1) Regression, correlation and hypothesis testing

1.1) Exponential models

Chapter CONTENTS

Worked example

The table shows some data collected on the temperature, in °C, of a colony of bacteria (t) and its growth rate (g).

Temperature, t (°C)	3	5	6	8	9	11
Growth rate, g	1.40	1.94	1.97	2.85	3.2	4.64

The data are coded using the changes of variable $x = t$ and $y = \log g$. The regression line of y on x is found to be $y = -0.0536 + 0.0637x$.

a) Find the initial growth rate
b) Given that the data can be modelled by an equation of the form $g = kb^t$ where k and b are constants, find the values of k and b.

Your turn

The table shows some data collected on the temperature, in °C, of a colony of bacteria (t) and its growth rate (g).

Temperature, t (°C)	3	5	6	8	9	11
Growth rate, g	1.04	1.49	1.79	2.58	3.1	4.46

The data are coded using the changes of variable $x = t$ and $y = \log g$. The regression line of y on x is found to be $y = -0.2215 + 0.0792x$.

a) Find the initial growth rate
b) Given that the data can be modelled by an equation of the form $g = kb^t$ where k and b are constants, find the values of k and b.

a) 0.6
b) $k = 0.6$, $b = 1.20$ (3 sf)

Worked example	Your turn
A rabbit population, P, is modelled with respect to time in years, t. An exponential model is proposed: $$P = kb^t$$ The data is coded using $x = t$ and $y = \log P$. The regression line of y on x is found to be $y = 3 + 0.2x$. Determine the values of k and b.	A rabbit population, P, is modelled with respect to time in years, t. An exponential model is proposed: $$P = kb^t$$ The data is coded using $x = t$ and $y = \log P$. The regression line of y on x is found to be $y = 2 + 0.3x$. Determine the values of k and b. $k = 100, b = 2.00$ (3 sf)

1.2) Measuring correlation

Worked example	Your turn

Calculate the product moment correlation coefficient for the following data:

x	y
1	3
2	4
3	5
4	8

Calculate the product moment correlation coefficient for the following data:

x	y
1	3
2	6
3	5
4	8

$r = 0.868$ (3 sf)

Worked example

From the large data set, the daily mean temperature, t °C, and the daily total rainfall, r mm, were recorded from 27th May to 5th June inclusive 1987 in Leuchars.

Day	1	2	3	4	5	6	7	8	9	10
t	8.5	9.0	10.3	12.8	13.5	12.8	9.8	8.8	10.0	10.4
r	0	2.4	8.1	0.2	0.4	tr	6.1	3.6	tr	31.8

a) State the meaning of tr in the table above.

b) Calculate the product moment correlation coefficient for the ten days, stating clearly how you deal with the 'tr' readings.

c) With reference to your answer to part b, comment on the suitability of a linear regression model for these data.

Your turn

From the large data set, the daily mean windspeed, w knots, and the daily maximum gust, g knots, were recorded for the first 10 days in September in Hurn in 1987.

Day	1	2	3	4	5	6	7	8	9	10
w	4	4	8	7	12	12	3	4	7	10
g	13	12	19	23	33	37	10	n/a	n/a	23

a) State the meaning of n/a in the table above.

b) Calculate the product moment correlation coefficient for the remaining 8 days.

c) With reference to your answer to part b, comment on the suitability of a linear regression model for these data.

a) Data on daily maximum gust is not available for these days

b) $r = 0.9533$ (4 sf)

c) r is close to 1 so there is strong positive correlation between daily mean windspeed and daily maximum gust. This means that the data points lie close to a straight line, so a linear regression model is suitable.

1.3) Hypothesis testing for zero correlation

Chapter CONTENTS

Worked example	Your turn
A scientist takes 19 observations of the masses of two reactants in an experiment. She calculates a product moment correlation coefficient of $r = 0.54$. The scientist believes there is a positive correlation between the masses of the two reactants. Test at the 1% level of significance, the scientist's claim, stating your hypotheses clearly.	A scientist takes 14 observations of the masses of two reactants in an experiment. She calculates a product moment correlation coefficient of $r = -0.45$. The scientist believes there is a negative correlation between the masses of the two reactants. Test at the 5% level of significance, the scientist's claim, stating your hypotheses clearly. $H_0: \rho = 0$ $H_1: \rho < 0$ ∴ One-tailed test. Sample size $= 14$ Significance level in tail $= 5\%$ Reject H_0 if $r < -0.4575$ $r = -0.45 > -0.4575$ The result is not significant. Insufficient evidence to reject H_0. Insufficient evidence to suggest there is a negative correlation between the masses of the two reactants.

Worked example	Your turn
A scientist takes 20 observations of the masses of two reactants in an experiment. She calculates a product moment correlation coefficient of $r = 0.54$. The scientist believes there is no correlation between the masses of the two reactants. Test at the 1% level of significance, the scientist's claim, stating your hypotheses clearly.	A scientist takes 30 observations of the masses of two reactants in an experiment. She calculates a product moment correlation coefficient of $r = -0.45$. The scientist believes there is no correlation between the masses of the two reactants. Test at the 10% level of significance, the scientist's claim, stating your hypotheses clearly. $H_0: \rho = 0$ $H_1: \rho \neq 0$ ∴ Two-tailed test. Sample size $= 30$ Significance level in each tail $= 5\%$ Reject H_0 if $r < -0.3061$ $r = -0.45 < -0.3061$ The result is significant. Sufficient evidence to reject H_0. Sufficient evidence to suggest there is a correlation between the masses of the two reactants.

Worked example	Your turn

Worked example

The table from the large data set shows the daily mean temperature, $t\ °C$, and the daily total rainfall, $r\ mm$, in Leuchars for a sample of nine days in October 1987.

t	11.4	10.5	6.5	8.3	8.2	5.7	7.6	12.1	11.2
r	0	1	3.9	16.3	7.9	4.1	15.2	0	tr

Test, at the 10% level of significance, whether there is evidence of a negative correlation between daily mean temperature and daily total rainfall.
State your hypotheses clearly.

Your turn

The table from the large data set shows the daily maximum gust, x knots, and the daily maximum relative humidity, $y\%$, in Leeming for a sample of eight days in May 2015.

x	31	28	38	37	18	17	21	29
y	99	94	87	80	80	89	84	86

Test, at the 10% level of significance, whether there is evidence of a positive correlation between daily maximum gust and daily maximum relative humidity. State your hypotheses clearly.

$H_0: \rho = 0$
$H_1: \rho > 0$ ∴ One-tailed test.
Sample size $= 8$
Significance level in each tail $= 10\%$
Reject H_0 if $r > 0.5067$
$r = 0.1149 < 0.5067$
The result is not significant.
Insufficient evidence to reject H_0.
Insufficient evidence to suggest there is a positive correlation between daily maximum gust and daily maximum relative humidity.