Problem 1

In a longitudinal study where we have repeated measurements on each subject we need special methods to account for the correlation within subjects. If we incorrectly assume independence and ignore correlation we can end up underestimating the probability of a type 1 error in a hypothesis test or overestimate coverage probability when constructing a confidence interval. Thus we require special statistical techniques to obtain valid statistical inferences.
Also, in order to achieve greater precision and more efficient estimates, special methods are needed to account for the within subject correlation. For example when estimating a mean, we could give each subject a weight in order to account for the within subject correlation. The optimal weight is 1/variance:
$\frac{n_i}{1+(n_i-1)\rho}$
where ρ is the within subject correlation and n_i is the number of observations for the i_{th} subject

Problem 2

$$P(Y=0)=.3 \quad P(Y=1)=.4 \quad P(Y=2)=.3$$
$$\mu=E(Y)=.3(0)+.4(1)+.3(2)=1$$
$$Var(Y)=E(Y^2)-E(Y)^2$$
$$E(Y^2)=0^2*.3+1^2*.4+2^2*.3=1.6 \quad E(Y)^2=1^2=1$$
$$Var(Y)=1.6-1=.6$$
$$\text{Y takes values in } [0,m] \text{ where } m=2$$
$$Y\text{exhibits extra binomial variation if: } Var(Y)>\mu(1-\mu/m)$$

If the random variable Y has a larger variance than its theoretical distribution, we have over dispersion
In this case we are comparing Y to a binomial distribution, $Binomial(m,p)$ where $m=2$, $p=\mu/m=1/2$

$$Var(Y)=.6 \quad \mu(1-\mu/m)=1(1-1/2)=.5 \quad .6>.5$$
$$\text{Since } Var(Y)>\mu(1-\mu/m), \text{ Y exhibits extra binomial variation}$$

Problem 3

(a) False

(b) True

(c) False

(d) True

(e) False

(f) False

Problem 4

Since the distribution of response is multivariate normal
and β and θ are functionally unrelated
We have orthogonal parameters which means that $\hat{cov}(\hat{\beta}_M, \hat{\theta}_M) = 0$ matrix
Which is why SAS does not computed this estimate

Problem 5

(a)

$$E(Y_{i2}) = .2 \quad E(Y_{i1}) = .1 \quad Var(Y_{i2}) = .2{*}.8 = .16 \quad Var(Y_{i1}) = .1{*}.9 = .09 \quad Corr(Y_{i1}, Y_{i2}) = .4$$
$$E(Y_{i2}|Y_{i1}) = \gamma_1 + \gamma_2 Y_{i1}$$
$$E(Y_{i2}) = E(E(Y_{i2}|Y_{i1})) = E(\gamma_1 + \gamma_2 Y_{i1}) = \gamma_1 + .1\gamma_2 = .2$$
$$\gamma_1 = .2 - .1\gamma 2$$
$$E(Y_{i2}|Y_{i1} = 1) = \gamma_1 + \gamma_2 \quad E(Y_{i2}|Y_{i1} = 0) = \gamma_1$$
$$Cov(Y_{i1}, Y_{i2}) = Corr(Y_{i1}, Y_{i2})\sqrt{Var(Y_{i1})Var(Y_{i2})} = .4/\sqrt{(}.09 * .16) = .048$$
$$E(Y_{i1}Y_{i2}) = P(Y_{i2} = 1, Y_{i1} = 1) = P(Y_{i2} = 1|Y_{i1} = 1)P(Y_{i1} = 1) = .1(\gamma_1{+}\gamma_2)$$
$$E(Y_{i1}Y_{i2}) = Cov(Y_{i1}, Y_{i2}) + E(Y_{i1})E(Y_{i2}) = .048 + .1(.2) = .068$$
$$.068 = .1(\gamma_1 + \gamma_2)$$
$$\gamma_1 + \gamma_2 = .068/.1 = .68$$
$$.68 = .2 - .1\gamma 2 + \gamma_2$$
$$\gamma_2 = .48/.9 = 8/15 \approx .5333$$
$$\gamma_1 = .68 - 8/15 \approx .1467$$
$$P(Y_{i2} = 1|Y_{i1} = 1) = .68 \quad P(Y_{i2} = 1|Y_{i1} = 0) \approx .147$$

(b)

$$Var(Y_{i2}) = .16 \quad P(Y_{i2} = 1|Y_{i1} = 1) = .68 \quad P(Y_{i2} = 1|Y_{i1} = 0) = .1467$$
$$Var(Y_{i2}|Y_{i1} = 0) = P(Y_{i2} = 1|Y_{i1} = 0){*}(1{-}P(Y_{i2}|Y_{i1} = 0)) = .68{*}.32 = .2176 > .16$$
$$Var(Y_{i2}|Y_{i1} = 1) = P(Y_{i2} = 1|Y_{i1} = 1){*}(1{-}P(Y_{i2}|Y_{i1} = 0)) = .1467{*}.8533 \approx .1252 < .16$$
$$Var(Y_{i2}|Y_{i1} = 1) < Var(Y) \text{ (reduction)}$$
$$Var(Y_{i2}|Y_{i1} = 0) > Var(Y) \text{ (inflation)}$$

Therefore conditioning on Y_{i1} is not guarenteed to reduce the variance of Y_{i2} since if $Y_{i1} = 1$ the conditional variance is larger than the marginal variance of Y_{i2}

(c)

From part a we know $P(Y_{i2} = 1, Y_{i1} = 1) = .068$

Using this and the marginal probabilities to fill out the following 2x2 table of probabilities

$.2 - .068 = .132 \quad .9 - .132 = .768 \quad .8 - .768 = .032$

	$Y_{i2} = 0$	$Y_{i2} = 1$	
$Y_{i1} = 0$	.768	.132	.9
$Y_{i1} = 1$	.032	.068	.1
	.8	.2	1

Using the 2x2 table to calculate the odds ratio that represents a measure of dependence between Y_{i1} and Y_{i2}

$$\text{Odds Ratio} \ = \frac{.768 * .068}{.032 * .132} = 12.36364 \approx 12.364$$