Linear regression 3.2

A Level: C C C Further A: C C C

An experiment is carried out to measure the resistance R of a semiconductor as a function of absolute temperature T. Theory suggests that above a certain temperature

$$R = R_0 \mathrm{e}^{\frac{b}{T}}$$

where R_0 and b are constants.

Part A Rearrange the equation

By taking the natural logarithms of both sides of the equation show that it can be written

$$\ln R = a + f(T)$$

where a is a constant and $f(T)$ is a function of T. Find expressions for a and $f(T)$.

Find an expression for a.

The following symbols may be useful: `R_0, T, a, b, e, f(T)`

Find $f(T)$.

The following symbols may be useful: `R_0, T, a, b, e, f(T)`

Part B Lines of best fit

In **Figure 1**, $\ln R$, where R is in $\mathrm{k\Omega}$, is plotted as a function of $\frac{1}{T}$, where $\frac{1}{T}$ is in $10^{-3}\,\mathrm{K}^{-1}$. Thus if $R = 1000\,\Omega$, then $\ln R = \ln(1\,\mathrm{k\Omega}) = 0$ and, if $T = 500\,\mathrm{K}$, then $\frac{1}{T} = 0.002\,\mathrm{K}^{-1} = 2 \times 10^{-3}\,\mathrm{K}^{-1}$.

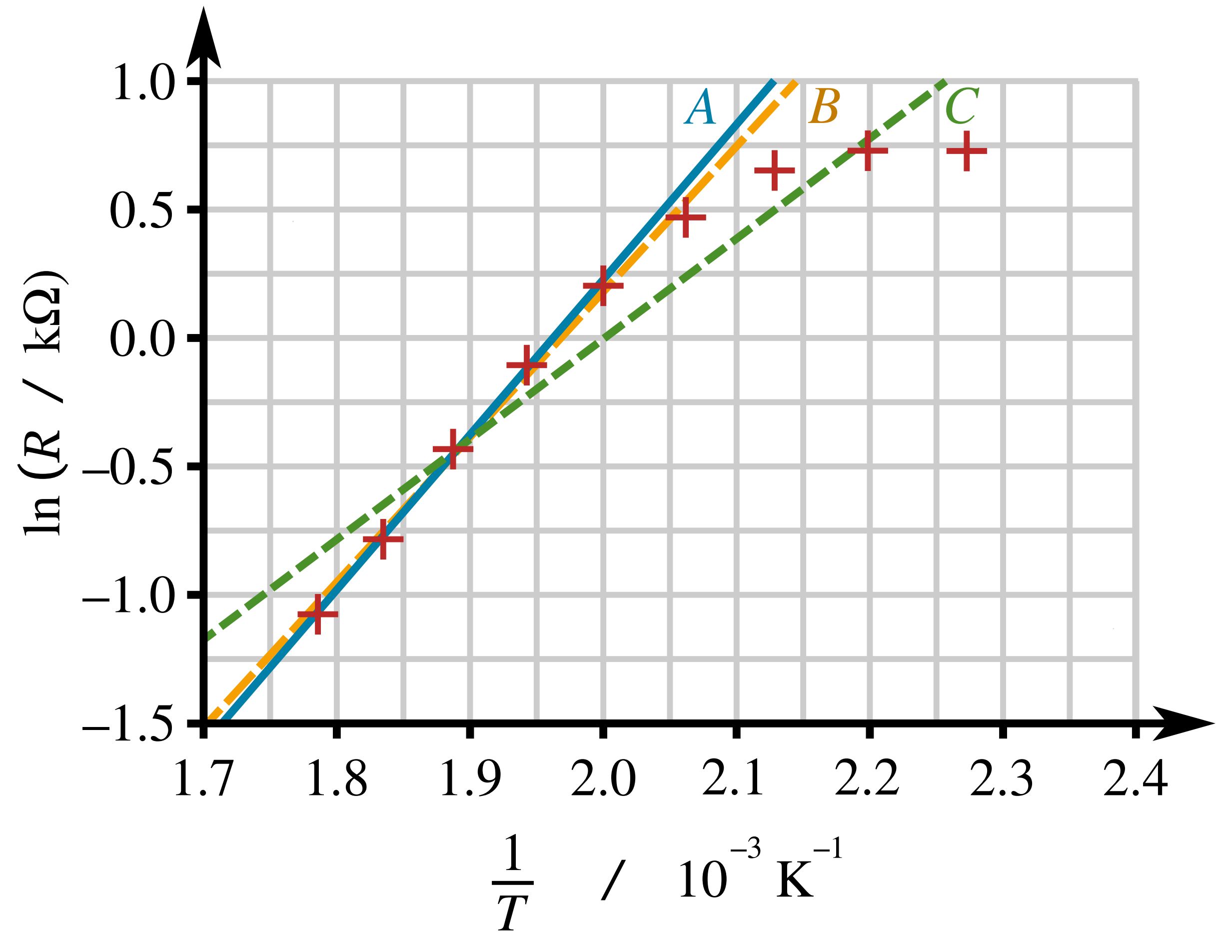

Figure 1: A plot of $\ln R$ against $\frac{1}{T}$; three lines of best fit, A (blue, solid), B (yellow, long-dashed) and C (green, short-dashed) fitted to different ranges of the data, are shown.

Three lines of best fit A, B and C are fitted to different ranges of the data as shown in **Figure 1**.

The parameters of the three fitted lines are:

$$\begin{array}{rllll} \mathrm{I}: & a = -11.197 & b = 5.686 & r = 0.997 & r^2 = 0.994 \\ \mathrm{II}: & a = -11.857 & b = 6.042 & r = 0.999 & r^2 = 0.998 \\ \mathrm{III}: & a = -7.816 & b = 3.906 & r = 0.955 & r^2 = 0.913 \end{array}$$

where a and b are as defined in Part A and the initial equation, and b has units of $10^3\,\mathrm{K}$.

Match the lines to the parameters.

I corresponds to line ☐.

II corresponds to line ☐.

III corresponds to line ☐.

Items:

A B C

Part C Deductions from the graphs

Theory suggests that above a certain temperature

$$R = R_0 \mathrm{e}^{\frac{b}{T}}$$

where R_0 and b are constants.

Using the information from the graphs in part B, suggest, to 1 sf, the temperature above which the theory is valid.

Part D Estimate the energy gap

According to the theory the energy gap between the insulating and the conducting energy bands in a semiconductor is $E_g = 2kb$, where k is the Boltzmann constant ($k = 1.4 \times 10^{-23}\,\mathrm{J\,K^{-1}}$). Select the line of best fit from Part B which best fits the theory and deduce the value of b; hence estimate E_g.

Part E Resistance when $T = 520\,\mathrm{K}$

Use the equation of the line of best fit from Part B to deduce the resistance at $520\,\mathrm{K}$.

Part F Resistance when $T = 450\,\mathrm{K}$

By looking at the graph in Part B, deduce the resistance when $T = 450\,\mathrm{K}$ giving your answer to 1 significant figure.

Linear regression 3.3

A Level C C C Further A C C C

A graph of Hubble's original data relating the recession velocity v of a galaxy to its distance D from us is shown in **Figure 1**; the velocity v is in $\mathrm{km\,s^{-1}}$ and the distance D is in Mpc. (Distances in astronomy are often measured in parsecs (abbreviated to pc), where $1\,\mathrm{pc} = 3.26$ light-years $= 3.09 \times 10^{16}\,\mathrm{m}$ and $1\,\mathrm{Mpc} = 10^6\,\mathrm{pc}$.)

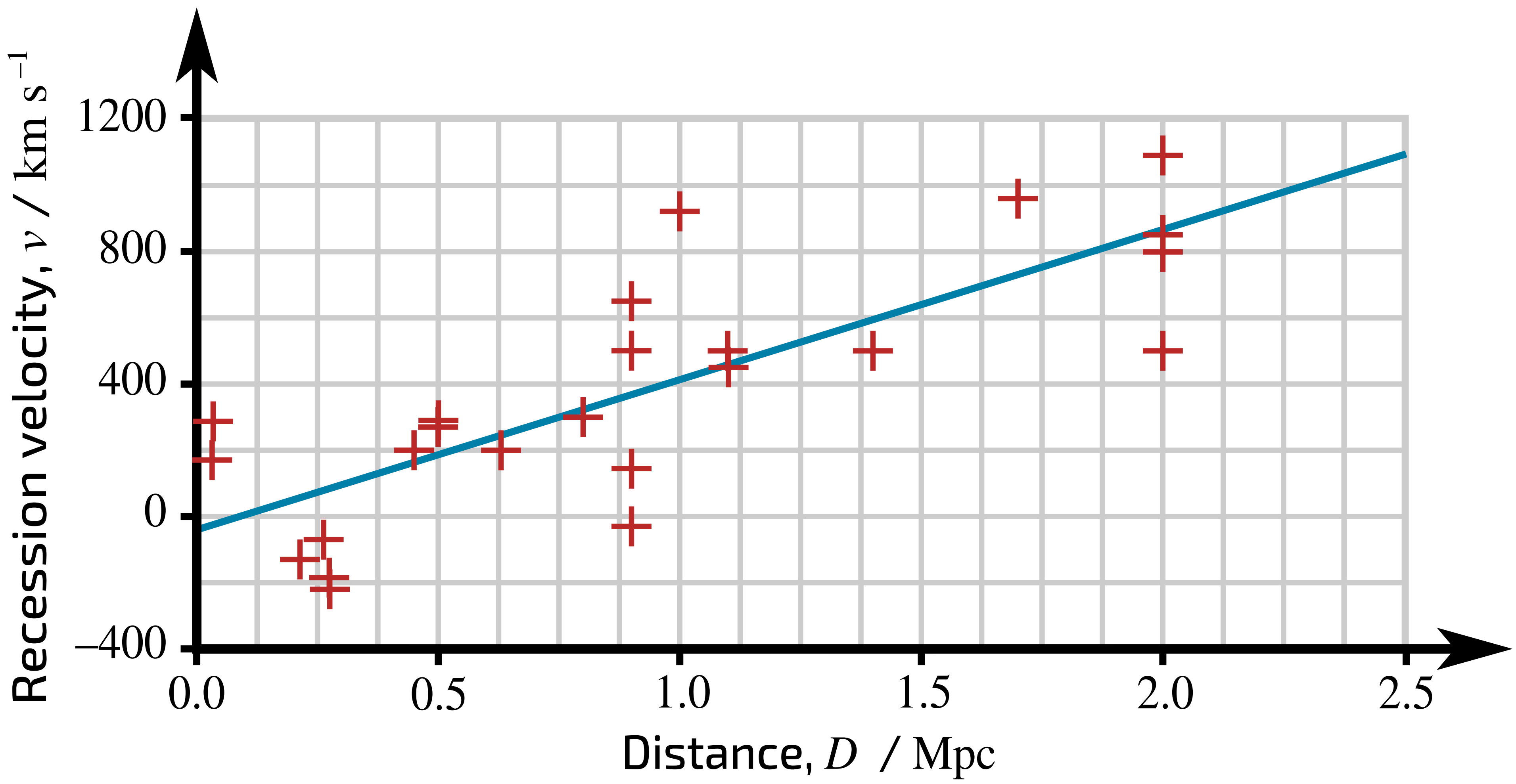

Figure 1: A graph of Hubble's original data relating the recession velocity v of a galaxy to its distance D from us. The regression line of best fit is shown.

The equation describing the best fit to the data is of the form $v = a + bD$ and has the following parameters

$$a = -40.8 \quad b = 454.2 \quad r = 0.790 \quad r^2 = 0.623.$$

Part A The units of a

What are the units of a?

- $\mathrm{km\,s^{-1}\,Mpc^{-1}}$
- $\mathrm{s\,km^{-1}}$
- $\mathrm{km\,s^{-1}/Mpc}$
- $\mathrm{Mpc/km\,s^{-1}}$
- $\mathrm{km\,s^{-1}}$
- $\mathrm{Mpc\,s\,km^{-1}}$

Part B The units of b

What are the units of b? (The quantity b is called the Hubble constant and is usually written H_0).

- $\mathrm{km\,s^{-1}/Mpc}$
- $\mathrm{Mpc\,s/km}$
- $\mathrm{Mpc\,km\,s^{-1}}$
- $\mathrm{Mpc/km\,s^{-1}}$
- $\mathrm{km\,s^{-1}\,Mpc^{-1}}$
- Mpc

Part C Recession velocity

Using the best fit equation above estimate the recession velocity of a galaxy at a distance of 6.0×10^6 light years; give your answer to 2 s.f.

Part D The age of the Universe using the original data

Nowadays the value of the Hubble constant is known to be close to 70 in the same units as b. (It is significantly smaller than that originally determined by Hubble because of a calibration error in Hubble's original data.) The equation describing the relationship between v and D in the same units as above is therefore

$$v = 70D.$$

It is straightforward to show that the age of the Universe is given by $\frac{1}{H_0}$, where H_0 is the Hubble constant.

Find the age of the Universe using the value of b estimated from Hubble's original data above. Give your answer in years and to 2 s.f.

Part E The age of the Universe using current data

Find the age of the Universe using the current value of $H_0 = 70$ (in the same units as b). Give your answer in years and to 2 s.f.

Created for isaacphysics.org by Julia Riley

Gameboard:
STEM SMART Double Maths 33 - Regression, Correlation & Hypothesis Tests

Correlation Hypothesis Testing 1

A Level
P P P

In each part, carry out a hypothesis test for the requested type of correlation at the stated significance level.

Part A Positive correlation

A sample of size $n = 17$ has a correlation coefficient of $r = 0.601$. Test at the 1% significance level whether the population from which the sample was taken has positive correlation, then fill in the blanks below.

The null hypothesis is that the population has no correlation. The alternative hypothesis is that the population exhibits positive correlation.

$$\mathrm{H}_0 : \rho = 0 \qquad \underline{\hspace{2cm}}$$

For a one-tailed test, the critical value of the correlation coefficient for a sample of size 17 is ______ at the 1% significance level.

The correlation coefficient for the sample is 0.601. This is ______ than the critical value. Hence, we ______ the null hypothesis. There is evidence that the population exhibits positive correlation.

Items:

0.5577 | 0.5742 | 0.6055 | $\mathrm{H}_1 : \rho < 0$ | $\mathrm{H}_1 : \rho \neq 0$ | $\mathrm{H}_1 : \rho > 0$ | **reject** | **do not reject** | **smaller** | **equal to** | **larger**

Part B Negative correlation

A researcher believes that times to run the $400 \mathrm{~m}$ at a particular track are slower if there has been a larger amount of rainfall earlier in the day. The researcher times one particular athlete at the same time every day on ten different autumn days, recording the depth of rainfall (in mm) beforehand. The researcher calculates that the correlation coefficient is -0.713. Test at the 5% significance level whether an athlete's time and the amount of rainfall are indeed negatively correlated at this track.

Available items

1. The null hypothesis is that there is no correlation. The alternative hypothesis is there is negative correlation.

1. The null hypothesis is that there is negative correlation. The alternative hypothesis is that there is no correlation.

2. $\mathrm{H}_0: \rho = 0$ $\mathrm{H}_1: \rho > 0$

2. $\mathrm{H}_0: \rho = 0$ $\mathrm{H}_1: \rho < 0$

3. For a one-tailed test, the critical value of the correlation coefficient for a sample of size 5 is 0.8054 at the 10% significance level.

3. For a one-tailed test, the critical value of the correlation coefficient for a sample of size 10 is 0.5494 at the 5% significance level.

4. The correlation coefficient for the sample is -0.713. This is negative, and has a magnitude greater than the critical value ($|-0.713| > 0.5494$).

4. The correlation coefficient for the sample is -0.713, and $-0.713 < 0.5494$.

5. Hence, we do not reject the null hypothesis. There is not significant evidence for negative correlation between an athlete's time and the amount of rainfall.

5. Hence, we reject the null hypothesis. There is evidence that an athlete's time and the amount of rainfall have negative correlation.

Part C Any (linear) correlation

An author wonders whether the amount of time their cat sits next to them is correlated with the number of words they write during the day. Over fifty-three days, the author records the number of words they write and for how long the cat sits nearby, and finds $r = 0.3300$. Test the data at the 1% significance level.

Choose from the options below to construct a complete hypothesis test.

- [ ] This question is looking for correlation in either direction. A two-tailed test is needed.

$$\mathrm{H}_0 : \rho = 0 \qquad \mathrm{H}_1 : \rho \neq 0$$

- [ ] This question is looking for positive correlation. A one-tailed test is needed.

$$\mathrm{H}_0 : \rho = 0 \qquad \mathrm{H}_1 : \rho > 0$$

- [ ] For a one-tailed test, the critical value of the correlation coefficient for a sample of size 53 is 0.3188 at the 1% significance level.
- [ ] For a two-tailed test, the critical value of the correlation coefficient for a sample of size 53 is 0.3509 at the 1% significance level.
- [ ] The correlation coefficient for the sample is 0.3300, and $0.3300 < 0.3509$.
- [ ] The correlation coefficient for the sample is 0.3300, and $0.3300 > 0.3188$.
- [ ] Hence, we do not reject the null hypothesis. There is not significant evidence that the number of words the author writes is correlated with the amount of time their cat sits near them.
- [ ] Hence, we reject the null hypothesis. There is evidence that the number of words the author writes is correlated with the amount of time their cat sits near them.

Created for isaacphysics.org by Jonathan Waugh

Gameboard:

STEM SMART Double Maths 33 - Regression, Correlation & Hypothesis Tests

Correlation Hypothesis Testing 2

A Level

P P P

A town planner believes that on summer weekday afternoons the amount of traffic into the centre of their town is higher when the temperature is higher. They want to test this hypothesis at the 1% significance level.

Every weekday (Monday to Friday) for six weeks they monitor the traffic on the main roads into town, and record the mean afternoon temperature, and they find that the correlation coefficient is -0.4517.

Part A Initial conclusion

Without doing any calculations, which of these statements can the town planner make? Choose all that apply.

- [ ] The correlation coefficient is negative, so there is no evidence that the amount of traffic is positively correlated with temperature.
- [ ] The correlation coefficient is negative. There may be a negative correlation between the amount of traffic and afternoon temperature.
- [ ] The correlation coefficient of their sample is negative, so there is a negative correlation between the amount of traffic on summer afternoons and temperature.
- [ ] The correlation coefficient is negative, so there is definitely no positive correlation between the amount of traffic and temperature.
- [ ] There is no evidence that the amount of traffic and afternoon temperature are correlated.

Part B Choosing a hypothesis test

Using the given data, which of the following hypothesis tests would it be most useful for the town planner to carry out?

- A hypothesis test to see if the amount of traffic and afternoon temperature are negatively correlated at the 1% significance level.
- A hypothesis test to see if the amount of traffic and afternoon temperature are negatively correlated at the 20% significance level.
- A hypothesis test to see if the amount of traffic and afternoon temperature are negatively correlated at the 50% significance level.

Part C Null and alternative hypotheses

The town planner carries out the most useful test listed in part B.

Drag and drop symbols into the spaces below to state the null and alternative hypotheses for this test, where ρ represents the population correlation coefficient and r represents the sample correlation coefficient.

H_0: $\square\square\square$

H_1: $\square\square\square$

Items:

ρ r $<$ $=$ $>$ 0 1

Part D Carrying out the test

Carry out the hypothesis test, and make a conclusion. Then fill in the blanks below.

The critical value of the correlation coefficient is ______.

Comparing the town planner's value to the critical value gives ______.

Therefore, ______ the null hypothesis. There ______ significant evidence that there is a ______ correlation between the amount of traffic in summer and afternoon temperature.

Items:

0.4226 | 0.8822 | 0.9343 | $|-0.4517| > 0.4226$ | $|-0.4517| < 0.9343$ | $|-0.4517| < 0.8822$ | **reject** | **do not reject** | **is** | **may be** | **is not** | **negative** | **positive** | **no**

Created for isaacphysics.org by Jonathan Waugh

Gameboard:

STEM SMART Double Maths 33 - Regression, Correlation & Hypothesis Tests

Spearman's Rank Test 1

Further A
P P P

In an examination consisting of 4 separate subjects, a group of 10 students obtained the following marks. The overall mark was out of 375, and the Physics and Chemistry marks were each out of 100.

Student	A	B	C	D	E	F	G	H	I	J
Overall	274	255	246	245	229	228	219	213	205	176
Physics	76	77	67	65	58	60	52	63	47	45
Chemistry	82	–	65	67	–	64	68	54	51	38

Find Spearman's rank correlation coefficients between each pair of data to test whether there is evidence of a positive correlation between them.

Part A Overall and Physics marks

Find Spearman's rank correlation coefficient r_s for the relationship between the overall mark and the Physics mark, giving your answer to 3 s.f.

Part B Overall and Physics significance

Test at the 1% significance level whether there is a positive association between the overall mark and that obtained in Physics.

Find the appropriate critical value for Spearman's rank correlation coefficient, giving your answer to 3 s.f.

What do you conclude about whether there is a positive association at the 1% significance level between the overall mark and that obtained in Physics?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Spearman's rank correlation coefficient is [ ] the calculated value; the null hypothesis that there is [ ] association between the two values can therefore be [ ] at the 1% level and there is [ ] evidence for a positive association at this level.

Items:

accepted | negative | no | positive | significant | rejected | less than | equal to | greater than

Part C Overall and Chemistry marks

Find Spearman's rank correlation coefficient r_s for the relationship between the overall mark and the Chemistry mark, giving your answer to 3 s.f.

Part D Overall and Chemistry significance

Test at the 1% significance level whether there is a positive association between the overall mark and that obtained in Chemistry.

Find the appropriate critical value for Spearman's rank correlation coefficient, giving your answer to 3 s.f.

What do you conclude about whether there is a positive association at the 1% significance level between the overall mark and that obtained in Chemistry?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Spearman's rank correlation coefficient is ______ the calculated value; the null hypothesis that there is ______ association between the two values cannot therefore be ______ at the 1% level, and there is no ______ evidence for a positive association at this level.

Items:

greater than **rejected** **significant** **less than** **equal to** **negative** **positive** **accepted** **no**

Part E Physics and Chemistry marks

Find Spearman's rank correlation coefficient r_s for the relationship between the Physics mark and the Chemistry mark, giving your answer to 3 s.f.

Part F Physics and Chemistry significance

Test at the 1% significance level whether there is a positive association between the Physics mark and that obtained in Chemistry.

Find the appropriate critical value for Spearman's rank correlation coefficient, giving your answer to 3 s.f.

What do you conclude about whether there is a positive association at the 1% significance level between the mark obtained in Physics and that obtained in Chemistry?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Spearman's rank correlation coefficient is ______ the calculated value; the null hypothesis that there is ______ association between the two values cannot therefore be ______ at the 1% level, and there is no ______ evidence for a positive association at this level.

Items:

negative | greater than | no | significant | rejected | equal to | less than | accepted | positive

---

Created for isaacphysics.org by Julia Riley

Gameboard:

STEM SMART Double Maths 33 - Regression, Correlation & Hypothesis Tests

Using Pearson's PMCC 1

Further A

Find Pearson's product moment correlation coefficient for the following sets of statistics.

Part A Data set 1

Find the correlation coefficient for the following set of statistics, giving your answer to 3 s.f.

$\Sigma x = 275$, $\Sigma x^2 = 10781$, $\Sigma y = 251$, $\Sigma y^2 = 8407$, $\Sigma xy = 7842$, $n = 10$

Part B Data set 2

Find the correlation coefficient for the following set of statistics, giving your answer to 3 s.f.

$\Sigma x = 907$, $\Sigma x^2 = 105944$, $\Sigma y = 289$, $\Sigma y^2 = 10916$, $\Sigma xy = 14929$, $n = 12$

Part C Data set 3

Find the correlation coefficient for the following set of statistics, giving your answer to 3 s.f.

$\Sigma(x-\bar{x})^2 = 1592$, $\Sigma(y-\bar{y})^2 = 2473$, $\Sigma(x-\bar{x})(y-\bar{y}) = -1280$, $n = 10$

Created for isaacphysics.org by Julia Riley

Using Pearson's PMCC 2

Further A

P P P

An electrical device has been designed and is being tested. The summary statistics for the relationship between the input and output voltages (V_i and V_o respectively) from the device are as follows.

$$\Sigma V_\mathrm{i} = 302,\ \Sigma V_\mathrm{i}^2 = 15532,\ \Sigma V_\mathrm{o} = 228,\ \Sigma V_\mathrm{o}^2 = 10980,\ \Sigma V_\mathrm{i} V_\mathrm{o} = 12296,\ n = 6$$

Part A The correlation coefficient

Find Pearson's product moment correlation coefficient for the set of statistics, giving your answer to 3 s.f.

Part B Significance of the correlation between input and output

Test at the 1% significance level whether there is a correlation between the input and output voltages.

Find the appropriate critical value for Pearson's product moment correlation coefficient, giving your answer to 3 s.f.

What do you conclude about whether there is a correlation at the 1% significance level between the input and output voltages?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Pearson's product moment correlation coefficient is ____ the calculated value; the null hypothesis that there is ____ correlation between the two values can therefore be ____ at the 1% level and there is ____ evidence for a correlation at this level.

Items:

rejected | equal to | accepted | less than | no | positive | greater than | significant | negative

Part C The regression line

Calculate the equation of the regression line $V_\mathrm{o} = a + bV_\mathrm{i}$ relating the output voltage (V_o) to the input voltage (V_i).

Find b giving your answer to 3 s.f.

Find a, giving your answer to 3 s.f.

Correlation - the Photoelectric Effect

A Level: C C C Further A: P P P

Millikan carried out experiments to test the photoelectric effect equation that Einstein proposed in 1905. He measured what was called the stopping potential V as a function of the frequency f of light falling on a photocathode. One set of data he obtained is as follows.

Frequency f $(10^{14}\mathrm{Hz})$	5.49	6.91	7.41	8.22	9.58
Voltage V (V)	0.46	1.04	1.16	1.61	2.19

Part A The correlation coefficient

Find Pearson's product moment correlation coefficient for these data, giving your answer to 3 s.f.

Part B The equation of the regression line

Calculate the equation of the regression line $V = a + bf$ relating the stopping potential V to the frequency f.

Find b, giving your answer to 4 s.f.

Find a, giving your answer to 4 s.f.

Part C The Planck constant and work function

Einstein's equation is commonly written in the form

$$eV = hf - \phi$$

where e is the charge on an electron ($1.6 \times 10^{-19}\,\mathrm{C}$), h is the Planck constant and ϕ is called the work function of the material used for the photocathode (in this case sodium). Given V is in volts and the frequency is in units of $10^{14}\,\mathrm{Hz}$, deduce from your answers in Part B a value for h in $\mathrm{J\,s}$ and ϕ in J.

Find the value of the Planck constant h, giving your answer to 2 s.f.

Find the value of the work function ϕ, giving your answer to 2 s.f.

Created for isaacphysics.org by Julia Riley

Gameboard:

STEM SMART Double Maths 33 - Regression, Correlation & Hypothesis Tests

Correlation - Attenuation of Gamma Rays

A Level C C C Further A P P P

An experiment to investigate the attenuation of gamma rays by lead was carried out by measuring the count rate as a function of the number of lead sheets placed in front of a gamma ray source. The following results were obtained.

Number of lead sheets	0	1	2	3	4	5	6
Gamma ray count rate	13106	4301	1469	474	163	63	17

Part A Rank correlation coefficient

Write down Spearman's rank correlation coefficient for these data.

Part B Linear product moment correlation coefficient

Find the Pearson product moment correlation coefficient between the two sets of data, giving your answer to 4 s.f.

Part C Significance of the linear product moment correlation coefficient

Test at the 1% significance level whether there is a negative correlation between the count rate and the number of lead sheets.

Find the appropriate critical value for Pearson's product moment correlation coefficient, giving your answer to 4 s.f.

What do you conclude about whether there is a negative correlation at the 1% significance level between the count rate and the number of lead sheets?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Pearson's product moment correlation coefficient is ______ the modulus of the calculated value; the ______ hypothesis cannot be rejected so there is ______ evidence for a correlation at the 1% level.

Items:

accepted **alternative** **equal to** **no significant** **null** **sufficient** **rejected** **greater than** **less than**

Part D Logarithmic product moment correlation coefficient

Consider now the relationship between the natural log of the counts and the number of lead sheets.

Find the Pearson product moment correlation coefficient between the natural log of the count rate and the number of lead sheets, giving your answer to 4 s.f.

Part E Significance of the logarithmic product moment correlation coefficient

Test at the 1% significance level whether there is a negative correlation between the natural log of the count rate and the number of lead sheets.

Find the appropriate critical value for Pearson's product moment correlation coefficient, giving your answer to 4 s.f.

What do you conclude about whether there is a negative correlation at the 1% significance level between the natural log of the count rate and the number of lead sheets?

Choose the correct words to fill in the following sentence.

The appropriate critical value for Pearson's product moment correlation coefficient is ______ the modulus of the calculated value; the ______ hypothesis can be rejected at the 1% level and there is ______ evidence for a negative correlation at this level.

Items:

alternative | greater than | rejected | significant | less than | accepted | no sufficient | null | equal to

---

Part F The equation of the regression line

Calculate the equation of the regression line

$$\ln N = a + bm$$

where N is the gamma ray count and m is the number of lead sheets.

Find the value of b, giving your answer to 4 s.f.

Find the value of a, giving your answer to 4 s.f.

Part G Attenuation constant

Theory predicts that the relationship expected between the number of counts N and the number of lead sheets m is given by

$$N = N_0 e^{-\alpha m}$$

where α is the attenuation constant and N_0 is a constant.

Deduce the value of α, giving your answer to 4 s.f.

Find the value of N_0, giving your answer to 4 s.f.

Created by for isaacphysics.org by Julia Riley