```
stat off.
                                                                    Stat 600 derivation
                                   We Start with:
                                                      F(B) = [- \frac{7}{2} \left(\frac{K+1}{2} \mathbb{L} (\frac{1}{2} \in K) \log \reft(\frac{1}{2} \log \reft(\frac{1}{2} \log \reft) \right)
                                                                                   + 2 5 7 BE
                                   we know that Dumped We butons method
                          is \beta^{(e+1)} = \beta^{(e)} - \Pi(\nabla^2 f(\beta^{(e)}))^{-1} \nabla f(\beta^{(e)})
                   Soit Suffices to Show D2 F(B(1)) = (XTWKX + AI)
                                              \nabla F(\beta^{(4)}) = \left[ X^T \{ P_K - 1(y=K) \} + \lambda \beta_K^{(4)} \right]
                          First let's do the gradient: Begularitation to Barbaritation to Barbaritation of Barbaritat
                                  A E BRUS JER
regulation.
    KRIM
                                 2 = Z Z B si leikeihood term: Discons
                                                                                            - Z 1 (2: e3 log Px (Xi, B) Je (B)
                                     X Z Z BK; j = -1 E Vi=K3 PR(XijB) · PR (XijB) · Xi
                                   λ Bece)
                                                                    + 5 1(y=2) Pm (X= 18) X=
            Which:
             FOR adividual
                                                          Summy overally sold [ Exi=k] Xi + Pk (Xi; P) Xi
                 Putting these together = [(Pr (xi j B) - 1 (yiex)) xi
                       we get the = = (PR-1 (PR-1 (PR-1 (PR-1 (PR-1)))
                   Dr (B(e)) = [x [(R-1 (4=k)) + | BK)]
        band: TO extend the want hours to Tive without
                                 a finite second moments we will truncate
                                       and then Use Cheby specify in centurity
```

Now let's get the Hessian
let's split up into two
Let's split up into two parts $\frac{\partial^{2}}{\partial \beta_{R}} \left(-\log P_{K}(X_{i};\beta)\right) = -P_{K}(X_{i};\beta) \left(1 - P_{K}(X_{i};\beta)\right) \times i \times i$ the form of Wik
A COLOR DE LA COLO
Now for the requestizing term:  \[ \frac{\partial 2}{\partial 2} \
we saw the gradient was
Jen JBK
DPK AI
which we can put together as
$(X^TW_KX+\lambda I)$
which is what we wanted to show.