概率论与数理统计习题课（6-9章）

第六章

4种重要分布： N、$\chi 2$、t、F

三个结构定理：$\chi 2$、t、F

两个抽样定理：单总体、双总体

基本内容与重要结论：

1. 总体与样本.

样本相互独立同分布

2. 常用的统计量:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k. \qquad M_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k.$$

注意：样本方差不是样本二阶中心矩。

3. 数理统计中常用的分布。

① 标准正态分布. $E[\chi^2(n)]=n, D[\chi^2(n)]=2n$

② χ^2分布. $X_i \sim N(0,1) \Rightarrow \chi^2=\sum_{i=1}^{n} X_i^2 \sim \chi^2(n).$

③ t分布. $X \sim N(0,1), Y \sim \chi^2(n) \Rightarrow T=\dfrac{X}{\sqrt{Y/n}} \sim t(n).$

④ F分布.

$$X \sim \chi^2(n_1), Y \sim \chi^2(n_2) \Rightarrow F=\frac{X/n_1}{Y/n_2} \sim F(n_1,n_2).$$

注: ①χ^2分布, t分布, F分布中的随机变量均相互独立.

②F、t分布上侧分位数的计算:

$$t_{1-\alpha}(n)=-t_\alpha(n) \quad F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}.$$

4. 两个抽样分布定理的重要结论:

***Th*6.2.4** (单个正态总体):

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}); \quad \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1); \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

***Th*6.2.5** (两个独立正态总体):

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1-1, n_2-1);$$

$$\sigma_1^2 = \sigma_2^2 = \sigma^2, \ T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_\varpi\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2).$$

$$S_\omega = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

第七章

- 点估计
  - 统计量构造
    - 矩法:令$A_k = \boldsymbol{\gamma_k}$
    - 极大似然法
  - 优良性：无偏、有效、相合
- 区间估计
  - 单总体
    - $\boldsymbol{\mu}$
    - $\boldsymbol{\sigma^2}$
  - 双总体
    - $\boldsymbol{\mu_1 - \mu_2}$
    - $\boldsymbol{\sigma_1^2 / \sigma_2^2}$

5. 参数的点估计方法.

①矩估计法: 由样本矩估计相应的总体矩.

②极大似然估计法:

第一步: 构造似然函数

$$L(\theta)=\prod_{i=1}^{n} p(x_i;\theta) \text{ 或 } L(\theta)=\prod_{i=1}^{n} f(x_i;\theta);$$

第二步: 对似然函数取对数 $\ln L(\theta)$;

第三步:对$\ln L(\theta)$求导并令其等于0, 得似然方程(组)

第四步: 求解似然方程.

注: 当似然方程无解的时候, 应直接寻求使似然函数达到最大的解求得极大似然估计。

6. 估计量的评选标准: 无偏性, 有效性, 一致性.

注意: 样本方差是总体方差的最优无偏估计;

样本均值是总体均值的最优无偏估计.

7. 区间估计: 置信区间, 置信度, 枢轴变量法.

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

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n).\qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \sim N(0,1);$$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_W \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2). \quad \sigma_1^2 = \sigma_2^2 = \sigma^2.$$

$$F = \frac{\frac{1}{n_1}\sum_{i=1}^{n_1}(X_i - \mu_1)^2 / \sigma_1^2}{\frac{1}{n_2}\sum_{i=1}^{n_2}(Y_i - \mu_2)^2 / \sigma_2^2} \sim F(n_1, n_2);$$

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1).$$

第八章

一般步骤：

1、提出假设

2、建立检验统计量

3、确定H_0的否定域

4、对H_0作判断

两类错误：弃真、纳伪

μ、σ的检验

U

t

F

χ^2

注意：区间估计与假设检验的区别与联系。

8.回归模型的引进

若Y关于X_1，X_2，$\cdots$ X_k的回归函数为

$$y=\mu(x_1,x_2,\cdots,x_k)$$

$$=E(Y\mid X_1=x_1,X_2=x_2,\cdots X_k=x_k)$$

设想：$Y=\mu(x_1,x_2,\cdots,x_k)$ +随机误差

其它未知的、未考虑的因素以及随机因素的影响所产生.

得数学模型:

$$Y=\mu(x_1,x_2,\cdots,x_k)+\varepsilon$$

$$\varepsilon=Y-\mu(x_1,x_2,\cdots,x_k)$$

可视为随机误差，通常要求:

1) $E(\varepsilon)=0$;

2) $D(\varepsilon)=E(\varepsilon^2)=\sigma^2$ 尽可能小.

注 意 到 $\sigma^2 = E[Y - \mu(x_1, x_2, \cdots, x_n)]^2$

σ^2 是用回归函数近似因变量 Y 产生的均方误差.

建立模型涉及三个问题:

1) 确定对因变量 Y 影响显著的自变量;

2) 确定回归函数 $\mu(x)$ 的类型;

3) 对参数进行估计.

9.一元线性回归模型

若回归函数是线性函数

$$\mu(x_1, x_2, \cdots, x_k) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

其中b_0 , b_1 , …,b_k是未知常数，称为线性回归问题.

若Y关于X 的回归函数为

$$\mu(x) = E(Y \mid X = x) = a + bx$$

有一元线性回归模型：

$$Y=a+bx+\varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

其中a、b、σ^2为未知参数，且

a —回归常数(又称截距)

b —回归系数(又称斜率)

ε — 随机误差（随机扰动项）

若随机误差$\varepsilon \sim N(0,\sigma^2)$, 称为一元线性正态回归模型.

ε_i是第 i 次观察时的随机误差，有

1) $E(\varepsilon_i)=0$， $D(\varepsilon_i)=\sigma^2$， $i=1,2$， …， n;

2) ε_1， …， ε_n相互独立.

用观察值（x_i，y_j）a、b、σ^2进行估计.

$$\boxed{\begin{cases} \hat{b} = \dfrac{l_{xy}}{l_{xx}} \\ \hat{a} = \bar{y} - \hat{b}\bar{x} \end{cases}}$$

10. 一元线性回归分析.

$$\hat{y}=\hat{a}+\hat{b}\,x. \qquad \hat{b}=\frac{l_{xy}}{l_{xx}},\ \hat{a}=\bar{y}-\hat{b}\bar{x}.$$

$$l_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_i y_i-n\bar{x}\,\bar{y}.$$

$$l_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2.$$

$$l_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-n\bar{y}^2.$$

$$\hat{\sigma}^2=\frac{1}{n-2}(l_{yy}-\hat{b}^2 l_{xx}).$$

相关系数检验法. $R=\dfrac{l_{xy}}{\sqrt{l_{xx}}\sqrt{l_{yy}}}.$

$|R|>R_\alpha(n-2),$

线性相关关系显著

第六章：8、9

9、设X_1、X_2为来自服从参数为λ的指数分布的总体的简单随机样本，$Z=\dfrac{\min\{X_1,X_2\}}{\max\{X_1,X_2\}}$，对任意给定的c,计算$P\{Z\le c\}$。

解： $P\{Z\le c\}=P\{X_1\ge X_2,Z\le c\}+P\{X_1<X_2,Z\le c\}$

$$=P\{X_1\ge X_2,\frac{X_2}{X_1}\le c\}+P\{X_1<X_2,\frac{X_1}{X_2}\le c\}$$

$$=\iint\limits_{x_1\ge x_2,x_2\le cx_1} f(x_1,x_2)dx_1dx_2+\iint\limits_{x_1<x_2,x_1\le cx_2} f(x_1,x_2)dx_1dx_2$$

$$=\int_0^{+\infty}(\int_0^{cx_1}\lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}dx_2)dx_1+\int_0^{+\infty}(\int_0^{cx_2}\lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}dx_1)dx_2$$

$$=\frac{c}{1+c}+(1-\frac{1}{1+c})$$

$$=\frac{2c}{1+c}$$

第七章：5、7、12、13、17

7、设总体X的分布律为 $X \sim \begin{pmatrix} 1 & 2 & 3 \\ 1-\theta & \theta-\theta^2 & \theta^2 \end{pmatrix}$，其中$\theta$未知。以$N_i$表示来自的样本中等于i的个数。求常数 $\alpha_1,\alpha_2,\alpha_3$ ，使$T=\sum_{i=1}^{3}\alpha_i N_i$为的$\theta$无偏估计量，并求$T$的方差。

解：$N_i \sim B(n, p_i)$, $\quad p_1 = 1-\theta, p_2 = \theta-\theta^2, p_3 = \theta^2$

$$E(T) = \sum_{i=1}^{3}\alpha_i E(N_i) = \sum_{i=1}^{3}\alpha_i n p_i = n(\alpha_1(1-\theta)+\alpha_2(\theta-\theta^2)+\alpha_3\theta^2)$$

$$= n\alpha_1 + (n\alpha_2 - n\alpha_1)\theta + (n\alpha_3 - n\alpha_2)\theta^2$$

$$要求E(T)=\theta,即要求\begin{cases} n\alpha_1=0 \\ n\alpha_2-n\alpha_1=1 \\ n\alpha_3-n\alpha_2=0 \end{cases} \Rightarrow \begin{cases} \alpha_1=0 \\ \alpha_2=1/n \\ \alpha_3=1/n \end{cases}$$

则：$T=\frac{1}{n}(N_2+N_3)=\frac{1}{n}(n-N_1)=1-\frac{N_1}{n}$

$$D(T)=\frac{1}{n^2}D(N_1)=\frac{1}{n^2}np_1(1-p_1)=\frac{(1-\theta)\theta}{n}$$

17、从某种型号的一批电子管中抽出容量为10的样本作寿命试验，算得标准差s=45(小时)，设整批电子管的寿命服从正态分布，试求这批电子管寿命标准差的单侧置信上界（置信度为0.95）

解：电子管的寿命为X~$N(\mu,\sigma^2)$, n,=10,s=45,

$\alpha=0.05$, 求σ^2区间估计，μ未知枢轴变量为：

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

$$P\{\chi^2 \geq \chi^2_{1-\alpha}(n-1)\} = 1-\alpha$$

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

$$\Rightarrow P\{\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\} = 1-\alpha$$

$\therefore \sigma$的上界：$\sqrt{\dfrac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}}$

第八章：5、6、9、11、12、

5、测试某溶液的水分，测得 10 个观测值，样本均值为 0.452%，标准差为 0.037%.设总体服从正态分布，试在显著性水平 0.05下，分别检验假设

$$(1) H_0: \mu \geq 0.5\%, H_1: \mu < 0.5\% \qquad (2) H_0: \sigma \geq 0.04\%, H_1: \sigma < 0.04\%$$

解：设$X_1, X_2 \cdots X_n$是总体$X \sim N(\mu, \sigma^2)$的样本，

其样本值为$x_1, x_2 \cdots x_n$

（1）当H_0成立时，

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}} = \frac{\bar{X} - 0.5\%}{S / \sqrt{n}} \sim t(n-1)$$

$\because P\{T < -t_\alpha(n-1)\} = \alpha$

$\therefore$ 拒绝域$(-\infty, -t_\alpha(n-1))$

代入数据对H_0做出判断

（2）当H_0成立时，

$$\chi^2 = \frac{n-1}{\sigma^2} S^2 = \frac{n-1}{0.04\%^2} S^2 \sim \chi^2(n-1)$$

$\because P\{\chi^2 < \chi^2_{1-\alpha}(n-1)\} = \alpha$

$\therefore$ 拒绝域$(-\infty, \chi^2_{1-\alpha}(n-1))$

代入数据对H_0做出判断

11、设总体$X \sim N(\mu, 2^2), X_1, X_2, \cdots X_{16}$为其样本，$\bar{X} = \frac{1}{n}\sum_{i=1}^{16} X_i$为样本均值，对假设

$H_0: \mu = 0, H_1: \mu \neq 0$

（1）试证：下述三个拒绝域具有相同的显著性水平α=0. 05

$\{2\bar{X} \leq -1.645\}$，　$\{1.5 \leq 2\bar{X} \leq 2.125\}$，　$\{2\bar{X} \leq -1.96$及$2\bar{X} \geq 1.96\}$

（2）在上述三个拒绝域中应选哪一个比较合理？为什么？

解：当H_0成立时，

$$U = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = 2\bar{X}$$

$$P\{2\bar{X} \leq -1.645\} = P\{U \leq -u_{1-0.95}\} = \Phi(u_{0.95}) = 0.05$$

$$P\{1.5 \leq 2\bar{X} \leq 2.125\} = P\{u_{1-0.9332} \leq 2\bar{X} \leq u_{1-0.9832}\}$$

$$= P\{U \leq u_{1-0.9832}\} - P\{U \leq u_{1-0.9332}\} = 0.9832 - 0.9332 = 0.05$$

$$P\{2\bar{X} \leq -1.96 \cup 2\bar{X} \geq 1.96\} = P\{|U| \geq u_{1-0.025}\} = 0.05$$

12、设总体$X \sim N(\mu,4^2)$, X_1,X_2,X_3,X_4为其样本，$\bar{X}$为样本均值，对假设

$H_0:\mu=5, H_1:\mu\neq 5$

（1）给出一个显著性水平为$\alpha=0.05$的拒绝域；

（2）若μ=6，试计算犯第二类错误的概率β。

解：

（2）β=$P\{$接受$H_0|H_0$不成立$\} = P\{|\dfrac{\bar{X}-5}{4/\sqrt{4}}|< u_{\alpha/2}\}$

$$= P\{-u_{\alpha/2}-\frac{1}{2}<\frac{\bar{X}-6}{2}<u_{\alpha/2}-\frac{1}{2}\}(H_0\text{不成立，} U\sim N(1/2,1))$$

$$=\Phi(u_{\alpha/2}-\frac{1}{2})-\Phi(-u_{\alpha/2}-\frac{1}{2})$$

相关例题:

例1. 证明如下等式:

① $$\sum_{i=1}^{n}(X_i-\overline{X})=0;$$

② $$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2.$$

③ $$\sum_{i=1}^{n}(X_i-A)^2=\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-A)^2;$$

点评: 以上公式极其简单, 却是统计学中常用公式.

例2、总体X的密度函数为

$$\therefore\quad f(x)=\begin{cases}(\alpha+1)x^{\alpha}, & 0<x<1;\\ 0, & \text{其他。}\end{cases}\qquad (\alpha>1)$$

求α的矩估计量和极大似然估计量.

解:设$X_1,X_2,\ldots,X_n$为总体X的一组样本，其样本值为$x_1,x_2,\ldots,x_n$，则

1) 矩估计法

$$\because\ E(X)=\int_0^1 x(\alpha+1)x^{\alpha}dx=\frac{x^{\alpha+2}}{\alpha+2}(\alpha+1)\Big|_0^1=\frac{\alpha+1}{\alpha+2},$$

令 $\dfrac{\alpha+1}{\alpha+2}=\overline{X}$, 可得θ的矩法估计量为

$$\hat{\alpha} = \frac{1-2\overline{X}}{\overline{X}-1} = \frac{1}{1-\overline{X}} - 2.$$

2) 极大似然估计法

1. 构造似然函数

$$L(x_1,\ldots,x_n;\alpha) = \begin{cases} (\alpha+1)^n \prod_{i=1}^{n} x_i^{\alpha}, & 0 < x_i < 1; \\ 0, & \text{其它} \end{cases}$$

2. 取对数：当 $0<x_i<1$, $(i=1,2,\ldots,n)$ 时

$$\ln L = n\ln(\alpha+1) + \alpha\sum_{i=1}^{n}\ln x_i$$

3. 建立似然方程 $\dfrac{d\ln L}{d\alpha}=\dfrac{n}{\alpha+1}+\sum_{i=1}^{n}\ln x_i=0,$

4. 由似然方程求解得极大似然估计值为

$$\hat{\alpha}=\frac{n}{-\sum_{i=1}^{n}\ln x_i}-1,$$

5. 极大似然估计量为

$$\hat{\alpha}=\frac{n}{-\sum_{i=1}^{n}\ln X_i}-1.$$

例3、按环境保护条例, 在排放的工业废水中, 某有害物质含量不得超过0.5‰，现在取5份水样，测定该有害物质含量，得如下数据：

0.530‰ 0.542‰ 0.510‰ 0.495‰ 0.510‰

能否据此抽样结果说明有害物质含量超过了规定？ ($\alpha=0.05$).

解：假定有害物质含量X服从正态分布，$X \sim N(\mu, \sigma^2)$，$X_1, X_2, \ldots, X_n$为总体X的一组样本，其样本值为$x_1, x_2, \ldots, x_n$

需检验假设

H_0: $\mu \geq 0.5‰$， H_1： $\mu < 0.5‰$

单侧
检验

非常
重要!

若H_0成立，检验统计量

$$T = \frac{\bar{X} - 0.0005}{S/\sqrt{n}} \sim t(n-1);$$

$$P\{T < -t_\alpha(n-1)\} = \alpha$$

n=5,查表得 $t_{0.05}(4)$=2.1318，拒绝域 (-∞, -2.1318,)

统计量T的观测值为

$$t = \frac{0.0005174 - 0.0005}{0.000018542/\sqrt{5}} = 2.0984 > -2.1318$$

故在显著性水平0.05下接受H_0，即认为排放的废水中该有害物质含量超过规定标准。

例4、已知某炼铁厂铁水含碳量服从正态分布 $N(4.55,、0.108^2)$，现在测定了9炉铁水，平均含碳量为4.484，如果方差没有变化，可否认为现在生产的铁水平均碳的含量仍为4.55?
(α=0.05).

解：已知$\sigma=\sigma_0=0.108$，需检验假设

$$H_0: \mu = 4.55 ，\mathrm{H}_1: \mu \neq 4.55$$

若H_0成立, 检验统计量 $$U=\frac{\bar{X}-4.55}{\sigma_0/\sqrt{n}} \sim N(0,1)$$

查表得 $u_{\frac{\alpha}{2}} = u_{0.025} = 1.96$

H_0的拒绝域为

$w = (-\infty, -1.96) \cup (1.96, \infty)$ 或 $|u| > 1.96$

计算统计量U的观测值

$$|u| = \left| \frac{\bar{x} - \mu_0}{\sigma_0 / \sqrt{n}} \right| = \frac{|4.484 - 4.55|}{0.108 / \sqrt{9}} = 1.833 < 1.96 \quad \Rightarrow \quad u \notin w$$

故在显著性水平0.05下接受H_0，即认为现在生产的铁水的平均含碳量仍为4.55.

例5、已知总体 $X \sim N(\mu,\sigma^2)$，抽取n=100的简单随机样本，方差已知。现确定的估计区间为(43.88,46.52)，试问随机估计区间的置信度是多少？ $\Phi(1.65)=0.95$，$\sigma=8$

解.对已知 σ 的正态总体 $N(\mu,\sigma^2)$，μ 的区间估计，选取枢轴变量为：$U=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

得到置信度 $1-\alpha$ 为的置信区间为：

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

代入样本值得到区间长度为： $2\frac{\sigma}{\sqrt{n}}u_{\alpha/2}$

则 $2\frac{\sigma}{\sqrt{n}}u_{\alpha/2} = 46.52 - 43.88 = 2.64$

由于 $\sigma = 8, \sqrt{n} = 10,$ 所以 $1.32 = \frac{u_{\alpha/2} \cdot 8}{10},$ $u_{\alpha/2} = 1.65$

反查正态分布函数表知：

$$1 - \frac{\alpha}{2} = 0.95, \text{即:} \alpha = 0.10, 1 - \alpha = 0.90$$

从而区间估计的置信度为0.90。

例6、为制定在服装标准, 调查了一组女青年的身高X与裤长Y的数据, 经计算得$(x_i, y_i), i=1,2,\cdots 30$

$$\sum_{i=1}^{30} x_i = 4797, \quad \sum_{i=1}^{30} y_i = 3068, \quad \sum_{i=1}^{30} x_i^2 = 767949,$$

$$\sum_{i=1}^{30} y_i^2 = 314112, \qquad \sum_{i=1}^{30} x_i y_i = 491124$$

试求：1.裤长Y对身高X的经验回归直线;

2. 用相关系数检验法，在显著性水平$\alpha=0.01$下检验回归方程的显著性.

解： 1. 由已知

$$\bar{x} = \frac{1}{30}\sum_{i=1}^{30} x_i = 159.9, \qquad \bar{y} = \frac{1}{30}\sum_{i=1}^{30} y_i = 102.3,$$

$$l_{xx} = \sum_{i=1}^{30} x_i^2 - \frac{1}{30}[\sum_{i=1}^{30} x_i]^2 = 908.7,$$

$$l_{xy} = \sum_{i=1}^{30} x_i y_i - \frac{1}{30}(\sum_{i=1}^{30} x_i)(\sum_{i=1}^{30} y_i) = 550.8$$

$$l_{yy} = \sum_{i=1}^{30} y_i^2 - \frac{1}{30}[\sum_{i=1}^{30} y_i]^2 = 357.87,$$

$$\hat{b} = \frac{l_{xy}}{l_{xx}} = \frac{550.8}{908.7} = 0.61,$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 102.3 - 0.61 \times 159.9 = 504$$

Y关于X的经验回归方程为

$$\hat{y} = \hat{a} + \hat{b}x = 5.4 + 0.61x$$

2. 样本相关系数 $R = \dfrac{l_{xy}}{\sqrt{l_{xx}}\sqrt{l_{yy}}} = 0.966$

查表得 $R_{0.01}(28)=0.463$, 因为$|R| > R_{0.01}(28)$,

可认为裤长Y与身高X之间的线性相关关系显著。

一、选择填空题：

1. 样本X_1, X_2, X_3, X_4来自正态总体$N(\mu,\sigma^2)$，μ已知,σ未知，下列随机变量中不能作为统计量的是___C___.

(A) $\overline{X}=\frac{1}{4}\sum_{i=1}^{4}X_i$

(B) $X_1+X_2-2\mu$

(C) $\chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{4}(X_i-\overline{X})^2$

(D) $S^2=\frac{1}{3}\sum_{i=1}^{4}(X_i-\overline{X})^2$

2. 总体$X\sim N(\mu,\sigma^2)$，$X_1, X_2, \ldots, X_{20}$是X的一个样本，令 $Y=3\sum_{i=1}^{10}X_i-4\sum_{i=11}^{20}X_i$，则$Y$服从分布 $N(-10\mu, 250\sigma^2)$.

3. $X_1, X_2, \ldots, X_5$是来自总体$X\sim N(0, 1)$的一个样本，若 $\dfrac{C(X_1+X_2)}{\sqrt{X_3^2+X_4^2+X_5^2}}$服从 t 分布，则常数

$C=\boxed{\sqrt{\dfrac{3}{2}}}$.

4. 设$X_1, X_2, \ldots, X_8$和$Y_1, Y_2, \ldots, Y_{10}$分别来自正态总体$N(-1, 2^2)$和$N(2, 5)$的样本，且相互独立，S_1^2和S_2^2分别表示两样本的样本方差，则服从$F(7, 9)$的统计量是__B__.

(A) $\dfrac{2S_1^2}{5S_2^2}$;　(B) $\dfrac{5S_1^2}{4S_2^2}$;　(C) $\dfrac{4S_2^2}{5S_1^2}$;

(D) $\dfrac{5S_1^2}{2S_2^2}$;

$$\because F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$$

5. 样本$X_1, X_2, \ldots, X_n (n>1)$来自总体$X \sim N(0, 1)$，$\overline{X}$ 与S分别是样本均值和样本标准差, 则有__C__.

(A) $\overline{X} \sim N(0,1)$;　　(B) $nX \sim N(0,1)$;

(C) $\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$;　　(D) $\dfrac{\overline{X}}{S} \sim t(n-1)$;

6. $X_1, X_2, \dots, X_n (n>1)$是来自$X \sim N(\mu, \sigma^2)$的一个样本，$\bar{X}$ 和S^2分别是样本均值和样本方差，则下列结论正确的是__B_____.

(A) $2X_1 - X_2 \sim N(\mu, \sigma^2)$;

(B) $\dfrac{n(\bar{X}-\mu)^2}{S^2} \sim F(1, n-1)$;

(C) $\dfrac{S^2}{\sigma^2} \sim \chi^2(n-1)$;

(D) $\dfrac{\bar{X}-\mu}{S}\sqrt{n-1} \sim t(n-1)$;

二、 $X_1, X_2, \ldots, X_n$是来自总体$X \sim N(\mu,\sigma^2)$ 的样本，S^2为样本方差, 求样本容量的最大值，使其满足不等式

$$P\{\frac{(n-1)S^2}{\sigma^2} \le 15\} \ge 0.95$$

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

$$P\{\frac{(n-1)S^2}{\sigma^2} \le 15\} \ge 0.95 \quad \Rightarrow P\{\frac{(n-1)S^2}{\sigma^2} > 15\} < 0.05$$

查表得 $\chi^2_{0.05}(7) = 14.067,\ \chi^2_{0.05}(8) = 15.507$

因此$n-1 \le 7$，即$n \le 8$ 时不等式成立，样本容量的最大值为 8 .

设事件 A、B 满足 $P(B|A)=1$ 且 $0<P(A)<1$。判断下列各式是否正确，并给出理由。（1）$P(A)\le P(B)$　（2）$A\subset B$

解：

$$P(B|A)=1\text{且}0<P(A)<1,\text{得}P(B|A)=\frac{P(AB)}{P(A)}=1,\text{即}P(AB)=P(A)$$

（2）错误。反例：在[0,1]区间上任意取一点，假设事件

A={该点落在[0,1/2]}, B={该点落在[0,1/2)}

则 $P(A)=P(B)=P(AB)=P(A\cup B)=\dfrac{1}{2},P(B|A)=1$ 但 $B\subset A$

2、一个班内有 20 位同学都想去参观一个展览会，但只有 3 张参观票，大家都同意通过这 20 位同学抽签决定 3 张票的归属。计算下列事件的概率：

（1）第二人抽到票的概率；

（2）第二人才抽到票的概率；

（3）第一人宣布抽到了票，第二人又抽到票的概率；

（4）前两人中至少有一人抽到票的概率。

解：设 A_i 表示第 i 个人抽到票，i=1, 2…, 20

(1) $P(A_2)=P(A_1A_2+\bar{A}_1A_2)=P(A_1)P\left(A_2\mid A_1\right)+P(\bar{A}_1)P\left(A_2\mid\bar{A}_1\right)=\dfrac{3}{20}$

抽签等概

(2) $P(\bar{A}_1A_2)=P(\bar{A}_1)P(A_2\mid\bar{A}_1)=\dfrac{51}{380}$

(3) $P(A_2\mid A_1)=\dfrac{2}{19}$

(4)

$$P(A_1A_2+A_1\bar{A}_2+\bar{A}_1A_2)=P(A_1)P(A_2\mid A_1)+P(A_1)P(\bar{A}_2\mid A_1)+P(\bar{A}_1)P(A_2\mid\bar{A}_1)=\frac{27}{95}$$

二维随机变量(X,Y)在由直线$y=x, x+y=1$及直线$y=0$所围成区域G上服从均匀分布。

（1） 求(X,Y)的联合概率密度函数；　（2）求条件概率密度函数$f_{X|Y}(x|y)$；（3）求条件概率$P(X<\frac{1}{3}|Y=\frac{1}{4})$。

解：区域G的面积为1/4，于是（X，Y）的联合概率密度为：

（1） $$f(x,y)=\begin{cases}4, & y<x<1-y, 0<y<1/2\\ 0, & 其他\end{cases}$$

（2） $$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\int_y^{1-y}4dx=4(1-2y), & 0<y<1/2\\ 0, & 其他\end{cases}$$

当$0<y<1/2, f_Y(y)=4(1-2y)>0,$有

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\begin{cases}\dfrac{1}{1-2y}, & y<x<1-y\\ 0, & 其他\end{cases}$$

（3）由 $$f_{X|Y}(x|1/4)=\begin{cases}2, & 1/4<x<3/4\\ 0, & 其他\end{cases}$$

得到$P(X<1/3|Y=1/4)=\int_{1/4}^{1/3}2dx=1/6$

1、（10 分）常数a与任意随机变量独立吗？给出数学证明过程.

答：将常数a看作随机变量X，即$P\{X=a\}=1$，其分布函数为

$$F_X(x)=\begin{cases}0, & x<a,\\ 1, & x\geq a.\end{cases}$$

设随机变量Y的分布函数为$F_Y(y)$，X, Y的联合分布函数为$F(x,y)$.

当$x<a$，对任意实数y，

$$0\leq F(x,y)=P\{X\leq x,Y\leq y\}=P\{X\leq x\}=0,$$

则$F(x,y)=0=F_X(x)F_Y(y)$.

当$x\geq a$，$P\{X\leq x\}=1$，此时$\{X\leq x\}=\Omega$，故对任意实数y，

$$F(x,y)=P\{X\leq x,Y\leq y\}=P\{Y\leq y\}=F_Y(y),$$

因为$F_X(x)=1$，所以$F(x,y)=F_X(x)F_Y(y)$.

综上，对任意实数x,y，总有$F(x,y)=F_X(x)F_Y(y)$，即常数a与任意随机变量独立.