第一章（旧书1，新书CH2）

一、知识点总结

1. $X_1, X_2, \cdots, X_n \sim N(\mu_1, \sigma_1^2)$

$Y_1, Y_2 \cdots Y_n \sim N(\mu_2, \sigma_2^2)$

① $A = X_1 + X_2 + \cdots + X_9 + Y_1 + Y_2$

$$\begin{cases} E(A) = 9\mu_1 + 2\mu_2 \\ Var(A) = 9\sigma_1^2 + 2\sigma_2^2 \end{cases}$$

$A \sim N(9\mu_1 + 2\mu_2,\ 9\sigma_1^2 + 2\sigma_2^2)$

② $B = \dfrac{X_1 + X_2 + \cdots + X_5}{5} + Y_1 + Y_2$

$E(B) = \frac{1}{5}(\underbrace{\mu_1 + \cdots + \mu_1}_{5个}) + \mu_2 + \mu_2$

$= \mu_1 + 2\mu_2$

$B \sim N(\mu_1 + 2\mu_2,\ \frac{1}{5}\sigma_1^2 + 2\sigma_2^2)$

$Var(B) = \frac{1}{25}(\sigma_1^2 + \cdots + \sigma_1^2) + \sigma_2^2 + \sigma_2^2$

$= \frac{1}{5}\sigma_1^2 + 2\sigma_2^2$

$$\begin{cases} E(aX + bY) = aE(X) + bE(Y) \\ Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) \end{cases}$$

$$\begin{cases} E(aX - bY) = aE(X) - bE(Y) \\ Var(aX - bY) = a^2 Var(X) + b^2 Var(Y) \end{cases}$$

二、典型例题分析.

1. **Example 4** **SKILLS** INTERPRETATION; CRITICAL THINKING

Bottles of water are delivered to shops in boxes containing 12 bottles each. The weights of bottles are normally distributed with mean weight 2 kg and standard deviation 0.05 kg. The weights of empty boxes are normally distributed with mean 2.5 kg and standard deviation 0.3 kg.

a Assuming that all random variables are independent, find the **probability** that a full box will weigh between 26 kg and 27 kg.

b Two bottles are selected at random from a box. Find the probability that they differ in weight by more than 0.1 kg.

c Find the weight m that a full box should have on its label so that there is a 1% chance that it weighs more than m.

a Let $W = X_1 + X_2 + \ldots + X_{12} + C$

$$\text{E}(W) = \text{E}(X_1 + X_2 + \ldots + X_{12} + C)$$
$$= \text{E}(X_1) + \text{E}(X_2) + \ldots + \text{E}(X_{12}) + \text{E}(C)$$
$$= 12 \times 2 + 2.5 = 26.5$$

$$\text{Var}(W) = \text{Var}(X_1 + X_2 + \ldots + X_{12} + C)$$
$$= \text{Var}(X_1) + \text{Var}(X_2) + \ldots + \text{Var}(X_{12}) + \text{Var}(C)$$
$$= 12 \times 0.05^2 + 0.3^2 = 0.12$$

$W \sim \text{N}(26.5, 0.12)$

$\text{P}(26 < W < 27)$

$$\text{P}\left(\frac{26 - 26.5}{\sqrt{0.12}} < Z < \frac{27 - 26.5}{\sqrt{0.12}}\right)$$

$\text{P}(-1.44 < Z < 1.44)$

$\text{P}(Z < 1.44) - \text{P}(Z < -1.44)$

$2\text{P}(Z < 1.44) - 1 = 2 \times 0.9251 - 1$
$= 0.850$ (3 s.f.)

b Let $Y = X_1 - X_2$

$$\text{E}(Y) = \text{E}(X_1 - X_2) = \text{E}(X_1) - \text{E}(X_2) = 0$$

$$\text{Var}(Y) = \text{Var}(X_1 - X_2) = \text{Var}(X_1) + \text{Var}(X_2) = 0.005$$

So $Y \sim \text{N}(0, 0.005)$

$$\text{P}(|Y| > 0.1) = 1 - \text{P}(-0.1 < Y < 0.1)$$
$$= 1 - (\text{P}(Y < 0.1) - \text{P}(Y < -0.1))$$
$$= 2 - 2\text{P}(Y < 0.1)$$
$$= 2 - 2\text{P}\left(Z < \frac{0.1 - 0}{\sqrt{0.005}}\right)$$
$$= 2 - 2\text{P}(Z < 1.41)$$
$$= 2 - 2 \times 0.9207$$
$$= 0.1586$$

c Find m such that $\text{P}(W > m) = 0.01$

So $\text{P}(W < m) = 0.99$

$$\text{P}\left(Z < \frac{m - 26.5}{\sqrt{0.12}}\right) = 0.99$$

$$m - \frac{26.5}{\sqrt{0.12}} = 2.3263$$

$$m = 2.3263 \times \sqrt{0.12} + 26.5$$
$= 27.3$ kg (3 s.f.)

2.

5. (i) The volume, B ml, in a bottle of *Burxton's* water has a normal distribution $B \sim \text{N}(325, 6^2)$ and the volume, H ml, in a bottle of *Hargate's* water has a normal distribution $H \sim \text{N}(330, 4^2)$.
Rebecca buys 5 bottles of *Burxton's* water and one bottle of *Hargate's* water.

Find the probability that the total volume in the 5 bottles of *Burxton's* water is more than 5 times the volume in the bottle of *Hargate's* water.

(5)

(ii) Two independent random samples X_1, X_2, X_3, X_4, X_5 and Y_1, Y_2, Y_3, Y_4, Y_5 are each taken from a normal population with mean μ and standard deviation σ.

(a) Find the distribution of the random variable $D = Y_1 - \overline{X}$

(3)

(b) Hence show that $\text{P}(Y_1 > \overline{X} + \sigma) = 0.181$ correct to 3 decimal places.

(2)

Ankit believes that $\text{P}(U_1 > \overline{U} + \sigma) = 0.181$ correct to 3 decimal places, for **any** random sample U_1, U_2, U_3, U_4, U_5 taken from a normal population with mean μ and standard deviation σ.

(c) Explain briefly why the result from part (b) should not be used to confirm Ankit's belief.

(1)

(d) Find, correct to 3 decimal places, the actual value of $\text{P}(U_1 > \overline{U} + \sigma)$.

(6)

Sol:

5 (i)	Let $R = B_1 + B_2 + B_3 + B_4 + B_5 - 5H$ so $E(R) = -25$ (o.e.)	B1
	$\text{Var}(R) = 5\times 6^2 + 5^2 \times 4^2$ $\quad R \sim N(-25, \sqrt{580}^2)$	M1A1
	$P(R > 0) = P(Z > \frac{0--25}{\sqrt{580}}) = P(Z > 1.04)$, $= 0.149619\ldots$(calc) or 0.1492 (tables)	dM1 A1 (5)
(ii)(a)	$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{5}\right)$	B1
	$\text{Var}(D) = \sigma^2 + "\frac{\sigma^2}{5}"\left[= \frac{6\sigma^2}{5}\right]$, so $\quad D \sim N\left(0, \frac{6\sigma^2}{5}\right)$	M1, A1 (3)
(b)	$P(Y_1 > \bar{X} + \sigma) = P(D > \sigma) = P\left(Z > \frac{\sigma}{\sqrt{\frac{6}{5}}\sigma}\right)$	M1
	$= P(Z > 0.912\ldots) = 0.181$ (3 dp) (*)	A1cso (2)
(c)	Since U_1 and $\bar{U}$ are not independent (so variance formula cannot be used)	B1
	Can be implied e.g. U_1 used to calculate $\bar{U}$, U_1 and $\bar{U}$ from same sample o.e.	(1)
(d)	Let $F = U_1 - \bar{U} = U_1 - \frac{(U_1+U_2+U_3+U_4+U_5)}{5}, = \frac{4U_1 - (U_2+U_3+U_4+U_5)}{5}$	M1, A1
	$\text{Var}(F) = \frac{4^2\sigma^2 + 4\sigma^2}{5^2} = 0.8\sigma^2$,so $\quad F \sim N(0, 0.8\sigma^2)$	dM1, A1
	$P(F > \sigma) = P\left(Z > \frac{\sigma}{\sigma\sqrt{0.8}}\right) = P(Z > 1.118..)$	M1
	$= 0.1314$ (tables) or $0.131776\ldots$(calc) **awrt 0.131~0.132**	A1cso (6)

第二章 sampling （新书CH1，旧书CH2）

一、知识点总结

1. 抽样过程（会背，用题目中的关键词替换术语）

Sampling method	PROCEDURE
Simple random sampling	➤ Label each member of the population 1 to n; ➤ Use the random number generator to generate the numbers.
Systematic sampling	➤ Make a list of every term (eg: 001-400); ➤ Calculate the length of regular interval $k = \frac{n}{N(\text{样本量})}$; ➤ Use the random number table (generator) to select a number between $1 \sim k$; ➤ Select every kth member.
Stratified sampling	➤ Label every strata(001-100， 001-150); ➤ Calculate the numbers of sample in every strata; ➤ Use the random number table (generator) to select k_1 k_2 values.

2. 优劣分析

Sampling method	advantage	disadvantage
Simple random sampling	➤ Free of bias; ➤ Easy and inexpensive to implement for small populations and small samples	➤ A sampling frame is needed; ➤ Not suitable when the population size or the sample size is large.
Systematic sampling	➤ Simple and easy to use; ➤ Suitable for large samples and large populations.	➤ A sampling frame is needed; ➤ It can introduce bias if the sampling frame is not random.
Stratified sampling	➤ Sample accurately reflects the population structure.	➤ Population must be clearly classified into separate strata; ➤ A sampling frame is needed.

3、Quota sampling — non-random sampling 最大特点：非随机

How to ···
1. Calculate the numbers of sample···
直接计算每个Quota的样本量
2. Survey until the quota reached; ← 抄写这句话
(具体看题目)
3. If the quota is full, the other is ignore.

Quota sampling	
Advantages	**Disadvantages**
• Allows a small sample to still be representative of the population • No sampling frame required • Quick, easy and inexpensive • Allows for easy comparison between different groups within a population	• Non-random sampling can introduce bias • Population must be divided into groups, which can be costly or inaccurate • Non-responses are not recorded

第三章：Estimate, confidence interval, test

分三个模块：① 估计 ② CI ③ test { 1. 2. 3.

第一部分：Estimate

一、知识点总结

1. Statistics (统计量) $T = f(X_1, X_2, \cdots X_n)$

unbiased estimator: $E(T) = \theta$ (θ - population parameter)

$$bias = E(T) - \theta$$

11 The random variable $X \sim U[-\alpha, \alpha]$.

a Find $E(X)$ and $E(X^2)$.

A random sample X_1, X_2, X_3 is taken and the statistic $Y = X_1^2 + X_2^2 + X_3^2$ is calculated.

b Show that Y is an unbiased estimator of α^2.

sol: a) $E(X) = \frac{-\alpha+\alpha}{2} = 0$ $\quad Var(X) = \frac{(\alpha+\alpha)^2}{12} = \frac{1}{3}\alpha^2$

b) $Var(X) = E(X^2) - (E(X))^2$ $\quad E(X^2) = \frac{\alpha^2}{3}$

$Y = X_1^2 + X_2^2 + X_3^2$

$E(Y) = E(X_1^2) + E(X_2^2) + E(X_3^2) = \frac{\alpha^2}{3} + \frac{\alpha^2}{3} + \frac{\alpha^2}{3} = \alpha^2$

So Y is an unbiased estimator of α^2

2. 特殊的无偏估计量 $\begin{cases} \bar{X} = \frac{\Sigma X}{n} \\ S^2 = \frac{\Sigma X^2 - n\bar{X}^2}{n-1} \end{cases}$

$\begin{cases} E(\bar{X}) = \mu \\ Var(\bar{X}) = \frac{\sigma^2}{n} \end{cases}$

standard error $= \frac{\sigma}{\sqrt{n}}$ or $\frac{S}{\sqrt{n}}$

$\Downarrow$

estimate the size of sample required to achieve a standard error of less than 0.15.

第二部分：confidence interval : CI

一、知识点总结

1、定义

- A confidence interval (C.I.) for a population parameter θ is a range of values defined so that there is a specific probability that the true value of the parameter lies within that range.

2、

confidence level	CI	width
90%	$\bar{X} \pm 1.6449 \cdot \frac{\sigma}{\sqrt{n}}$	$2 \times 1.6449 \cdot \frac{\sigma}{\sqrt{n}}$
95%	$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$	$2 \times 1.96 \cdot \frac{\sigma}{\sqrt{n}}$
99%	$\bar{X} \pm 2.5758 \frac{\sigma}{\sqrt{n}}$	$2 \times 2.5758 \frac{\sigma}{\sqrt{n}}$

$\Downarrow$ 可根据 C_1, C_2 求 σ

$\Downarrow$ $\bar{X}$未知时 用 width 求 σ

3、CI 与 Hypothesis test

90% CI $\left(\bar{x} - 1.6449 \cdot \frac{\sigma}{\sqrt{n}} , \quad \bar{x} + 1.6449 \frac{\sigma}{\sqrt{n}} \right)$

$\Leftarrow$ test: $H_0: \mu = m$ $\quad H_1: \mu \neq m$

level of significance $= 1 - 90\% = 10\%$

if $\mu \in$ CI, accept H_0

if $\mu \notin$ CI. reject H_1

Example 11 SKILLS PROBLEM-SOLVING

A random sample of size 25 is taken from a normal distribution with standard deviation 2.5. The mean of the sample is 17.8.

a Find a 99% C.I. for the population mean μ.

b What size sample is required to obtain a 99% C.I. with a width of at most 1.5?

c What confidence level would be associated with the interval based on the above sample of 25 but of width 1.5, i.e. (17.05, 18.55)?

a 99% confidence limits are:

$\bar{x} \pm 2.5758 \times \frac{\sigma}{\sqrt{n}} = 17.8 \pm 2.5758 \times \frac{2.5}{\sqrt{25}}$

So a 99% confidence interval is (16.51, 19.09)

b Width of 99% C.I. is $2 \times 2.5758 \times \frac{2.5}{\sqrt{n}}$

so you require $1.5 > \frac{12.879...}{\sqrt{n}}$

i.e. $n > 73.719...$

so you need $n = 74$

c A width of 1.5 $\Rightarrow 1.5 = 2 \times z \times \frac{2.5}{\sqrt{25}}$

$z = 1.5$

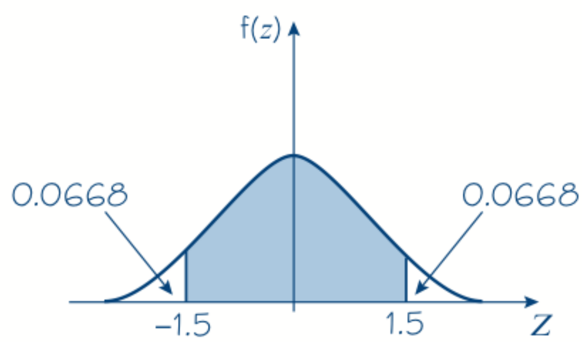
From the table on page 135 you find that

$P(Z < 1.5) = 0.9332$

and so $P(Z > 1.5) = P(Z < -1.5)$

$= 1 - 0.9332$

$= 0.0668$



So the confidence level is

$100 \times (1 - 2 \times 0.0668) = 86.6\%$

第三部分：test 假设检验 (在正态分布中找 CR)

- The central limit theorem says that if $X_1, X_2, \ldots, X_n$ is a random sample of size n from a population with mean μ and variance σ^2, then $\bar{X}$ is approximately $\sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

1, 单变量的假设检验

$H_0: \mu = m \qquad H_1: \mu \neq m \; (\alpha/2) \;/\; \mu > m \; (\alpha) \;/\; \mu < m \; (\alpha)$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = a \qquad CR - \text{critical region}$$

if $a \in CR$ reject H_0

$a \notin CR$ accept H_0

☆ 作答

Ans:

1. (14R.4)

A manufacturing company produces solar panels. The output of each solar panel is normally distributed with standard deviation 6 watts. It is thought that the mean output, μ, is 160 watts.

A researcher believes that the mean output of the solar panels is greater than 160 watts. He writes down the output values of 5 randomly selected solar panels. He uses the data to carry out a hypothesis test at the 5% level of significance.

He tests $H_0: \mu = 160$ against $H_1: \mu > 160$

On reporting to his manager, the researcher can only find 4 of the output values. These are shown below

168.2 157.4 173.3 161.1

Given that the result of the hypothesis test is that there is significant evidence to reject H_0 at the 5% level of significance, calculate the minimum possible missing output value, α. Give your answer correct to 1 decimal place.

(6)

Test statistic, $z = \frac{132+\frac{\alpha}{5}-160}{\frac{6}{\sqrt{5}}}$	M1A1ft
Critical z values is 1.6449	B1
Therefore the test statistic is significant if	
$\frac{132+\frac{\alpha}{5}-160}{\frac{6}{\sqrt{5}}} > 1.6449$	M1
Therefore	
$132+\frac{\alpha}{5}-160 > 1.6449\times\frac{6}{\sqrt{5}}$	
$\alpha > 5\left(1.6449\times\frac{6}{\sqrt{5}}+28\right)$	
$\alpha > 162.0686493\ldots$	A1
Accept awrt 162.1	
	(6)

2, 双变量的假设检验 X, Y

$H_0: \mu_1 - \mu_2 = m \qquad H_1: \mu_1 - \mu_2 \neq m \; (\alpha/2) \qquad \mu_1 - \mu_2 > m \; (\alpha) \qquad \mu_1 - \mu_2 < m \; (\alpha)$

$$z = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = a$$

if $a \in CR$ reject H_0

$a \notin CR$ accept H_0

☆ 必须根据题意再作答

Merchandise is sold at concerts. The manager of a concert claims that the mean value of merchandise sold to premium ticket holders is more than £6 greater than the mean value of merchandise sold to standard ticket holders.

(a) Given that all the tickets for the next concert have been sold, describe how a stratified sample should be taken at the concert.

The mean value of merchandise sold to a random sample of 60 standard ticket holders at the concert is £15 with a standard deviation of £10.

The mean value of merchandise sold to a random sample of 55 premium ticket holders at the concert is £23 with a standard deviation of £8.

(b) Test the manager's claim at the 5% level of significance. State your hypotheses clearly.

(8)

(c) For the test in part (b), state whether or not it is necessary to assume that values of merchandise sold have normal distributions. Give a reason for your answer.

(2)

Ans :

(a)	**Record / List** all ticket numbers of **standard** and **premium** tickets	B1
	Use **random numbers** to select a sample of standard and a sample of premium ticket holders i.e. **within strata.**	B1
	Sample sizes in proportion to the no of standard and no of premium ticket holders at the concert.	B1
(b)	$\mathrm{H}_0 : \mu_p - \mu_s = 6$ oe [p = premium s = standard]	B1
	$\mathrm{H}_1 : \mu_p - \mu_s > 6$ oe	B1
	Standard error = $\sqrt{\frac{10^2}{60} + \frac{8^2}{55}} = \left[\sqrt{2.83030...}\right] = [1.682\ldots]$	M1
	$z = \dfrac{\pm(23-15-6)}{"\sqrt{\frac{10^2}{60}+\frac{8^2}{55}}"}$	dM1
	$= \pm 1.1888\ldots$ awrt ± 1.19	A1
	cv 5% one tailed = 1.6449	B1
	Not significant, insufficient evidence to reject H_0	dM1
	Insufficient evidence to support the **manager's claim** <u>or</u> the mean value of merchandise sold to premium ticket holders is NOT more than £6 greater than the mean value of merchandise sold to standard ticket holders.	A1cso
(c)	**Sample size is large** so **Central Limit Theorem (CLT) applies** so	B1
	do not need to assume merchandise sold has a normal distribution.	dB1

3, table — large sample

If the population is not normal, by assuming that s is a close approximation to σ, then for large samples, $\dfrac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}}$ can be treated as having an approximate N(0, 1^2) distribution.

一般是双变量的假设检验（注意关键词，并判定单边 or 双边）

$H_0: \mu_1 - \mu_2 = m$ $H_1: \mu_1 - \mu_2 \neq m$ $\mu_1 - \mu_2 > 0$ $\mu_1 - \mu_2 < 0$

Example 10 SKILLS REASONING/ARGUMENTATION

As part of a study of the health of young children, a random sample of 220 children from area A and a second, independent random sample of 180 children from area B were weighed.
The results are given in the table below.

	n	$\bar{x}$	s
Area A	220	37.8	3.6
Area B	180	38.6	4.1

a Test, at the 5% level of significance, whether there is evidence of a difference in the mean weight of children in the two areas. State your hypotheses clearly.

b State an assumption you have made in carrying out this test.

c Explain the significance of the central limit theorem to this test.

a $H_0: \mu_A = \mu_B \qquad H_1: \mu_A \neq \mu_B$

Test statistic $z = \dfrac{\bar{x}_A - \bar{x}_B - (\mu_A - \mu_B)}{\sqrt{\dfrac{\sigma_A^2}{n_A} + \dfrac{\sigma_B^2}{n_B}}}$

$$z = \frac{38.6 - 37.8}{\sqrt{\dfrac{3.6^2}{220} + \dfrac{4.1^2}{180}}}$$

$z = 2.0499\ldots$
$= 2.05$ (3 s.f.)

Two-tail 5% critical values are $z = \pm 1.96$

Since $2.05 > 1.96$, the result is significant so reject H_0.

There is evidence that the mean weight of children in the two areas is different.

b The test statistic requires σ so you have to assume that $s^2 = \sigma^2$ for both samples.

c You are not told that the populations are normally distributed but the samples are both large and so the central limit theorem enables us to assume that $\bar{X}_A$ and $\bar{X}_B$ are both normal.

第四章 Goodness of fit & Contingency tables

本章重点为假设检验，需要求解期望值，自由度，需要查找卡方分布查找拒绝域。