参数估计 $\begin{cases} \text{点估计} \\ \text{区间估计} \end{cases}$

1°

引例 1. 已知某学校高数成绩成正态分布，现在大一有 3000 人参加高数考试，从中随机抽取 100 名同学观察其成绩，计算得到这 100 人的高数期末成绩平均值为 70 分，标准差为 8，试估计这所学校高数成绩的具体分布形式。

$$\left.\begin{array}{l} N\sim(\mu,\sigma^2) \\ \hat{\mu}=\bar{x}=70 \\ \sigma^2=s^2=8^2 \end{array}\right\} N(70,8^2)$$

2°

引例 2. 某事故高发路段，一天中发生交通事故的次数 X 是一个随机变量，假设根据以往经验判断 X 服从参数为 $\lambda(\lambda>0)$ 的泊松分布，但是 λ 未知，现通过调查和整理 250 天的数据得到以下样本值，试估计参数 λ.

一天事故次数 k	0	1	2	3	4	5	6
发生 k 次的天数	75	90	54	22	6	2	1

$$\left.\begin{array}{l} X\sim\pi(\lambda) \\ \lambda=E(X) \end{array}\right\} X\sim\pi(1.22)$$

$$\hat{\lambda}=\bar{x}=\frac{75\times0+90\times1+54\times2+22\times3+4\times6+2\times5+6\times1}{250}$$

$$=1.22$$

点估计 估计量 $\hat{\lambda}=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

估计值 $\hat{\lambda}=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

方法一：矩估计

$$\hat{\mu}=\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad (X_1,\dots X_n\text{为样本})$$

思路：① 总体矩 $\mu=E(X)=g(\theta)$

② 令 $\mu=g(\theta)=\bar{x} \Rightarrow \hat{\theta}=h(\bar{x})$

③ 代入 $\bar{x}$，$\hat{\theta}=h(\bar{x})$，得到结果

1°

例 1．设总体 X 的概率分布为

X	0	1	2	3
p	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中 $\theta(0<\theta<\frac{1}{2})$ 是未知参数．利用总体的如下样本值：3,1,3,0,3,1,2,3,

求 θ 的矩估计值．

离散

① $E(X)=0\times\theta^2+1\times2\theta(1-\theta)+2\times\theta^2+3(1-2\theta)$

$=3-4\theta$

② 令 $E(X)=3-4\theta=\bar{x} \Rightarrow \hat{\theta}=\frac{3-\bar{x}}{4}$

③ $\bar{X} = \dfrac{3+1+3+0+3+1+2+3}{8} = 2$

$\therefore \hat{\theta} = \dfrac{3-2}{4} = 0.25$

2°

例 2 设总体 X 的概率密度函数为

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1}, & 0<x<1\\ 0, & \text{其他}\end{cases},$$

其中 $\theta(\theta>0)$ 为未知参数，$X_1,X_2,\cdots,X_n$ 是 X 的一个样本，求 θ 的矩估计量.

连续

① $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x \cdot \theta x^{\theta-1}\,dx$

$= \int_0^1 \theta x^{\theta}\,dx = \dfrac{\theta}{\theta+1} x^{\theta+1}\Big|_0^1 = \dfrac{\theta}{\theta+1}$

② $E(X) = \dfrac{\theta}{\theta+1} = \bar{X}$, $\hat{\theta} = \dfrac{\bar{x}}{1-\bar{x}}$

$\underbrace{(\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i)}_{\text{样本均值.}}$

方法2：最大似然估计

概率最大 ⇒ 估计值

思路：1）离散

①似然函数

$$L(\theta) = p(x_1) * p(x_2) * \cdots p(x_n)$$

②取对数

$$\ln L(\theta) = \sum_{i=1}^{n} \ln p(x_i, \theta)$$

③对θ求导，令 $\ln L(\theta) d\theta = 0$

得出 θ

1°

例 1．设总体 X 的概率分布为

X	0	1	2	3
p	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中 $\theta(0<\theta<\frac{1}{2})$ 是未知参数．利用总体的如下样本值：3,1,3,0,3,1,2,3,

求 θ 的矩估计值．

$$① L(\theta) = (1-2\theta)^4 * [2\theta(1-\theta)]^2 * (\theta^2)^2$$

$$= 4\theta^6 (1-\theta)^2 (1-2\theta)^4$$

② $\ln L(\theta) = \ln[4\theta^6(1-\theta)^2(1-2\theta)^4]$

$= \ln 4 + 6\ln\theta + 2\ln(1-\theta) + 4\ln(1-2\theta)$

③ $\ln L(\theta) d\theta = \frac{6}{\theta} - \frac{2}{1-\theta} - \frac{8}{1-2\theta} = \frac{24\theta^2-28\theta+6}{\theta(1-\theta)(1-2\theta)}$

令其为0 得 $\theta = \frac{7\pm\sqrt{13}}{12}$

又 $0<\theta<\frac{1}{2}$

$\therefore \hat{\theta} = \frac{7-\sqrt{13}}{12}$

2) 连续性

思路: ① $L(\theta) = f(x_1) * f(x_2) \cdots f(x_n)$

② $\ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i, \theta)$

③ $\ln L(\theta) d\theta = 0$ 得 $\hat{\theta}$

2°

例 4. 设总体 X 的概率密度函数为

$$f(x;\lambda)=\begin{cases}\lambda\alpha x^{\alpha-1}e^{-\lambda x^{\alpha}}, & x>0\\ 0, & 其他\end{cases},$$

其中 $\lambda(\lambda>0)$ 为未知参数，$\alpha(\alpha>0)$ 为已知常数，$X_1,X_2,\cdots,X_n$ 是 X 的一个样本，求 λ 的最大似然估计量.

① $L(\alpha)=f(x_1)f(x_2)\cdots f(x_n)$

$$=(\lambda\alpha)^n\prod_{i=1}^{n}x_i^{\alpha-1}e^{-\lambda\sum_{i=1}^{n}x_i^{a}}$$

② $\ln L(\lambda)=n\cdot\ln\lambda\alpha+(\alpha-1)\sum_{i=1}^{n}\ln x_i$

$$+\left(-\lambda\sum_{i=1}^{n}x_i^{a}\right)$$

③ $\frac{d}{d\lambda}\ln L(\lambda)=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i^{\alpha}=0$

$\therefore \hat{\lambda}=\dfrac{n}{\sum_{i=1}^{n}x_i^{\alpha}}$

换为估计量

即 $\hat{\lambda}=\dfrac{n}{\sum_{i=1}^{n}X_i^{\alpha}}$

无偏估计：$E(\hat{\theta})=\theta$

例 5. 已知总体 X 的期望 $E(X)=0$，方差 $D(X)=\sigma^2$，$X_1,X_2,\cdots,X_n$ 是来自 X 的一个样本，样本均值为 $\overline{X}$，样本方差为 S^2，则以下是 σ^2 的无偏估计的是（ ）

A. $n\overline{X}^2+S^2$　B. $\frac{1}{2}n\overline{X}^2+\frac{1}{2}S^2$　C. $\frac{1}{3}n\overline{X}^2+S^2$　D. $\frac{1}{4}n\overline{X}^2+\frac{1}{4}S^2$

$E(\quad) = \sigma^2$

$E(\bar{X}^2) = D(\bar{X}) + [E(X)]^2 = \frac{\sigma^2}{n}$

$E(\bar{S}) = \sigma^2$

区间估计以假设检验.

正态总体常用分布：

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

② $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

③ $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

上α分位点：$P\{X > X_\alpha\} = \alpha \quad (0<\alpha<1)$

区间估计

方法一：双侧区间估计

$$P\{\underline{\theta} < X < \overline{\theta}\} > 1-\alpha.$$

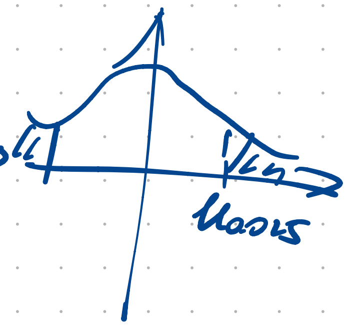
1. σ^2 已知，估计 μ

例 1. 已知某银行储户平均账户余额服从正态分布，且总体标准差为 1500 元，现在银行经理随机抽取了一个容量为 36 的储户样本，观察到样本账户的余额均值为 4500，试估计平均账户余额的置信度为 95%的置信区间。($\Phi(1.96)=0.975$)

$$\sigma = 1500,\quad n=36,\quad \bar{x}=4500,\quad 1-\alpha = 95\%$$

$$\downarrow$$

$$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \rightarrow$$



求 $P\{\,?<\mu<\,?\,\}=95\% \Rightarrow P\left\{\,?<\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}<\,?\,\right\}=95\%$

$$P\left\{-u_{0.025} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < u_{0.025}\right\} = 95\%$$

$$P\left\{\bar{x}-\frac{\sigma}{\sqrt{n}}u_{0.025} < \mu < \bar{x}+\frac{\sigma}{\sqrt{n}}u_{0.025}\right\} = 95\%$$

解题：

① $\sigma=1500$, $n=36$, $\bar{x}=4500$, $1-\alpha=95\%$

② 选择统计量 $\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

③ 估计区间：$\left(\bar{x}-\dfrac{\sigma}{\sqrt{n}}u_{0.025},\ \bar{x}+\dfrac{\sigma}{\sqrt{n}}u_{0.025}\right)$

④ 代入参数 $\left(4500-\dfrac{1500}{\sqrt{36}}\times1.96,\ 4500+\dfrac{1500}{\sqrt{36}}\times1.96\right)$

$\Rightarrow (4010, 4990)$

2. σ^2 未知，估计 μ

例 2. 某公司规定其销售人员每周提交一份与客户联系的情况报告，已知销售人员每周平均与客户联系次数服从正态分布，现在经理随机抽取了 36 份报告察看，这些报告显示销售人员每周平均与客户联系 22.4 次，样本标准差为 5 次，试估计平均联系客户次数的置信度为 95%的置信区间。($t_{0.025}(36)=2.0281, t_{0.025}(35)=2.0301$)

① $n=36$, $\bar{x}=22.4$, $S=5$, $1-\alpha=95\%$

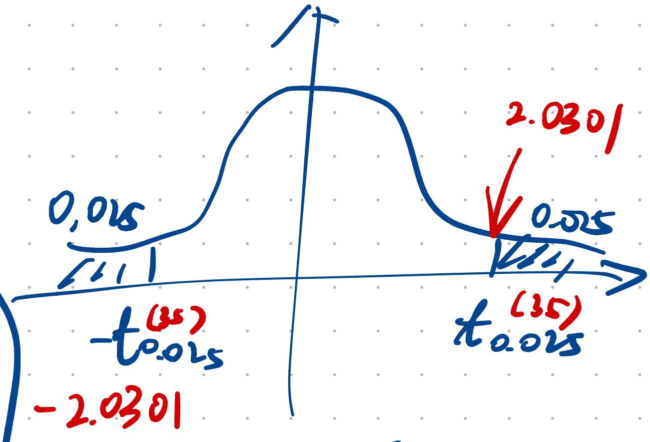
$P\{?<\mu<?\}=95\%$

② $\hookrightarrow P\left\{?<\dfrac{\bar{x}-\mu}{S/\sqrt{n}}<?\right\}\sim t(n-1)$

③ $P\{ ? < \frac{\bar{x}-\mu}{S/\sqrt{n}} < ? \} = 95\%$

↓

$P\{-2.0301 < \frac{\bar{x}-\mu}{S/\sqrt{n}} < 2.0301\} = 95\%$



★注意自由度 $n=36 \Rightarrow t(35)$

$\therefore P\{\bar{x} - \frac{S}{\sqrt{n}} t_{0.025}^{(35)} < \mu < \bar{x} + \frac{S}{\sqrt{n}} t_{0.025}^{(35)}\} = 95\%$

置信区间

④代入 $(22.4 - \frac{5}{\sqrt{36}} \times 2.0301,\ 22.4 + \frac{5}{\sqrt{36}} \times 2.0301)$

$\Rightarrow (20.71,\ 21.\quad)$

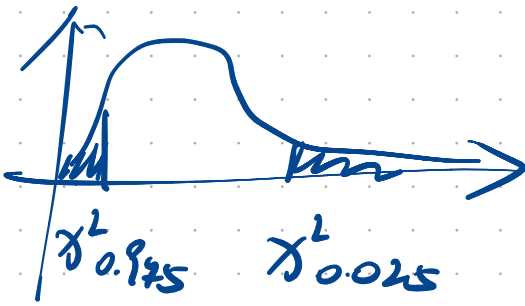
3. 估计 σ^2

例 3. 假设学生身高服从正态分布，现在某学校随机抽取了 25 名同学测量其身高数据，算得平均身高 170cm，样本标准差为 12cm，试求该学校学生身高的方差 σ^2 的置信度为 95% 的置信区间。（$\chi^2_{0.975}(24) = 12.401, \chi^2_{0.025}(24) = 39.364$）

$P\{ ? < \sigma^2 < ? \} = 95\%$

① $n=25,\ \bar{x}=170,\ S=12,\ 1-\alpha=95\%$

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



③ $P\left\{\chi^{2}_{0.975}(24) < \dfrac{(n-1)s^2}{\sigma^2} < \chi^{2}_{0.025}(24)\right\} = 95\%$

$\Rightarrow$ 区间 $\left(\dfrac{(n-1)s^2}{\chi^{2}_{0.025}(24)}, \dfrac{(n-1)s^2}{\chi^{2}_{0.975}(24)}\right)$

④ 代入参：$\left(\dfrac{(25-1)\times12^2}{39.364}, \dfrac{(25-1)\times12^2}{12.401}\right)$

方法2：单侧区间估计

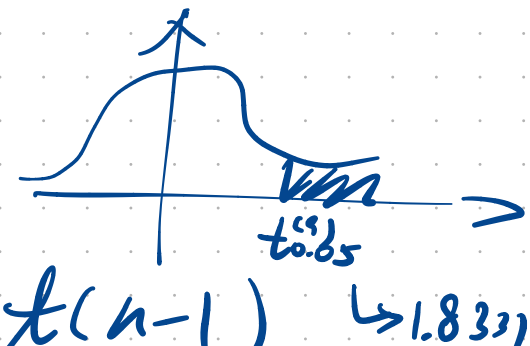
例 4. 已知某电子元件的寿命服从正态分布，现在从一批电子元件中抽取了 10 只作寿命试验，观察到平均寿命为 1500 小时，样本方差为 $\frac{10}{3}$，试估计平均寿命的 95%的置信下限。

$(t_{0.05}(9)=1.8331)$

① $n=10$，$\bar{x}=1500$，$s^2=\frac{10}{3}$，$s=\sqrt{\frac{10}{3}}$，$1-\alpha=95\%$

$P\{\mu > ?\} = 95\%$

② 选统计量：$\dfrac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$



$\Rightarrow P\left\{\dfrac{\bar{x}-\mu}{s/\sqrt{n}} < ?\right\} = 95\%$

$\therefore P\left(\frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{0.05}^{(9)}\right) = 95\%$

$P\left(\mu > \bar{x} - \frac{s}{\sqrt{n}} t_{0.05}^{(9)}\right)$

③置信下限：$\bar{x} - \frac{s}{\sqrt{n}} t_{0.05}^{9}$

④代参数：$1500 - \sqrt{\frac{10}{3}} / \sqrt{10} * 1.8331$

假设检验

显著性水平 $\alpha(=0.05)$

2. 关于总体的未知参数 θ 的假设检验类型：

类型		H_0	H_1
双边检验		$\theta=\theta_0$	$\theta\neq\theta_0$
单边检验	右边	$\theta\leq\theta_0$	$\theta>\theta_0$
	左边	$\theta\geq\theta_0$	$\theta<\theta_0$

3.假设检验的步骤

(1)据题意写出原假设 H_0 和备择假设 H_1；

(2)选择检验方法，写出检验统计量及其分布；⇒ 选择统计量

(3)根据给定的显著性水平确定拒绝域；

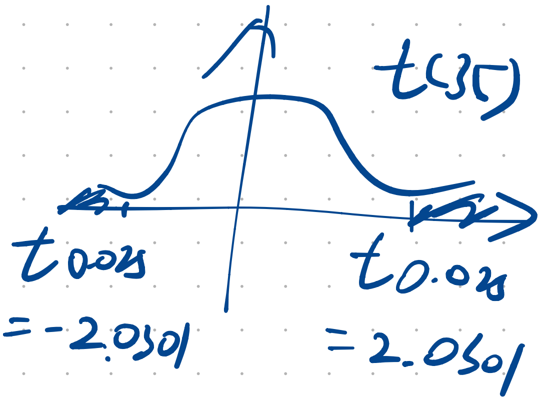
(4)计算检验统计量的值，做出推断.

例 1. 设某次高数考试成绩服从正态分布，现在从考试的考生中随机地抽取 36 位观察其成绩，算得平均成绩为 66.5 分，标准差为 15 分。问在显著性水平 0.05 下，是否可以认为这次考试全体考生的平均成绩为 70 分？ $t_{0.025}(35) = 2.0301$

$n=36$, $\bar{x}=66.5$, $s=15$, $\alpha=0.05$

① $H_0: \mu=70$, $H_1: \mu \neq 70$

② $t=\dfrac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$



③ 拒绝域：⇒ 超过则拒

$|t| > 2.0301$

④: $|t| = \left|\dfrac{66.5-70}{15/\sqrt{36}}\right| = |-1.4| = 1.4 < 2.0301$

原假设无法拒绝

⇒可认为均分70

例 2. 要求一种元件使用寿命不得低于 1000 小时，现在某工厂从一批元件中随机抽取 25 件，测得其寿命的平均值为 950 小时，已知该种元件寿命服从标准差$\sigma=100h$的正态分布，试在显著性水平 0.05 下，是否可以认为这批元件是合格的？

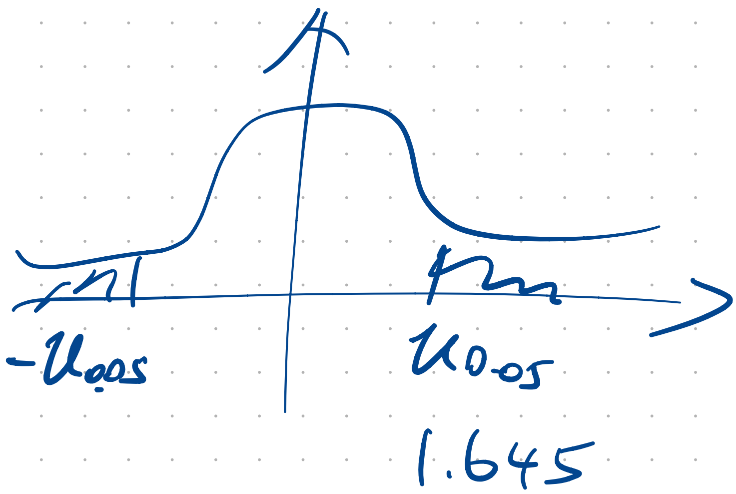
$n=25$，$\bar{x}=950$，$\sigma=100$

① $H_0: \mu \geq 1000$，$H_1: \mu < 1000$

② $\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

③ 拒绝域

$$P\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq -\mu_{0.05}\right\} \leq 0.05$$



④代入 $\because \dfrac{950-1000}{100/\sqrt{25}} = -2.5 < -1.645$

拒绝域内

$\therefore$ 产品不合格

STATISTICS – Exercises SESSION N° 1

About Chapter :

- Chapter 1 : DESCRIPTIVE STATISTICS

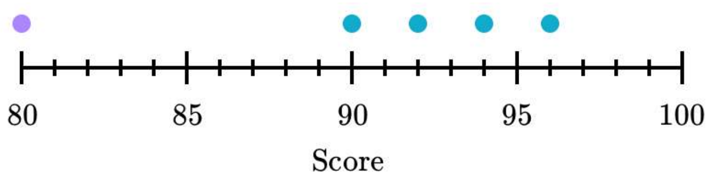
TD1 – 1 :

1. What is the median of the following statistical series? 1, 2, 4, 6, 4
2. The following table gives the distribution of the number of points scored by each player during a Quidditch season.

Player	**Fred**	**George**	**Olivier**	**Angelina**	**Harry**
Number of points	11	12	7	3	?

If the team scored on average 8 points, how many points did Harry score?

3. Anna played 5 rounds of golf this year. During the final round she obtained the lowest score of 80 points.



What is the impact of this last game on the mean and on the median of the scores of the previous games? Interpret the results.

TD1 – 2 :

Calculate the mean, median and mode for the following samples:

1) (3 ; 5 ; 2 ; 6 ; 5 ; 9 ; 5 ; 2 ; 8 ; 6)

2) (1.28 ; 2.16 ; 0.75 ; 1.44 ; 2.05 ; 0.65 ; 1.26 ; 1.73 ; 1.81 ; 0.92)

TD1 – 3 :
The table below represents the distribution of the intelligence quotient (IQ) of 100 students. This distribution is grouped into 9 classes of width 10.

Class mark (mid-value)	59.5	69.5	79.5	89.5	99.5	109.5	119.5	129.5	139.5
frequency	1	2	9	22	33	22	8	2	1
	1	3	12	34	67	89	97	99	100

1) Calculate the mean and standard deviation of this distribution

2) Determine the values of the 1st and 3rd quartile of the distribution. Calculate the Semi-interquartile range.

$Q_x = (n+1) \times \frac{x}{4}$

3) Calculate the coefficient of symmetry $\gamma_1 = \frac{\mu_3}{\sigma^3}$ and the coefficient of kurtosis $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$. Compare these values to the standard normal distribution.

$$\frac{Q_1 - 84.5}{10} = \frac{13.25}{22} \qquad \frac{Q_3 - 104.5}{10} = 0.4$$

$$Q = \frac{132.5}{22} + 84.5$$

STATISTICS – Exercises SESSION N° 2

About chapters :

- Chapter 2 : Statistical theory of estimation

TD2 – 1 :

Let $x_1,\ldots,x_n$, be a sample of size n, results of independent observations of a random variable X whose probability law depends on a parameter. Determine the estimators of this parameter by the Maximum Likelihood method and calculate their precision for the following situations:

最大似然

1) X follows an exponential law $f_X(x,\lambda) = \lambda e^{-\lambda x}$ for $x \geq 0$ and 0 otherwise.
2) Let us define $a = 1/\lambda$. Calculate the minimum variance (Cramer - Rao bound) for an estimator of the parameter a. Draw a conclusion.

2) X follows a Bernoulli law with parameter p

Remark : Here you must conclude by indicating whether the estimator is efficient and with minimum variance for the two situations.

TD2 – 2 :

An urn contains an unknown proportion of red balls and white balls. A sampling with replacement of size 60 gave a percentage of 70% for the red balls.

Calculate the 95% and 99.73% confidence limits of the proportion of red balls in the urn.

TD2 – 3 :

We assume that the random variable X that represents the weight of a given object follows a Gaussian distribution of mean m and standard deviation σ.

The following table shows the results of 10 weighings of the same object (in grams):

72.20	72.24	72.26	72.30	72.36	72.39	72.42	72.48	72.50	72.54

1) Calculate the point estimates of the mean m and of the standard deviation σ of the variable X.

2) The standard deviation is unknown but you can consider the estimated value in question 1), calculate a confidence interval at the threshold of 5% of the mean m.

置信区间 $a=0.05$

Last update on October 27th, 2022

知

Now, consider the variance is known and equal to the value calculated in 1), how does the interval change? Calculate and compare the 2 intervals.

3) The standard deviation of the balance, measured from many previous studies, is indeed, for an object of about this mass, 0.08. In this question, we consider then $\sigma = 0.08$. Determine the level of significance α so that the resulting confidence interval is $[72.31; 72,43]$.

TD2 – 4 :

Consider the following sample:

Sample : (1.28 ; 2.16 ; 0.75 ; 1.44 ; 2.05 ; 0.65 ; 1.26 ; 1.73 ; 1.81 ; 0.92)

Calculate the sample mean and sample variance. Deduce the confidence intervals for the mean, variance and standard deviation of the population. We will assume that the random variable follows a Gaussian law.

STATISTICS – Exercises SESSION N° 3

About Chapters :

- Chapter 3 : Hypothesis testing

TD3 – 1 :

During an extrasensory perception experiment, a person is isolated in a room and tasked to guess the color (red or blue) of a card chosen randomly from a 50-card deck. The cards were previously well shuffled. The person does not know the total number of blue or red cards in the deck.

Suppose the person correctly guessed the color of 32 cards. Determine if it is possible to conclude that the person possesses extrasensory perception (sixth sense) at the 0.05 and 0.01 significance level.

TD3 – 2 :

A manufacturer delivers a batch of bulbs with an average lifetime specification of 1600 hours. After inspection on a sample of 100 bulbs, the average lifespan is 1,578 hours with a standard deviation of 120 hours.

1) Can the delivery be considered to meet specification (that is, the average lifetime is 1600 hours) at the 5% significance level?

2) In fact, what the buyer is interested in is whether the lifetime is equal to or greater than the specification. Can the delivery be considered to meet this condition at the 5% significance level?

TD3 – 3:

You are hired by TOYOTA for a project to design a low-consumption car model. First of all, it is necessary to ensure that fuel consumption is the primary factor behind the purchase of this type of car. After performing a survey of 120 owners of this type of car, 40 considered that the criterion "consumption" was the decisive criterion of purchase.

1) Let *p be* the proportion of owners who favor the “consumption” criterion. You are tasked to test the following hypotheses at a given level of significance α :

$H_0: p = 0.4$
$H_1: p \neq 0.4$

Suppose that the proportion of owners who favored the criterion " consumption" within the total population is exactly 40%. What is the probability of concluding that H_0 should be rejected? Test these hypotheses by considering the level of significance $\alpha = 0.05$.

2) Build a confidence interval of 95% for the proportion of owners whose choice was motivated mainly by the fuel consumption.

3) The company claims that, for this model of cars, the average fuel consumption μ is equal to 7 liters per 100 km on urban roads (50 km/h). On a random sample of 5 cars (the population is assumed to be normal), the consumption was (in liters per 100 km): 8.1; 8.5; 8.9; 9.7; 9.8

At the threshold $\alpha = 0.05$, do you confirm the consumption announced by the manufacturer?

TD3 – 4 *:

Two groups A and B are each composed of 100 people with the same disease. A new drug was given to group A and placebo treatment to group B (referred to as a control group). It was observed that 75 patients from group A and 65 from group B were cured.

1) Test the hypothesis that the new drug is effective in the cure of the disease by considering a significance level of 1%.

2) Answer the previous question by considering groups of 300 people, and the fact that there were 225 people healed in group A and 195 in group B. Any Comments? Compare both results.

$$-\frac{n}{p} - \frac{n}{1-p}$$

$$\frac{-n}{p} - \frac{n}{1-p}$$

$$\frac{-n(1-p) - np}{}$$

TD1-1:

1° 1, 2, 4, 4, 6 ⇒ the median is 4.

2° Harry's score is x.

$$\frac{11+12+7+3+x}{5}=8 \Rightarrow x=7$$

3° Scores are: 96, 94, 92, 90, 80

① for the 4 rounds:

mean: $\bar{x}_1=\frac{96+94+92+90}{4}=93$

for with the las rounds:

mean: $\bar{x}_2=\frac{96+94+92+90+80}{5}=90.4$

② $X_{m1}=\frac{94+92}{2}=93$

$X_{m2}=92$

so the mean and the median are decreases, and the mean is decrease more.

TD1-2

1) 2, 2, 3, 5, 5, 5, 6, 6, 8, 9

① the mean = $\frac{2+2+3+5+5+5+6+6+8+9}{10} = 5.1$

② the median = 5

③ the mode = 5

TD1-3

1) the mean = $\frac{59.5+69.5\times2+79.5\times9+89.5\times22+99.5\times33+109.5\times22+119.5\times8+129.5\times2+139.5}{100}$

$= 99.3$

the standard =

$$\sqrt{\left[(59.5-99.3)^2+(69.5-99.3)^2\times2+(79.5-99.3)^2\times9+(89.5-99.3)^2\times22+(99.5-99.3)^2\times33+(109.5-99.3)^2\times22+(119.5-99.3)^2\times8+(129.5-99.3)^2\times2+(139.5-99.5)^2\right]\times\frac{1}{100}}$$

$$=\sqrt{\left[1584.04+1776.08+3528.36+2112.88+1.32+2288.88+3264.32+1824.08+1600\right]\times\frac{1}{100}}$$

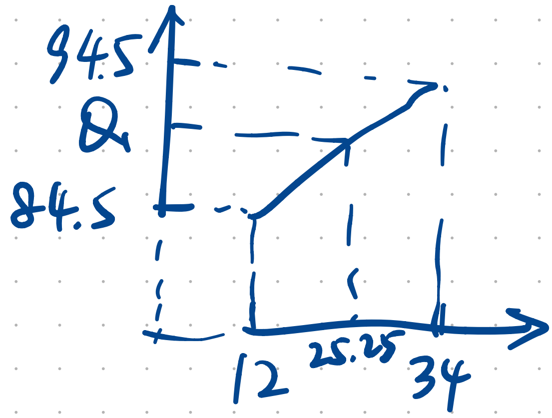
$$=\sqrt{179.7992} = 13.409$$

2) 位置 $Q_1 = (100+1)\times\frac{1}{4} = 25.25$

$Q_3 = (100+1)\times\frac{3}{4} = 75.75$

表中值为 mid-value

$\therefore Q_1 \in [84.5, 94.5]$



$$\frac{Q_1-84.5}{94.5-84.5} = \frac{25.25-12}{34-12}$$

$\Rightarrow Q_1 = 90.659 \approx 90.7$

$Q_3 \in [104.5, 114.5]$

$$\frac{Q_3-104.5}{114.5-104.5} = \frac{75.75-67}{22} \Rightarrow Q_3 = 108.5$$

semi-IR: $\frac{Q_3-Q_1}{2} = \frac{108.5-90.7}{2} = 8.9$

3) skewness 偏斜

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \begin{cases} \mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3 \\ \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \end{cases}$$

$\gamma_1 = 0$ 对称

$\gamma_1 > 0$ 右偏

$\gamma_2 < 0$ 左偏

kurtosis 峰度

$$\begin{cases} r_2 = \frac{\mu_4}{\sigma_4} - 3 \\ \mu_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4 \end{cases}$$

$r_2 > 0$ leptokurtic 高峰度 (尖)

$r_2 < 0$ platykurtic 低峰度 (偏)

$$\mu_3 = \frac{1}{100}\sum_{i=1}^{100}(x_i - \bar{x})^3$$

$$\sigma_3 = \left[\sqrt{\frac{1}{100}\sum_{i=1}^{100}(x_i - \bar{x})^2}\right]^3$$

$$r_2 = \frac{\frac{1}{100}\sum_{i=1}^{100}(x_i - \bar{x})^4}{\left[\sqrt{\frac{1}{100}\sum_{i=1}^{100}(x_i - \bar{x})^2}\right]^4} - 3$$

TD 2-1.

1) $f_X(x,\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{other} \end{cases}$

① $L(\lambda) = f(x_1) \cdot f(x_2) \cdots f(x_n)$

$= \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdots \lambda e^{-\lambda x_n}$

$= \lambda^n \cdot e^{-\lambda(x_1+x_2+\cdots+x_n)}$

$= \lambda^n \cdot e^{-\lambda \cdot \sum_{i=1}^{n} x_i}$

② $\ln L(\lambda) = n\ln\lambda + \ln e^{-\lambda \cdot \sum_{i=1}^{n} x_i}$

$= n\ln\lambda - \lambda \sum_{i=1}^{n} x_i$

③ $\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$

$\therefore \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$ ✓

$\therefore a = \frac{1}{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}$ ✓

$E(\hat{a})=a$

$I(a)$ { 代入 $a=\frac{1}{\lambda}$
求二阶倒
求期望后取负

对比

$\frac{1}{I(\hat{a})} \neq Var(\varphi)$

$$L(a)=\left(\frac{1}{a}\right)^n \cdot e^{-\frac{1}{a}\cdot\sum_{i=1}^{n}x_i}$$

$$\ln L(a)=n\ln\frac{1}{a}-\frac{1}{a}\sum_{i=1}^{n}x_i=-n\ln a-\frac{1}{a}\sum_{i=1}x_i$$

$$\frac{d}{da}\ln L(a)=-\frac{n}{a}+(a^{-2})\sum_{i=1}^{n}x_i$$

$$\frac{d}{da}\ln L'(a)=\left[-n\cdot a^{-1}+(a^{-2})\sum_{i=1}^{n}x_i\right]'$$

$$=n\cdot a^{-2}+(-2)(a^{-3})\sum_{i=1}^{n}x_i$$

$$=n\cdot a^{-2}-2a^{-3}\sum_{i=1}^{n}x_i$$

$$=\frac{n}{a^2}-\frac{2\sum_{i=1}^{n}x_i}{a^3}$$

$$I(a)=-E\left[\frac{d}{da}\ln L'(a)\right]=E\left[\frac{n}{a^2}-\frac{2\sum_{i=1}^{n}x_i}{a^3}\right]$$

$$=\left[\frac{n}{a^2}-\frac{2\cdot n\cdot \cancel{a}}{a^{\cancel{3}2}}\right]=\frac{n}{a^2}$$

$$Var(\hat{a})=\frac{1}{n\lambda^2}=\frac{a^2}{n}$$

$$\frac{1}{I(a)}=Var(\hat{a}) \Rightarrow \text{无偏}$$

$Var(\hat{a})$：X总体方差 $\frac{1}{\lambda^2}$，样本方差为 $\frac{1}{n\lambda^2}$

2) 伯努利：$P(X=k)=p^k(1-p)^{1-k}$

$E(X)=p$

① $L(p)=\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

② $\ln L(p)=\ln p^{\sum_{i=1}^{n} x_i}+\sum_{i=1}^{n}(1-x_i)\ln(1-p)$

$$=\sum_{i=1}^{n} x_i\cdot\ln p+\sum_{i=1}^{n}(1-x_i)\ln(1-p)$$

③ $\frac{d}{dp}\ln L(p)=\frac{1}{p}\sum_{i=1}^{n} x_i+(-p)'\cdot\frac{1}{1-p}\sum_{i=1}^{n}(1-x_i)$

$$=\frac{1}{p}\sum_{i=1}^{n} x_i-\frac{1}{1-p}\sum_{i=1}^{n}(1-x_i)=0$$

$\therefore p=\frac{\sum_{i=1}^{n} x_i}{n}$

④ 对 $I(a)$:

$$\frac{d}{dp}\ln' L(p) = \left[p^{-1}\sum_{i=1}^{n}x_i - (1-p)^{-1}\sum_{i=1}^{n}(1-x_i)\right]'$$

$$= -p^{-2}\sum_{i=1}^{n}x_i + (-p)'(1-p)^{-2}\sum_{i=1}^{n}(1-x_i)$$

$$= -\frac{1}{p^2}\sum_{i=1}^{n}x_i - \frac{1}{(1-p)^2}\sum_{i=1}^{n}(1-x_i)$$

$$E\left(\frac{d^2\ln L(p)}{d^2p}\right) = -\frac{1}{p^2}n\cdot p - \frac{1}{(1-p)^2}\cdot(1-p)\cdot n$$

$$= -\frac{n}{p} - \frac{n}{1-p}$$

$$= \frac{-n(1-p)-n\cdot p}{p(1-p)}$$

$$= \frac{-n}{p(1-p)}$$

取负得 $\frac{n}{p(1-p)}$

$var(p) = p(1-p) \Rightarrow$ 样本方差 $\frac{p(1-p)}{n}$

$Var(p) = \frac{1}{I(p)} \Rightarrow$ 无偏差.

TD2-3.

1) $m = \frac{72.2+72.24+72.26+72.30+72.36+72.39+72.42+72.48+72.50+72.54}{10}$

$= 72.369$.

$\sigma = \sqrt{\frac{(72.2-m)^2+(72.24-m)^2+\cdots}{9}}$

总体方差 $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$

样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

2)