

### Question 5

a)  $\log L = n_1 \cdot \log p_1 + n_2 \cdot \log p_2 \rightarrow$

OD 1: 5 times route 1, 5 times route 2  $\rightarrow \log L = 5 \cdot \log(0.4) + 5 \cdot \log(0.6) = -3.10$

OD 2: 8 times route 1, 2 times route 2  $\rightarrow \log L = 8 \cdot \log(0.9) + 2 \cdot \log(0.1) = -2.37$

b) Maximum likelihood: likelihood of obtaining these observations given the best model (with the maximum likelihood) :

OD 1:  $p_1 = 5/10 = 0.5$ ,  $n_2 = 5/10 = 0.5 \rightarrow \log L_{max} = 5 \cdot \log(0.5) + 5 \cdot \log(0.5) = -3.01$

OD 2:  $p_1 = 8/10 = 0.8$ ,  $n_2 = 2/10 = 0.2 \rightarrow \log L_{max} = 8 \cdot \log(0.8) + 2 \cdot \log(0.2) = -2.17$

c) Model reliability according to likelihood ratio  $\rightarrow \log L_{max} - \log L$

OD 1:  $\log L_{max} - \log L = -3.01 - (-3.10) = 0.09$

OD 2:  $\log L_{max} - \log L = -2.17 - (-2.37) = 0.20$

Conclusion: likelihood ratio is closer to 1 for OD 1 (model result is closer to best possible model), and therefore this model result is more reliable

d) Observed: number of travellers per OD is fixed (40 for OD 1 and 60 for OD 2)

observed: OD 1  $\rightarrow n_{obs,OD1,1} = 15$ ,  $n_{obs,OD1,2} = 25$

observed: OD 2  $\rightarrow n_{obs,OD2,1} = 50$ ,  $n_{obs,OD2,2} = 10$

model: OD 1: 40 travellers  $\rightarrow n_{model,OD1,1} = 40 \cdot 0.4 = 16$ ,  $n_{model,OD1,2} = 40 \cdot 0.6 = 24$

model: OD 2: 60 travellers  $\rightarrow n_{model,OD2,1} = 60 \cdot 0.9 = 54$ ,  $n_{model,OD2,2} = 60 \cdot 0.1 = 6$

Frequencies (special case of chi-square)  $\rightarrow$

$$\chi^2 = \sum_{\text{all frequency bins}} ((n_{obs} - n_{model})^2 / n_{model}) =$$

$$(15 - 16)^2 / 16 + (25 - 24)^2 / 24 + (50 - 54)^2 / 54 + (10 - 6)^2 / 6 = 3.07$$

Degrees of freedom =  $4 - 2 = 2$ , because 4 frequency bins, but for each OD pair, the total number of travellers is fixed, thus 1 degree of freedom per OD pair

95% confidence level  $\rightarrow$  critical value = 5.99  $\rightarrow 3.07 < 5.99 \rightarrow$  The model is not rejected

### Question 6

a)  $q_{model} = q_0 [1 + a \sqrt{t_p}]$

$$err_x(q_{model}) = (\partial q_{model} / \partial x) \cdot err(x).$$

$$\rightarrow err_{q_0}(q_{model}) = [1 + a \sqrt{t_p}] \cdot err(q_0) = q_0 [1 + a \sqrt{t_p}] \cdot err(q_0) / q_0$$

$$\rightarrow err_a(q_{model}) = q_0 a \sqrt{t_p} \cdot err(a) = q_0 a \sqrt{t_p} \cdot err(a) / a$$

b) Error in qmodel expressed as relative errors in q0 and a respectively (see also a)

$$\rightarrow err_{q_0}(q_{model}) = q_0 [1 + a \sqrt{t_p}] \cdot err(q_0) / q_0 = [q_0 + q_0 a \sqrt{t_p}] \cdot err(q_0) / q_0$$

$$\rightarrow err_a(q_{model}) = q_0 a \sqrt{t_p} \cdot err(a) / a$$

If the relative errors in  $q_0$  and  $a$  are the same, say  $100 \cdot x\%$  ( $> 0$ ), then the error in  $q_0$  always has the largest influence on the total error, because:

$$err_a(q_{model}) = q_0 a \sqrt{t_p} \cdot x$$

$$err_{q_0}(q_{model}) = [q_0 + q_0 a \sqrt{t_p}] \cdot x = err_a(q_{model}) + q_0 \cdot x > err_a(q_{model})$$

c) Uncorrelated  $\rightarrow err(q_{model}) = \sqrt{(err_{q0}(q_{model})^2 + err_a(q_{model})^2)}$   
 $= \sqrt{([1 + a\sqrt{t_p}] * err(q_0))^2 + (q_0\sqrt{t_p} * err(a))^2}$

d) Negatively correlated  $\rightarrow err(q_{model}) = err_{q0}(q_{model}) - err_a(q_{model})$   
 $= [1 + a\sqrt{t_p}] * err(q_0) - q_0\sqrt{t_p} * err(a)$   
 $= [q_0 + q_0a\sqrt{t_p}] * err(q_0)/q_0 - q_0a\sqrt{t_p} * err(a)/a$

Note: minus sign because individual errors in  $q_0$  and  $a$  are negatively correlated. In the expression, the errors in  $q_0$  and  $a$  are absolute “mean” errors (usually the standard deviations of the respective distributions of individual errors), both with a positive value.

e) In c) the errors add up quadratically, while in d) the errors are cancelling each other out due to the negative correlation between (the individual errors in)  $q_0$  and  $a$ , and due to the fact that both the  $err(q_0)$  and  $err(a)$  have a positive influence on  $err_{q0}(q_{model})$  and  $err_a(q_{model})$  respectively (a higher value of  $q_0$  or  $a$  will result in a higher value of  $q_{model}$ ). The exception is when there is no rainfall, because then  $err_a(q_{model}) = 0$ . Thus, in general, the error is larger in c) than in d), except when there is no rainfall. If  $t_p=0$ , both errors are equally large  $\rightarrow err(q_{model}) = err(q_0)$ .

f)  $err(q_{model}) = [q_0 + q_0a\sqrt{t_p}] * err(q_0)/q_0 - q_0a\sqrt{t_p} * err(a)/a$   
 $= [q_0 + (1/3)*q_0\sqrt{t_p}] * 0.05 - (1/3)q_0\sqrt{t_p} * 0.1 = 0.05q_0[1 - (1/3)\sqrt{t_p}]$   
 $t_p = 0 \rightarrow err(q_{model}) = 0.05*q_0$   
 $t_p = 1 \rightarrow err(q_{model}) = 0.05*(1 - (1/3))*q_0 = 0.033*q_0$   
 $t_p = 9 \rightarrow err(q_{model}) = 0.05*((1 - 3*(1/3))) * q_0 = 0$

According to c),  $err_{q0}(q_{model})$  and  $err_a(q_{model})$  have opposite effects on  $err(q_0)$  because of the negative correlation between the (individual errors) in  $q_0$  and  $a$ , i.e. the errors in  $q_0$  and  $a$  weaken each other. Because  $err(a)/a > err(q_0)/q_0$ , the weakening effect of  $err_a(q_{model})$  on  $err(q_{model})$  increases with  $t_p$ . For  $t_p = 9$ , the errors  $err_{q0}(q_{model})$  and  $err_a(q_{model})$  cancel each other out completely.

g)  $err(q_{model}) = [q_0 + q_0a\sqrt{t_p}] * err(q_0)/q_0 - q_0a\sqrt{t_p} * err(a)/a$   
 $= [q_0 + q_0a\sqrt{t_p}] * x - q_0a\sqrt{t_p} * x = x * q_0 = err(q_0)$

According to c),  $err_{q0}(q_{model})$  and  $err_a(q_{model})$  have opposite effects on  $err(q_0)$  because of the negative correlation between the (individual errors) in  $q_0$  and  $a$ , i.e. the errors in  $q_0$  and  $a$  weaken each other. However, because  $err(a)/a = err(q_0) / q_0$ , the weakening effect is not as strong as in f) and also does not depend on  $t_p$ . In fact, for the rainfall part of the equation, the errors  $err_{q0}(q_{model})$  and  $err_a(q_{model})$  cancel each other out completely, leaving the part without rainfall:  $err(q_{model}) = err(q_0)$ .