# Statistics exercises 1

## 1. Stem and Leaf

| Stem | Leaf |
|---|---|
| 6 | 2, 5, 8 |
| 7 | 0, 3, 5, 5, 8 |
| 8 | 1, 3, 4, 5, 7, 9 |
| 9 | 2, 5, 6, 8 |
| 10 | 0 |

## 2. Box Plot

Median : 70

Q1 : 63

Q3 : 78

Minimum : 55

Maximum : 88

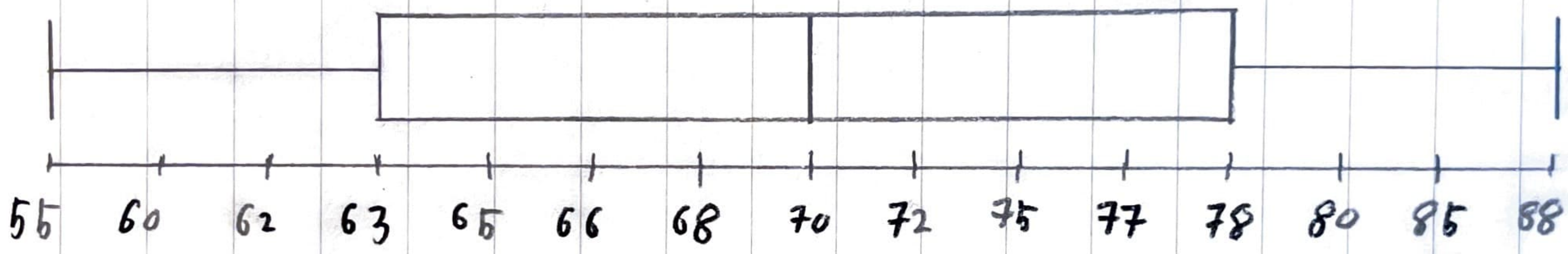
IQR = 78 - 63 = 15

H-spread

Lower : 63 - 1.5 • 15 = 40.5

Upper : 78 + 1.5 • 15 = 100.5

No outliers


55
60
62
63
65
66
68
70
72
75
77
78
80
85
88

Statistics exercises 2

1. Calculate trimean

$Q_1 : 18$
$M = 30$
$Q3 = 42$

eyeball dat shit

$= (Q_1 + (\text{median} \cdot 2) + Q3)/4$

$= (18 + (30 \cdot 2) + 42)/4$

$= 30$

2. Geometric mean

$1.05 \cdot 1.10 \cdot 0.97 \cdot 1.06 = 1.181841$

$= \sqrt[4]{1.181841}$

$= 1.0426$

$= 1.0426 - 1 = 0.0426$ or $4.26\%$

3. Trimmed mean (10% trim)

$10\% \cdot n$

$n = 10$

$10\% \cdot 10 = 1$

$(70+72+75+80+85+90+92+95)/8$

$= 82.375$

statistics exercise 3

1. Permutation : All arrangements of 8 people sitting in 4 spots

$n=8$

$r=4$ $\quad {}_8P_4 = \frac{8!}{(8-4)!} = 1680$

2. $n=7$ $\quad {}_7C_4 = \frac{7!}{4!(7-4)!} = \cancel{840}$ 35 ways

$r=4$

3. combination formula

Red; 10 $\quad$ Select 3 Red and 2 Blue

Blue: 15 $\quad$ or use this

$n: 5$ $\quad \left[ {}_nC_r = \frac{n!}{(n-r)!\,r!} \right]$

${}_{10}C_3 = 120$ $\quad \frac{120 \cdot 105}{53130} = 0.237$

${}_{15}C_2 = 105$

${}_{25}C_5 = 53130$ $\quad = 23.7\%$

Does the order of selection matter?

Yes → perm / No → comb

Grouping → comb

Arranging → perm

Statistics exercise 4.

1. Percentage returns / compound returns

Year 1 : 1.10 $(1+0.10)\cdot(1+0.15)\cdot(1-0.05)\cdot(1+0.08)$

Year 2 : 1.15 $\cdot(1+0.12)$

Year 3 : 0.95 $= 1.422$

Year 4 : 1.08 $= \sqrt[5]{1.422} \approx 1.073$

Year 5 : 1.12 $\hookrightarrow (1.073-1)\cdot 100 = 7.3\%$

2.

| | A | B |
|---|---|---|
| median | 13 | 10 |
| Q1 | 9 | 7 |
| Q3 | 15 | 15 |
| minimum | 7 | 5 |
| maximum | 16 | 18 |
| IQR | 6 | 8 |

A H-spread

Lower : $9-1.5\cdot6=0$

Upper : $15+1.5\cdot6=24$

No outliers

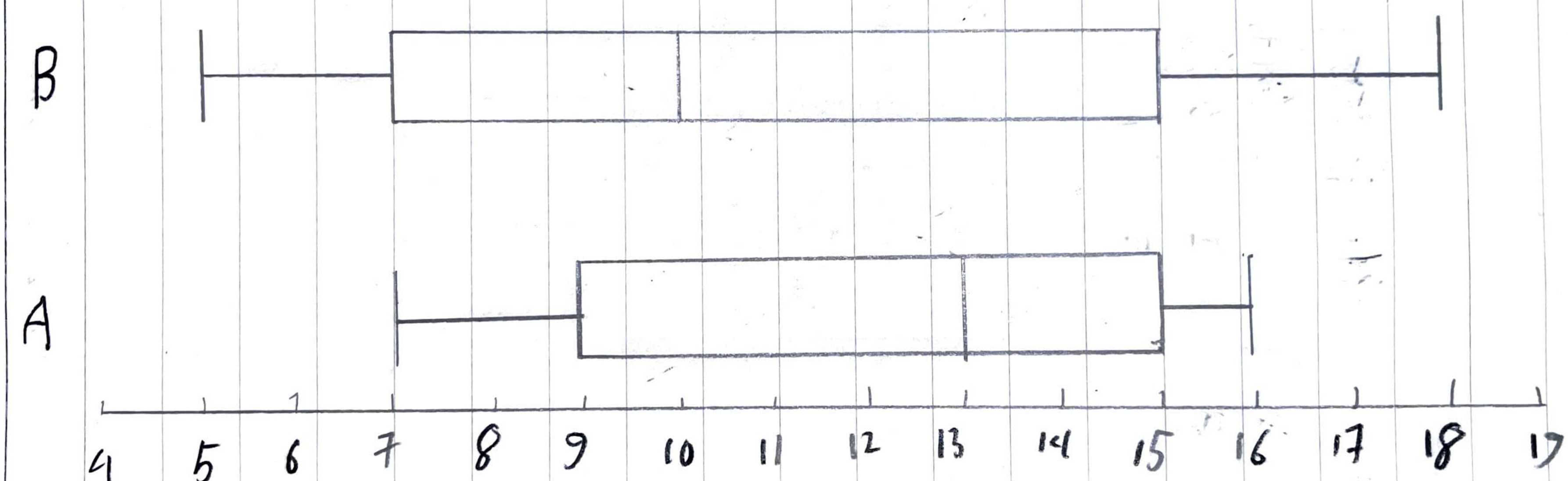
B H-spread

$7-1.5\cdot8=-5$

$15+1.5\cdot8=27$

No outliers

A has a highet median than B.
No outliers



3. $(1/2)\cdot(4/52) = 1/26$

$= 0.0385$

$= 3.85\%$

4.

| leaf X | stem | leaf Y |
|---|---|---|
| 2 4 7 9 | 1 | 3 6 8 |
| 1 4 6 8 | 2 | 0 3 5 7 9 |
| 0 2 | 3 | 1 3 |

~~5. $2^5 = 32$~~ ~~$10/32 = 5/16$~~ ~~coin flipped 5 times~~

$5C_3 = 10$ find prob of 3 heads

~~6. 80% chance of success, 15 attempts, chances of 12 successful attempts (at least)~~

~~$p = 0.8$~~

~~$n = 15$~~

~~$k = 12, 13, 14, 15$~~

~~If k12~~

~~$15C_{12} = 455$~~

~~$455 \cdot (0.8^{12}) \cdot (0.2^{3}) = 0.2501$~~

4.

5.

$$P = \frac{N!}{X!(N-X)!} \cdot \pi^{X}(1-\pi)^{N-X}$$

$\pi$ = Probability of success

$N$ = total trials

$X$ = successes

$$P = \frac{5!}{3!(5-3)!}\left(\frac{1}{2}\right)^{3}\left(1-\frac{1}{2}\right)^{5-3}$$

$$= \frac{5!}{3!2!}\left(\frac{1}{8}\right)\left(\frac{1}{2}\right)^{2}$$

$$= \frac{5 \cdot 4 \cdot 3!}{3!2!} \cdot \frac{1}{32}$$

$$= \frac{10}{32} = \frac{5}{16}$$

4

6.

$$P(x=12) = \frac{N!}{x!\,(N-x)!} \cdot \pi^{x} (1-\pi)^{N-x}$$

$$= \frac{15!}{12!\,3!} \cdot 0.8^{12} \cdot 0.2^{3}$$

$$= 0.227$$

$$P(x=13) = \frac{15!}{13!\,2!} (0.8)^{13} (0.2)^{2}$$

$$= 0.236$$

$$P(x=14) = \frac{15!}{14!\,1!} (0.8)^{\cancel{13}\,14} (0.2)^{2}$$

$$= 0.\cancel{236}\;137$$

$$P(x=15) = \frac{15!}{15!} \cdot 0.8^{15} \cdot 0.2^{\cancel{3}\,0}$$

$$= 0.035$$

$$P(x \geq 12) = P(x=12) + P(x=13) + P(x=14) + P(x=15)$$

$$= 0.635$$

7. Pearson correlation

$n = 5$

$\Sigma X = 2+4+6+8+10$

$= 30$

$\Sigma Y = 10+15+20+25+30$

$= 100$

$\Sigma X^2 = 4+16+36+64+100$

$= 220$

$\Sigma Y^2 = 100+225+400+625+900$

$= 2250$

$\Sigma XY = 20+60+120+200+300$

$= 700$

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$= \frac{5 \cdot 700 - (30)(100)}{\sqrt{[5 \cdot 220 - (30)^2][5 \cdot 2250 - (100)^2]}}$$

$= 1$

Statistics exercise 5

1. Standard deviation

$$\frac{70+85+78+90+88}{5} = \cancel{80.2}\ 82.2$$

$70-82.2 = -12.2 \rightarrow (-12.2)^2 = 148.84$

$85-82.2 = 2.8 \rightarrow (2.8)^2 = 7.84$

$78-82.2 = -4.2 \rightarrow (-4.2)^2 = 17.64$

$90-82.2 = 7.8 \rightarrow (7.8)^2 = 60.84$

$88-82.2 = 5.8 \rightarrow (5.8)^2 = 33.64$

$$\frac{148.84+7.84+17.64+60.84+33.64}{5} = 53.76$$

$\sqrt{53.76} = 7.33$

$$\boxed{SD = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}}$$ use this!

5.

2.

① Check condition :

$\mu = n \cdot p$

$= 100 \cdot 0.30$

$= 30$

$30 > 5$

$n(1-p) = 100(1-0.3)$

$= 70$

$70 > 5$

If larger than 5, normal distribution. If its less than 5, use binomial distribution

② $SD = \sqrt{n \cdot p(1-p)}$

$= \sqrt{100 \cdot 0.3 \cdot 0.7}$

$= 4.58$

③ $P(x < 25) \rightarrow P(Y \leq 25)$

$\hookrightarrow 24.5$

$= P(Y \leq 24.5)$

(If question asks less than do 24.5, if not 25.5)

④ $z = \dfrac{24.5 - 30}{4.58} = -1.20$

⑤ ~~Look at z table = 0.1151~~

~~= 11.51%~~

~~(negative z-table)~~

5.

(5) $P(Z \leq -1.20) = 0.1151$

$P(x < 25) = 11.51\%$

Since question asks for less than 25, do $P(x<25)$
if arrow goes ">", minus the ~~result~~ result from
Z-table by 1.

3.

(1) Check condition

$$\left.\begin{array}{l} n \cdot p = 100 \cdot 0.4 = 40 \\ n(1-p) = 100 \cdot 0.6 = 60 \end{array}\right\} \text{greater than } 5$$

mean $= 40$

$\sigma = \sqrt{100 \cdot 0.4 \cdot 0.6} = 4.899$

$P(x \geq 45) \Rightarrow P(x \geq 44.5)$

$Z = \dfrac{44.5 - 40}{4.899} = 0.919$

$P(Z \leq 0.919) = 0.81594$

$= 1 - 0.81594$

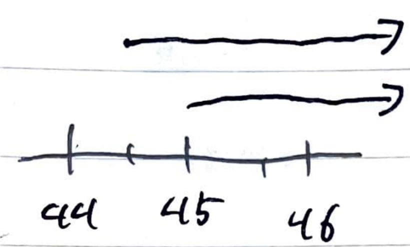
5.

3) $= 0.1788$

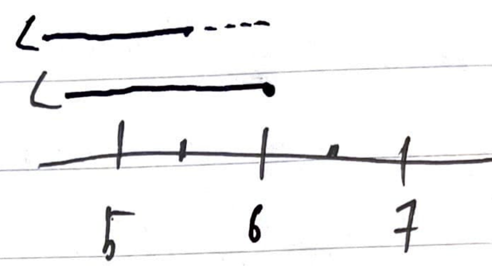
$= 17.88\%$

Note.

$P(x \geq 45)$




$P(x < 6)$




44.5 can since rounds up to 45.
But cannot 45.5 since goes too high to 46 instead of 45

5.5 can since it rounds up to 6

6.

1. Two-tail T-test $\alpha = 0.05$

$H_0$: Mean lifespan of bulbs have no significant difference from 1000 hours

$H_1$: Mean lifespan of bulbs have a significant difference from 1000 hours

Sample Mean $= (950 + 960 + \dots + 995)/10 = 990.5$

Standard deviation $= 25.868256$

Standard error $= \dfrac{25.868256}{\sqrt{10}} = 8.180260806$

T-value $= \dfrac{990.5 - 1000}{8.180260806} = -1.161332166$

$df = 10 - 1 = 9$

P-value $= 0.138$

Critical t-value $= \pm 2.262$

$\Big\}$ go online for table & calculator

$-1.161$ within $\pm 2.262$, fail to reject $H_0$.

2. Null hypothesis: There is no difference in weight before and after the training program (left tail)

mean of differences: $\frac{-3-3-5-2-3-3-3-3}{8} = -3.125$

sd of $\bar{x}$ diff: 0.83452296

t statistic: $-3.125 / (0.8345 / \sqrt{8}) = -10.59148183$

df = 8 - 1
= 7

Critical value: ~~2.365~~ 1.895 (get in online calculator)

Null hypothesis rejected, sufficient evidence to prove considerable difference.

$-10.59148183 < \cancel{2.365}\ 1.895$

tail

Left: If something decreases in value

Right: If something increases in value

Neither: Two-tail, if told to find diff of something (decrease & increase)

3.

$H_0$: No difference with new diet plan. (Right Tail)

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{sd^2_A}{n_A} + \frac{sd^2_B}{n_B}}}$$

$$t = \frac{8-6}{\sqrt{\frac{2^2}{25} + \frac{2.5^2}{25}}} \rightarrow \sqrt{0.16+0.25} \rightarrow \sqrt{0.41} \approx 0.64$$

$$t = \frac{8-6}{0.64} = \frac{2}{0.64} = 3.13$$

$df = 25+25-2$
$= 48$

Critical $T = 1.6779$

$p = 0.00002$ or $\cancel{4.71440}$ 3.13 $> 1.6779$

P-value is less than significance level, we reject the null hypothesis

T-value larger than critical value, $H_0$ rejected.

7.

1. One-way ANOVA

grand mean $= \frac{15+16+\dots+24}{15} = 20.6$

$\bar{X}_A = \frac{15+16+14+15+17}{5} = 15.4$

$\bar{X}_B = \frac{20+22+19+21+20}{5} = 20.4$

$\bar{X}_C = \frac{25+27+26+28+24}{5} = 26.0$

$SS_{total} = (15-20.6)^2 + (16-20.6)^2 + \dots + (24-20.6)^2 = 89.2$

$SS_{between} = 5 \cdot ((15.4-20.6)^2 + (20.4-20.6)^2 + (26.0-20.6)^2)$
$= 71.6$

$SS_{within} = SS_{total} + SS_{between} = 89.2 - 71.6 = 17.6$

$df_{between} = k-1 = 3-1 = 2$
$df_{within} = N-k = 15-3 = 12$

$MS_{between} = SS_{between} / df_{between} = 71.6/2 = 35.8$
$MS_{within} = SS_{within} / df_{between} = 17.6/12 = 1.47$

F-Statistic = $\frac{MS_{between}}{MS_{within}} = \frac{35.8}{1.47} = 24.35$

Critical F-value = 3.89

↳ $df_{between}$ numerator

↳ $df_{within}$ denominator

F > Crit-F, reject null hypothesis

or

P-value = < 0.001

$P \leq \alpha$, reject null hypothesis

3. Chi-square test for independence

$\alpha$ : 0.05

$H_0$ : Independent

$H_1$ : not independent

$df = (\text{rows}-1)(\text{columns}-1)$

$= (3-1)(3-1)$

$= 2 \cdot 2$

$= 4$

critical value ~~Decision rule~~: 9.48773 (~~X~~ Chi-square distribution table)

Expected values

$XA = (40 \cdot 30) / 90 = 13.333$

$XB = (40 \cdot 35) / 90 = 15.556$

$XC = (40 \cdot 25) / 90 = 11.111$

$YA = (30 \cdot 30) / 90 = 10$

$YB = (30 \cdot 35) / 90 = 11.667$

$YC = (30 \cdot 25) / 90 = 8.333$

$ZA = (20 \cdot 30) / 90 = 6.667$

$ZB = (20 \cdot 35) / 90 = 7.778$

$ZC = (20 \cdot 25) / 90 = 5.556$

7

$$x^2 \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$x^2 = \frac{(10-13.333)^2}{13.333} + \frac{(20-15.556)}{15.556} \cdots + \frac{(10-5.556)}{5.556}$$

$x^2 =$ ~~0.555049 6106~~ 11.25

~~$x^2 >$ Ho, Null hypothesis is rejected. Not Independent~~

$x^2 >$ critical value, reject null hypothesis.

or use p-value (chi-square p-value calculator)

p-value : 0.23893

p-value $<$ $\alpha$, reject Ho since p is smaller than alpha. significant result.

Statistics exercise 7

3.

Python self $= (78+82+85)/3 = 81.667$

Python instructor $= (90+88+92)/3 = 90$

Java self $= (72+75+74)/3 = 73.667$

Java instructor $= (85+80+84)/3 = 83$

C++ Self $= (65+68+70)/3 = 67.667$

C++ instructor $= (78+75+80)/3 = 81$

Self $= (81.667+73.667+67.667)/3 = 74.334$

instructor $= (90+83+81)/3 = 84.667$

Use samples directly not means

Python $= (81.667+90)/2 = 85.834$

Java $= (73.667+83)/2 = 78.334$

C++ $= (67.667+81)/2 = 74.334$

grand mean $= (74.334+84.667+85.834+78.334+74.334)/5$

$= 79.5006$

$SS_{tot} = (78-79.5006)^2 + (82+79.5006)^2 \ldots + (80+79.5006)^2$

$= 990.512$

$df_{tot} = n \cdot p \cdot q - 1$ n = number of samples / category

$= 3 \cdot 3 \cdot 2 - 1$ A & P = Language

$= 17$ B & q = lesson method

7

$$\sigma^2_{tot} = 990.512/17 = 58.265$$

$$SS_{btw} = ((81.667 - 79.5006)^2 + (90 - 79.5006)^2 + \dots$$
$$+ (81 - 79.5006)^2) \cdot 3$$
$$= 910.469$$

$$df_{btw} = p \cdot q - 1$$
$$= 3 \cdot 2 - 1$$
$$= 5$$

$$\sigma^2_{btw} = 910.469/5 = 182.094$$

$$SS_A = ((85.834 - 79.5006)^2 + (78.334 - 79.5006)^2$$
$$+ (74.334 - 79.5006)^2) \cdot \underset{n}{3} \cdot \underset{q}{2}$$
$$= 409.000001$$

$$df_A = p - 1$$
$$= 3 - 1$$
$$= 2$$

$$\sigma^2_A = 409.000001/2 = 204.5000001$$

$SS_B = ((74.334 - 79.5006)^2 + (84.667 - 79.5006)^2) \cdot B$
$\cdot 3)$

$= 480.4690007$

$df_B = q - 1$
$= 2 - 1$
$= 1$

$\sigma^2_B = 480.469 / 1 = 480.469$

$SS_{AB} = SS_{btw} - SS_A - SS_B$
$= 910 - 469.000101 - 480.4690007$
$= 21.0000032$

$df_{AB} = (p-1) \cdot (q-1)$
$= 2 \cdot 1$
$= 2$

$\sigma^2_{AB} = 21.0000032 / 2 = 10.5000016$

7

$SS_{err} = (78-81.067)^2 + (90-90)^2 + (72-73.667)^2$

$\dots (80-81)^2$

$= 110$

$df_{err} = (n-1) \cdot p \cdot q$

$= (3-1) \cdot 3 \cdot 2$

$= 12$

$\sigma^2_{err} = 110/12 = 9.16666675$

$F_A = 204.5000001 / 9.16666675 = 22.30909072$

$F_B = 480.469 \quad / 9.16666675 = 52.4247996$

$F_{AB} = 10.5000016 / 9.16666675 = 1.14545471$

$df_A = 2 \qquad p_A = < 0.001$

$df_B = 1 \qquad p_B = < 0.001$

$df_{AB} = 2 \qquad p_{AB} = 0.351$

$df_{err} = 12$

Reject null hypothesis for factor A (Programming language) and factor B (Lesson method) since smaller than $\alpha = 0.05$. This means the programming language and lesson method does have a significant effect on student test scores.