Probability 5.1

A Level

Find the following probabilities.

Part A $\mathsf{P}(\mathsf{Y})$ and $\mathsf{P}(\mathsf{Y}|\mathsf{X})$

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

Find $\mathsf{P}(\mathsf{Y})$.

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

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

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

Find $\mathsf{P}(\mathsf{C} \cap \mathsf{D})$.

Find $\mathsf{P}(\mathsf{C}|\mathsf{D}')$.

Created by for isaacphysics.org by Julia Riley

Probability 5.2

A Level

Consider the situation in which $\mathsf{P}(\mathsf{X}) = 0.3$, $\mathsf{P}(\mathsf{X} \cup \mathsf{Y}) = 0.7$ and $\mathsf{P}(\mathsf{Y}) = k$. Find the value of k in the following situations.

Part A X and Y mutually exclusive

Find the value of k if X and Y are mutually exclusive.

Part B X and Y independent

Find the value of k if X and Y are independent. Give your answer as an exact fraction.

Created by for isaacphysics.org by Julia Riley

All materials on this site are licensed under the **Creative Commons license**, unless stated otherwise.

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 $\mathsf{P}(\mathsf{T})$

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

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

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

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, $\mathrm{P}(\mathrm{E}|\mathrm{T})$ = ______, i.e. the probability of being unemployed is ______ irrespective of age.

Using the values in the table, $\mathrm{P}(\mathrm{E}|\mathrm{T})$ = ______ and ______. Therefore T and E ______ independent events.

Items:

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

All materials on this site are licensed under the **Creative Commons license**, unless stated otherwise.

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

All materials on this site are licensed under the **Creative Commons license**, unless stated otherwise.

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 3.2

A Level

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 $\frac{1}{T} = 0.002\,\mathrm{K}^{-1} = 2 \times 10^{-3}\,\mathrm{K}^{-1}$.

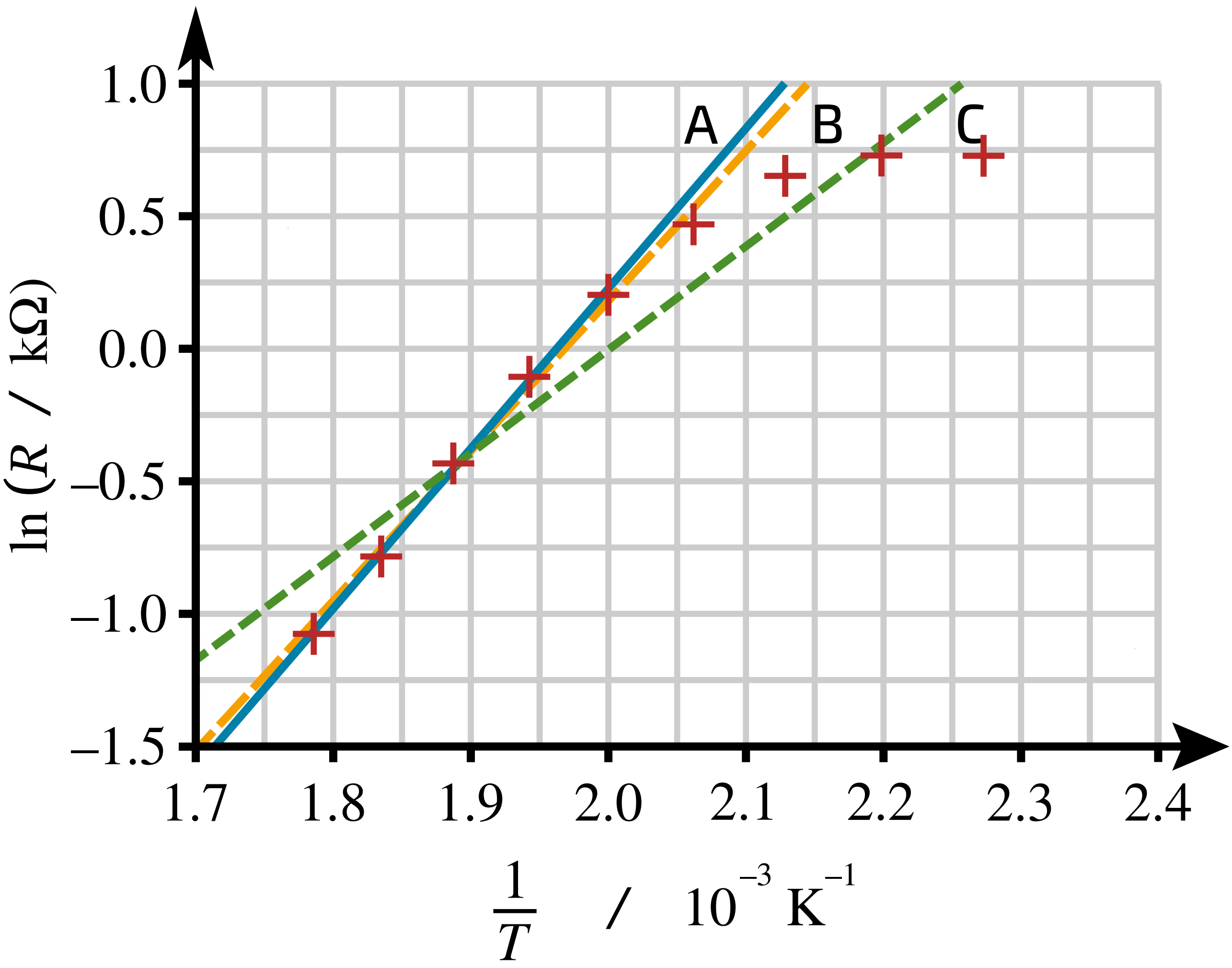

Figure 1: A plot of $\ln R$ against $\frac{1}{T}$; three best fit lines, A (blue, solid), B (yellow, long-dashed) and C (green, short-dashed) fitted to different ranges of the data, are shown.

Three best fit lines 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. Give your answer in the order $\mathrm{I}, \mathrm{II}, \mathrm{III}$; for example, if you think $\mathrm{I} \equiv$ A, $\mathrm{II} \equiv$ B and $\mathrm{III} \equiv$ C, type A,B,C with no spaces.

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

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

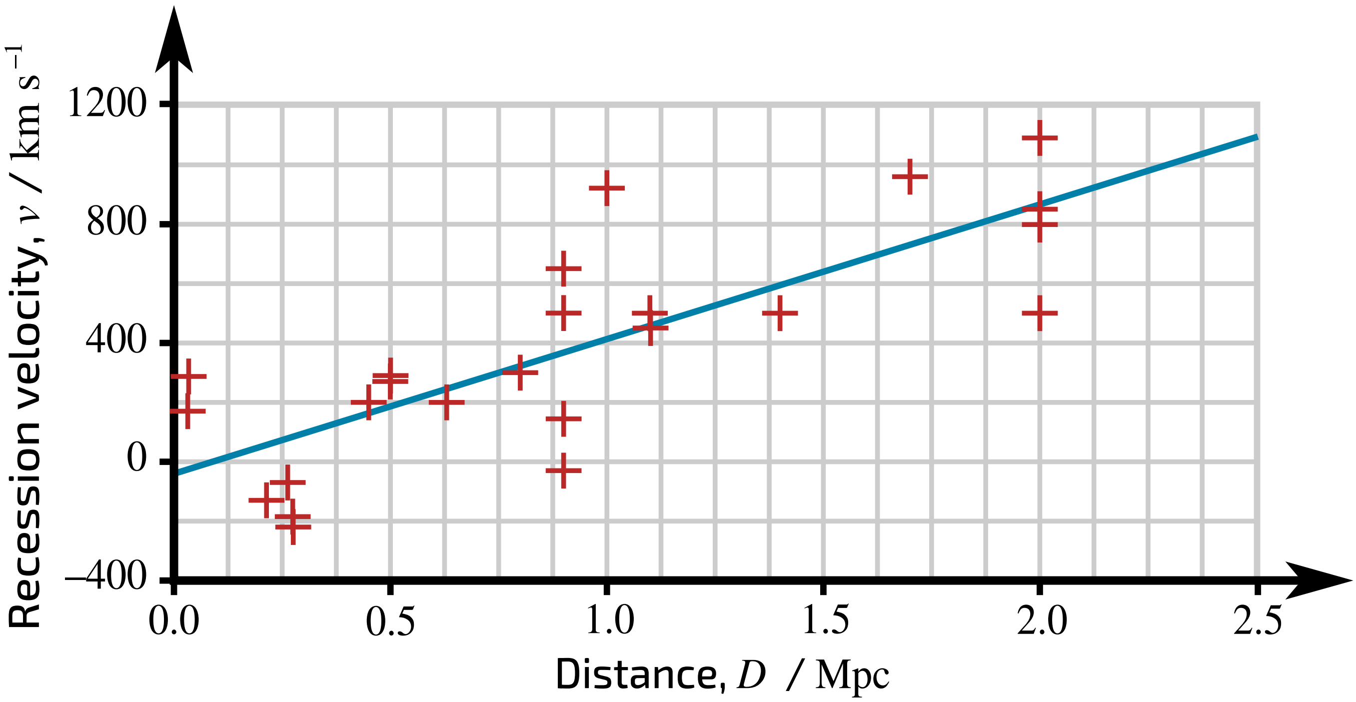

Figure 1: A graph of Hubble's original data relating the recession velocity v of a galaxy to its distance D from us. The regression line of best fit is shown.

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

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

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

All materials on this site are licensed under the **Creative Commons license**, unless stated otherwise.

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 ____ 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:

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

All materials on this site are licensed under the **Creative Commons license**, unless stated otherwise.

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. 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.
- [ ] There is no evidence that the amount of traffic and afternoon temperature are correlated.
- [ ] The correlation coefficient is negative, so there is definitely no positive correlation between the amount of traffic and temperature.
- [ ] The correlation coefficient is negative, so there is no significant evidence that the amount of traffic is positively correlated with temperature.

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.

H_0: | | |

H_1: | | |

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 ______ correlation between the amount of traffic in summer and afternoon temperature.

Items:

0.8822 | 0.4226 | **may be** | **positive** | **is not** | $|-0.4517| < 0.8822$ | **reject**

$|-0.4517| > 0.4226$ | **negative** | 0.9343 | **do not reject** | **is** | **no** | $|-0.4517| < 0.9343$

Created for Isaac Physics by Jonathan Waugh

All materials on this site are licensed under the **Creative Commons license**, unless stated otherwise.