, but with the data, what is also very important

notes

summary

12m 52s



1.3.8 Importance and Risks

► Importance :

- Categorizing data or creating visual representations helps in identifying patterns, trends, and relationships.
- Graphs make complex data more accessible and easier to interpret.
- Visual summaries provide a quick overview of data distribution and key features.

► Risks of Bias :

- Categorization can oversimplify or obscure important variations within data.
- Choice of graph type, scale, or binning can mislead or distort the true nature of data.
- Visual representations might exaggerate or minimize effects, leading to misinterpretations.

► Key Takeaway :

- While categorization and graphing are powerful tools, analysts must remain mindful of biases that can affect data interpretation.

is to make graph. So okay, data classification, but also to make graphic is the way of getting inside of what is in your data. So you see in red, the, now perhaps in blue is the real average measurement, I didn't check, I think picture is worth, is worth a thousand words. It's very important to make graphic of your data and make the analysis of your data but also the graphics let you discover when you are very, what's happened is a trend that you can also see. provide a quick overview of data distribution and key features. Nevertheless, there are a risk of bias. Categorization can over simplify and obscure important variation within the data. Choice of rough type scale. You know that we can play with the scale, where is your region? Probably you want to talk to you in this discussion. Can mislead and distort the true nature of the data. Visual representation might exaggerate and minimize effects leading and misinterpretation. So make that categorization.

notes

summary

14m 8s



1.3.9 Pareidolia



Our brain has the tendency to see motives in random sets as well as a tendency to make a script, which means to tell a story, and give a meaning to an image. Statistics offers tools, like statistical tests, to pass the perception.

Make graphic, but don't fall in the trap. You will see that the only thing that you have to see. So take away while categorization and graphing are powerful tool. Analyst must remain mindful of bias that can affect data in reputation. The story of this slide. So I have this in mind. I write it and I ask Chagipiti, what do you think about it? I'm really curious.

notes

summary

16m 49s



- https://fr.wikipedia.org/wiki/Cydonia_Malus

17m 27s



1.3.11 Survivorship bias



The damaged portions of returning planes show locations where they can sustain damage and still return home; those hit in other places presumably do not survive. (Image shows hypothetical data.) The error was to consider that the planes are *more probably* hit in those places and then reinforce them.

It was just a question of the resolution of the picture. A few years later, when we were able to make better picture of Mars, the statue has disappeared. Opax immersion has erased when they see that the human way. They think that they are amazing. So this type of bias exists also inside. It's not only for making fabric and another reform for laughing, but it exists also in the reality. We have the impression of a curve say something or show us something. And it's better to check better and to see the resolution and things like that.

notes

summary

18m 48s



1.3.12 Pitfalls in data analysis

Cognitive biases

- ▶ illusions
- ▶ of attention
- ▶ of memory
- ▶ of judgment
- ▶ of reasoning

Examples

- ▶ Pareidolia
- ▶ Confusion correlation/causation
- ▶ Bias of confirmation (Cherry picking)
- ▶ Bias of belief
- ▶ Error of attribution
- ▶ Barnum effect (experience of Forer)

https://en.wikipedia.org/wiki/List_of_cognitive_biases

Another bias is all the survivor bias. In the second World War, British aircraft makers looked at the aircraft they come back from the battlefield and tried to see where they get peace. And the problem that they tried to make a better protection of those zones. But the product was about the thing that didn't come back. And it was not necessary to make those regions stronger. It's a region where no planes come back. I mean, here is a place. So OK, this is what we call the survivor bias. I remember when I made the interview for being the main deputy director of mechanical engineering. So the question was about women. They wanted to have more women in mechanical engineering. And I thought, what is the case today? But at my time, they are think or treat like in physics also for girls. And so in the interview, they asked me what you would do for understanding what is the program or you do not have more women in mechanical engineering. And I say we have to ask it's a girl who have got costs. Because if you ask the question, you're the guy here, we're the guy here. You don't have the answer why other and not come. Survivor bias.

notes

summary

19m 48s



1.3.12 Pitfalls in data analysis

Cognitive biases

- ▶ illusions
- ▶ of attention
- ▶ of memory
- ▶ of judgment
- ▶ of reasoning

Examples

- ▶ Pareidolia
- ▶ Confusion correlation/causation
- ▶ Bias of confirmation (Cherry picking)
- ▶ Bias of belief
- ▶ Error of attribution
- ▶ Barnum effect (experience of Forer)

https://en.wikipedia.org/wiki/List_of_cognitive_biases

OK, there are a lot of biases. As a researcher, you have to be aware of these type of bias. So they can happen with your analysis. So we have illusions. We have bias of attention that we could more attention and something we know than something that we don't know. Of memory. Yes, I remember that. But a lot of students don't remember that it can be as important. Of judgment. And the specialist of judgment. This is important. This is not important. Even specialists and experts make errors and we can get elements of reasoning. All the problem of causation and correlations. And we mix of judgment and correlation.

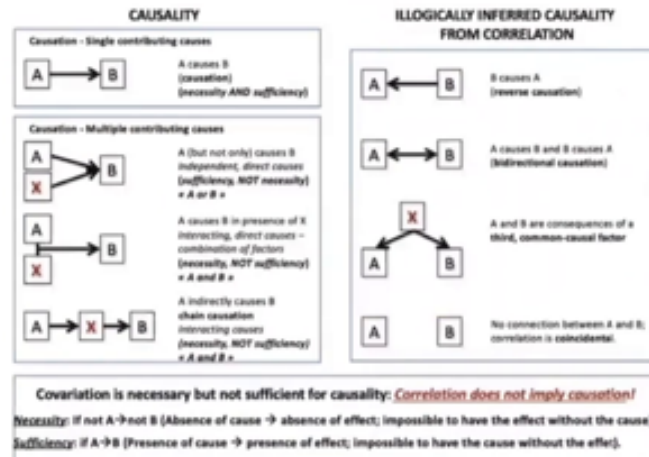
notes

summary

21m 46s



1.3.13 Causality



So here I put a few examples. Parahidolia, confusion, correlation, causation is a very classical bias. It's not because you see a current evolving at the same time. It's your effect that's really there. I think this is a cause. It's really these factors that you are checking. Perhaps both are evolving. I don't know if you ever see this correlation between the numbers of movie or Mr. Cage and the number of deaths by hanging in the United States. You see like that. So it's really possible to put correlation absolutely stupid. But why? Because they depend about the same factor. And so it had been named Barnum effect, because Barnum was a circus in the United States. If you are interested by that, a long list of committees are viable. A lot are quite closed, and sometimes the name doesn't make a big difference.

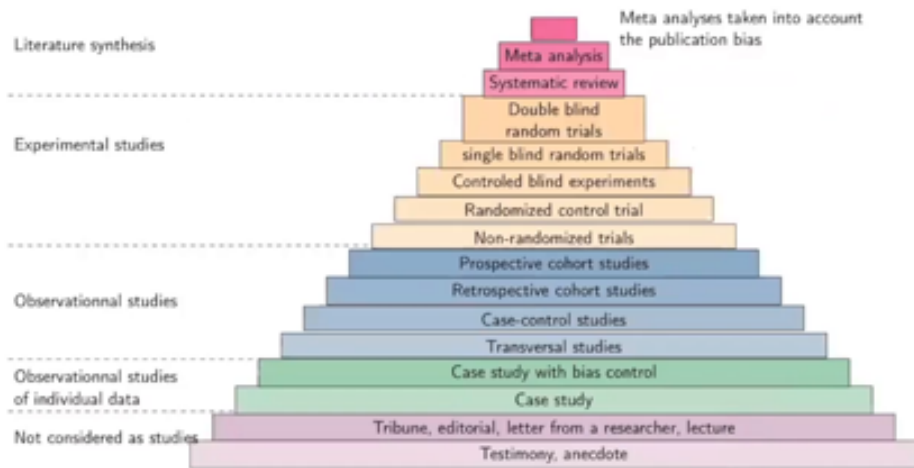
notes

summary

22m 37s



1.3.14 Scientific consensus



A is provoking another event B. And it can happen different things. It can happen that several events can also provoke the same thing. So be careful when you go to the US wide, because you can have different causes of something. It's not because you have proved something from A to B that all B to A will be the same. You can have several factors that have to be in coherence for provoking something. Or you can have some intermediary paintings or what's the X. After the bad thing that you can admit between cause and error, typically when there is feedback loop, it will be quite complicated to identify what is the cause and what is the consequence. You can have big directional causation when you have really an important loop. And you can have two events that seems to be related because they are depending on another. What I explained about the movies of Mr. Cage and the Death by Hanging. And sometimes things are, they look correlated, but in fact, they have no relation. It's wrong. And it could happen that something seems to be related. But in fact, there are no way of rating them. And finally, they are perhaps not related.

notes

summary

24m 49s



1.3.15 Falsifiability (Popper)

- ▶ A statement is said refutable (falsifiable) if and only if it can be logically contradicted by an empirical test. ... more precisely if and only if it exists a possible observation statement (true or false) that would logically contradict the theory.
- ▶ To be accepted as scientific a statement, a theory must be falsifiable.

Causality. I want to be quick. And this is just a stairs about what we call the scientific consensus. Usually in science, like engineering and physics, we have not too much problem with that. In life science, in medicine, it became quite complicated. I will not talk so much, but you can understand that between the testimony of somebody, they say, oh, I see that. And having a meta analysis of different papers and all the literature in the domain, you have level of consensus. And when somebody threatens something, it's really true. It's quite interesting to understand that which level of the consensus. We are.

notes

summary

26m 44s



1.3.15 Falsifiability (Popper)

- ▶ A statement is said refutable (falsifiable) if and only if it can be logically contradicted by an empirical test. ... more precisely if and only if it exists a possible observation statement (true or false) that would logically contradict the theory.
- ▶ To be accepted as scientific a statement, a theory must be falsifiable.

Poor being, science, a proposition needs to be passively able. The passively ability. That means that you need to be able to make an experiment that will accept or refute the proposition. If you are not able to do that, in fact, it's not science. So cosmology, perhaps, is not a science. Because there are limits of things.

notes

summary

27m 38s



1.3.16 Simpson paradox

The **fraction of healing** of 700 patients with and without the use of a given drug. 350 individuals with the medication (group "with"), 350 without the medication (group "without").

	With	Without
Men	81/87	234/270
Women	192/263	55/80
All	273/350	289/350

	With	Without
Men	93%	87%
Women	72%	69%
All	78%	83%

Is the drug to be proposed or not?

And no problem of accepting cosmology as a science, but you understand. Some people say, typically, psychology is not really a science. Because all psychology, all the element of psychology, that's all. You cannot put people in an experiment to see if they became here. It's not accepted. So to be accepted as a scientific statement, a theory must be passively able. That means that you need to be able to make an experiment depending on the reason of the experiment between the two rules.

notes

summary

28m 13s



1.3.18 Analysis

- ▶ If **taking the drug** appears less effective than **doing nothing**, it is because if we randomly choose a person who has taken the drug, the probability is high that this person is a woman and therefore that his recovery is less easy than a person chosen at random from those who did not use the drug, in which case the probability is greater that it is a man and that he recovered more easily.
- ▶ In the example, **being a woman** is a common cause **to take the medicine** and **to heal less easily**.
- ▶ In this case, the aggregated data does not allow to really test the effectiveness of the drug.
- ▶ To test the efficacy of the drug a **homogeneous group** is needed so that the difference in cure rate is attributable to the drug alone and not to the effect of estrogen.

It comes from Popper. Simpson Paradox. It's also not a bias. A paradox is something that seems very, very particular. So you have 1700 patient, and you have a selection of 350 that have taken the drive. And 3002 have not taken the drive. So let's see, theory can be agreed with that. Let's see if the drug is efficient. see what is the ratio of creation for people who have taken the medicine in relation to people who have not taken the medicine. So on the 350 people who have taken the medicine, 273 that have taken the medicine have cured, have been cured. And when you look at the people who have not taken the medicine, you have even more, you have 289 that have been cured. So you say, finally the medicine is not really, when you look at that, you say, the medicine is not effective. And after you look at the man, and you see that in the man, if you look the one who have taken the medicine, in fact, you can look at the ratio if you want, you have 93% of the man who have taken the medicine that have been cured in comparison of only 87% of the man who have not taken the medicine that have been cured. And when you look at the woman, you see that you have 72% of the woman who have taken the medicine that have been cured, again only 69. So if you know the gender of your patient to give the medicine, but if you don't know the gender, it's a paradox, it's exactly a paradox. And the paradox is coming, if you check, because you have not an adequate numbers of person in each subgroups. And nevertheless, the question arises, do you have to put all the groups together or you analyze the groups in between? So if you want to solve

notes

summary

28m 50s



1.3.18 Analysis

- ▶ If **taking the drug** appears less effective than **doing nothing**, it is because if we randomly choose a person who has taken the drug, the probability is high that this person is a woman and therefore that his recovery is less easy than a person chosen at random from those who did not use the drug, in which case the probability is greater that it is a man and that he recovered more easily.
- ▶ In the example, **being a woman** is a common cause **to take the medicine** and **to heal less easily**.
- ▶ In this case, the aggregated data does not allow to really test the effectiveness of the drug.
- ▶ To test the efficacy of the drug **a homogeneous group** is needed so that the difference in cure rate is attributable to the drug alone and not to the effect of estrogen.

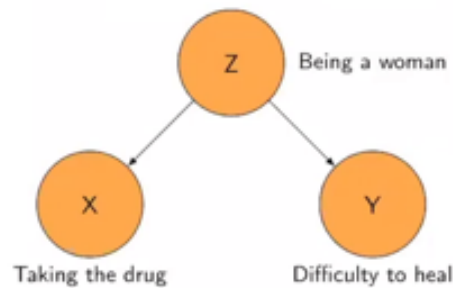
that, it's very important to make the calculations step by step. It's what I call a causal analysis, a cause analysis. How many people so we make for women and we compare them. We have an hypothesis, self-healing against medicine. So let's count really the same. It's what I'm doing is this type of practice. And when you do that, it's became clear that the situation, the story for making the long story short, in fact, the illness is related to this different movement for man, and it was not correct to make a group of all of that because it was different. But if you don't analyze carefully, you have the impression that you have a paradox. So in the exercise, you will have another situation, not between man and woman, but between pressure, the measurement of pressure, the same case that related to the measurement of pressure, how people, depending on the pressure, are cured or not by the

notes

summary

1.3.19 Confounding factor

A confounding factor, (or extraneous determinant or lurking variable) is a variable that influences both the dependent variable and independent variable, causing a spurious association. Confounding is a causal concept, and as such, cannot be described in terms of correlations or associations



medicine. So here you have an analysis. If taking the drug appears less effective than doing nothing, it's because if we randomly choose a person who has taken the drug, the probability is high that this person is a woman and therefore that the recovery is less easy than a person shooting at random from those who did not use the drug, which in this case would be man. So this is the problem, the groups who are not of the same sex. In the example, being a woman is a common cause of taking the medicine and to heal. The two things, the probability to take the medicine was related to the probability to heal. In this case, the aggregated data does not allow to really test the effectiveness of the drug. To test the efficacy of the drug, an homogenous group is needed so that the difference in curing rate is attributable to the drug alone and not to the effect of being a woman. Now this is a case of checking the data correctly and be sure when you are making groups or not making groups what you are doing and why you are making groups or not making groups.

notes

summary

33m 49s



1.3.20 Data about scientific publications

Country	Code	Region	Income	Level	2 000	2 008	pop2000	pop2008
1 Afghanistan	AFG	South Asia	Low income	1	4	112	20 779 957	37 171 922
2 Angola	AGO	Sub-Saharan Africa	Lower middle income	2	7	30	16 395 477	30 809 787
4 Albania	ALB	Europe & Central Asia	Upper middle income	3	23	180	3 089 027	2 866 376
5 Andorra	AND	Europe & Central Asia	High income	4	-	4	65 390	77 008
6 United Arab Emirates	ARE	Middle East & North Africa	High income	4	330	3 145	3 134 067	9 630 966
7 Argentina	ARG	Latin America & Caribbean	Upper middle income	3	4 386	9 811	36 870 796	44 494 102
8 Armenia	ARM	Europe & Central Asia	Upper middle income	3	346	521	3 069 597	2 915 741
9 Antigua and Barbuda	ATG	Latin America & Caribbean	High income	4	0	6	76 007	96 282
10 Australia	AUS	East Asia & Pacific	High income	4	23 276	53 650	19 153 000	24 982 488
11 Austria	AUT	Europe & Central Asia	High income	4	6 577	12 362	8 015 566	8 840 521
12 Azerbaijan	AZE	Europe & Central Asia	Upper middle income	3	155	761	8 048 600	9 939 771
13 Burundi	BDI	Sub-Saharan Africa	Low income	1	2	21	6 378 871	11 175 379
14 Belgium	BEL	Europe & Central Asia	High income	4	9 723	15 688	10 255 250	11 427 054
15 Benin	BDN	Sub-Saharan Africa	Lower middle income	2	43	228	6 885 948	11 485 035
16 Burkina Faso	BFA	Sub-Saharan Africa	Low income	1	48	352	11 607 951	19 751 486
17 Bangladesh	BGD	South Asia	Lower middle income	2	440	9 139	127 463 862	161 376 713
18 Bulgaria	BGR	Europe & Central Asia	Upper middle income	3	1 453	9 311	8 170 173	7 025 037
19 Bahrain	BHR	Middle East & North Africa	High income	4	83	322	464 610	1 569 440
20 Bahamas, The	BHS	Latin America & Caribbean	High income	4	2	20	298 043	385 635
21 Bosnia and Herzegovina	BHN	Europe & Central Asia	Upper middle income	3	75	704	3 755 176	3 323 929
22 Belarus	BLR	Europe & Central Asia	Upper middle income	3	1 170	1 180	9 979 610	9 483 489
23 Belize	BIZ	Latin America & Caribbean	Lower middle income	2	1	9	247 310	383 071
24 Bolivia	BOL	Latin America & Caribbean	Lower middle income	2	39	109	8 418 270	11 353 140
25 Brazil	BRA	Latin America & Caribbean	Upper middle income	3	12 783	60 148	174 790 339	209 469 320

<https://data.worldbank.org/indicator/IP.JRN.ARTC.SC?end=2018&start=2000>

And there are all these ideas of a confounding factor. This is the concept in this situation

notes

summary

35m 6s



1.3.21 Loading data

Matlab has specialized routines to load data in the workspace

MATLAB

- ▶ `X=0 :1 :20` creates the vector $X = [0, 1, 2, \dots, 20]$
- ▶ `X=[...]` creates a vector with specific values
- ▶ `Y= repmat(X,h,w)` creates a matrix copying X, h times vertically w times horizontally
- ▶ `T = table(var1, ..., varN, Name, Value)` creates a table
- ▶ `T = readtable(FileName.xls', 'Sheet', SheetName)` loads a table from Excel

to be considered. So I'm proposing you a database about scientific publication and to check which countries are participating more in the science and technology.

notes

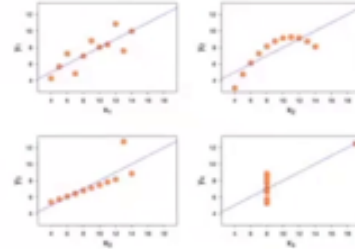
summary

35m 20s



1.3.22 ANSCOMBE'S QUARTET

- ▶ Four data sets that have nearly identical simple descriptive statistics, yet have very different distributions and appear very different when graphed.
- ▶ Each dataset consists of eleven (x,y) points
- ▶ Constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data when analyzing it
- ▶ Intended to counter the impression among statisticians that "numerical calculations are exact, but graphs are rough."



And I want to present you a few elements. So first thing is about loading data in MATLAB. You have some routine for loading data and you have here a summary, a cheat sheet for

notes

summary

35m 33s

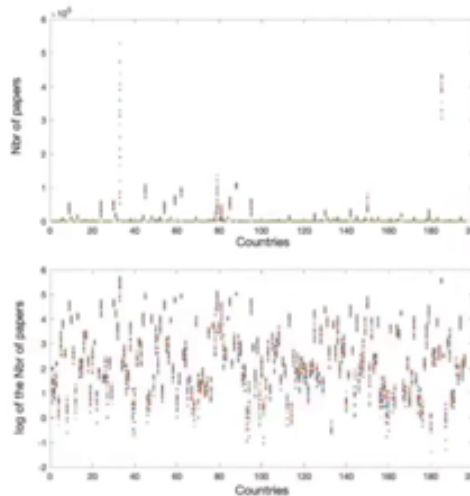


1.3.23 Visual analysis

- ▶ To detect patterns, aberrations, etc.
- ▶ The change of the metric is also interesting

Matlab

- ▶ `plot(x, y)`
`plot(x, y, LineSpec)`
`plot(x1, y1, ..., xn, yn)`
`plot(..., Name, Value)`
- ▶ `bar(x, y)`
`bar(..., Width)`
`bar(..., Style)`
`bar(..., Name, Value)`



loading data in MATLAB. When you are making graphics, this is an argument also about buyers. I was hesitating where to put these slides. I would just let you know about something we call the Atcon Quartet. It's not a play of music. It's the fact that we need graphics and we need data and we need sense. So you see here four series of data. One is really for the way in which we are aligned. Another is in fact, so don't agree with relation. One has aligned with one point and is one point which is very different. And the other was a vertical line is one point very different. If you do the calculation only and never graphic them, they have the same risk, the same linear regression factors, the same quantity, you have the impressions, they are the same. This is what's just an argument for saying, make calculation, make graphics. That would not be the case. What is the first interpretation you've got when you see this is the real answer is you have a resonance at one point you have a resonance if I don't know what

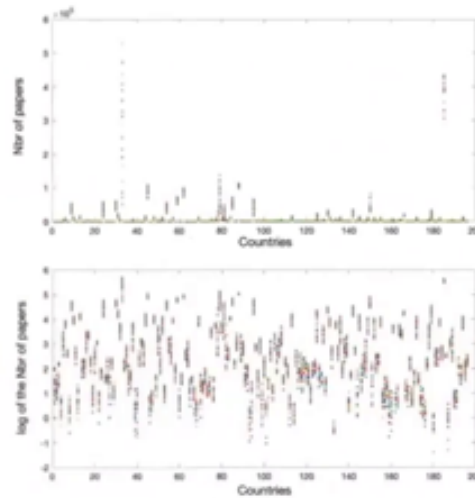
notes

summary

35m 55s



patterns,
s, etc.
ge of the metric is
esting



notes

37m 55s

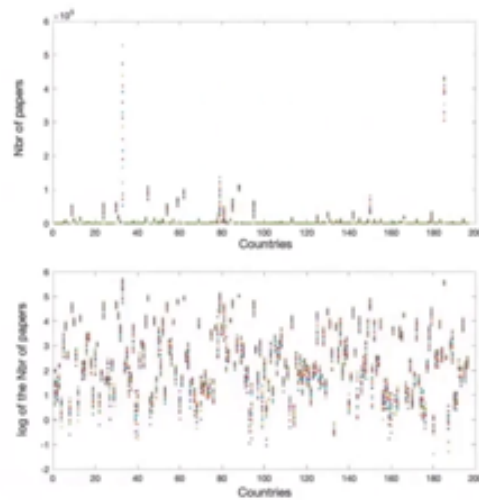


Qual analysis

patterns,
 etc.
 of the metric is
 ing

neSpec)
 $\dots, x_n, y_n)$
 ne, Value)

lth)
 e)
 ne, Value)



I look between this data about the publication okay you see to China and United States there

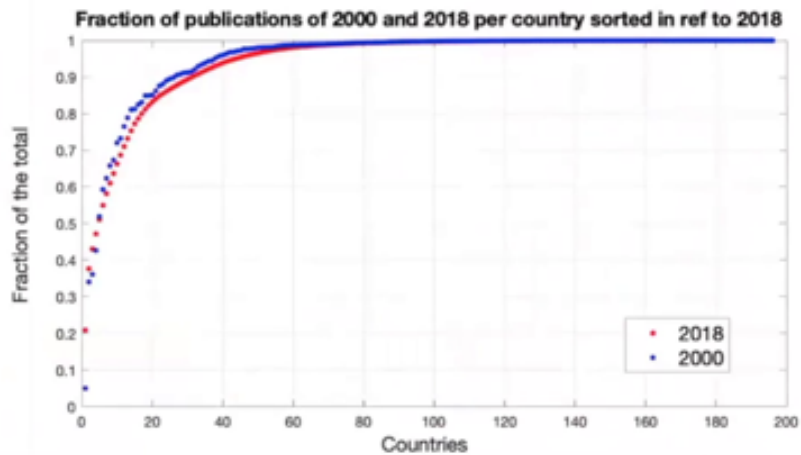
notes

summary

38m 25s



1.3.25 Plot of the sorted data



were really a lot of publication and but if you sort your data typically it was sorting by

notes

summary

38m 37s



1.3.24 Sorting data

$$x_1 \ x_2 \ \dots \ x_N \Rightarrow x_{(1)} < x_{(2)} < \dots < x_{(N)}$$

MATLAB

- ▶ `[B,Index]=sort(A,dim,direction)`
 - ▶ `[B,Index]=sortrows(A,col,direction)`
 - ▶ `[tblB,Index]=sortrows(tblA,col,direction)`
- A, B : matrices*
tblA, tblB : table
dim : dimension to realize the sorting (1,2, ...)
col : column of reference for sorting
direction : 'ascend' or 'descend'

by what to do I sorting sorting the country from the from the highest one of the data we are checking from the highest value to the lowest value you start to see to see something that happened so you can tell us history here between 2000 and 2008 an evolution that you cannot see

notes

summary

38m 39s



1.3.24 Sorting data

$$x_1 \ x_2 \ \dots \ x_N \Rightarrow x_{(1)} < x_{(2)} < \dots < x_{(N)}$$

MATLAB

- ▶ `[B,Index]=sort(A,dim,direction)`
 - ▶ `[B,Index]=sortrows(A,col,direction)`
 - ▶ `[tblB,Index]=sortrows(tblA,col,direction)`
- A, B : matrices*
tblA, tblB : table
dim : dimension to realize the sorting (1,2, ...)
col : column of reference for sorting
direction : 'ascend' or 'descend'

when you are putting your data without this sorting of the of the values where all the country are sorted I don't know by continent or by letter or by the first letter of the thing so sorting is quite very important but again it could be also a bias

notes

summary

39m 8s



Introduction

Course organization
Why learning DOE?
A few basic statistical concepts
Weighing three objects

ta

$$x_{(1)} < x_{(2)} < \dots < x_{(N)}$$

direction)

col,direction)

(tblA,col,direction)

alize the sorting (1,2, ...)

nce for sorting

'descend'

Marie Fürbringer Modelling and design of experiments

and here you see a few a few routines about the function for sorting so don't in my lab don't don't make between sorting it will sort your array each column for itself and sort row where you can sort the row and the column you choose one and

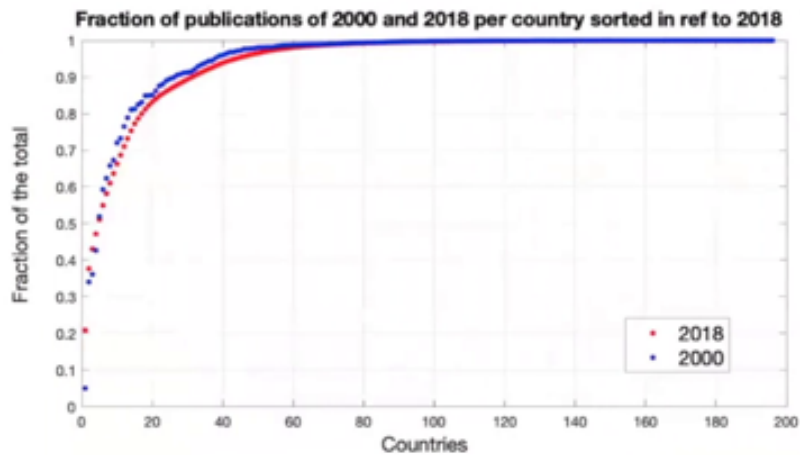
notes

summary

39m 25s



1.3.25 Plot of the sorted data



you sort all the other was with this one okay Pareto principle so it's also very important

notes

summary

40m 1s



1.3.27 Dealing with categories

Defining a column of a table as *categorical* allows you to perform some computation by categories

MATLAB

```

▶ B=categorical(A)
tbl.var = categorical(tbl.var)
▶ statarray = grpstats(tbl,group,stats)
stats=grpstats(tbl,'var1',{ 'min','max' },'tblVars','var2')
▶ gscatter(x,y,group)
A,B : array or column of a table
tbl : table
group : the variable(s) of the groupe
stats : statistics to compute such as min, max, etc.

```

engineering to identify what are the minimum of factors that have the maximum of consequence so sometimes people say the 20 80 percent rule these ideas that have a 20 percent of the cause at 80 percent of the contiguous it's never exactly like that but sometimes it is presented like that what are the few causes that have most of the attack and so the idea is ordering your consequence from the one having the most consequences and one less consequence like that and after you can even add up the different participation and eventually you see where you get 80 percent of the consequence and how many effect and how many factors you have in 80 percent of the consequence and Pareto principle Pareto live in Los Angeles that's why we like it so it's why you have a liceo Pareto very important statistician too especially for the

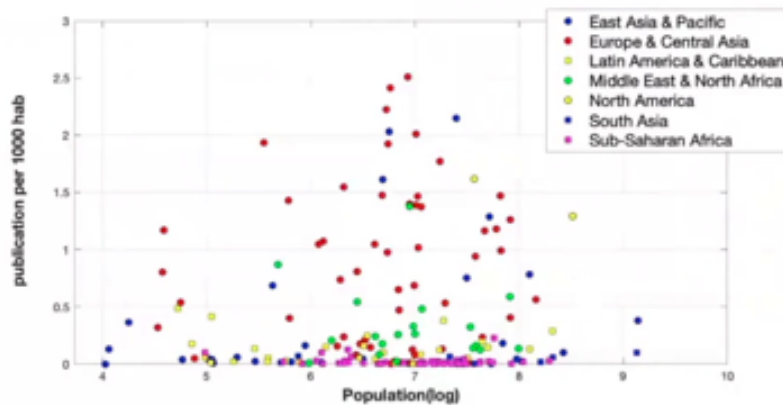
notes

summary

40m 2s



1.3.28 Scatter plot



Dr Jean-Marie Fürbringer

Modelling and design of experiments

economy I would like to let you know that in my class and eventually in Python even if I don't know you have function for working in this category so you can define a variable as a category with the category command and after you can make group statistics group scatter it's helped you quite a lot for winning time when you are making this element you have to analyze that

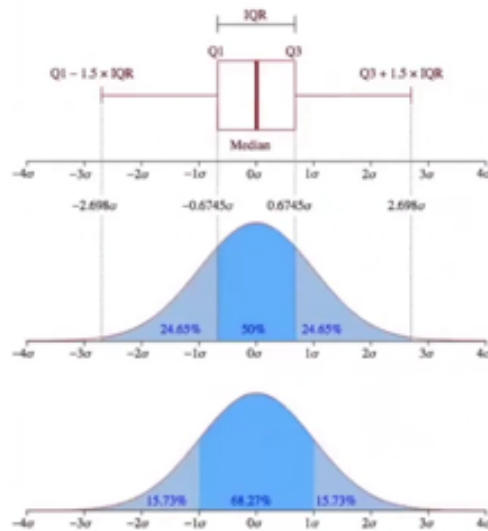
notes

summary

41m 18s



1.3.29 Box-plot (definition)



- *Outliers* are points placed at more than 1.5 IQR at the left of Q_1 or at the right of Q_3
- *Whiskers* are drawn at the minimum or maximum of the data points after the exclusion of the outliers

scatter plot are one of the basic elements for understanding that means that having two distributives that are was two two factors you can make it by groups this have been made by groups

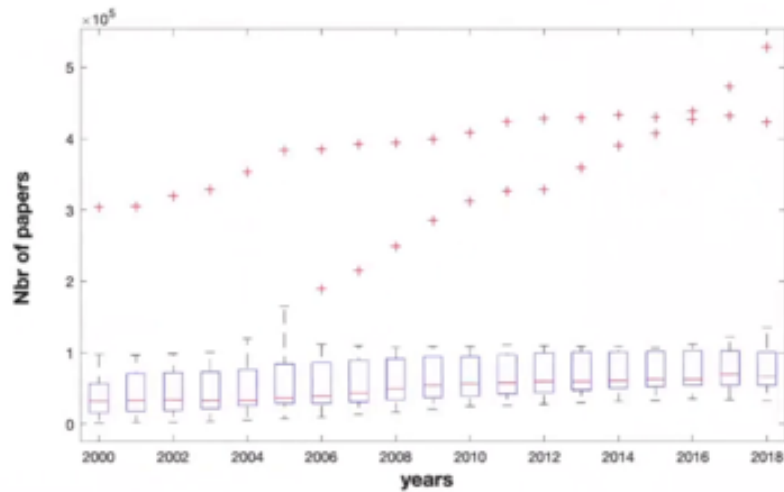
notes

summary

41m 55s



1.3.30 Box-plot



so very very practical box plot and provided it's not the first time you will learn about box plot is a way of summarizing several series and compare them you have in box plot you have the median so the value was the probability of 50 percent of the probability and right and left side after you divide your distribution in quarter of probability so around the median the left 20 percent probability left is 25 percent at right so it give you the box and after you have the moustache I don't know okay and so it's not exactly so it's the quartile one minus 1.5 the interquartile range so the interquartile range is the length of your box so it's not representing exactly the normal distribution because the normal distribution has the plus minus sigma is bigger than the quartile but it's close to so it's a very easy way to summarize when you have a lot of series and you want rapidly compare them box plot is a very interesting so it's also a way of this of the germinating outlays so if you are really data that are outside of this moustache we usually say that they are outlays

notes

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

summary

.....

.....

.....

.....

.....

.....

.....

.....

.....

42m 14s



1.3.31 Location and range

$$X_{(1)}, X_{(2)}, \dots, X_{(N/4)}, \dots, X_{(N/2)}, \dots, X_{(3N/4)}, X_{(N-1)}, X_{(N)}$$

- ▶ Median
- ▶ Average
- ▶ Range
- ▶ Standard deviation
- ▶ Variance
- ▶ Quartile
- ▶ Percentile

MATLAB

$$M = \text{median}(A)$$

$$M = \text{mean}(A)$$

$$R = \text{range}(A)$$

$$S = \text{std}(A)$$

$$v = \text{var}(A)$$

$$Y = \text{quantile}(X, p)$$

so this was for the publication a way of rapidly seeing what's happened and we see exceptions so exception was China and United States each time and you see the rest of the country which quite it's really interesting to see the evolution of things over time over a factor

notes

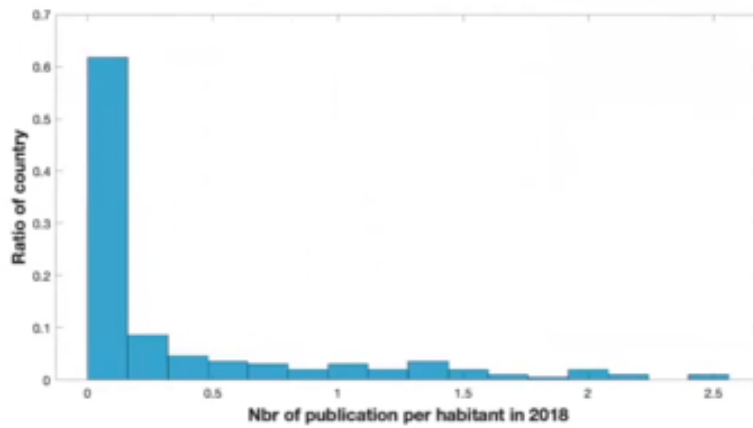
summary

44m 6s



1.3.32 Quartiles and histogram

	Mean	Deviation	Variance	Q25	Median	Q75
2000	0.18	0.36	0.13	0.0025	0.018	0.11
2018	0.37	0.57	0.33	0.0128	0.08	0.48



Dr Jean-Marie Fürbringer Modelling and design of experiments

etc location and range so you have different function media mean range standard deviation quartile so just remembering you that MATLAB is doing that location and range a few calculation

notes

summary

44m 30s



1.3.34 Data vs distributions

- ▶ Original data : $Y_i \sim N(\mu_i, \sigma_i)$ avec $0 \leq i \leq n$
- ▶ Linear : $a_j = \sum_i x_{ij} Y_i \sim N\left(\mu = \sum x_{ij} \mu_i, \sigma = \sqrt{\sum x_{ij}^2 \sigma_i^2}\right)$
- ▶ Average : $\sqrt{n-1} \left(\frac{\bar{Y} - \mu}{s}\right) \sim T(\nu)$
- ▶ Quadratic function : $(a_j)^2 \sim \chi^2(w_j)$
- ▶ Quadratic function quotient : $\frac{(a_j)^2}{(a_i)^2} \sim F(w_j, w_i)$

I go rapidly and this quartile and histogram let you just describe what easily in first approximation distribution yeah this I will come back next I will you don't need that you don't need that for today so I will stop my course

notes

summary

44m 48s

