Statistical Inference

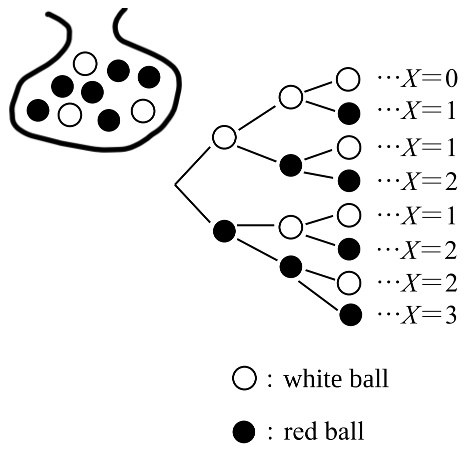
Hereafter, all root events are assumed to be equally certain.

1

When 3 balls are taken out of a bag containing 6 red balls and 3 white balls, find the probability distribution of the number X of red balls taken out. Also, find the probability $P(X \leqq 2)$.

solution

The way to take out the balls is as shown in the tree diagram on the right, and the possible values of X are 0, 1, 2, and 3.



$$P(X=0)=\frac{{}_3C_3}{{}_9C_3}=\frac{1}{\frac{9\cdot 8\cdot 7}{3\cdot 2\cdot 1}}=\frac{1}{84}$$

$$P(X=1)=\frac{{}_6C_1\cdot{}_3C_2}{{}_9C_3}=\frac{6\cdot\frac{3\cdot 2}{2\cdot 1}}{\frac{9\cdot 8\cdot 7}{3\cdot 2\cdot 1}}=\frac{18}{84}$$

$$P(X=2)=\frac{{}_6C_2\cdot{}_3C_1}{{}_9C_3}=\frac{\frac{6\cdot 5}{2\cdot 1}\cdot 3}{\frac{9\cdot 8\cdot 7}{3\cdot 2\cdot 1}}=\frac{45}{84}$$

$$P(X=3)=\frac{{}_6C_3}{{}_9C_3}=\frac{\frac{6\cdot 5\cdot 4}{3\cdot 2\cdot 1}}{\frac{9\cdot 8\cdot 7}{3\cdot 2\cdot 1}}=\frac{20}{84}$$

The probability distribution of X is shown in the table on the right.

X	0	1	2	3	total
P	$\frac{1}{84}$	$\frac{18}{84}$	$\frac{45}{84}$	$\frac{20}{84}$	1

Also, $P(X \leqq 2)=\frac{1}{84}+\frac{18}{84}+\frac{45}{84}=\frac{64}{84}=\mathbf{\frac{16}{21}}$.

2

There are 30 lots with prizes as shown in the table on the right.
Find the average of the prize money when 1 of these lots is drawn.

	prize	number of lots
1st class	10000 yen	1 lot
2nd class	1000 yen	4 lots
3rd class	100 yen	25 lots

solution

If the prize for 1 draw of this lottery is X yen,
the probability distribution of the random variable X
is as shown in the table on the right.

X	10000	1000	100	total
P	$\frac{1}{30}$	$\frac{4}{30}$	$\frac{25}{30}$	1

Thus, the average $E(X)$ of the prize money is

$$E(X) = 10000 \times \frac{1}{30} + 1000 \times \frac{4}{30} + 100 \times \frac{25}{30} = \mathbf{550(yen)}\ .$$

3

When 3 balls are removed from a bag containing 6 red balls and 3 white balls, find the mean, variance, and standard deviation of the number *X* of red balls removed.

solution

1, the probability distribution of *X* is as shown in the table on the right.

X	0	1	2	3	total
P	$\frac{1}{84}$	$\frac{18}{84}$	$\frac{45}{84}$	$\frac{20}{84}$	1

Therefore, **the average *E*(*X*) of *X*** is

$$E(X) = 0 \times \frac{1}{84} + 1 \times \frac{18}{84} + 2 \times \frac{45}{84} + 3 \times \frac{20}{84}$$

$$= \frac{0 + 18 + 90 + 60}{84} = \mathbf{2}\ (\textbf{pieces}).$$

The variance *V*(*X*) of *X* is

$$V(X) = E(X^2) - \{E(X)\}^2 = 0^2 \times \frac{1}{84} + 1^2 \times \frac{18}{84} + 2^2 \times \frac{45}{84} + 3^2 \times \frac{20}{84} - 2^2$$

$$= \frac{0 + 18 + 180 + 180}{84} - 4 = \frac{\mathbf{1}}{\mathbf{2}}.$$

The standard deviation $\boldsymbol{\sigma(X)}$ of *X* is $\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{1}{2}} = \frac{\sqrt{\mathbf{2}}}{\mathbf{2}}$ **(pieces)**.

Alternative solution for the variance *V*(*X*)

$$V(X) = E((X - m)^2) = (0-2)^2 \times \frac{1}{84} + (1-2)^2 \times \frac{18}{84} + (2-2)^2 \times \frac{45}{84} + (3-2)^2 \times \frac{20}{84}$$

$$= \frac{4 + 18 + 0 + 20}{84} = \frac{1}{2}$$

4

Let X be the number of times the front shows up when 1 coin is tossed 4 times. Find the mean, variance and standard deviation of the random variable X.

Also, find the mean, variance and standard deviation of the random variable Y defined by $Y=3X+2$.

solution

The random variable X follows a probability distribution as shown in the table on the right.

X	0	1	2	3	4	total
P	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{6}$	1

$P(X=0)=\left(\frac{1}{2}\right)^4$

$P(X=1)={}_4C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^3$

$P(X=2)={}_4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$

$P(X=3)={}_4C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)$

$P(X=4)={}_4C_4\left(\frac{1}{2}\right)^4$

Therefore, **the average $E(X)$ of X** is

$$E(X)=0\times\frac{1}{16}+1\times\frac{4}{16}+2\times\frac{6}{16}+3\times\frac{4}{16}+4\times\frac{1}{16}=\frac{32}{16}=\mathbf{2}.$$

The variance $V(X)$ of X is

$$V(X)=E(X^2)-\{E(X)\}^2$$

$$=0^2\times\frac{1}{16}+1^2\times\frac{4}{16}+2^2\times\frac{6}{16}+3^2\times\frac{4}{16}+4^2\times\frac{1}{16}-2^2$$

$$=\frac{0+4+24+36+16}{16}-4=\frac{80}{16}-4=\mathbf{1}.$$

The standard deviation $\sigma(X)$ of X is $\sigma(X)=\sqrt{V(X)}=\sqrt{1}=\mathbf{1}$.

Also, **the average $E(Y)$ of Y** is

$$E(Y)=E(3X+2)=3E(X)+2=3\cdot2+2=\mathbf{8}.$$

The variance $V(Y)$ of Y is $V(Y)=V(3X+2)=3^2V(X)=9\cdot1=\mathbf{9}$.

The standard deviation $\sigma(Y)$ of Y is $\sigma(Y)=\sigma(3X+2)=|\,3\,|\,\sigma(X)=3\cdot1=\mathbf{3}$.

5

Let X be the number of times that a die is thrown 4 times repeatedly, and the number of times that a die with a roll of 2 or less is thrown. Find the probability distribution of X.

Also, find the probability that a die with 2 or less will be thrown 3 or more times.

solution

The probability that the roll of the die is 2 or less in 1 trial is $\frac{2}{6}=\frac{1}{3}$.

Therefore, the random variable X follows a binomial distribution $B\left(4,\ \frac{1}{3}\right)$.

Since the probability that a die with a roll of 2 or less is thrown r times is

$$P(X=r)={}_4\mathrm{C}_r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{4-r}\qquad (r=0,\ 1,\ 2,\ 3,\ 4),$$

the probability distribution of X is as shown in the following table.

X	0	1	2	3	4	total
P	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{24}{81}$	$\frac{8}{81}$	$\frac{1}{81}$	1

Also, the probability $P(X\geqq 3)$ that a die with a roll of 2 or less will appear 3 or more times is

$$P(X\geqq 3)=\frac{8}{81}+\frac{1}{81}=\frac{9}{81}=\mathbf{\frac{1}{9}}.$$

6

Find the mean, variance, and standard deviation of the number of times X that a 5 appears when 1 die is thrown 360 times.

solution

The random variable X follows a binomial distribution $B\left(360,\ \frac{1}{6}\right)$.

Therefore, **the average $E(X)$ for X** is $E(X) = 360 \cdot \frac{1}{6} = \mathbf{60}$ **(times)**.

The variance $V(X)$ of X is $V(X) = 360 \cdot \frac{1}{6} \cdot \frac{5}{6} = \mathbf{50}$.

The standard deviation $\sigma(X)$ of X is $\sigma(X) = \sqrt{50} = \mathbf{5\sqrt{2}}$ **(times)**.

7

If the range of possible values of x for the random variable X is $0 \leqq x \leqq 2$ and

its probability density function is expressed as $f(x) = \frac{1}{2}x \ (0 \leqq x \leqq 2)$,

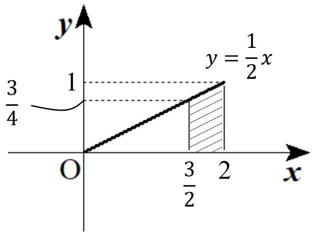
find the probability $P\left(\frac{3}{2} \leqq X \leqq 2\right)$.

solution

From the figure on the right, it is

$$P\left(\frac{3}{2} \leqq X \leqq 2\right) = \frac{1}{2}\left(\frac{3}{4} + 1\right) \cdot \left(2 - \frac{3}{2}\right)$$

$$= \mathbf{\frac{7}{16}}.$$



8

Answer the following questions.

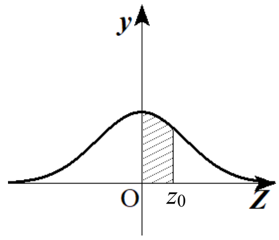
(1) If the random variable Z follows a standard normal distribution $N(0,\ 1)$, find the probability that

① $P(-1 \leqq Z \leqq 1)$

② $P(Z \geqq -0.5)$

(2) Find $P(2 \leqq X \leqq 9)$ when the random variable X follows a normal distribution $N(1,\ 4^2)$.

(3) When 100 white-fronted butterflies were collected, the mean body length was 19.6 mm, and the standard deviation was 0.5 mm. If the body lengths of the butterflies follow a normal distribution, how many of them are more than 20 mm?



normal distribution table

z_0	0	~	3	~	5	6	7	8
0.2	0.0793		0.0910		0.0987	0.1026	0.1064	0.1103
~								
0.5	0.1915		0.2109		0.2088	0.2123	0.2157	0.2190
~								
0.8	0.2881		0.2967		0.3023	0.3051	0.3078	0.3106
~								
1.0	0.3413		0.3485		0.3531	0.3554	0.3577	0.3599
~								
1.2	0.3849		0.3907		0.3944	0.3962	0.3980	0.3997
~								
1.6	0.4452		0.4484		0.4505	0.4515	0.4525	0.4535
~								
1.9	0.4713		0.4732		0.4744	0.4750	0.4756	0.4761
2.0	0.4772		0.4788		0.4798	0.4803	0.4808	0.4812
~								
2.3	0.4893		0.4901		0.4906	0.4909	0.4911	0.4913
~								
2.5	0.4938		0.4943		0.4946	0.4948	0.4949	0.4951

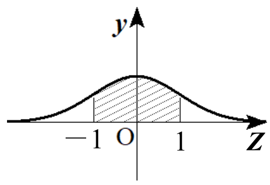
The table above is an excerpt of only those portions relevant to this section.

solution

(1) ① $P(-1 \leqq Z \leqq 1) = 2 \times P(0 \leqq Z \leqq 1)$

$= 2 \times 0.3413$

$= \mathbf{0.6826}$

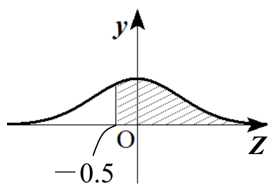


② $P(Z \geqq -0.5) = P(-0.5 \leqq Z \leqq 0) + P(Z \geqq 0)$

$= P(0 \leqq Z \leqq 0.5) + P(Z \leqq 0)$

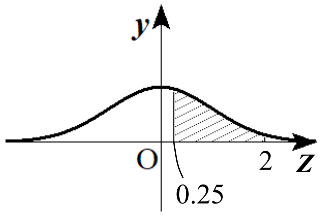
$= 0.1915 + 0.5$

$= \mathbf{0.6915}$



(2) If $Z = \dfrac{X-1}{4}$, the random variable Z follows $N(0,\ 1)$.

Since $Z = \dfrac{2-1}{4} = 0.25$ when $X = 2$ and $Z = \dfrac{9-1}{4} = 2$ when $X = 9$,



$P(2 \leqq X \leqq 9) = P(0.25 \leqq Z \leqq 2)$

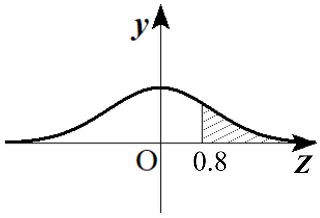
$= P(0 \leqq Z \leqq 2) - P(0 \leqq Z \leqq 0.25)$

$= 0.4772 - 0.0987$

$=$ **0.3785** .

(3) If the body length is X mm , X follows $N(19.6,\ 0.5^2)$, so let $Z = \dfrac{X-19.6}{0.5}$ and Z follows $N(0,\ 1)$.

When $X = 20$, $Z = \dfrac{20-19.6}{0.5} = 0.8$, so



$P(X \geqq 20) = P(Z \geqq 0.8)$

$= P(Z \geqq 0) - P(0 \leqq Z \leqq 0.8)$

$= 0.5 - 0.2881$

$= 0.2119$.

Therefore, the number of white-fronted butterflies with a body length of 20 mm or more is

$100 \times 0.2119 = 21.19$.

Thus, there are approximately **21 butterflies**.

9

There is a person who has a 96% success rate in Kendama.

Find the probability that the number of successes is less than 140 when this person plays Kendama 150 times.

solution

Let X be the number of times the Kendama succeeds, X follows a binomial distribution $B\left(150,\ \frac{24}{25}\right)$.

Therefore, the mean m of X is $m = 150 \cdot \frac{24}{25} = 144$ (times)

and the standard deviation σ is $\sigma = \sqrt{150 \cdot \frac{24}{25}\left(1 - \frac{24}{25}\right)} = \sqrt{\frac{144}{25}} = \frac{12}{5}$ (times) .

Since 150 is sufficiently large, we may assume that X follows a normal distribution $N(144,\ 2.4^2)$.

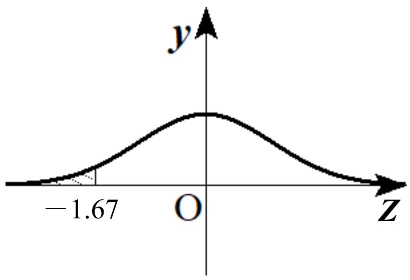
From this, by setting $Z = \frac{X - 144}{2.4}$, we can assume that Z follows a standard normal distribution $N(0,\ 1)$.

From the above, the probability to be found is

$$
\begin{aligned}
P(X \leqq 140) &= P(Z \leqq -1.67) \\
&= P(Z \geqq 0) - P(0 \leqq Z \leqq 1.67) \\
&= 0.5 - 0.4525 \\
&= \mathbf{0.0475}\ .
\end{aligned}
$$

When $X = 140$,

$Z = \frac{140 - 144}{2.4} = -\frac{5}{3} \fallingdotseq -1.67$.



１０

Answer the following questions.

(1) Let X_1, X_2, and X_3 be the rolls of a single die thrown 3 times, and let $\bar{X}=\dfrac{X_1+X_2+X_3}{3}$ be the average of the rolls. Find the mean $E(\bar{X})$ and standard deviation $\sigma(\bar{X})$ of $\bar{X}$.

(2) It is known that sweet potatoes from a certain field follow a normal distribution with mean 270g and standard deviation 30g.
Find the probability that the sample mean $\bar{X}$ is less than or equal to 264g when 36 randomly selected.

solution

(1) The probability distribution of the roll X obtained by throwing 1 die once is shown in the table on the right.

X	1	2	3	4	5	6	total
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Therefore, the average $E(X)$ of X is

$$E(X)=1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{21}{6}=\frac{7}{2}.$$

The standard deviation $\sigma(X)$ of X is

$$\sigma(X)=\sqrt{E(X^2)-\{E(X)\}^2}=\sqrt{1^2\times\frac{1}{6}+2^2\times\frac{1}{6}+3^2\times\frac{1}{6}+4^2\times\frac{1}{6}+5^2\times\frac{1}{6}+6^2\times\frac{1}{6}-\left(\frac{7}{2}\right)^2}$$

$$=\sqrt{\frac{91}{6}-\frac{49}{4}}=\sqrt{\frac{182-147}{12}}=\sqrt{\frac{35}{12}}=\frac{\sqrt{105}}{6}.$$

Thus, the mean $E(\bar{X})$ of the sample mean $\bar{X}$ for a sample of size 3 is $\boldsymbol{E(\bar{X})}=(\text{population mean})=\boldsymbol{\frac{7}{2}}$.

The standard deviation $\sigma(\bar{X})$ is $\boldsymbol{\sigma(\bar{X})}=\dfrac{\sigma(X)}{\sqrt{3}}=\dfrac{1}{\sqrt{3}}\cdot\dfrac{\sqrt{105}}{6}=\boldsymbol{\dfrac{\sqrt{35}}{6}}$.

(2) Since the population distribution is normal distribution $N(270,\ 30^2)$, the sample mean $\bar{X}$ follows a normal distribution $N\left(270,\ \dfrac{30^2}{36}\right)$. Therefore, the standardized random variable

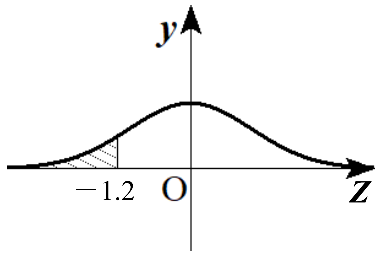
$$Z=\frac{\bar{X}-270}{\frac{30}{\sqrt{36}}}=\frac{\bar{X}-270}{5}$$ follows a standard normal distribution $N(0,\ 1)$.

Since $Z=-1.2$ when $\bar{X}=264$, the probability of finding is

$P(\bar{X}\leqq 264)=P(Z\leqq -1.2)=P(Z\geqq 0)-P(0\leqq Z\leqq 1.2)$

$=0.5-0.3849$

$=\mathbf{0.1151}$.



1 1

Answer the following questions.

(1) It is known that the body lengths of ayu migrating up a certain river follow a normal distribution with a population standard deviation of 2.0cm, although the mean body length differs from year to year. In one year, we examined the length of 16 ayu in this river and found that the sample mean was 8.0cm. Estimate the mean length m of ayu at 95% confidence level.

(2) We randomly selected 100 people who had won a prize in a crane game and asked them how much they spent. The sample mean was 1000 yen and the sample standard deviation was 500 yen.
Estimate the average cost of taking a prize in the crane game at the 95% confidence level.

solution

(1) Since the population standard deviation $\sigma=2.0$, the sample size $n=16$, and the sample mean $\bar{x}=8.0$,

the 95% confidence interval for m is $$8.0-1.96\times\frac{2.0}{\sqrt{16}}\leqq m\leqq 8.0+1.96\times\frac{2.0}{\sqrt{16}}.$$

That is, $7.02\leqq m\leqq 8.98$.

Thus, the average length of ayu in that year can be estimated to be **between 7.0cm and 9.0cm**.

(2) Let m be the average cost.

Since the sample size of 100 is large, we use the sample standard deviation $s=500$ instead of the population standard deviation σ.

Since the sample mean $\bar{x}=1000$, the 95% confidence interval of m is

$$1000-1.96\times\frac{500}{\sqrt{100}}\leqq m\leqq 1000+1.96\times\frac{500}{\sqrt{100}}.$$ That is, $902\leqq m\leqq 1098$.

Thus, we can estimate that the average cost of taking a prize in the crane game is **between 902 yen and 1098 yen**.

１２

When a high school basketball player's free-throw record was randomly sampled for 144 free throws, he was successful 92 times.

Estimate the free-throw success rate of the basketball player at a 95% confidence level for the interval.

solution

The sample proportion $\bar{p}$ is $\dfrac{92}{144} \approx 0.64$, so the 95% confidence interval for the population proportion p is

$$0.64 - 1.96\sqrt{\frac{0.64 \times 0.36}{144}} \leqq p \leqq 0.64 + 1.96\sqrt{\frac{0.64 \times 0.36}{144}}.$$

That is, $0.5616 \leqq p \leqq 0.7184$.

Thus, the free throw success rate of the basketball player can be estimated to be **between 56% and 72%**.

１３

A 10-item nationwide disaster prevention questionnaire was conducted, in which respondents were asked whether they had prepared emergency rations and other items on a scale of ○ or ×. The average number of ○ was 7.00 and the standard deviation was 1.00. When this questionnaire was administered to 400 randomly selected households in one city, the average number of ○ was 6.88. In this case, answer the following questions.

(1) Can the city's results be considered comparable to the national level? Test at 5% level of significance. Also, test at the 1% level of significance.

(2) The mayor of one city has a stronger interest in whether these results are less than the national ones. Can we conclude that a city's result is less than the national result? Test at 1% level of significance.

solution

(1) Hypothesis is that the results are comparable to the national level.

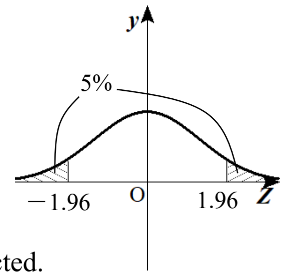
Let X be the mean of ○, X follows a normal distribution $N(7.00,\ 1.00^2)$, so the sample mean $\bar{X}$ follows $N\left(7.00,\ \frac{1.00^2}{400}\right)$.

In detail, the alternative hypothesis is that the results are not comparable to the national level, and the null hypothesis is that the results are comparable to the national level.

Let $Z=\frac{\bar{X}-7.00}{\frac{1.00}{\sqrt{400}}}$, then the random variable Z follows a standard normal distribution $N(0,\ 1)$.

Since the rejection region at the 5% significance level is $Z \leqq -1.96$, $1.96 \leqq Z$,

$$\bar{X} \leqq 7.00-1.96\times\frac{1.00}{\sqrt{400}},\qquad 7.00+1.96\times\frac{1.00}{\sqrt{400}} \leqq \bar{X}.$$



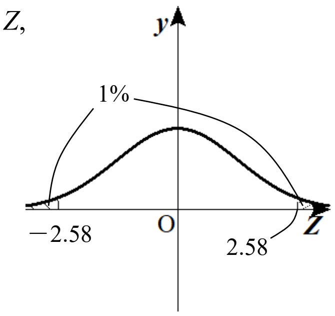
That is, $\bar{X} \leqq 6.902$, $7.098 \leqq \bar{X}$.

The observed sample mean $\bar{x}=6.88$ falls within this interval, so the hypothesis is rejected.

Therefore, at the **5% level of significance**, we can conclude that the results for **one city is not comparable to the national level**.

Also, since the rejection region at the 1% significance level is $Z \leqq -2.58$, $2.58 \leqq Z$,

$$\bar{X} \leqq 7.00-2.58\times\frac{1.00}{\sqrt{400}},\qquad 7.00+2.58\times\frac{1.00}{\sqrt{400}} \leqq \bar{X}.$$



That is, $\bar{X} \leqq 6.871$, $7.129 \leqq \bar{X}$.

Since the observed sample mean $\bar{x}=6.88$ does not fall into this interval, the hypothesis is not rejected.

Thus, at the **1% level of significance**,

it cannot be said that the results for one city are not comparable to the national level.

(2) Hypothesis is that the results are greater than or equal to the national level.

Let X be the mean of ○, we can examine when X follows a normal distribution $N(7.00,\ 1.00^2)$.

In this case, the sample mean $\bar{X}$ follows

$N\left(7.00,\ \frac{1.00^2}{400}\right)$.

Let $Z=\frac{\bar{X}-7.00}{\frac{1.00}{\sqrt{400}}}$, then the random variable Z

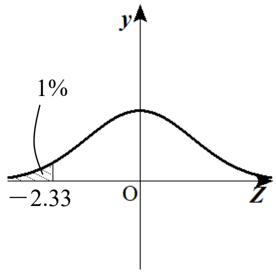
follows a standard normal distribution $N(0,\ 1)$.

> In detail,
> the alternative hypothesis is that the result is less than the national;
> the null hypothesis is that the result is greater than the national.
> To see if the hypothesis that the results are more than the nation can be rejected, we can examine the distribution when the results are the same as for the nation.

The rejection region at the 1% level of significance for "Results are comparable to the national level" is $Z \leqq -2.33$, so

$$\bar{X} \leqq 7.00-2.33\times\frac{1.00}{\sqrt{400}}.$$

That is, $\bar{X} \leqq 6.8835$.



The observed sample mean $\bar{x}=6.88$ falls within this interval, and so the hypothesis is rejected.

Thus, at the 1% level of significance,

we can conclude that the result for **one city is less than the national result**.