

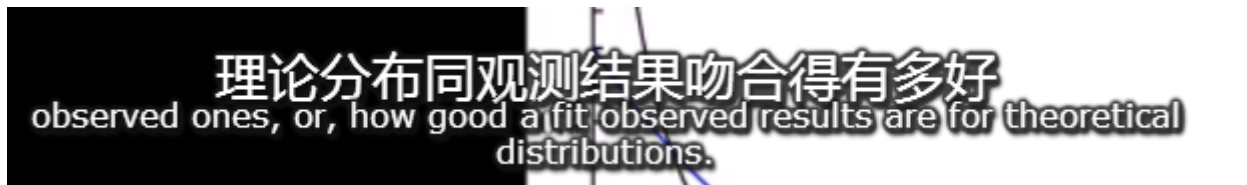
χ²分布介绍

χ²分布是概率论与统计学中常用的一种概率分布。k个独立的标准正态分布变量的平方和服从自由度为k的卡方分布。卡方分布常用于假设检验和置信区间的计算。若来自正态总体的k个随机变量、.....、相互独立，且数学期望为0、方差为1（即服从标准正态分布），则随机变量 $X=\sum Zi^2$ ，被称为服从自由度为k的χ²分布，记作 $X\sim\chi^2(k)$ 。

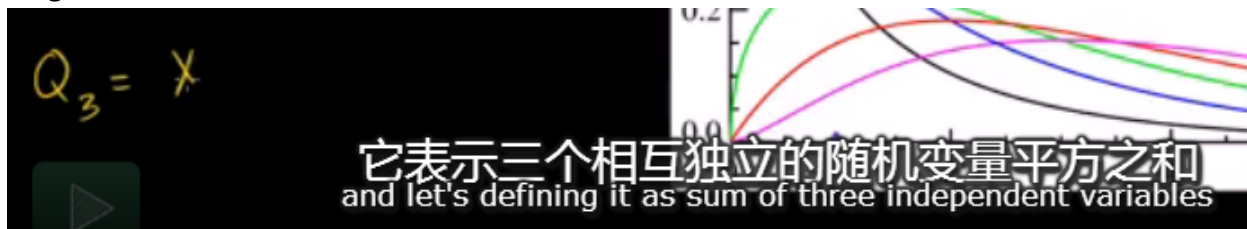
n '	P									
	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	
1	0.02	0.1	0.45	1.32	2.71	
2	0.01	0.02	0.02	0.1	0.21	0.58	1.39	2.77	4.61	
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	
4	0.21	0.3	0.48	0.71	1.06	1.92	3.36	5.39	7.78	
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	3
6	0.68	0.87	1.24	1.64	2.2	3.45	5.35	7.84	10.64	3
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	3
8	1.34	1.65	2.18	2.73	3.4	5.07	7.34	10.22	13.36	3
9	1.73	2.09	2.7	3.33	4.17	5.9	8.34	11.39	14.68	3
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	3
11	2.6	3.05	3.82	4.57	5.58	7.58	10.34	13.7	17.28	3
12	3.07	3.57	4.4	5.23	6.3	8.44	11.34	14.85	18.55	2
13	3.57	4.11	5.01	5.89	7.04	9.3	12.34	15.98	19.81	2
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	2
15	4.6	5.23	6.27	7.26	8.55	11.04	14.34	18.25	22.31	
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	
17	5.7	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	2
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	25.99	2
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.2	3
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	3
21	8.03	8.9	10.28	11.59	13.24	16.34	20.34	24.93	29.62	3
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	3
23	9.26	10.2	11.69	13.09	14.85	18.14	22.34	27.14	32.01	3
24	9.89	10.86	12.4	13.85	15.66	19.04	23.34	28.24	33.2	3
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	3
26	11.16	12.2	13.84	15.38	17.29	20.84	25.34	30.43	35.56	3
27	11.81	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	4
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	4
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	4
30	13.79	14.95	16.79	18.49	20.6	24.48	29.34	34.8	40.26	4
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.8	!
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	

60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.4	81.9
70	43.28	45.44	48.76	51.74	55.33	61.7	69.33	77.58	85.53	94.4
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	105.9
90	59.2	61.75	65.65	69.13	73.29	80.62	89.33	98.64	107.56	117.5
100	67.33	70.06	74.22	77.93	82.36	90.13	99.33	109.14	118.5	129.6

chi square distribution 卡方分布



degree of freedom



73课

皮尔逊 χ^2 检验

这一节以一个简单的餐厅一周每日顾客量预计和观测值的例子，使用 χ^2 检验进行了假设检验。 χ^2 检验由皮尔逊重新发现，运用很广泛。

Day:	M	T	W	T	F	S	網
Expected %:	10	10	15	20	30	15	
Observed:	30	14	34	45	57	20	

举个例子，假设你有一万块钱放在2个钱包里，如果你知道其中一个钱包里有多少钱，那另一个钱包里有多少钱你就知道了， $x_1 + x_2 = 10000$ 。两个钱包相当于两个变量，一个已知，另一个也会相应知道，而且另一个的钱数取决于第一个钱包的钱数，第一个多它就少，第一个少它就多。所以另一个钱包的“自由取决于第一个钱包。这就是自由度为1的情况。同样，如果换成三个钱包，只有知道了两个钱包的钱数，你才能知道第三个钱包的钱数，所以第三个钱包的自由度为2。

我不太懂，卡方分布的条件是变量服从正太分布，这里显然没有说正太分布，难道又是中心极限定理吗？抽取总体中的样本？近似于正太分布？进而平方和后服从正太分布？但是我又不懂为什么要相减完了又除？晕了呢？求指教

这是皮尔逊卡方检验，统计量的构造是 $\chi^2 = \sum \{ [(实际频数 - 理论频数)^2 / 理论频数] \}$ 。视频中也说到，皮尔逊卡方检验统计量近似服从卡方分布（最严格的卡方检验构造就是若干个正态分布之和了）。可以上网查相关资料。

74课

列联表 χ^2 检验

列联表是以列表方式表示两个或多个变量或属性共同出现的频率。这一节使用一个列联表的例子，再一次练习了 χ^2 检验。

contingency table

75~77课

方差分析1：计算总平方和

方差分析（ANOVA），是用于两个及两个以上样本均数差别的显著性检验。这一节从计算总平方和SST，总平方和可以理解为计算方差时，不除以n的那部分。

the total sum of squares 简称SST

the grand mean(总平均值)

自由度可否理解为独立分布的变量个数？

76 方差分析中，由于各种因素的影响，研究所得的数据呈现波动状，这种波动可以分为组间波动和组内波动两种情况。这一节讲解了两者的差异和联系。

sum of squares within SSW组内平方和

这里要计算的是组间平方和SSB B表示between(组间)
we'll call it sum of squares between, the B stands for between.

组内平方和+组间平方和=总平方和
is that the sum of squares within plus the sum of squares between

组内自由度 + 组间自由度 = 总平方和的自由度

m-1 m(n-1) mn-1

77课

方差分析3：F统计量假设检验

F检验，是指一种统计学意义上服从F-分布的零假设的检验。这一节继续前两节的内容，对特定例子进行了F检验

F统计量是 组间平方和同除以其自由度...
Sum of Squares between divided by, our degrees of freedom between

有时这被称为组间均方(MSB)
and this is sometimes called the mean squares between, MSB,
F-statistic = $\frac{SSB}{m-1}$

F表示费希尔 是这位生物学家兼统计学家提出了此分布
F stands for Fisher who is the biologist and statistician who came up with this.

78课

相关性和因果性

相关性是指两个或多个事物同时发生，具有关联，而因果性是指因为A所以B，两者具有明显的差异。这一节通过实际例子讲解这一问题。