1.1) $y = \mu + \beta^T x + \epsilon/\sqrt{w}$

We write the data in following way.

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \mu \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} / \sqrt{w}$$

$n \times 1 \qquad n \times \overline{p+1} \qquad \overline{p+1} \times 1 \qquad n \times 1$

$\Rightarrow y = X^* \beta^* + \underset{\sim}{\epsilon}\, w^{-1/2}$

$\Rightarrow w^{1/2} y = w^{1/2} X^* \beta^* + \underset{\sim}{\epsilon}$

$\Rightarrow y^* = X_1 \beta^* + \underset{\sim}{\epsilon}$

where $y^* = w^{1/2} y$, $X_1 = w^{1/2} X^*$, $\beta^* = \begin{pmatrix} \mu \\ \beta \end{pmatrix}$

We know to get OLS estimate of β^* we have to minimize $\underset{\sim}{\epsilon}'\underset{\sim}{\epsilon} = (y^* - X_1 \underset{\sim}{\beta}^*)'(y^* - X_1 \beta^*)$

$$\frac{\partial}{\partial \underset{\sim}{\beta}} (y^* - X_1 \underset{\sim}{\beta}^*)'(y^* - X_1 \underset{\sim}{\beta}^*) = 0$$

$\Rightarrow X_1' X \underset{\sim}{\beta}^* = X_1' y^*$

$\Rightarrow \hat{\underset{\sim}{\beta}}^* = (X_1' X_1)^{-1} X_1' y^*$

Putting the values of x_1^*, y^*, β^* we get

$$\begin{pmatrix} \hat{\mu} \\ \hat{\underset{\sim}{\beta}} \end{pmatrix} = (x^{*\prime} W x^*)^{-1} (x^{*\prime} W y)$$

We know an unbiased estimator of σ^2 is $RSS/n-p-1$

$$\hat{RSS} = (Y^* - X_1 \hat{\beta}_*)'(Y^* - X_1 \hat{\beta}_*)$$
$$= (Y - X^* \hat{\beta}_*)' W^{1/2} W^{1/2} (Y - X^* \hat{\beta}_*)$$
$$= (Y - X^* \hat{\beta}_*) W (Y - X^* \hat{\beta}_*)$$
$$\therefore \hat{\sigma}^2 = (Y - X^* \hat{\beta}_*)' W (Y - X^* \hat{\beta}_*)$$

Proof: Model: $Y = X \underset{\sim}{\beta} + \epsilon$ $\quad Y^{n\times 1}, X^{n\times p}, \beta^{p\times 1}$

$$E(Y - X\hat{\beta})'(Y - X\hat{\beta})$$
$$= E(Y'Y - \hat{\beta}'X'X\hat{\beta}) = E(Y'Y - Y'X(X'X)^{-1}X'Y)$$

since $X'X\hat{\beta} = X'Y$

$$= E(Y'(I_n - X(X'X)^{-1}X')Y)$$
$$= E(Y'AY) \text{ (say)}$$
$$= E(tr(Y'AY))$$
$$= tr\, E(Y'AY) = tr(A\, E(Y'Y))$$

$$\boxed{E(Y^2) = V(Y) + E^2(Y)}$$

$$= tr[A(D(Y) + (X\underset{\sim}{\beta})(X\underset{\sim}{\beta})')]$$
$$= tr[\sigma^2 I_n A] + tr[(X\underset{\sim}{\beta})' A X \underset{\sim}{\beta}]$$
$$= \sigma^2 tr[I_n - X(X'X)^{-1}X'] + \underset{\sim}{\beta}'[X'X - X'X]\underset{\sim}{\beta}$$
$$= \sigma^2(n - tr(X(X'X)^{-1}X')) = \sigma^2(n-p)$$
$$\therefore E\left(\frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n-p}\right) = \sigma^2 \text{ (Proved)}$$

2) $\underset{\sim}{\epsilon} \sim N_n(0, \sigma^2 I_n)$ for simplicity $X^* \to X$, $y^* \to Y$

$\beta_* \to \beta$

$Y \sim N_n(X\underset{\sim}{\beta}, \sigma^2 I_n)$

$$f_Y(\underset{\sim}{y}) = \frac{1}{\sqrt{|I_n\sigma^2|}(\sqrt{2\pi})^n} \exp\{-(Y - X\underset{\sim}{\beta})' I_n (Y - X\underset{\sim}{\beta})\} \frac{1}{2\sigma^2}$$

to get a ML estimate of $\underset{\sim}{\beta}$ we have to minimize $(Y - X\underset{\sim}{\beta})'(Y - X\underset{\sim}{\beta})$ w.r.t. $\underset{\sim}{\beta}$

hence this reduces to the problem of OLS.

for σ

$$\ln f_Y(\underset{\sim}{y}) = \ln\sigma + K - (Y - X\underset{\sim}{\beta})' I_n (Y - X\underset{\sim}{\beta}) \times \frac{1}{2\sigma^2}$$

$$\frac{\partial}{\partial\sigma} \ln f_Y(\underset{\sim}{y}) = \frac{n}{\sigma} - \frac{1}{\sigma^3}(Y - X\beta')(Y - X\underset{\sim}{\beta}) = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}(Y - X\underset{\sim}{\beta}')(Y - X\underset{\sim}{\beta})$$

3) $X|W \sim N(\mu_x, W^{-1}\Sigma)$ $\quad \underset{p\times 1}{x_i} \in \mathbb{R}^p$

let $W^{-1}\Sigma = \Sigma^*$

$$f(\underset{\sim}{X}|W) = f(x_1, \_\_ x_n | W)$$

$$= \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma_*|^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(\underset{\sim}{x_i}-\underset{\sim}{\mu_x})'\Sigma_*^{-1}(\underset{\sim}{x_i}-\underset{\sim}{\mu_x})}$$

now

$$Q(\underset{\sim}{x}) = \sum_{i=1}^{n}(\underset{\sim}{x_i}-\underset{\sim}{\mu_x})'\Sigma_*^{-1}(\underset{\sim}{x_i}-\underset{\sim}{\mu_x})$$

$$= \sum_{i=1}^{n}(\underset{\sim}{x_i}-\underset{\sim}{\bar{x}}+\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x})'\Sigma_*^{-1}(\underset{\sim}{x_i}-\underset{\sim}{\bar{x}}+\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x})'$$

$$= \sum_{i=1}^{n}(\underset{\sim}{x_i}-\bar{x})\Sigma_*^{-1}(\underset{\sim}{x_i}-\bar{x}) + n(\bar{x}-\mu_x)\Sigma^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x})$$

$\left[\text{since } \frac{1}{n}\sum_{i=1}^{n}\underset{\sim}{x_i} = \underset{\sim}{\bar{x}}\right]$

$$= tr\sum_{i=1}^{n}(\underset{\sim}{x_i}-\bar{x})'\Sigma_*^{-1}(x_i-\bar{x}) + n(\underset{\sim}{\bar{x}}-\mu_x)'\Sigma^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x})$$

$$= tr\,\Sigma_*^{-1}\sum_{i=1}^{n}(\underset{\sim}{x_i}-\bar{x})(x_i-\bar{x})' + n(\underset{\sim}{\bar{x}}-\mu_x)'\Sigma^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x})$$

$$= tr\,\Sigma_*^{-1}A + n(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x'})\Sigma^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x})$$

$\Sigma_*^{-1}A$ is a pd matrix so $tr\,\Sigma^{-1}A > 0$

hence

$Q(\underset{\sim}{x})$ will minimize if $n(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x'})\Sigma_*^{-1}(\underset{\sim}{\bar{x}}-\underset{\sim}{\mu_x}) = 0$

which means

$\underset{\sim}{\bar{x}}'$ is the MLE of $\underset{\sim}{\mu_x}'$

Let $C = B'\Sigma^{-1}B$ such that $A = BB'$ since A is PD

$\Rightarrow |C| = |B'B|\,|\Sigma|^{-1}$

$\Rightarrow |\Sigma|^{-1} = \frac{|C|}{|A|}$

also $tr\, \Sigma^{-1}A = tr\, C$ ← property of trace

We have

$$f(\underset{\sim}{X}|W) \leq \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma_*|^{\frac{np}{2}}} e^{-\frac{1}{2} tr\, \Sigma^{-1}A} \quad \text{——— (1)}$$

putting $\hat{\mu}_{x\,mle} = \bar{x}$

RHS of (1)

$$\frac{1}{(2\pi)^{\frac{np}{2}}} \frac{|C|}{|A|} e^{-\frac{1}{2} tr\, C} \quad \text{——— (2)}$$

Since C is pd $\exists$ a lower triangular matrix T with $t_{ii} > 0$ $\forall i$ such that $C = TT'$

$$\therefore |C| = \prod_{i=1}^{p} t_{ii}^2$$

$$tr\, C = tr\, TT' = \sum_{i=1}^{p} t_{ii}^2 + \sum\sum_{i \neq j} t_{ij}^2 \geq \sum_{i=1}^{p} t_{ii}^2$$

hence neumerator of (2)

$$\leq \left(\prod_{i=1}^{p} t_{ii}^2\right)^{n/2} e^{-\frac{1}{2}\sum_i t_{ii}^2} \quad \text{——— (3)}$$

log (RHS of 3)

$$= \frac{1}{2}\sum_{i=1}^{p} (n \log t_{ii}^2 - t_{ii}^2)$$

$$\leq \frac{1}{2}\sum_{i=1}^{p} (n \log n - n)$$

$$= \frac{np}{2} \log n - \frac{np}{2}$$

hence RHS of (3) is maximum iff $t_{ii}^2 = n$

$\Rightarrow T = \sqrt{n}\, I$

$\therefore C = TT' = nI$

$\Rightarrow B'\Sigma^{-1}B = nI$

$\Rightarrow \Sigma_*^{-1} = n(BB')^{-1} = nA^{-1}$

$$\hat{\Sigma}_{*mle} = \frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{x_i} - \underset{\sim}{\bar{x}})(\underset{\sim}{x_i} - \underset{\sim}{\bar{x}})'$$

$$\Rightarrow \hat{\Sigma}_{mle} = W\,\frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{x_i} - \underset{\sim}{\bar{x}})(\underset{\sim}{x_i} - \underset{\sim}{\bar{x}})'$$

4) Let us consider the Model

$$Y = \mu + X\beta + \epsilon W^{-1/2}$$

Now $E_W(Y|X,W) = \mu + X\beta$

$$V(Y|X) = V(\mu + \beta X + \epsilon/\sqrt{W})$$

$$= V_W(\epsilon/\sqrt{W})$$

$$= V_W[E_W(\epsilon/\sqrt{W})] + E_W(V(\epsilon/\sqrt{W}))$$

$$= \sigma^2 E(1/W)$$

We know $E(1/W) \geq 1/E(W)$

$= 1$ since $E(W) = 1$

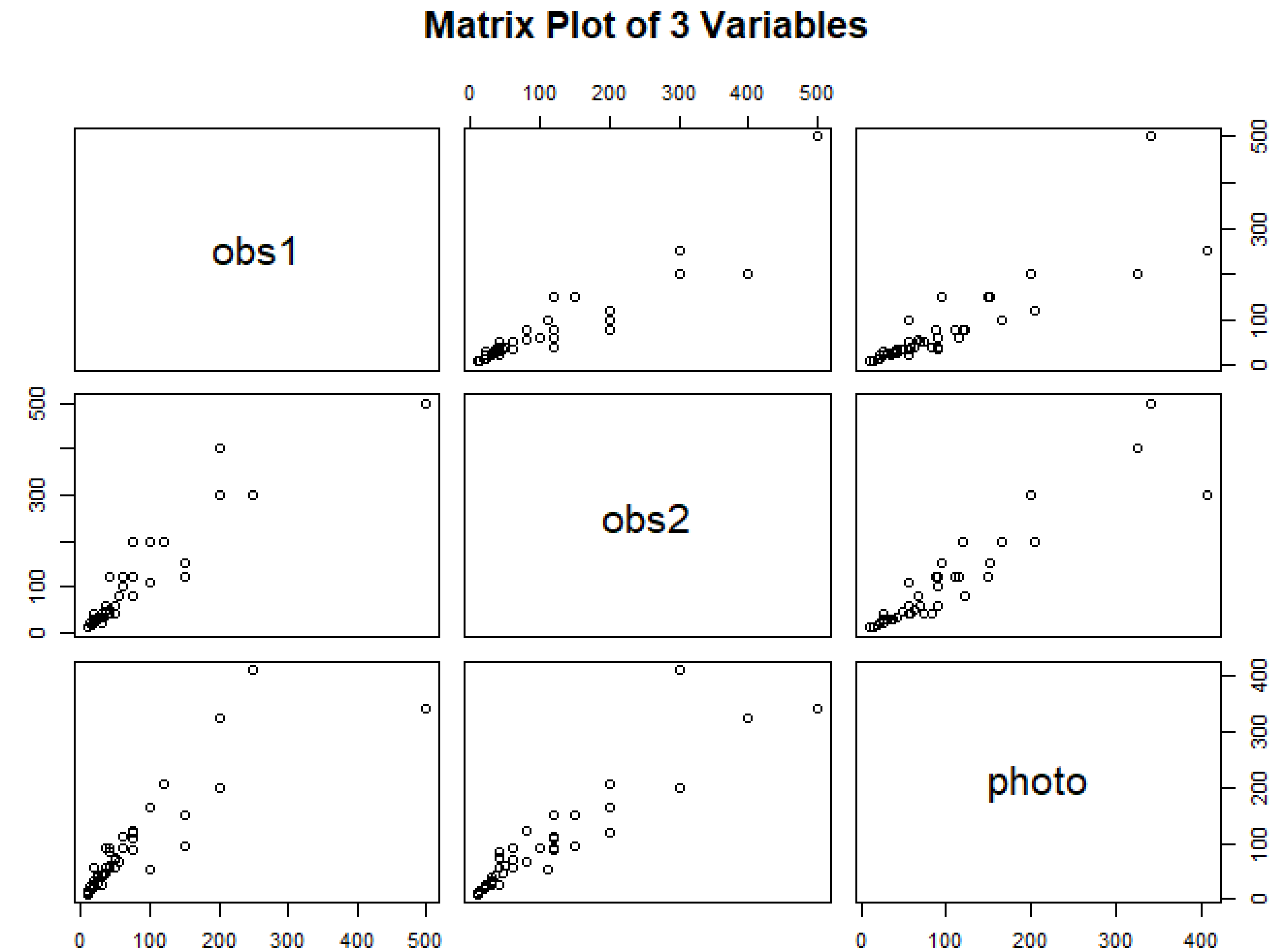
$$\Rightarrow Var(Y|X) \geq \sigma^2$$

if $E(W) = w$ then our estimates of μ & β are useful.

1.

2.

2.1.



Yes, the plots suggest a linear model of PHOTO on both OBS1 and OBS2 would be appropriate, since both the plots suggests an upward linear trend along the x=y diagonal line. Also it is clear that two predictors OBS1 and OBS2 are also linearly related as well.

2.2. The model with PHOTO as response and OBS1 as predictor is as follows

$$PHOTO = \beta_0 + \beta_1 * OBS1 + \in$$

$$=> \in = PHOTO - \hat{\beta}_0 + \hat{\beta}_1 * OBS1$$

So the error ∈ is the difference is the no of birds in the flock counted from the PHOTO and the estimated no of birds in the flock using 1st and 2nd Observer's count.

2.3. The variable photo is the no of birds counted from the photo taken. Photos were taken so that a more or less **exact** count of the number of birds in the flock could be made. OBS1 is the aerial count by observer one. In real life it is not quite an easy job to count the birds while airborne no matter how experienced the person is while counting from the Photo is a significantly easy task and requires no specialization. So clearly PHOTO is more close to the actual no of birds in a flock than OBS1 will ever be.

So, it is t appropriate to fit the regression of photo on obs1 rather than the regression of obs1 on photo.

2.4. OLS regression summary of PHOTO on OBS1

```
> summary(lm(photo~obs1))

Call:
lm(formula = photo ~ obs1)

Residuals:
     Min       1Q   Median       3Q      Max 
-125.928  -18.713   -9.033   11.699  161.711 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 26.64957    8.61448   3.094  0.00347 ** 
obs1         0.88256    0.07764  11.367 1.54e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44.41 on 43 degrees of freedom
Multiple R-squared:  0.7503,    Adjusted R-squared:  0.7445 
F-statistic: 129.2 on 1 and 43 DF,  p-value: 1.537e-14
```

Testing of hypothesis:

```
> y1=lm(photo~obs1)
> RSS1=sum(y1$residuals^2);RSS1
[1] 84790.18
> RSS2=sum((photo-obs1)^2);RSS2
[1] 104390
> F=((RSS2-RSS1)/2)/(RSS1/43);F
[1] 4.969868
> pvalue=1-pf(F,2,43);pvalue
[1] 0.01143572
```

At any standard level of significance, we reject the Null hypothesis in support of the alternative.

2.5. The meaning of hypothesis in question 4 is $\beta_0 = 0 \; and \; \beta_1 = 1$, in simple words the response no of birds obtained from PHOTO is equal the no of bird obtained from predictor OBS1.
From the p_value (approx. 0.011) corresponding to predictor OBS1 in the regression summary in Q4 indicates that the coefficient is significant and not equals 0. Which means that OBS1 is reliable in predicting PHOTO.

2.6.

2.6.1. The interpretation of ∈ remains the same as 2.2 but further we can add the variance of ∈ is increasing is not fixed rather increasing with “x”, the predictor. Since the variance isn’t constant for all the error term we have to deal with the problem of heteroscedasticity to estimate the regression coefficients. In order to get maximum likelihood estimates of regression coefficients we have to use WLS (Weighted Least Squared) method instead of using OLS (Ordinary Least Square) method.

2.6.2.

```
> y1=lm(photo~obs1,weights=1/sqrt(obs1))
> RSS1=sum(y1$residuals^2);RSS1
[1] 89587.65
> RSS2=sum(((photo-obs1)^2));RSS2
[1] 104390
> F=((RSS2-RSS1)/2)/(RSS1/43);F
[1] 3.552394
> pvalue=1-pf(F,2,43);pvalue
[1] 0.03733949
```

At level 0.05 we reject H_0 and conclude that the model In y=x is not significant.

2.7.

OLS regression summary of PHOTO on OBS1

```
> summary(lm(photo~obs1))

Call:
lm(formula = photo ~ obs1)

Residuals:
     Min       1Q   Median       3Q      Max
-125.928  -18.713   -9.033   11.699  161.711

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 26.64957    8.61448   3.094  0.00347 **
obs1         0.88256    0.07764  11.367 1.54e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 44.41 on 43 degrees of freedom
Multiple R-squared:  0.7503,    Adjusted R-squared:  0.744
F-statistic: 129.2 on 1 and 43 DF,  p-value: 1.537e-14
```

OLS Regression Summary of PHOTO on OBS2

```
> summary(lm(photo~obs2))

Call:
lm(formula = photo ~ obs2)

Residuals:
    Min      1Q  Median      3Q     Max
-58.698 -12.007  -3.071   7.692 162.116

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.16428    6.82973   2.367   0.0225 *
obs2         0.76907    0.04835  15.905   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.87 on 43 degrees of freedom
Multiple R-squared:  0.8547,    Adjusted R-squared:  0.8513
F-statistic:   253 on 1 and 43 DF,  p-value: < 2.2e-16
```

OLS Regression summary of PHOTO on average of OS1 and OS2

```
> obs=(obs1+obs2)/2
> summary(lm(photo~obs))

Call:
lm(formula = photo ~ obs)

Residuals:
     Min       1Q   Median       3Q      Max 
-104.020  -13.665   -3.782    8.331  155.474 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 18.25450    7.12483   2.562    0.014 *  
obs          0.85553    0.05714  14.973   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.65 on 43 degrees of freedom
Multiple R-squared:  0.8391,    Adjusted R-squared:  0.8353 
F-statistic: 224.2 on 1 and 43 DF,  p-value: < 2.2e-16
```

As we can see from the picture 1 the R_Sq value with predictor OS1 is **0.7503**, in picture 2 the value with predictor OS2 is **0.8547** and in picture 3 the R_sq value of both predictors combined is **0.8391**.

We know more the value of R_sq is the better is the model. So we can conclude that the combined predictor is certainly an improvement over OS1 but not over OS2.