假设检验

讲师：Jeary

目录

目标

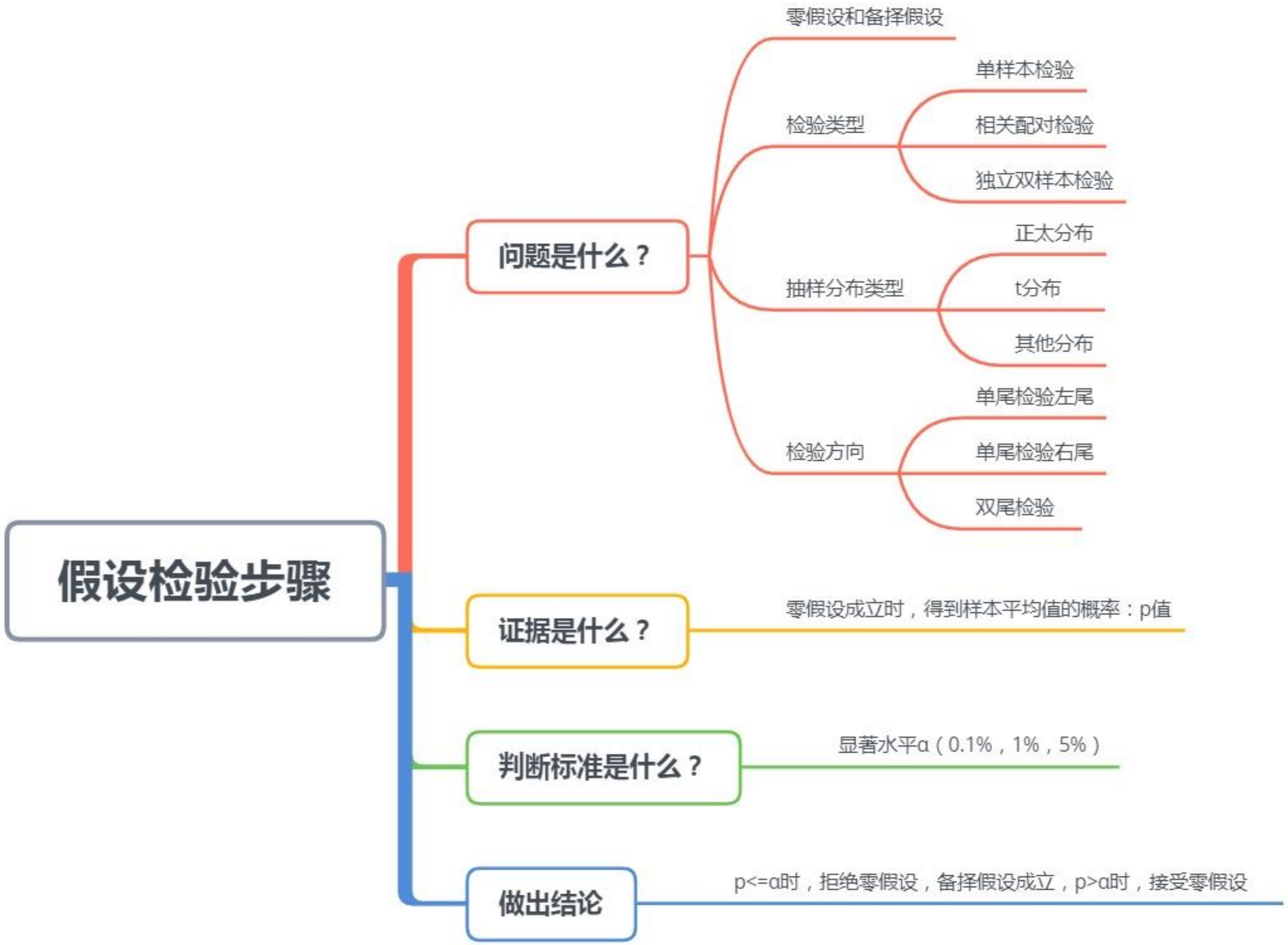
通过本章课程的学习，您将能够：

- 掌握假设检验的步骤
- 掌握t检验、χ^2检验和F检验的计算方法
- 了解t检验、χ^2检验和F检验的应用场景

01

简介

基本步骤



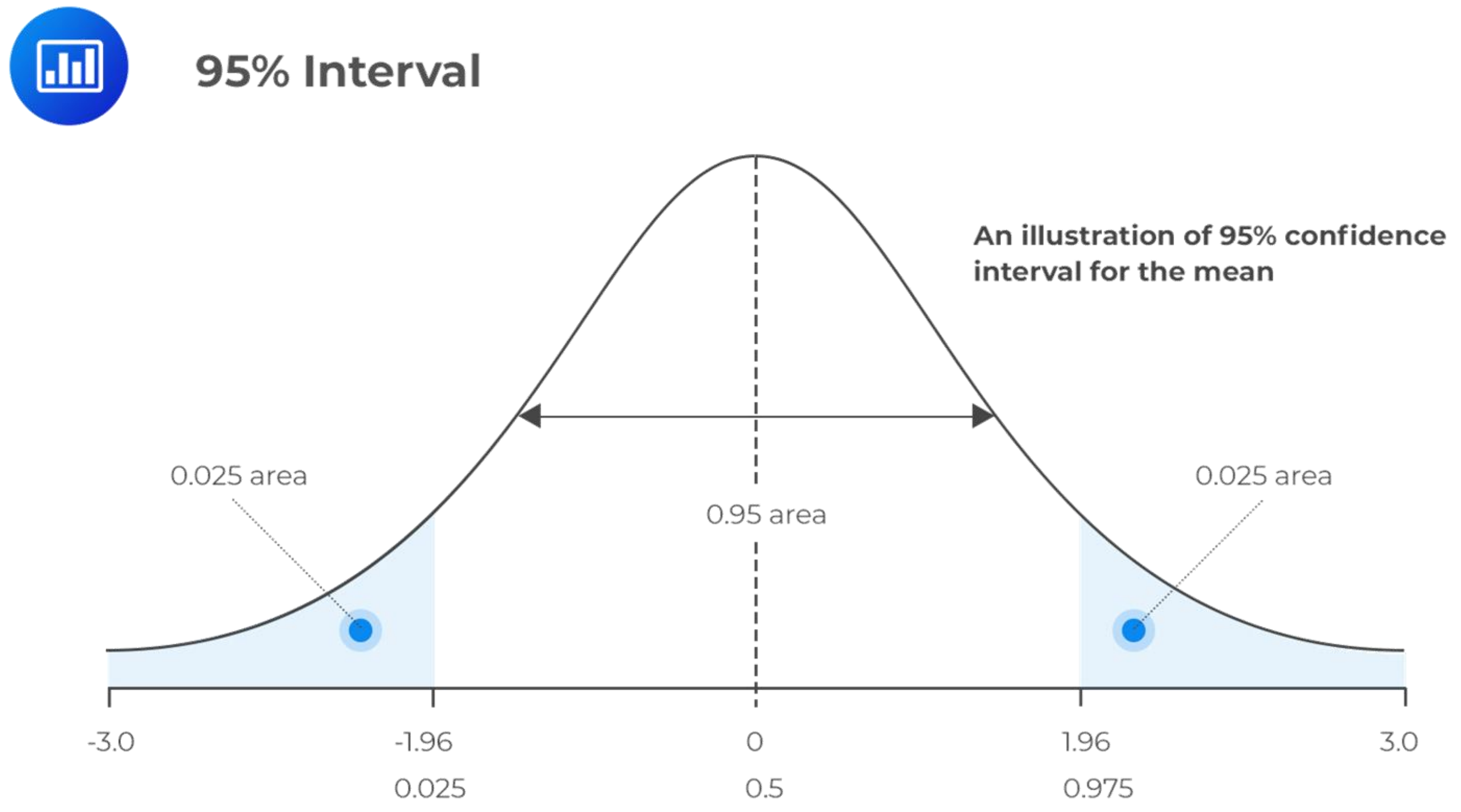
假设检验步骤
问题是什么？
零假设和备择假设
检验类型
单样本检验
相关配对检验
独立双样本检验
抽样分布类型
正太分布
t分布
其他分布
检验方向
单尾检验左尾
单尾检验右尾
双尾检验
证据是什么？
零假设成立时，得到样本平均值的概率：p值
判断标准是什么？
显著水平α（0.1%，1%，5%）
做出结论
p<=α时，拒绝零假设，备择假设成立，p>α时，接受零假设

基本概念

置信区间：指由样本统计量所构造的总体参数的估计区间

显著性水平：是在原假设为真时拒绝原假设的概率，一般由 α 表示

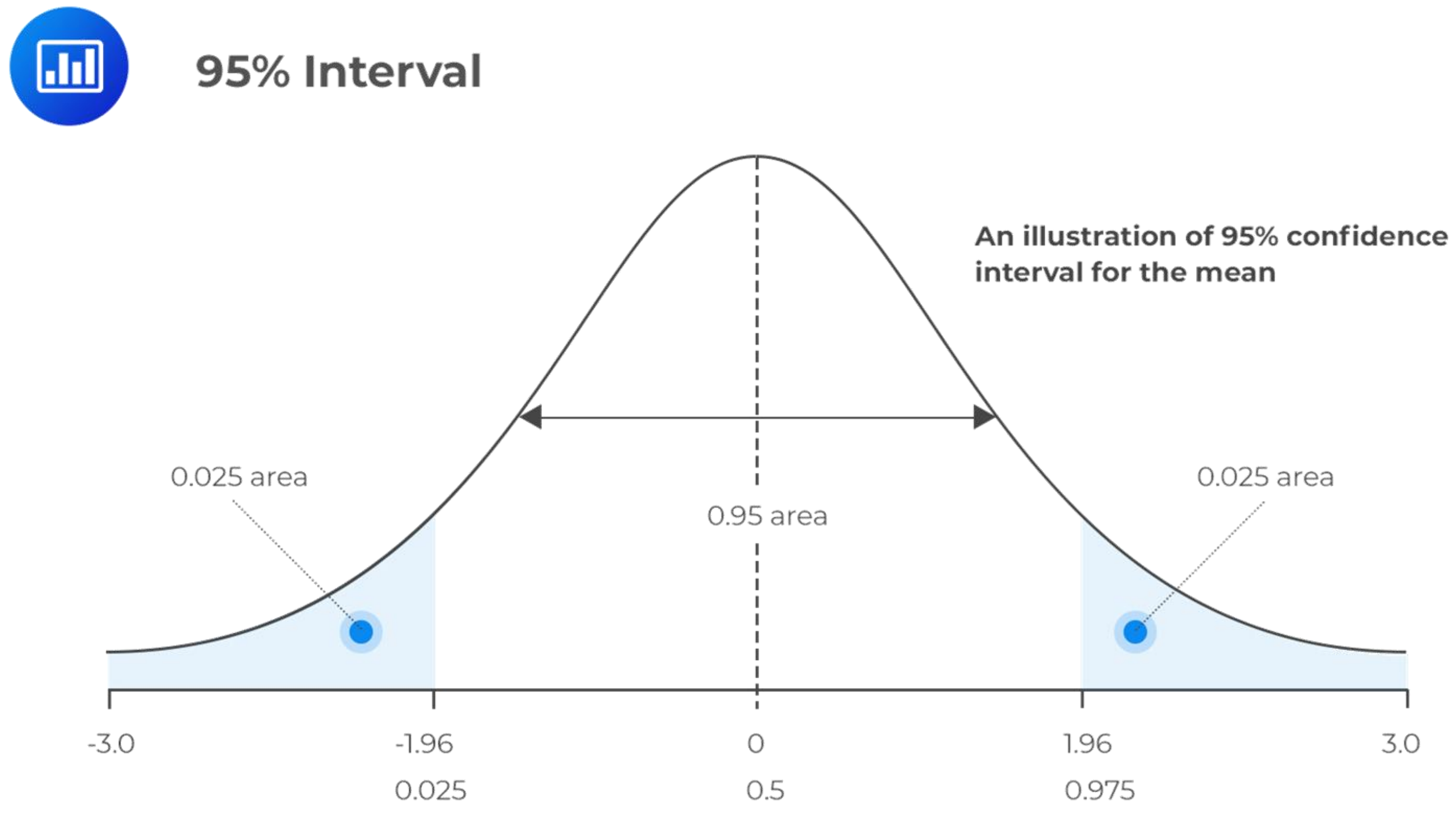
置信水平：将构造置信区间的步骤重复很多次，置信区间包含总体的参数真值的次数所占的比例，也称为置信度，表示为$1-\alpha$



基本概念

自由度：以样本的统计量来估计总体的参数时，样本中独立或能自由变化的数据的个数

P值：检验统计量之外的尾部面积



定义

概念

对总体的某种规律提出假设，通过样本数据推断、决定是否拒绝这一假设

基本定义

原假设：假定总体参数未发生变化

备择假设：假定总体参数发生变化，与原假设相对立

检验统计量：根据样本数据计算的随机变量，用在假设检验中

分类

- 假设检验
  - 正态总体均值的假设检验
    - Z检验
    - t检验
  - 正态总体方差的假设检验
    - χ^2检验
    - F检验

02

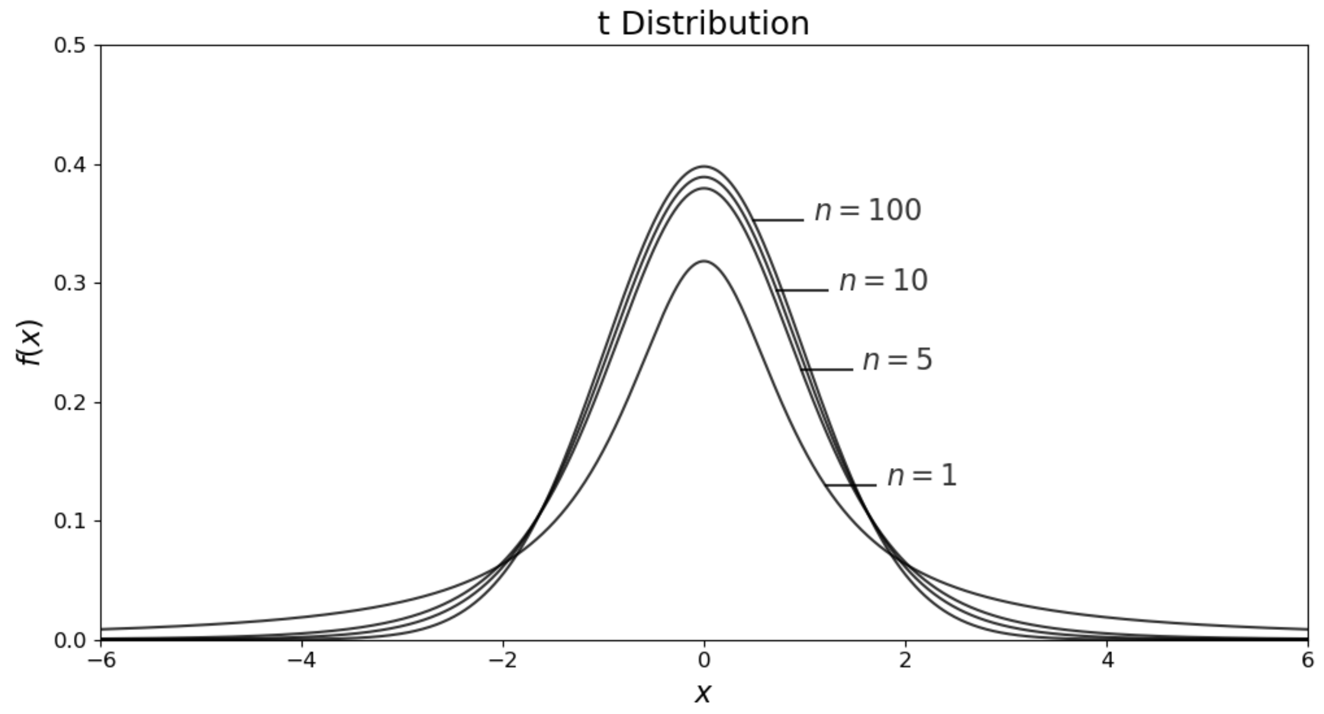
t检验

t分布

定义

设$X \sim N(0,1), Y \sim \chi^2(n)$，且$X,Y$相互独立，则称随机变量$t=\frac{X}{\sqrt{Y/n}}$，服从自由度为$n$的$t$分布，记为$t\sim t(n)$，$t$分布又称学生分布

概率密度函数：$h(t)=\frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2}, -\infty < t < \infty$



t检验

概念

用t分布理论来推论差异发生的概率，从而比较两个平均数的差异是否显著

使用条件

- 样本量较小，一般小于30，记作：$n < 30$
- 总体标准差σ未知
- 样本服从正态分布

02

t检验-单总体

简介

单总体t检验

检验一个**样本均值**与已知的**总体均值**的差异是否显著

t检验量计算

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

其中，样本均值为 $\bar{X}$，样本标准差为 s，样本数量为 n，总体均值为 μ

案例

背景

已知在线教育行业内学员对课程的满意度平均为3.5分， Y公司对已开课的《数据分析》学员做满意度调查得出平均分为3.71分，其满意度是否超过业内的平均分呢?

3.9	3.5	3.6	3.7	3.8	3.5	4.1	4.8	5.0	3.2	3.5	2.9	3.6	3.5	4.2
4.1	3.5	3.7	3.9	4.2	2.8	3.5	3.7	3.3	4.2	3.3	3.5	2.8	3.5	4.4

提出假设

原假设 H_0：$\bar{X} = \mu$

备择假设 H_1：$\bar{X} > \mu$

检验水准： $\alpha = 0.05$

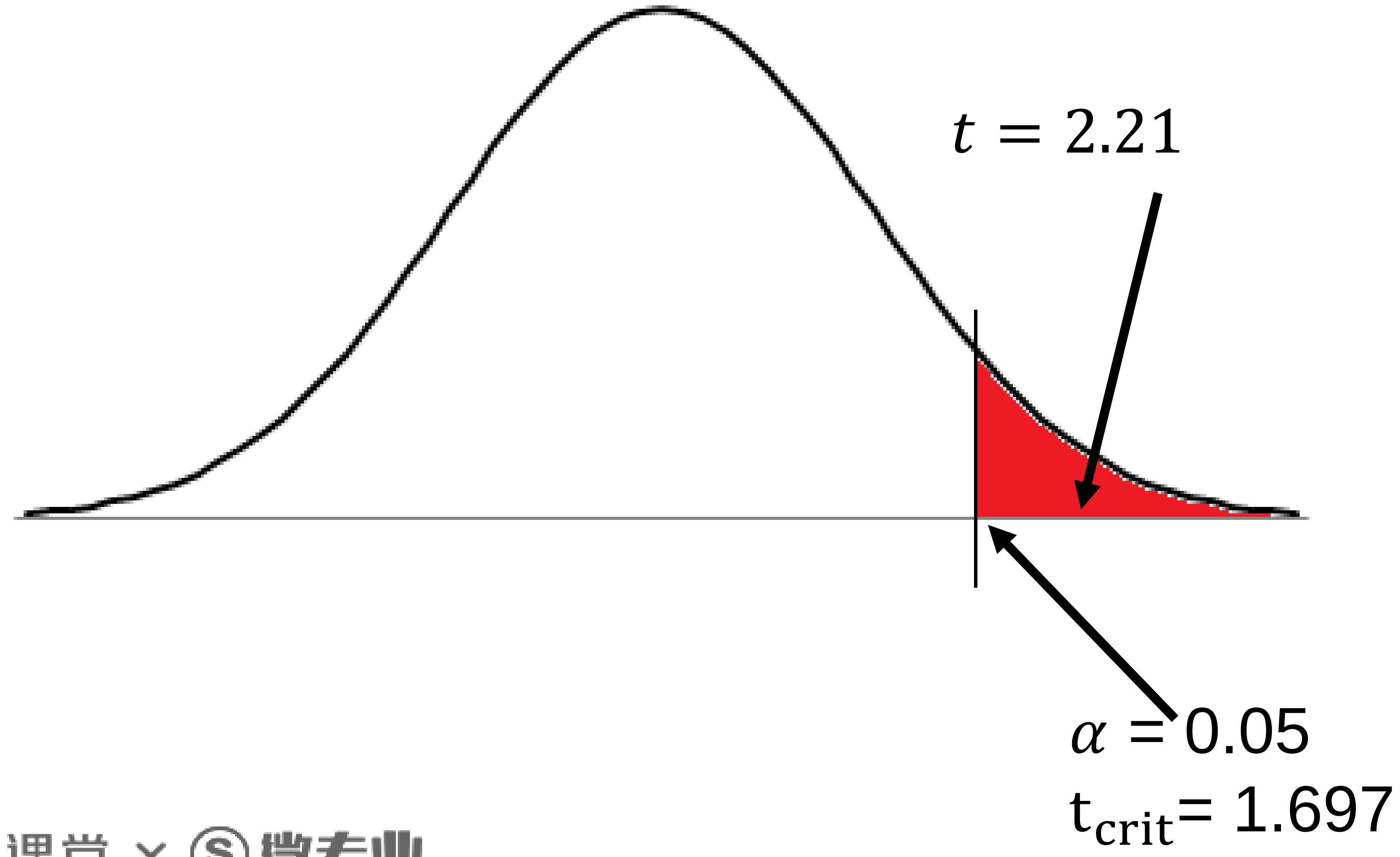
单双侧： 单侧检验，右边检验

单总体t检验-计算

计算t值、P值

$$t=\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}=\frac{3.71-3.5}{\frac{0.52}{\sqrt{30}}}=2.21$$

$$t_{crit}=2.21>t_{0.05,29}=1.699, p<0.05$$



查表

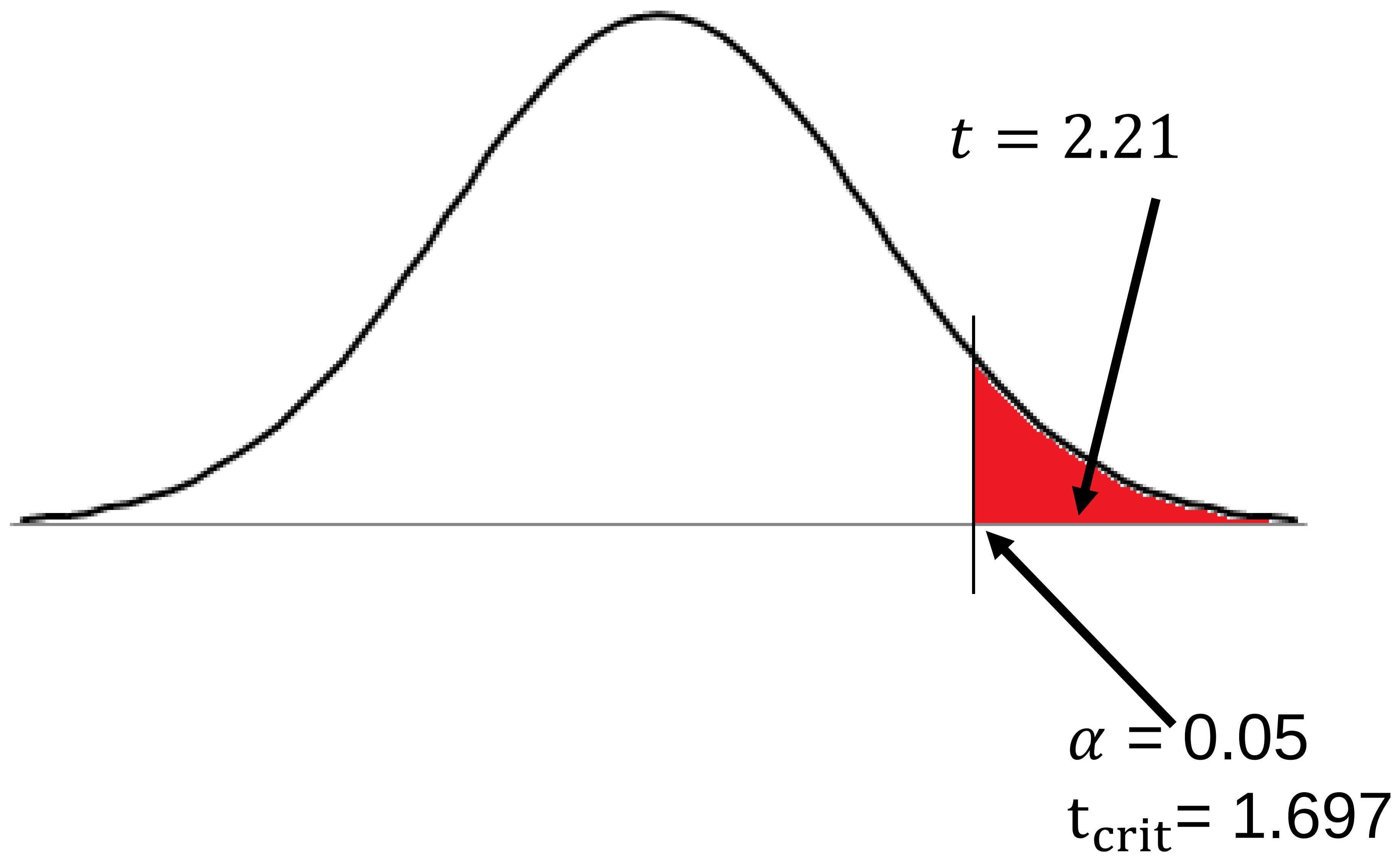
cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05
df							
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021

结论

■ 得出结论

拒绝原假设H_0，接受备择假设H_1

即《数据分析》课程的学员满意度是高于业内标准的！



02

t检验-配对

简介

定义

与单样本t检验相比，样本为配对样本的中观测值的差

两个样本的样本量相同，样本先后的顺序一一对应的

t检验量计算

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

其中，$\bar{d}$是配对样本差值的均值，s_d 是配对样本差值的标准差，n是配对样本量

案例

为了验证维生素E对鼠体内维生素A含量的影响，将大白鼠配成8对，每对分别饲以正常的饲料和缺乏维生素E的饲料，检测两组大白鼠肝中维生素A的含量，比较两组大白鼠中维生素A的含量是否有差别

大白鼠配对号	正常饲料组	维生素E缺乏组	差数d
1	3550	2450	1100
2	2000	2400	-400
3	3000	1800	1200
4	3950	3200	750
5	3800	3250	550
6	3750	2700	1050
7	3450	2500	950
8	3050	1750	1300
Mean	3318.75	2506.25	812.5

假设

提出假设

原假设 H_0： $\mu_0 = 0$

备择假设 H_1： $\mu_0 \neq 0$

计算

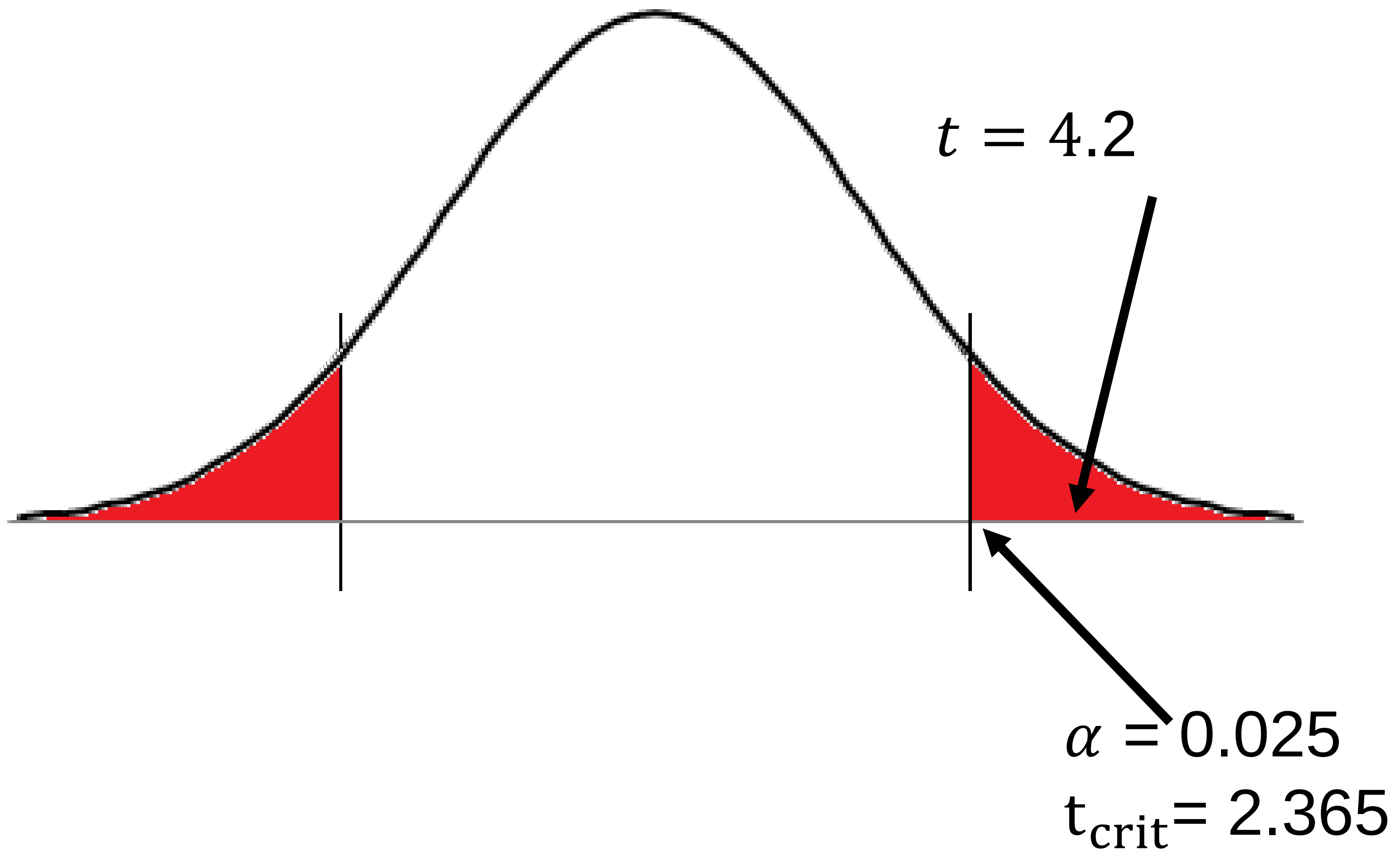
■ 计算t值、确定P值

$$t_{crit} = 4.2 > t_{0.05/2,7} = 2.365, p < 0.05$$

$$\bar{d} = \frac{\sum d}{n} = \frac{6500}{8} = 812.5\ (\text{U/g})$$

$$S_{\bar{d}} = \frac{S_d}{\sqrt{n}} = \sqrt{\frac{7370000-(6500)^2/8}{8\times(8-1)}} = 193.1298(U/g)$$

$$t = \frac{\bar{d}-\mu_d}{S_d/\sqrt{n}} = \frac{812.5 - 0}{193.1298} = 4.2070, \quad \nu = 8-1 = 7$$



查表

■ 计算t值、确定P值

$$t_{crit} = 4.2 > t_{0.05/2,7} = 2.365, p < 0.05$$

$$\bar{d} = \frac{\sum d}{n} = \frac{6500}{8} = 812.5\ (\text{U/g})$$

$$S_{\bar{d}} = \frac{S_d}{\sqrt{n}} = \sqrt{\frac{7370000 - (6500)^2/8}{8 \times (8-1)}} = 193.1298(U/g)$$

$$t = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}} = \frac{812.5 - 0}{193.1298} = 4.2070, \quad \nu = 8 - 1 = 7$$

df	单尾概率					
	0.25	0.10	0.05	0.025	0.01	0.005
	双尾概率					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977

结论

结论

拒绝原假设H_0，接受备择假设H_1

两种饲料喂养的两组大白鼠中维生素A的含量有显著差别，正常饲料组比缺乏维生素 E 饲料组的含量要高，即维生素E的摄入会引起体内维生素A含量的变化

02

t检验-双总体

简介

概念

检验两个样本平均数与其各自所代表的总体的差异是否显著

双总体t检验统计量计算

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

S_1^2和 S_2^2为两样本方差，n_1 和 n_2 为两样本容量

案例

背景

为了研究两个品牌的咖啡因含量是否有显著差异，从A、B两个咖啡店中随机抽取同一品类的大杯咖啡，并测量每杯咖啡中的咖啡因含量

	A	B
均值	164mg	170mg
标准差	5.1mg	3.1mg
样本量	37	35

提出假设

■ 提出假设

原假设 H_0： $\mu_0 = 0$

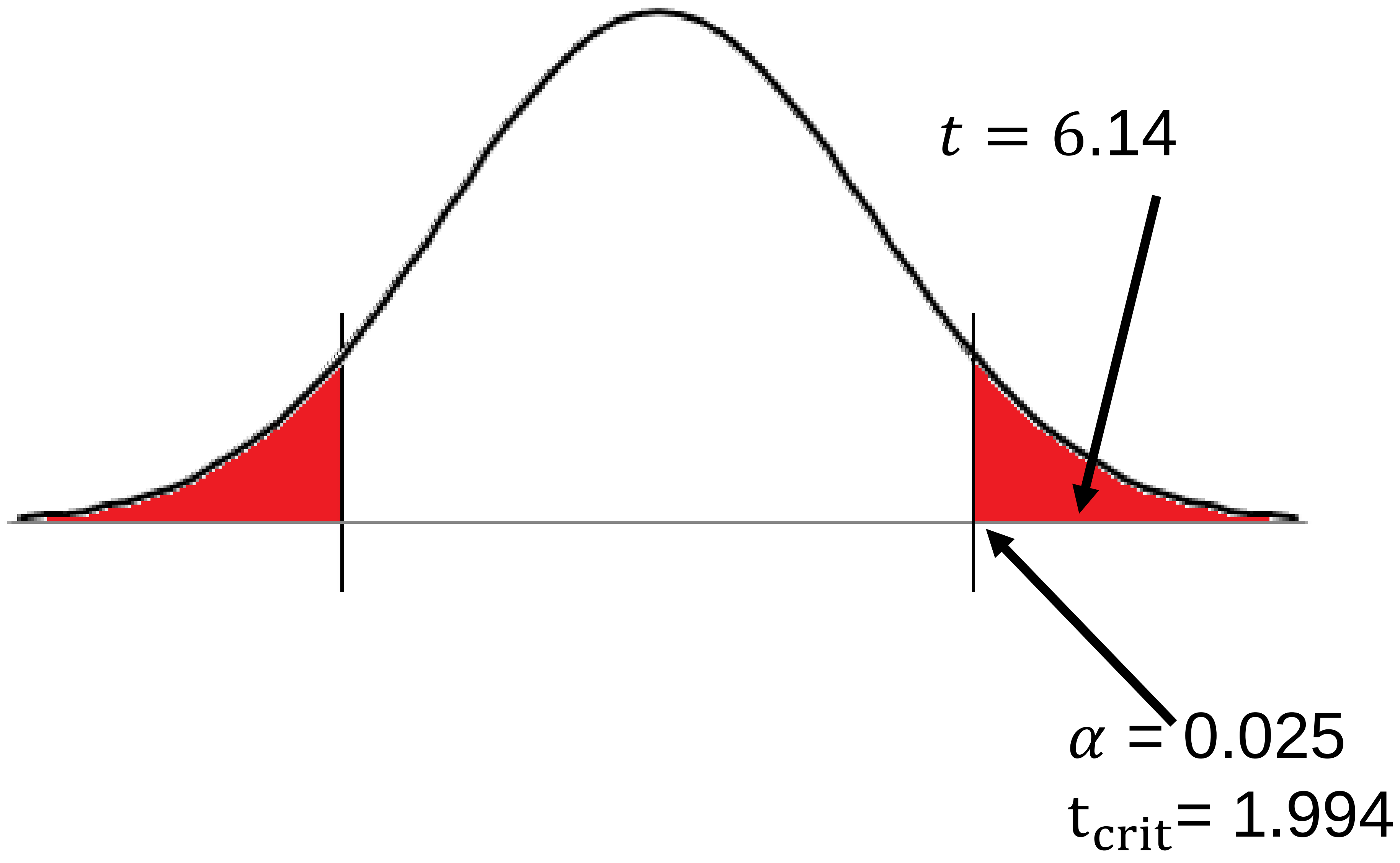
备择假设 H_1： $\mu_0 \neq 0$

计算

- 计算t值、确定P值

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{(170 - 164) - 0}{\sqrt{\frac{3.1^2}{35} + \frac{5.1^2}{37}}}$$



查表

	P(2):	0.5	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
n'	P(1):	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1		1	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2		0.816	1.886	2.92	4.303	6.965	9.925	14.089	22.327	31.599
3		0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4		0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.61
5		0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6		0.718	1.44	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7		0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8		0.706	1.397	1.86	2.306	2.896	3.355	3.833	4.501	5.041
9		0.703	1.383	1.833	2.262	2.821	3.25	3.69	4.297	4.781
10		0.7	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11		0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12		0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.93	4.318
13		0.694	1.35	1.771	2.16	2.65	3.012	3.372	3.852	4.221
14		0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.14
15		0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16		0.69	1.337	1.746	2.12	2.583	2.921	3.252	3.686	4.015
17		0.689	1.333	1.74	2.11	2.567	2.898	3.222	3.646	3.965
18		0.688	1.33	1.734	2.101	2.552	2.878	3.197	3.61	3.922
19		0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20		0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.85
21		0.686	1.323	1.721	2.08	2.518	2.831	3.135	3.527	3.819
22		0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23		0.685	1.319	1.714	2.069	2.5	2.807	3.104	3.485	3.768
24		0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25		0.684	1.316	1.708	2.06	2.485	2.787	3.078	3.45	3.725
26		0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27		0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.69
28		0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29		0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30		0.683	1.31	1.697	2.042	2.457	2.75	3.03	3.385	3.646
31		0.682	1.309	1.696	2.04	2.453	2.744	3.022	3.375	3.633
32		0.682	1.309	1.694	2.037	2.449	2.738	3.015	3.365	3.622
33		0.682	1.308	1.692	2.035	2.445	2.733	3.008	3.356	3.611
34		0.682	1.307	1.091	2.032	2.441	2.728	3.002	3.348	3.601
35		0.682	1.306	1.69	2.03	2.438	2.724	2.996	3.34	3.591
36		0.681	1.306	1.688	2.028	2.434	2.719	2.99	3.333	3.582
37		0.681	1.305	1.687	2.026	2.431	2.715	2.985	3.326	3.574
38		0.681	1.304	1.686	2.024	2.429	2.712	2.98	3.319	3.566
39		0.681	1.304	1.685	2.023	2.426	2.708	2.976	3.313	3.558
40		0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50		0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60		0.679	1.296	1.671	2	2.39	2.66	2.915	3.232	3.46
70		0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.436

得出结论

■ 结论

拒绝原假设H_0，接受备择假设H_1

A、B两个品牌的咖啡中咖啡因的含量有显著差别，B品牌的咖啡因含量显著高于A

03

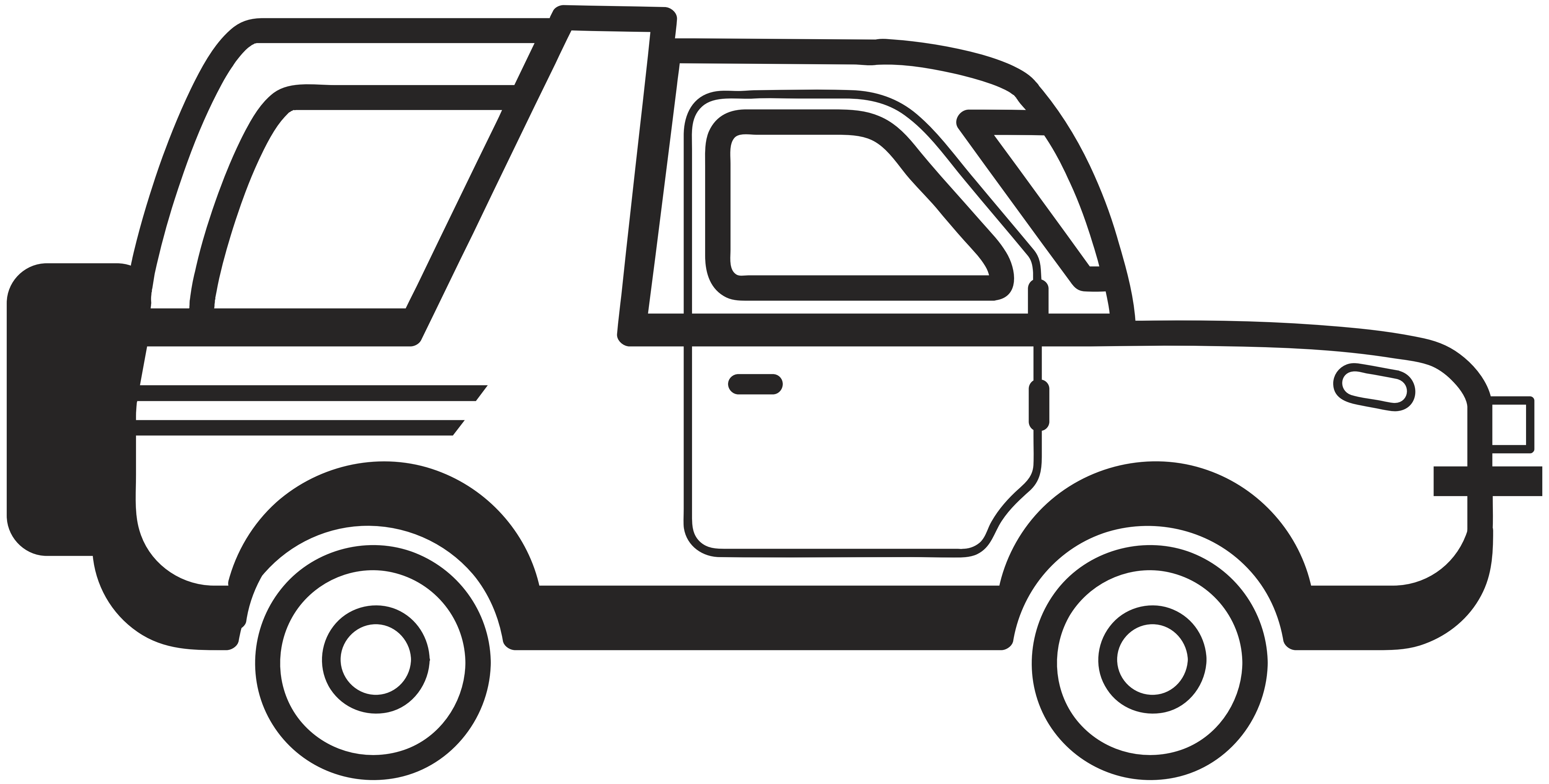
χ^2检验

引例

运营：“为什么学员的性别没有考虑”

分析师：“这个变量我做了卡方检验了，不显著，所以我没有放进模型”

应用场景

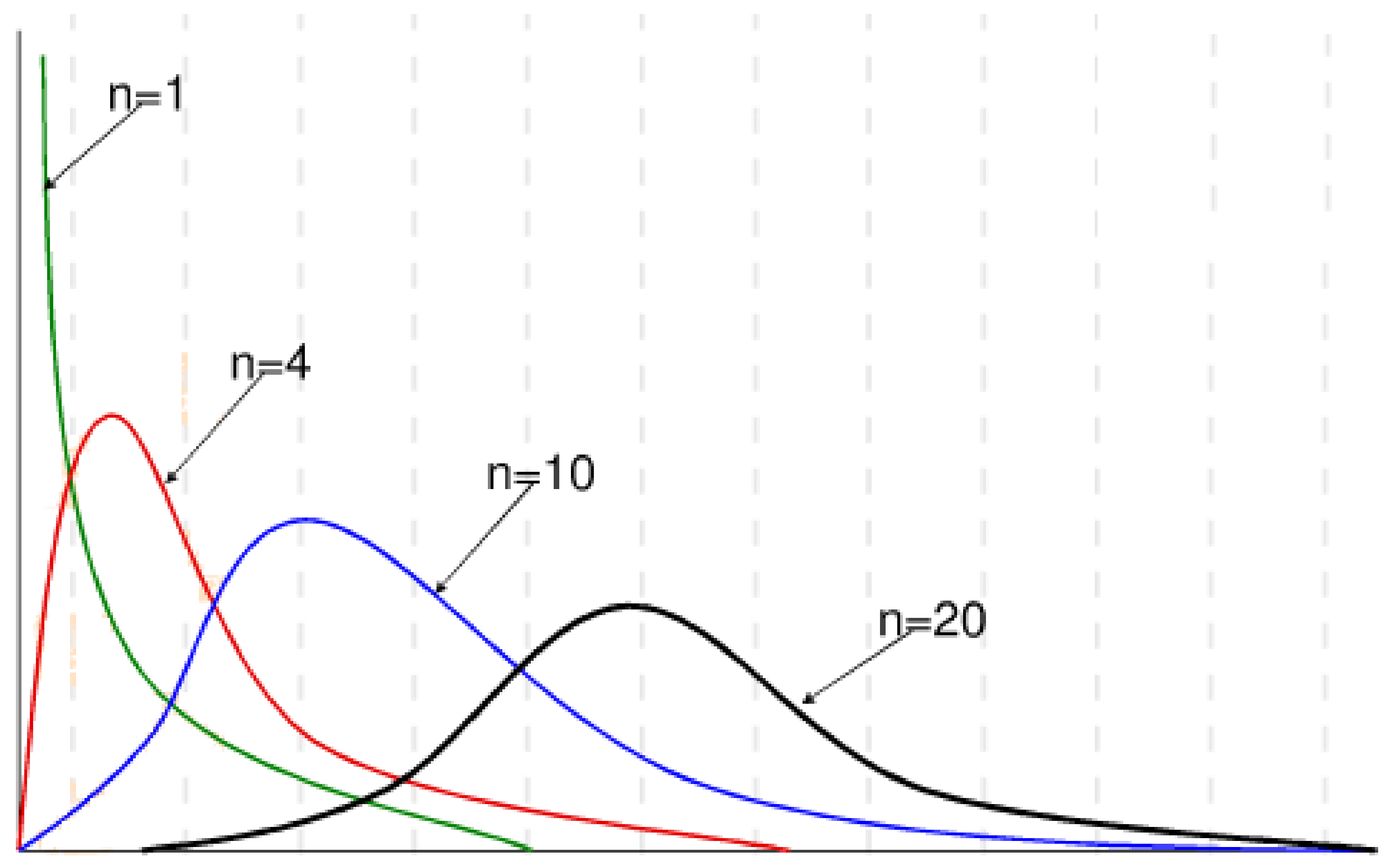


χ^2分布

定义

设X_1, X_2，… X_n是来自总体$N(0,1)$的样本，则称统计量 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 服从自由度为 n 的 χ^2 分布，记为：

$$\chi^2 \sim \chi^2(n)$$



χ^2分布

概率密度：

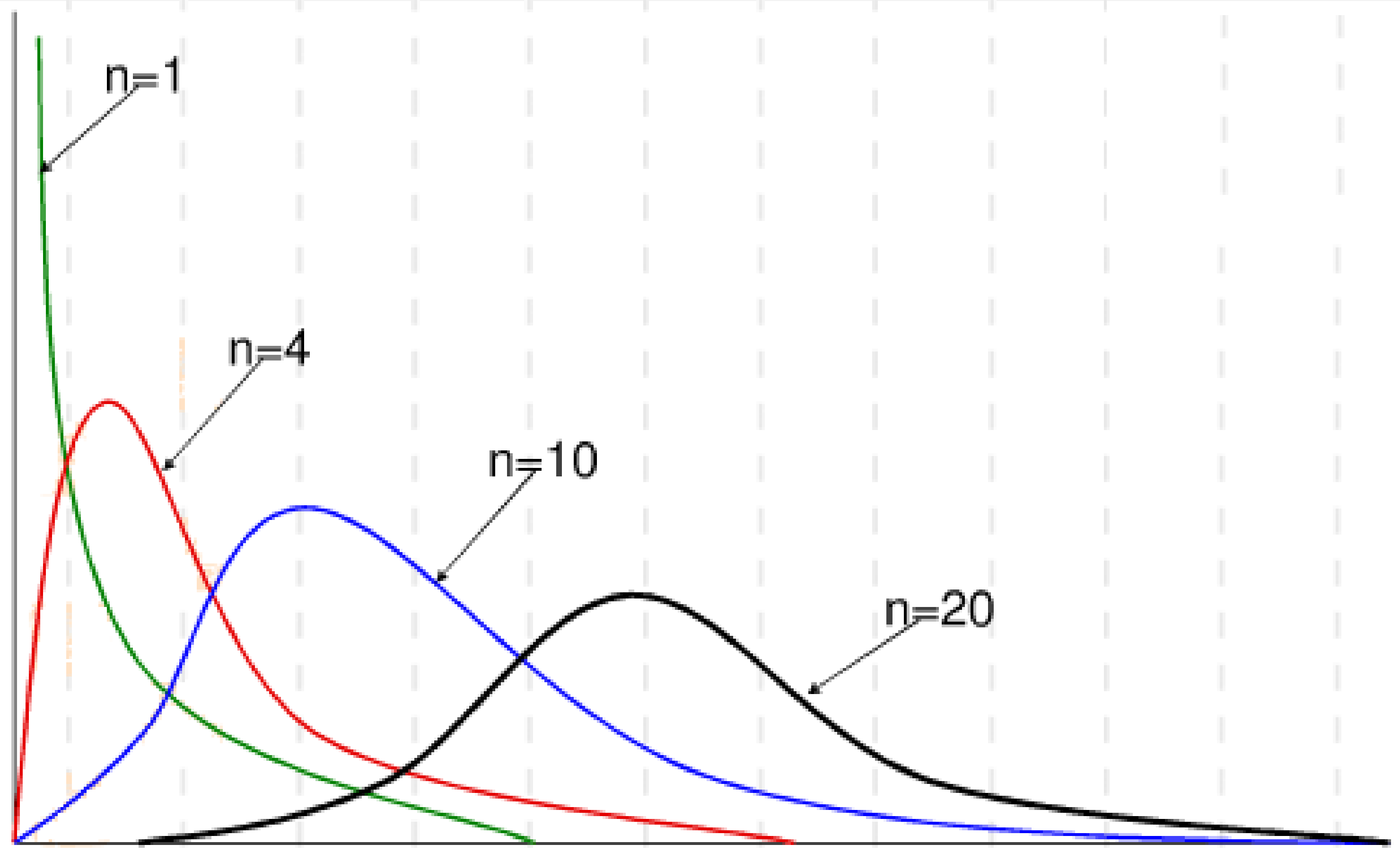
$$f(x)=\begin{cases}\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{1}{2}x}, & x>0\\ 0, & x\le 0\end{cases}$$

数学期望：

$$E(\chi^2)=n$$

方差：

$$D(\chi^2)=2n$$



χ^2检验

概念

卡方检验是统计样本的实际观测值与理论推断值之间的偏离程度，实际观测值与理论推断值之间的偏离程度决定卡方值的大小

χ^2检验量计算

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

设总体$X \sim N(\mu, \sigma^2)$，S为σ的无偏估计，样本数量为n

案例

案例

某种型号的电池寿命（以h计）服从方差$\sigma^2 = 5000$的正态分布，现有一批这种电池，从生产情况来看，寿命的波动性有所改变。现随机取26只电池，测出其寿命的样本方差$S^2 = 9200$。根据这一数据能否推断这批电池的寿命的波动性较以往的有显著的变化（取$\alpha = 0.02$）？

提出假设

原假设 H_0：$\sigma^2 = 5000$

检验水准：$\alpha = 0.02$

备择假设 H_1：$\sigma^2 \neq 5000$

计算

- 计算χ^2值

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(26-1)\times 9200}{5000} = 46$$

$$\chi^2_{\alpha/2}(n-1) = \chi^2_{0.01}(25) = 44.314$$

$$\chi^2_{1-\alpha/2}(25) = \chi^2_{0.99}(25) = 11.524$$

拒绝域：

$$\frac{(n-1)S^2}{\sigma^2} \geq 44.314 \text{或} \frac{(n-1)S^2}{\sigma^2} \leq 11.524$$

由于

$$\chi^2 = 46 > 44.314$$

结论

得出结论

拒绝原假设H_0，接受备择假设H_1

认为这批电池寿命的波动性较以往有显著的变化！

04

F检验

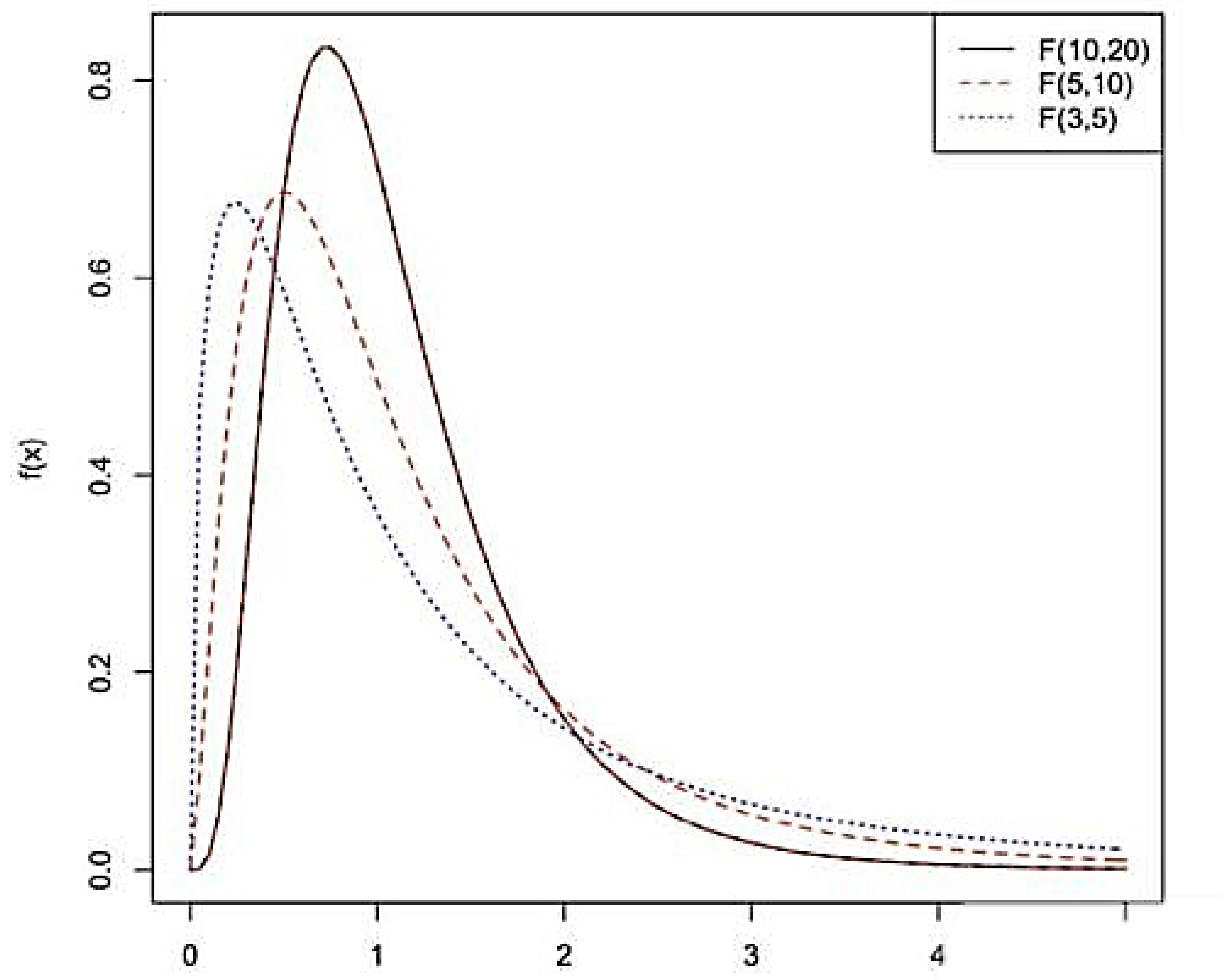
F分布

定义

设$U \sim \chi^2(n_1), V \sim \chi^2(n_2)$,且$U, V$相互独立，则称随机变量：

$$F = \frac{U/n_1}{V/n_2}$$

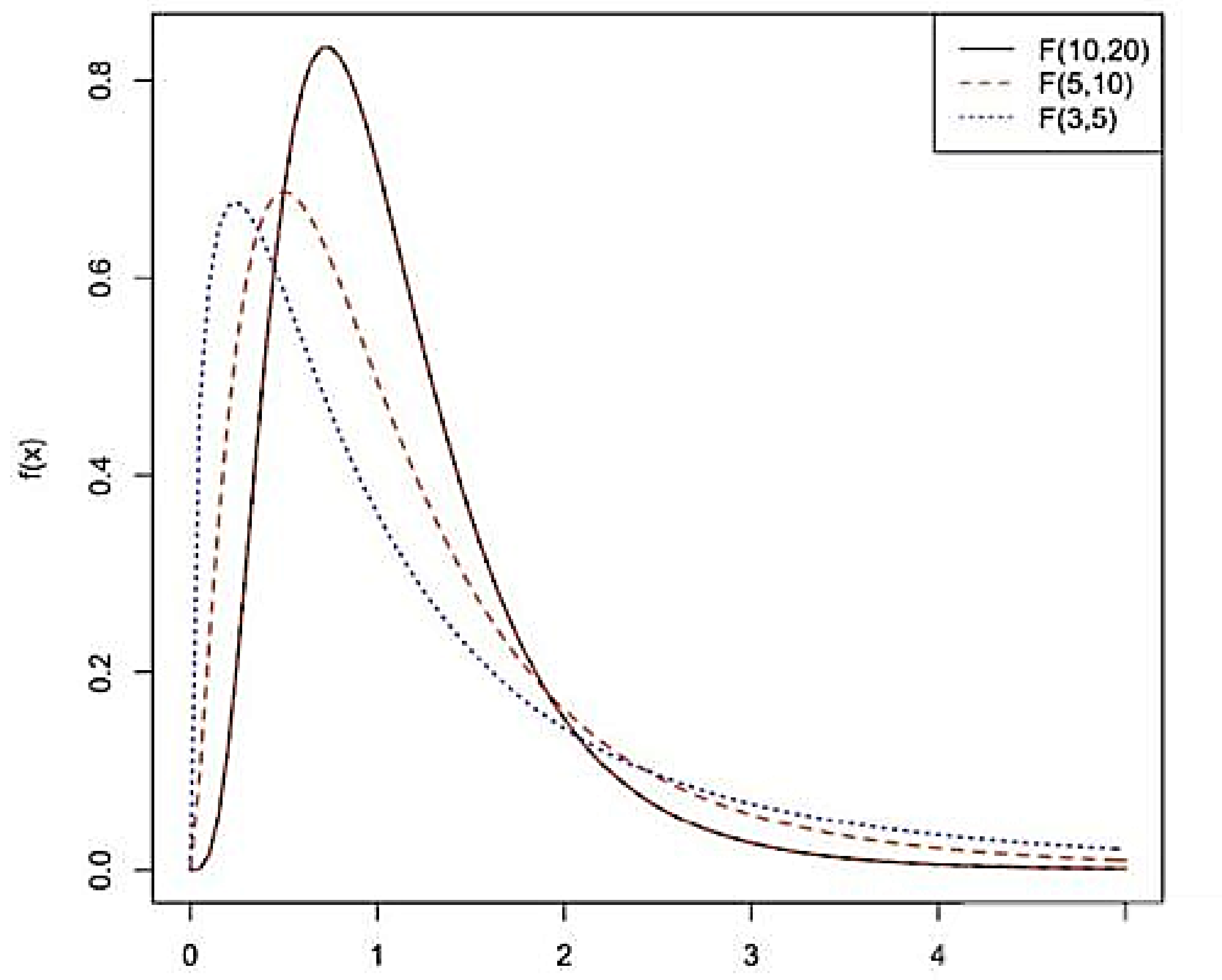
服从自由度为(n_1, n_2)的F分布，记为$F \sim F(n_1, n_2)$



F分布

■ **概率密度**

$$\psi(y)=\begin{cases}\dfrac{\Gamma\left[\dfrac{(n_1+n_2)}{2}\right](\dfrac{n_1}{n_2})^{n_1/2}y^{(\frac{n_1}{2})-1}}{\Gamma\left(\dfrac{n_2}{2}\right)\Gamma\left(\dfrac{n_1}{2}\right)[1+(\dfrac{n_1 y}{n_2})]^{(n_1+n_2)/2}}, y>0\\ 0, \qquad \text{其他}\end{cases}$$



两个总体的情况 – F检验

■ 概念

F检验最常用的别名叫做联合假设检验，也称方差比率检验、方差齐性检验。通常是用来分析用了超过一个参数的统计模型，以判断该模型中的全部或一部分参数是否适合用来估计母体。

■ F检验量计算

$$F = \frac{S_1^2}{S_2^2}$$

设总体 $X \sim N(\mu_1, \sigma_1^2)$， $Y \sim N(\mu_2, \sigma_2^2)$ 两样本独立，其样本方差分别为 S_1^2 , S_2^2

案例

案例

设两家银行储户的年存款余额均服从正态分布，经市场调查，分别抽取容量为21和16的样本，得样本均值分别为650元和800元，样本方差分别为80^2和70^2，能否认为第二家银行储户的平均年存款余额显著高于第一家银行（取$\alpha = 0.10$）？

提出假设

原假设 H_0：$\sigma_1^2 = \sigma_2^2$

检验水准： $\alpha = 0.10$

备择假设 H_1：$\sigma_1^2 \neq \sigma_2^2$

计算

■ 计算F值

$$F=\frac{S_1^2}{S_2^2}=\frac{80^2}{70^2}=1.306$$

$$F_{\frac{\alpha}{2}}(n_1-1,n_2-1)=F_{0.05}(21-1,16-1)=F_{0.05}(20,15)=2.33$$

$$F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)=0.4545$$

拒绝域为：

$$\frac{S_1^2}{S_2^2}\geq 2.33 \text{ 或 } \frac{S_1^2}{S_2^2}\leq 0.4545$$

由于：

$$0.4545<F<2.33$$

结论

■ 得出结论

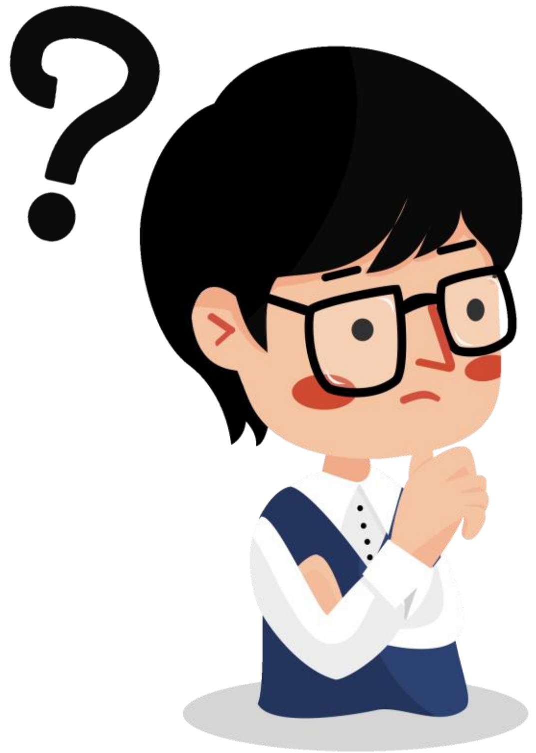
接受原假设H_0，拒绝备择假设H_1

认为两家银行储户的年存款余额的方差无显著性差异!

05

总结

思考题



01

如何选择t检验和χ^2检验?

总结

本章包含五小节内容：

第一节

- 假设检验的一般步骤、基本概念

第二、三、四节

- χ^2、t、F检验的概念的检验方法

第五节

- 本章总结

谢谢观看

参考书目：概率论与数理统计·第四版（浙江大学） 高等教育出版社