Generally: - Simple calculation mistake: -0.5 pts
unless it makes the result obviously wrong
e.g. probability too high, test conclusion
wrong etc.

1a) $P(A \cap \bar{B}) = P(A) \cdot P(\bar{B} \mid A)$

$$= P(A) \cdot [1 - P(B \mid A)]$$

$$= 0.2 \cdot [1 - 0.6]$$

$$= 0.2 \cdot 0.4$$

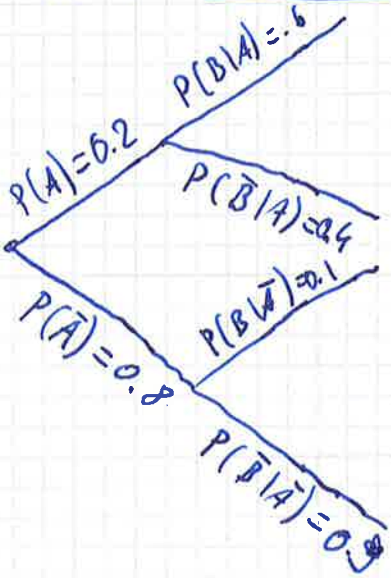
$$= 0.08 = 8\%$$

1b) $P(A) = 0.2$ $\quad P(B|A) = 0.6$

$\underline{P(\bar{A}) = 0.8}$, $\quad \underline{P(\bar{B}|A) = 0.4}$ $\quad P(\bar{A} \cap \bar{B}) = 0.72$

$$\left.\begin{aligned} &P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B}|\bar{A}) \\ &\Rightarrow \underline{\underline{P(\bar{B}|\bar{A})}} = \frac{P(\bar{A} \cap \bar{B})}{P(\bar{A})} = \frac{0.72}{0.8} = \underline{\underline{0.9}} \end{aligned}\right\} 1pt.$$

$\underline{P(B|\bar{A})} = 1 - P(\bar{B}|\bar{A}) = \underline{0.1}$



1c) $$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A) \cdot P(A)}{P(B \mid A) \cdot P(A) + P(B \mid \bar{A}) \cdot P(\bar{A})}$$

$$= \frac{0.6 \cdot 0.2}{0.6 \cdot 0.2 + 0.1 \cdot 0.8} = \frac{0.12}{0.12 + 0.08}$$

$$= \frac{0.12}{0.2} = \underline{\underline{0.6}} = 60\%$$

$\mu = 150 \quad \sigma = 10$

1 d) $P(R < 160) = \Phi\left(\frac{160 - \mu}{\sigma}\right) = \Phi\left(\frac{160-150}{10}\right) = \Phi(1)$

$\rightsquigarrow$ z-table $\Phi(1) = 0.841 \approx \underline{\underline{84\%}}$

1) e) $P(130 \leq R < 140) = \cancel{\Phi} \; P(R < 140) - P(R < 130)$

$$= \Phi\left(\frac{140-150}{10}\right) - \Phi\left(\frac{130-150}{10}\right) =$$

$$= \Phi(-1) - \Phi(-2)$$

$$= [1 - \Phi(1)] - [1 - \Phi(2)]$$

$$= [1 - 0.841] - [1 - 0.9772]$$

$$= 0.159 - 0.0228$$

$$= 0.1362$$

$$\underline{\underline{\approx 0.14 \quad \text{or} \approx 14\%}} \quad \text{or} \; = 13.62\%$$

$N_s = 25$

1f) $P(\bar{R} < 155) = \Phi\left(\frac{155-\mu}{\sigma/\sqrt{N_s}}\right)$

$= \Phi\left(\frac{155-150}{10/\sqrt{25}}\right)$

$= \Phi\left(\frac{5}{2}\right) = \Phi(2.5)$

$\underline{\underline{= 0.9938}} \approx 0.99$

$\underline{\underline{\approx 99\%}}$

0pts if they did not use $\sigma/\sqrt{N_s}$

$$\bar{x} = \sum_i x_i \Big/ \sum_i 1 = \sum_i x_i / N_s$$

1 g) $\bar{x} = \dfrac{21 + 19 + 22 + 26}{4} = \dfrac{88}{4} = \underline{\underline{22}}$ } 0.5 pt.

$$s^2 = \frac{1}{N_s - 1} \sum_i (x_i - \bar{x})^2$$

$$= \frac{1}{4-1} \left((-1)^2 + (-3)^2 + (0)^2 + (4)^2 \right)$$

$$= \frac{1}{3} (1 + 9 + 0 + 16) = \underline{\underline{\frac{26}{3}}} \approx 8.7$$

or 8.67

if calculating s (std.) instead of s^2 (variance): Lose 0.5 pts.

1h) A one-sided t-test. (because σ unknown
0.5pt 1.5pt. & small sample
& comparing to given value).

if not mentioned but done
correctly in 1i) -> still get the half point.

$\nu = N_s - 1 = 4 - 1 = 3$ (degrees of freedom, df)

0.5 pt

1c)

$$t = \frac{\bar{x} - 20}{\sqrt{S^2/N_s}} = \frac{22-20}{\sqrt{\frac{26}{3}/4}} = \frac{2}{\sqrt{2.16}} = 1.35873\ldots$$

2pt

either: $qt_{99\%}(\nu=3) = 4.541$ 0.5 pt

as $t \leq qt_{99\%}(\nu=3)$

1pt (one statement is enough.)

→ Cannot reject null hypothesis (the mean is larger than 20K only because of random fluctuations) ⟺ Cannot prove "the mean is larger than 20K at 99% confidence".

2 TOTAL: 25 PTS

2 a) $P(\text{"odd"}) = \frac{3}{5} = 0.6 = P(A)$

1

$P(x>3) = \frac{2}{5} = 0.4 = P(B)$

---

2 b) Independent if $P(A \cap B) = P(A) \cdot P(B)$.

2

$P(A \cap B) = \frac{1}{5} = 0.2$ (only "5")

$P(A) \cdot P(B) = \frac{3}{5} \cdot \frac{2}{5} = \frac{6}{25} = \cancel{0.25}$ 0.24

$\Rightarrow$ Not equal $\Rightarrow$ Not independent.

ALTERNATIVE SOLUTIONS:

I) Independent if $P(A|B) = P(A)$

$P(A|B) = \frac{1}{2} \neq \frac{3}{5}$

$\Rightarrow$ not independent.

II) Independent if $P(B|A) = P(B)$

$P(B|A) = \frac{1}{2} \neq \frac{2}{5}$ $\Rightarrow$ not independent

2 c) $\mathbb{E}(X) = \sum_i x_i p_i = \frac{1}{5} + \frac{2}{5} + \frac{3}{5} + \frac{4}{5} + \frac{5}{5}$

1

$= \frac{15}{5} = 3 = \mu$

2d) $\mathbb{E}(X^2 - 10) = \sum_i (x_i^2 - 10) p_i$

2

$= \frac{(1^2 + 2^2 + 3^2 + 4^2 + 5^2 - 5 \cdot 10)}{5}$

$= \frac{55 - 50}{5} = +1$

gain of 1 per round.

0 pts if they took $E(X)^2 - 10 = \mu^2 - 10 = 9 - 10 = -1$

2 e) $\bar{x} = \frac{2+5+2+5+1}{5} = 3$ (Mean) 0.5pt

2

$\tilde{x}_{50\%} = 2$ 0.5pt (Median. ordered list: 1, 2, 2, 5, 5)

$$s^2 = \frac{1}{N_s - 1} \sum_i (x_i - \bar{x})^2$$

$$= \frac{1}{4} \left((-2)^2 + (-1)^2 \cdot 2 + (2)^2 \cdot 2 \right)$$

$$= \frac{14}{4} = 3.5$$

1pt (0.5 if s, not s^2 was computed)

2 f) Fair die; expect $N_S \cdot p_i = 200 \cdot \frac{1}{5} = \underline{40}$ for each i

1

Result:	1	2	3	4	5
$N_{occurrences}$:	48	35	38	33	46
$N_{expected}$:	40	40	40	40	40
Difference :	+8	−5	−2	−7	+6

2 g) A χ^2 (chi-2, chi-squared) test.

1

(Because we compare absolute (or relative) frequencies to ~~prob~~ expected values (or probabilities)).

2 h) (4)

$$\underline{\underline{\chi^2}} = \sum_i \frac{(\overset{\text{Abs. freq.}}{n_i} - \overset{\text{expected}}{N_s p_i})^2}{N_s p_i} = N_s \sum_i \frac{\left(\overset{\text{rel. freq. } \hat{p}_i}{\frac{n_i}{N_s}} - p_i\right)^2}{p_i}$$

$$= \frac{(+8)^2 + (-5)^2 + (-2)^2 + (-7)^2 + (+6)^2}{200 \cdot \frac{1}{5}}$$

$$= \frac{178}{40} = \underline{\underline{4.45}}$$ 1pt

Degrees of freedom: $df = \underline{\underline{\nu}} = K - 1 = 5 - 1 = \underline{\underline{4}}$ ← Number of possible outcomes 1pt

$q\chi^2_{\nu=4} = 9.488$ 0.5pt

$\chi^2 = 4.45 < 9.488 = q\chi^2_{\nu=4}$

1.5pt → χ^2 is small enough that at the 95% confidence level we can say the die is fair.
(we cannot reject the null hypothesis "the die is fair")
(we cannot say the die is unfair).

2 i) We need to consider the expected absolute frequency for the least probable result.

1

$N_S \cdot \min(p_i)$. Here all $p_i = \frac{1}{5}$, so

$N_S \cdot \min(p_i) = N_S \cdot p_i = \frac{200}{5} = 40$.

This has to be large enough, and typically $N_S \cdot \min(p_i) \geq 5$ is considered sufficient.

No points for general statement.

Must mention $\underbrace{\text{expected value}}_{0.5}$ of $\underbrace{\text{least likely}}_{0.5}$ outcome.

2 j) Sorted Array: 0 pt.

3

[1 2 2 5 5] 0 pt

Length of array:

5 0.5 pt

Position of the median:

3 0.5 pt

The median:

5 1pt

) no subtraction for wrong syntax ([] or ,,, ...)

2 k) median = sortedArray[3] which is
2 the 4th element in the array - Python starts counting from 0. 1pt

Possible solutions: I) median = sortedArray [medianPosition -1]
1pt
II) medianPositionFloat = arrayLength/2 -0.5
III) Other similar solution

Just saying myMedian = np.median(myArray)
gives only 0.5pt.

2 e) possible solution: I) myMedian = ~~3~~ np.median(myArray)

3

II): Use an if-clause to check if arrayLength is even or odd. (e.g. by arrayLength % 2 == 0)

1

0.5 If odd → code as above (with correction)

1.5 If even -> take ~~arrayLength~~ (sortedArray[arrayLength/2] + sortedArray[arrayLength−1]) / 2

---

2 m) Proof by counter example: {−2, −2, 0, 1, 3}

2

has mean = 0 and median = 0 but it is not symmetric around 0.

0.5pt for answer w/o proof.

1pt for "handwavy" proof.

$I=2 \quad J=2$

3 a)

2

1pt

$$\bar{X}_{1,\cdot,\cdot} = \frac{1}{J}\sum_j \bar{X}_{1,j,\cdot} = \frac{67+83}{2} = 75$$

$$\bar{X}_{2,\cdot,\cdot} = \frac{139+211}{2} = 175$$

$$\bar{X}_{\cdot,1,\cdot} = \frac{1}{I}\sum_i \bar{X}_{i,1,\cdot} = \frac{67+139}{2} = 103$$

$$\bar{X}_{\cdot,2,\cdot} = \frac{83+211}{2} = 147$$

$$\bar{X}_T = \bar{X}_{\cdot,\cdot,\cdot} = \frac{67+83+139+211}{4} = 125 \quad \left(\text{or via } \frac{75+175}{2} \text{ or } \frac{103+147}{2}\right)$$

optional

1pt.

J \ I	Pt	Mo	all
RT	67	139	103
LN	83	211	147
all	75	175	125

optional

Transpose is ok.

3 b) $SS_{E,i,j} = S_{i,j}^2 \cdot (N_s - 1) = S_{ij}^2 \cdot 10$

$\uparrow$ 11

2

0.5 pt

$SS_{E\,1,1} = 11^2 \cdot 10 = \underline{1210}$

$SS_{E\,1,2} = 8^2 \cdot 10 = \underline{640}$

$SS_{E\,2,1} = 14^2 \cdot 10 = \underline{1960}$

$SS_{E\,2,2} = 17^2 \cdot 10 = \underline{2890}$

(1p if correct without explicit formula)

1pt $\underline{\underline{SS_{E(TOT)}}} = \sum_i \sum_j SS_{E\,i,j} = 1210 + 640 + 1960 + 2890$

$= \underline{\underline{6700}}$

3 c) $SS_{B,I} = J \cdot N_s \cdot \sum_i (\bar{X}_{i,\cdot,\cdot} - \bar{X}_T)^2$ $\quad N_s = 11$

1

$= 2 \cdot 11 \cdot [(75-125)^2 + (175-125)^2] = 2 \cdot 11 \cdot 5000$

$= \underline{\underline{110'000}}$

$SS_{B,J} = 2 \cdot 11 [(103-125)^2 + (147-125)^2] =$

$= 2 \cdot 11 \cdot 968$

$= \underline{\underline{21'296}}$

~~3~~ 3 d) for this we have to compare the

(1) MEAN SQUARE ERROR, $\frac{SS_{B,A}}{\nu_B} = MS_B$

$\nu_{B,I} = \nu_{B,J} = 2-1 = 1$ so $SS_B = MS_B$ here.

$MS_{B,I} > MS_{B,J}$

So factor I ("type of metal") is dominant.

0.5pt if only compared $SS_{B,I}$ to $SS_{B,J}$

$$\nu_{B, I:J} = (I-1)(J-1) = 1 \cdot 1 = \underline{1} \quad 0.5\,pt$$

3e) $$\nu_{B,I} = \nu_{B,J} = 2 - 1 = \underline{1} \quad 0.5\,pt$$

5

$$\nu_E = I \cdot J \cdot (N_s - 1) = 2 \cdot 2 \cdot (11-1) = \underline{40} \quad 0.5\,pt$$

$$\underline{MS_{B,I}} = \frac{SS_{B,I}}{\nu_{B,I}} = \underline{110'000}$$

$$\underline{MS_{B,J}} = \frac{SS_{B,J}}{\nu_{B,J}} = \underline{21'296}$$

$$\underline{MS_E} = \frac{SS_E}{\nu_E} = \frac{6700}{40} = \underline{167.5}$$

0.5 pt

$$\underline{SS_{B,I:J}} = N_s \sum_i \sum_j \left(\bar{x}_{i,j,\cdot} - \bar{x}_{i,\cdot,\cdot} - \bar{x}_{\cdot,j,\cdot} + \bar{x}_T\right)^2$$

$$= 11 \cdot \Big[(67 - 75 - 103 + 125)^2 + (83 - 75 - 147 + 125)^2 + (139 - 175 - 103 + 125)^2 + (211 - 175 - 147 + 125)^2\Big]$$

$$= 11\left[14^2 + (-14)^2 + (-14)^2 + 14^2\right]$$

all same as interaction has only one df

$$= 11 \cdot 4 \cdot 196$$

$$= \underline{8624}$$

$$\underline{MS_{B,I:J}} = \frac{SS_{B,I:J}}{\nu_{B,I:J}} = \frac{8624}{1} = \underline{8624}$$

2 pt

$$F_I = MS_{B,I}/MS_E = \frac{110'000}{167.5} = 656.7\ldots \approx \underline{657}$$

$$F_J = 21'296/167.5 \approx \underline{127}$$

$$F_{I:J} = 8624/167.5 \approx 51.50.. \approx \underline{52}$$

0.5 pt

	ν	SS	MS	F
FACT I	1	110'000	110'000	657
FACT J	1	21296	21'296	127
INTER I:J	1	8'624	8'624	52
ERROR	40	6700	167.5	/

0.5 pt [for table]

3 f) $F_{OBS, I:J} = 51$

2 $qF_{\nu_1=1, \nu_2=40}(P=95\%) = 4.085$

1 pt.

$F_{OBS\ I:J} > 4.085 \Rightarrow$

We reject the no-interaction hypothesis at the ~~[illegible]~~ 5% significance level.

1 pt

3 g) ③

2pt

$$\frac{\partial SS_E}{\partial \hat{a}} = \frac{\partial}{\partial \hat{a}} \sum_{i=1}^{N_T} \left[K_i - (\hat{a} + \hat{b}_1 T_i) \right]^2$$

$$= (-1) 2 \sum_i^{N} \left[K_i - (\hat{a} + \hat{b}_1 T_i) \right]$$

$$= \cancel{-2} \left(\cancel{N_T} \bar{K} - \cancel{N_T} \cdot \hat{a} - \cancel{N_T} \cdot \hat{b}_1 \cdot \bar{T} \right) \stackrel{!}{=} 0$$

$$\Rightarrow \quad \underline{\underline{\hat{a} = \bar{K} - \hat{b}_1 \bar{T}}}$$

1pt

This shows it is an extremum. To show it is a <u>minimum</u> we also need to show $\frac{\partial^2 SS_E}{\partial \hat{a}^2} > 0$ at $\hat{a} = \bar{K} - \hat{b}_1 \bar{T}$

$$\frac{\partial^2 SS_E}{\partial \hat{a}^2} = \frac{\partial}{\partial \hat{a}} \left[-2 N_T \left(\bar{K} - \hat{a} - \hat{b}_1 \bar{T} \right) \right]$$

$$\underline{\underline{= 2 N_T > 0}} \quad \checkmark$$

$$= N_{TOT} - 1 - \nu_m$$

3 h) $\nu_{E,m1} = 7 - 1 - \underset{\uparrow \text{1 slope paramete}}{1} = \underline{\underline{5}} = \nu_{E,\text{restricted}}$

1

$\nu_{E,m2} = 7 - 1 - 2 = \underline{\underline{4}} = \nu_{E,\text{full}}$

3 i) $\underline{\underline{F_{m1 vs m2}}} = \dfrac{\Delta SS_E / \Delta \nu_m}{SS_{E,\text{full}} / \nu_{E,\text{full}}}$

2

$$= \frac{(1255 - 405)/(2-1)}{405/4}$$

$$= \frac{850}{101.25} = 8.39506\ldots \approx \underline{\underline{8.4}}$$

1

$qF_{\Delta\nu, \nu_{E,\text{full}}}(p=95\%) = qF_{1,4}(p=95\%) = \underline{7.709}$ 0.5

$F_{OBS} > qF \Rightarrow$ The second modell is statistically significantly (at 5% significance) better. 0.5

3 j) calibratedData = measuredData * 1.2

1 (numpy arrays multiply element-wise)

3 k) [68.4 72.1]

1 (for 0:2 you get elements 0 and 1).

0.5 pts for [68.4 72.1 1200.2]

3 l) I) print(np.max (calibratedData))

2 II) print(calibratedData [calibratedData > 1000])

III) for i in calibratedData:

```
        if i > 1000:
            print("spike")
        else:
            pass
```

etc...

3 m) I) use

print(np.vectorize(myFilt)(calibratedData))

(ok. in words)

II) run a loop:

for i in range(calibratedData):

calibratedData[i] = myFilt(calibratedData[i])

etc...