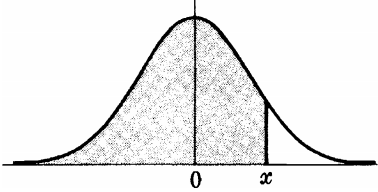
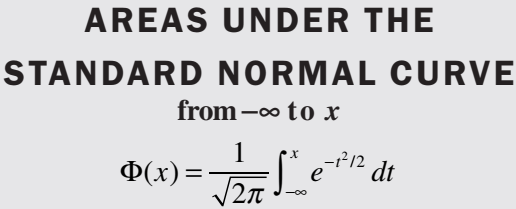
**\* 关于标准正态分布的查表问题：**

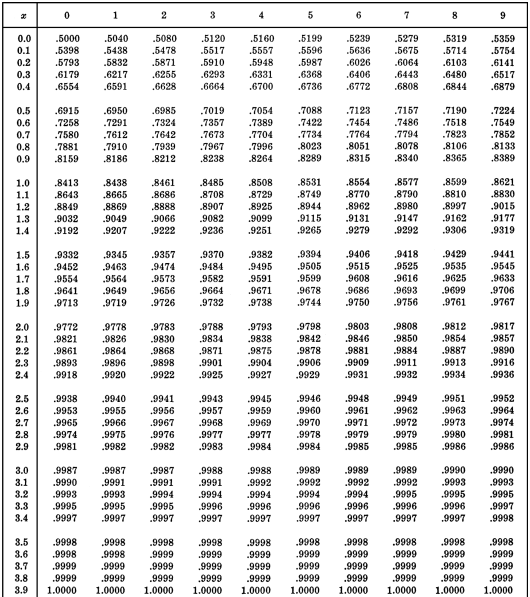
**见下面的图和表。钟形曲线下的总面积为1，α称为显著性水平（指达到这个水平就可以拒绝），所以1-α就称为置信水平。由于图形对称，所以两端阴影面积各站α/2。表的作用是根据Φ(x)的值来查x的值，Φ(x)就是曲线下面，横坐标从-∞到x进行积分所得到的面积（见右边那个图）。**

**例如：要通过查表求得在显著性水平α=0.03下的置信区间**

**首先，因α/2=0.015，故1-α/2=0.985（即Φ(x)=0.985），现在就需要通过查表来求x。**

**其次，在表中查与0.985最接近的数值，发现这个值在第2.1行、第7列上，将行、列值串起来得2.17，此即使Φ(x)=0.985的x值，即需要求得的临界值，根据图形的对称性，中心点是0，故置信区间是（-2.17, 2.17）。**





**\* 假设检验问题：用一个例子来说明。**

**某食品公司生产一种罐头，按标准每罐净重为227克，根据以往生产经验罐头重量的标准差为3克。现随机抽查该公司产品100罐（即n），测得平均净重为228克，判断这批罐头是否符合标准？（设显著性水平α=0.05 ）**

**解：**

**根据题意：μ0=227，σ=3，n=100，=228，α=0.05**

**现在要检验的是参数μ是否与μ0一致（即在显著性水平α=0.05下可接受）。**

**（1）提出：原假设H0：μ＝μ0，备择假设H1：μ≠μ0**

**（2）选取和计算检验统计量。可知=3.33，z服从标准正态分布N(0, 1)。**

**（3）根据显著性水平确定置信区间。 由α=0.05，得α/2=0.025，1-α/2=0.975，查上面的正态分布表得置信区间（即接受域）为（-1.96，1.96）。**

**（4）进行比较并做出决策。 上面计算出的z=3.33>1.96，在拒绝域内（即在接受域外），故拒绝H0，结论是这批罐头不符合标准。**

**\* 关联规则的Apriori算法：用一个例子来说明。**

**首先，见下面的交易记录，假设“最小支持度阈值”为0.5，“最小置信度阈值”是0.8。要求：**

**（1）按照Apriori算法生成所有合法的频繁集。**

**（2）只从3-项频繁集中生成右侧只包含一个商品的所有关联规则。**

|  |  |
| --- | --- |
| **序号** | **商品** |
| **1** | **A，C，D** |
| **2** | **B，C，E** |
| **3** | **A，B，C，E** |
| **4** | **B，E** |

**解答：**

**（1）第一步，生成1-项集及其支持度，如图所示。其中：支持度= 包含某商品的交易数/总交易数。其中的{D}因支持度小于“最小支持度阈值”，不属于频繁集。**

|  |  |
| --- | --- |
| **候选项集C1** | **支持度** |
| **{A}** | **50%** |
| **{B}** | **75%** |
| **{C}** | **75%** |
| **{D}** | **25%** |
| **{E}** | **75%** |

**第二步，生成2-项集及其支持度。**

**【生成方法：若有两个k-1项集，每个项集按字母顺序进行排序。如果两个k-1项集的前k-2个项相同，而最后一个项不同，则说明它们是可连接的，即可连接生成k项集。**

**例如有两个3项集：｛A, B, C｝和{A, B, D}，这两个3项集就是可连接的，它们可以连接生成4项集｛A, B, C, D｝。**

**又如两个3项集｛A, B, C｝和｛A, D, E｝，这两个3项集是不能连接生成4项集的。】**

**得到下面的2-项集。注意，{A，B}的支持度=同时购买了A和B的交易数/交易总数。**

**因{A，B}和{A，E}的支持度小于“最小支持度阈值”，不属于频繁集。**

|  |  |
| --- | --- |
| **候选项集C2** | **支持度** |
| **{A，B}** | **25%** |
| **{A，C}** | **50%** |
| **{A，E}** | **25%** |
| **{B，C}** | **50%** |
| **{B，E}** | **75%** |
| **{C，E}** | **50%** |

**第三步，生成3-项集及其支持度**

**得到下面的3-项集。其中{A，B，C}，{A，C，E}支持度小于“最小支持度阈值”，不是频繁集。**

**只剩一行，不能生成更大的频繁集了。**

|  |  |
| --- | --- |
| **候选项集C3** | **支持度** |
| **{A，B，C}** | **25%** |
| **{A，C，E}** | **25%** |
| **{B，C，E}** | **50%** |

**第四步，生成关联规则：**

**综上，所有可能的频繁集是：{A}，{B}，{C}，{E}，{A，C}，{B，C}，{B，E}，{C，E}，{B，C，E}**

**（2）根据题意，只从3-项频繁集{B，C，E}生成右侧只包含一个商品的关联规则，结果如下所示：**

**{B，C} → {E} 置信度（{B，C} → {E}）= 2/2=1 > “最小置信度阈值”，故该关联规则成立。**

**{B，E} → {C} 置信度（{B，E} → {C}）= 2/3=0.67 < “最小置信度阈值”，故抛弃该关联规则。**

**{C，E} → {B} 置信度（{C，E} → {B}）= 2/2=1 > “最小置信度阈值”，故该关联规则成立。**

**置信度的计算方法为：同时包含左、右侧商品的交易数/只包含左侧商品的交易数。**

**最终得到两个关联规则：{B，C} → {E}和{C，E} → {B}。**