Probability 5.1

A Level

Find the following probabilities.

Part A $P(Y)$ and $P(Y|X)$

It is given that $P(X) = 0.3$, $P(X \cup Y) = 0.6$ and $P(X \cap Y) = 0.2$.

Find $P(Y)$.

Find $P(Y|X)$, giving your answer as an exact fraction.

Part B $P(C \cap D)$ and $P(C|D')$

It is given that $P(C) = 0.6$, $P(D) = 0.5$ and $P((C \cup D)') = 0.3$.

Find $P(C \cap D)$.

Find $P(C|D')$.

Probabilities: Employment

A Level

Data about employment of people in their thirties and forties in a small rural area are shown in the following table.

	Unemployed	Employed
Thirties	206	412
Forties	358	305

A person from this area in these age groups is chosen at random. Let T be the event that the person is in their thirties and let E be the event that the person is employed.

Part A $\mathrm{P}(\mathrm{T})$

Find $\mathrm{P}(\mathrm{T})$.

Part B $\mathrm{P}(\mathrm{T} \text{ and } \mathrm{E})$

Find $\mathrm{P}(\mathrm{T} \text{ and } \mathrm{E})$.

Part C Independent events?

Are T and E independent events? Fill in the blanks below to complete the argument.

If T and E are independent, P(E|T) = [ ], i.e. the probability of being unemployed is [ ] irrespective of age.
Using the values in the table, P(E|T) = [ ] and [ ]. Therefore T and E [ ] independent events.

Items:

$\frac{412}{717}$ | $P(E) = \frac{206}{427}$ | $P(E) = \frac{239}{427}$ | **the same** | **are not** | $\frac{412}{1281}$ | P(T) | $\frac{2}{3}$ | P(E) | **are** | **random**

Part D Unemployed and in their thirties

Given that the person chosen is unemployed, find the probability that the person is in their forties.

Probability 5.3

A Level

Part A Substandard samples

A laboratory has two devices A and B which produce samples for an experiment. Device A has produced 100 samples of which 5% are substandard. Device B has produced 25 of which 4% are substandard. An experimenter has found a substandard sample. Assuming that samples are chosen at random, what is the probability that it was produced by device B? Give your answer to 3 significant figures.

Part B Equipment failure

In hot weather the antiquated air-conditioning system in Professor A's laboratory may break down. On any given hot day, there is a 5% chance that the air-conditioning system breaks down. If the air-conditioning breaks down, the probability this will lead to the Professor's equipment failing by the end of the day is 0.3. If the air-conditioning does not break down, the probability that the equipment fails by the end of the day is only 0.05.

One hot day the Professor checks their lab first thing in the morning and the air-conditioning and equipment are both working. When the Professor gets ready to leave at the end of the day, they notice that their equipment has failed. What is the probability that the failure was not due to a breakdown of the air-conditioning system?

Created by for isaacphysics.org by Julia Riley

Probability 5.4

A Level

The probability of a randomly selected person in a population having a particular genetic trait is 0.00001. A test for this trait successfully detects it, if present, 99.9% of the time, and only returns a false positive 0.1% of the time.

Part A Probability after one test

A person tests positive for the trait. Find the probability that they actually have the genetic trait. Give your answer to 3 significant figures.

Part B Probability after two tests

A second test is carried out on an individual who tested positive for the trait after one test.

What is the probability that an individual who takes two tests receives a positive result from both tests? Give your answer to 3 significant figures.

Find the probability, given that they have tested positive a second time, that they actually have the genetic trait. Give your answer to 3 significant figures.

Created by for isaacphysics.org by Julia Riley

Linear regression 3.3

A Level

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 $1\,\mathrm{Mpc} = 10^6\,\mathrm{pc}$.)

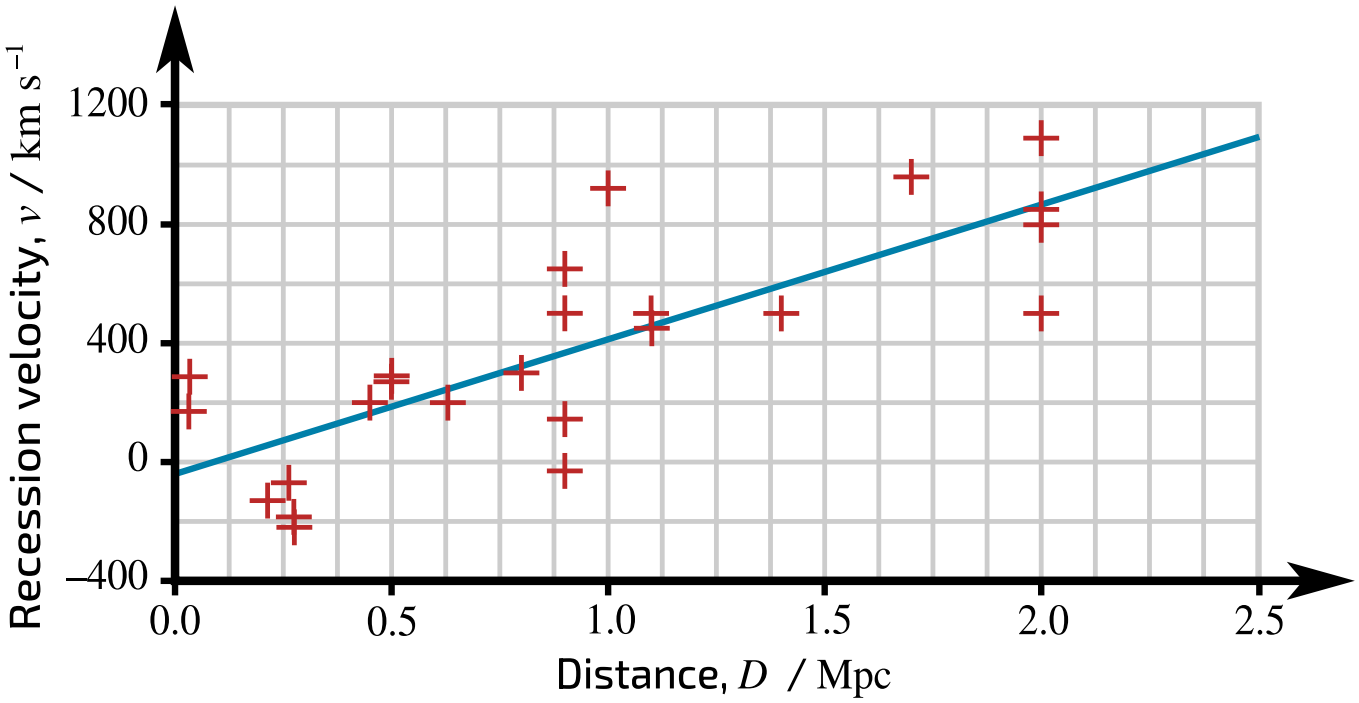

Figure 1: A graph of Hubble's original data relating the recession velocity v of a galaxy to its distance D from us. The regression line of best fit is shown.

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?

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

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).

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

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

Correlation Hypothesis Testing 1

A Level

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.

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

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

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

Items:

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

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

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

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 Isaac Physics by Jonathan Waugh

Spearman's Rank Test 1

Further A

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 correlation 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 correlation 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 hypothesis that there is [ ] correlation between the two values can therefore be [ ] at the 1% level and there is [ ] evidence for a positive correlation at this level.

Items:

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

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 correlation 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 correlation 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 hypothesis that there is [ ] correlation between the two values cannot therefore be [ ] at the 1% level, and there is no [ ] evidence for a positive correlation at this level.

Items:

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

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 correlation 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 correlation 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 hypothesis that there is [ ] correlation between the two values cannot therefore be [ ] at the 1% level, and there is no [ ] evidence for a positive correlation at this level.

Items:

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

Created for isaacphysics.org by Julia Riley

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.

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

Using Pearson's PMCC 2

Further A

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.

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

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 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:

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

Part C The regression line

Calculate the equation of the regression line $V_\text{o} = a + bV_\text{i}$ 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.

Created for isaacphysics.org by Julia Riley

Correlation - Attenuation of Gamma Rays

A Level Further A

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:

less than **alternative** **equal to** **null** **accepted** **sufficient** **greater than**

no significant **rejected**

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:

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

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