Assignment 01: Preliminary Mathematics
2018 Spring

1 Fundamental Knowledge

1.1 Definitions

- Scalars
- Matrix: order, dimensions
  - Rectangular matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$
  - Square matrix $\mathbf{A}, m = n$
  - Symmetric matrix $\mathbf{A^T} = \mathbf{A}$
  - Diagonal matrix $\mathbf{A}, a_{ij} = 0$ for $i \neq j$
  - Identity matrix $\mathbf{I}$
  - Null matrix $\mathbf{0}$
  - All-One matrix $\mathbf{J} = \begin{pmatrix}1\end{pmatrix}$
  - Upper/Lower Triangular matrix
  - Tridiagonal matrix
- Vectors $\mathbf{v} \in \mathbb{R}^n$
- Transposition $a_{ij} = b_{ji}$
- Diagonal a_{ii}

1.2 Matrix operations

- Addition $\mathbf{A} + \mathbf{B}$
- Multiplication $\mathbf{AB}$
  - Transpose of product $(\mathbf{ABC})^\mathbf{T} = \mathbf{C^T B^T A^T}$
  - Idempotent matrix: $\mathbf{AA} = \mathbf{A}$
  - Nilpotent matrix: $\mathbf{BB} = \mathbf{0}$
  - Orthogonal matrix: $\mathbf{UU^T} = \mathbf{U^T U} = \mathbf{I}$
- Direct sum:

$$\mathbf{H_1} \oplus \mathbf{H_2} \oplus \cdots \oplus \mathbf{H_n} = \begin{pmatrix} H_1 & 0 & \dots & 0 \\ 0 & H_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & H_n \end{pmatrix}.$$

- Kronecker product:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix}$$

- Hadamard product:

$$\mathbf{A} \odot \mathbf{B} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix}$$

1.3 Norm

- L_p-norm: $\|\mathbf{A}\|_\mathbf{p} = \|\text{vec}(\mathbf{A})\|_\mathbf{p}$
- Frobenius norm: $\|\mathbf{A}\| = \left[\mathbf{tr}(\mathbf{A}\mathbf{A}^\mathbf{T})\right]^{\mathbf{1/2}} = \left(\sum_{\mathbf{i=1}}^{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{a}_{\mathbf{ij}}^{\mathbf{2}}\right)^{\mathbf{1/2}} = \left(\sum_{\mathbf{i=1}}^{\mathbf{n}} \sigma_{\mathbf{i}}^{\mathbf{2}}(\mathbf{A})\right)^{\mathbf{1/2}}$ where $\sigma_\mathbf{i}(\mathbf{A})$ is the singular value of $\mathbf{A}$.
- Nuclear/trace norm: $\|\|_* = \sum_{\mathbf{i=1}}^{\min\{\mathbf{m,n}\}} \sigma_\mathbf{i}(\mathbf{A})$
- Max norm: $\|\mathbf{A}\|_{\mathbf{max}} = \max_{\mathbf{i,j}} |\mathbf{a_{ij}}|$

1.4 Matrix inversion

- General inverse $\mathbf{A}\mathbf{A}^{-\mathbf{1}} = \mathbf{I}$
- Generalized inverse $\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$
- Moore-Penrose inverse satisfies:
  - $\mathbf{A}\mathbf{A}^{-}\mathbf{A} = \mathbf{A}$
  - $\mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-} = \mathbf{A}^{-}$
  - $(\mathbf{A}^{-}\mathbf{A})^\mathbf{T} = \mathbf{A}^{-}\mathbf{A}$
  - $(\mathbf{A}\mathbf{A}^{-})^\mathbf{T} = \mathbf{A}\mathbf{A}^{-}$

1.5 Matrix properties

- Determinant $\det(\mathbf{A}) = |\mathbf{A}|$
- Rank: $\text{rank}(\mathbf{A})$
- Trace: $\text{tr}(\mathbf{A})$
- Linear independence

2 Statistics Background

2.1 Discrete random variables

- Probability mass function $f(Y = y) = p(y)$
- Expected value: $E(Y) = \sum_y y f(Y = y)$
- $E(g(Y)) = \sum_y g(y) f(Y = y)$
- $Var(Y) = E[(Y - E(Y))^2] = E(Y^2) - [E(Y)]^2$
- Binomial distribution $Y \sim Bin(n, p)$: $f(Y = y; q) = \binom{n}{y} q^y (1-q)^{n-y}, E(Y) = nq, Var(Y) = nq(1-q)$
- Poisson distribution $Y \sim Pois(\lambda)$: $f(Y = y; \lambda) = \frac{\lambda^y}{y!}\exp(-\lambda), E(Y) = Var(Y) = \lambda$

2.1.1 Variance-Covariance matrix

- $Var(Y) = E(Y^2) - E(Y)E(Y) = E[(Y - E(Y))^2]$ represented as σ_Y^2
- $Cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$ represented as $\sigma_{Y_1Y_2}$
- The matrix form:

$$
\begin{aligned}
V &= \text{Var}(Y) \\
&= E(YY^T) - E(Y)E(Y^T) \\
&= \begin{pmatrix} \sigma_{Y_1}^2 & \sigma_{Y_1Y_2} & \cdots & \sigma_{Y_1Y_n} \\ \sigma_{Y_1Y_2} & \sigma_{Y_2}^2 & \cdots & \sigma_{Y_2Y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{Y_1Y_n} & \sigma_{Y_2Y_n} & \cdots & \sigma_{Y_n}^2 \end{pmatrix}.
\end{aligned}
$$

- A VCV matrix should always be postive definite or positive semi-definite.
- *a.k.a* **dispersion matrix**
- The VCV matrix of a matrix product:

$$
\begin{aligned}
Var(AY) &= E(AYY^TA^T) - E(AY)E(Y^TA^T) \\
&= AE(YY^T)A^T - AE(Y)E(Y^T)A^T \\
&= A(E(YY^T) - E(Y)E(Y^T))A^T \\
&= A\text{Var}(Y)A^T \\
&= AVA^T
\end{aligned}
$$

- $Cov(AY, BY) = AVB^T$
- $Cov(AY, MZ) = ACM^T$, where $Cov(Y, Z) = C$

2.2 Continuous random variables

- Cumulative distribution function (cdf): $F(y) = P(Y \leq y)$
- Probability density function (pdf): $f(y) = \frac{d(F(y))}{dy} = F'(y)$
- $\int_{-\infty}^{\infty} f(y)dy = 1$
- $F(y) = \int_{\infty}^{y} f(t)dt$
- Expect: $E(Y) = \int_{-\infty}^{\infty} yf(y)dy$
- $E(g(Y)) = \int_{-\infty}^{\infty} g(y)f(y)dy$
- $Var(Y) = E(Y^2) - [E(Y)]^2$

2.2.1 Uniform distribution

- The density function:

$$
f(y) = \begin{cases} 1/(\theta_2 - \theta_1) & \theta_1 \leq y \leq \theta_2 \\ 0 & \text{otherwise.} \end{cases}
$$

- $E(Y) = (\theta_1 + \theta_2)/2$
- $Var(Y) = (\theta_2 - \theta_1)^2/12$

2.2.2 Normal distribution

- The density function:
$$f(y) = (2\pi\sigma^2)^{-1/2} \exp(-\frac{(y-\mu)^2}{2\sigma^2})$$
- $E(Y) = \mu$
- $Var(Y) = \sigma^2$

2.2.3 Chi-square distribution

- If $\mathbf{y} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$, then $y^T y \sim \chi^2_n$ is a central chi-square distribution with $df = n$.
- The mean of a central chi-square distribution is n.
- The variance of a central chi-square distribution is $2n$.
- If $\mathbf{y} \sim \mathbf{N}(\mu, \mathbf{I})$, then $\mathbf{y^T y} \sim \chi^2_{n,\lambda}$ with noncentrality parameter $\lambda = \mu^T \mu/2$. The mean of a noncentral chi-square distribution is $n + 2\lambda$ and the variance is $2n + 8\lambda$.
- If $\mathbf{y} \sim \mathbf{N}(\mu, \mathbf{V})$, then $\mathbf{y^T Q y}$ has a noncentral chi-square distribution only if $\mathbf{QV}$ is idempotent. The noncentrality parameter $\lambda = \mu^{\mathbf{T}} \mathbf{QVQ}\mu$ and the mean and variance are $tr(QV) + 2\lambda$ and $2tr(QV) + 8\lambda$.

2.2.4 The t-distribution

The t-distribution is based on the ratio of two independent random variables. The first is from a univariate normal distribution, and the second is from a central chi-square distribution.
Let $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \mathbf{1})$ and $u \sim \chi^2_n$ with x and u being independent, then

$$\frac{x}{(u/n)^{1/2}} \sim t_n.$$

- The mean of a t-distribution is the mean of the $\mathbf{x}$ variable.
- The variance is $n/(n-2)$, and n is the degrees of freedom of the chi-square distribution.

2.2.5 The F-distribution

The central F-distribution is based on the ratio of two independent central chi-square variables. Let $u \sim \chi^2_n$ and $w \sim \chi^2_m$ with u and w being independent, then

$$\frac{(u/n)}{(w/m)} \sim F_{n,m}.$$

- The mean of the F-distribution is $m/(m-2)$
- The variance is $\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$.

2.3 Quadratic forms and Bilinear forms

2.3.1 Quadratic forms

A quadratic form is a sum of squares of elements of a vector. The general form is $\mathbf{y^TQy}$, where $\mathbf{y}$ is a vector of random variables, and $\mathbf{Q}$ is a regulator matrix. The regulator matrix can take on various forms and values depending on the situation. Usually $\mathbf{Q}$is a symmetric matrix. Examples of different $\mathbf{Q}$ matrices are as follows:

- $\mathbf{Q} = \mathbf{I}$, then $\mathbf{y'Qy} = \mathbf{y^Ty}$ which is a total sum of squares of the elements in $\mathbf{y}$.
- $\mathbf{Q} = \mathbf{J}(1/n)$, then $\mathbf{y^TQy} = \mathbf{y^TJy}(1/n)$ where n is the length of $\mathbf{y}$. Note that $\mathbf{J} = \mathbf{11^T}$, so that $\mathbf{y^TJy} = (\mathbf{y^T1})(\mathbf{1^Ty})$ and $(\mathbf{1^Ty})$ is the sum of the elements in $\mathbf{y}$.
- $\mathbf{Q} = (\mathbf{I} - \mathbf{J}(1/n))/(n-1)$, then $\mathbf{y^TQy}$ gives the variance of the elements in $\mathbf{y}$.

The expected value of a quadratic form is

$$E(\mathbf{y^TQy}) = E(tr(\mathbf{y^TQy})) = E(tr(\mathbf{Qyy^T})) = tr(\mathbf{Q}E(\mathbf{yy^T})).$$

3 Exercises

3.1 Matrix Calculus

1. For $\mathbf{y} = \mathbf{x^TAx}; \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{n\times n}$.

   (1) Prove that $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A^T})\mathbf{x}$

   Solution:

$$\begin{aligned} \frac{\partial x^TAx}{\partial x} &= \begin{bmatrix} \frac{\partial x^TAx}{\partial x_1} \\ \vdots \\ \frac{\partial x^TAx}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial a_{ij}x_ix_j}{\partial x_1} \\ \vdots \\ \sum_{i=1}^n \sum_{j=1}^n \frac{\partial a_{ij}x_ix_j}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j + \sum_{i=1}^n a_{i1}x_i \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j + \sum_{i=1}^n a_{in}x_i \end{bmatrix} \\ &= (A + A^T)x \end{aligned}$$

   (2) Compute $\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T}$

   Solution:

$$\frac{\partial y}{\partial x^T} = x^T(A + A^T)$$

2. For $\mathbf{z} = \mathbf{x^TAy}$, where $\mathbf{x} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}$

   (1) Compute $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$

Solution:

$$\frac{\partial z}{\partial x} = Ay$$

(2) Compute $\frac{\partial \mathbf{z}}{\partial \mathbf{y}}$

Solution:

$$\frac{\partial z}{\partial y} = x^T A$$

3.2 Linear regression

3. A linear regression problem $\mathbf{y} = \mathbf{X}\beta$ can be solved either using least squares or maximum likelihood. Can you write down the two objective functions and reach the normal equations.

Solution: For least squares approach:

$$\hat{\beta} = \arg\min_\beta (y - X\beta)^T (y - X\beta)$$

while for maximum likelihood approach is based on the assumption that

$$e = y - X\beta \sim N(0_n, \sigma^2 I)$$

and the maximum log-likelihood estimator should be:

$$\begin{aligned} \hat{\beta} &= \arg\max_\beta (y - X\beta)^T (\sigma^2 I)^{-1} (y - X\beta) \\ &= \arg\max_\beta (y - X\beta)^T (y - X\beta) \end{aligned}$$

Therefore the objective function are same:

$$\max L = (y - X\beta)^T (y - X\beta)$$

which can be solved using

$$\frac{\partial L}{\partial \beta} = 0$$

Therefore, the final normal equation becomes:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

depending on an invertible matrix.

3.3 Trace of a matrix

4. Here is definition of the trace for a square matrix **X**:

$$\mathrm{tr}(X) = |X| = \sum_{i=1}^{n} x_{ii}$$

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times n}$ are two square matrices, and $c \in \mathbb{R}$. Prove that:

(1) $tr(A \pm B) = tr(A) + tr(B)$

(2) $tr(A^T) = tr(A)$

(3) $tr(cA) = c\mathrm{tr}(A)$

(4) $tr(AB) = tr(BA)$

(5) $tr(AA^T) = tr(A^T A)$

3.4 Determinant of a matrix

5. We know that the determinant of a square matrix $A \in \mathbb{R}^{n\times n}$ is defined by

$$\|A\| = \sum_{i=1}^{n}(-1)^{i+j} a_{ij}\|M_{ij}\|, M_{ij} = A_{-i,-j}$$

Prove that

(1) If A is diagonal or triangular, then $\|A\| = \prod_{i=1}^{n} a_{ii}$

3.5 Spectral Decomposition

6. The spectral decomposition of a square matrix $A \in \mathbb{R}^{n\times n}$ can be defined as

$$\mathbf{A} = \mathbf{\Gamma\Lambda\Gamma^T}$$

where Γ are orthonormal matrix such that $\Gamma^T\Gamma = I$ and $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues of $\mathbf{A}$. Write down the form of $\mathbf{A}^m$.

7. Eigen-decomposition can be associated with singular value decomposition (SVD).

(1) What are positive-definite (PD) matrix and positive semi-definite (PS) matrices?

Solution:

- $\mathbf{Q}$ is positive-definite if $\mathbf{y^TQy}$ is always greater than zero for all vectors, $\mathbf{y} \neq 0$.
- $\mathbf{Q}$ is positive semi-definite if $\mathbf{y'Qy}$ is greater than or equal to zero for all vectors $\mathbf{y}$, and for at least one vector $\mathbf{y}$, then $\mathbf{y'Qy} = 0$.
- $\mathbf{Q}$ is non-negative definite if $\mathbf{Q}$ is either positive definite or positive semi-definite.

(2) Can you use eigen-decomposition to determine if a square matrix is positive definite or not?

Solution: If all the eigenvalues are greater than zero, then the matrix is positive definite. If all the eigenvalues are greater than or equal to zero and one or more are equal to zero, then the matrix is positive semi-definite.

3.6 Other Matrix Decomposition Techniques

8. For matrix, the decomposition techniques are important knowledge to grasp.

   (1) What is QR-decomposition and Cholesky-decomposition?

   Solution: **QR-decomposition** of a real square matrix A can be decomposed as

   $$\mathbf{A} = \mathbf{QR}$$

   where $\mathbf{Q}$ is an orthogonal matrix $\mathbf{Q^TQ} = \mathbf{QQ^T} = \mathbf{I}$, and $\mathbf{R}$ is an upper triangular matrix. If $\mathbf{A}$ is invertible, then the factorization is unique if we require that the diagonal elements of $\mathbf{R}$ to be positive.
   The **Cholesky decomposition** of a matrix, say $\mathbf{V}$, is a lower triangular matrix such that

   $$\mathbf{V} = \mathbf{TT}',$$

   and $\mathbf{T}$ is lower triangular matrix.

   (2) Write some comments on the applications of the above matrix decomposition techniques.

   (3) `base::qr` and `base::chol` can be used to conduct the two decompositions. Give an example to illustrate the usage of the two decompositions?

3.7 Sufficient statistic

9. Sufficient statistics, minimal sufficient statistics are the important technique for data reduction.

   (1) What is sufficient statistic?

   Solution: A sufficient statistic is a statistic $T(X)$ from a sample w.r.t. a statistical model and its associated unknown parameter if

   > no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter.

   (2) How to prove that a statistic is a sufficient statistic for a parameter?

   Solution: Neyman-Fisher Factorization Theorem.

   (3) How to prove that a statistic is a minimal sufficient statistic for a parameter?

Solution: A sufficient statistic $T(X)$ is minimal if and only if for two arbitrary samples X and Y

$$\frac{f_\theta(X)}{f_\theta(Y)} \text{ is constant } \equiv T(X) = T(Y).$$

10. Suppose that ϵ_1 and ϵ_2 are independent standard normal random variables.
    (1) Identify the distribution of $Y = 3\epsilon_1 - 4\epsilon_2$, specifying any relevant parameter values.

    Solution: Since $\epsilon_1 \sim N(0,1)$ and $\epsilon_2 \sim N(0,1)$, therefore

    $$3\epsilon_1 - 4\epsilon_2 \sim N(3 \times \mu - 4 \times \mu, 3^2 \times \sigma^2 + 4^2 \times \sigma^2) = N(0, 25)$$

    (2) Identify the distribution of $Z = \epsilon_1^2 + \epsilon_2^2$, specifying any relevant parameter values.

    Solution: The sum of k square independent standard normal distribution will follow χ_k^2 distribution. Therefore, $Z \sim \chi_2^2$.

11. Suppose $X \in \mathbb{R}^{15\times 10}$ consisting of linearly independent columns.
    (1) Show that $A = X(X^TX)^{-1}X^T$ is a symmetric, idempotent matrix.

    Solution: Since $A^T = (X^T)^T[(X^TX)^{-1}]^TX^T = X(X^TX)^{-1}X^T = A$, so A is symmetric.
    And $A^2 = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T = A$, so A is idempotent.

    (2) Find the trace of A.

    Solution: Since $tr(AB) = tr(BA)$, we can get

    $$tr(X(X^TX)^{-1}X^T) = tr(X^TX(X^TX)^{-1}) = tr(I) = 10.$$

    (3) List all 15 eigenvalues of A.

    Solution: The eigenvalues are $0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1$.

    (4) Suppose ϵ is a vector consisting of 15 independent stanard normal random variables. Show that if $B = \epsilon^TA\epsilon$ and $C = \epsilon^T(I - A)\epsilon$, then B and C are independent.

    Solution: Since $A(I-A) = A - A^2 = A - A = 0$, and $(I-A)^2 = I - 2A + A^2 = I - A$, so B and C must be independent.

    (5) Identify the distributions of B, C and $D = 9B/C$, specifying any relevant parameter values.

Solution:

$$B \sim \chi^2(10), C \sim \chi^2(5), D = \frac{B/10}{C/5} \sim F_{10,5}$$

12. Suppose x and y are column vectors of length n. Show, from first principle, that $tr(x^T y) = tr(yx^T)$.

Solution:

$$tr(x^T y) = tr(\sum x_i y_i) = \sum x_i y_i = \sum y_i x_i = tr(yx^T)$$

since $y_i x_i$ is the i-th diagonal element of yx^T.

3.8 Discrete distributions

13. Write down some properties for negative binomial distribution.
    (1) What is negative-binomial distribution?
    (2) Can you write down the expectation and variance of the distribution?
    (3) Both Poisson and negative binomial distributions can be used to fit the counts. What are the differences between these two distributions?

3.9 Beta-Binomial model

14. Beta distribution is the conjugate prior for binomial distribution, since combination of them can produce a posterior probability of beta-form.

```r
y <- 13
N <- 35

a <- 1
b <- 1

theta <- seq(0, 1, len = 100)
likelihood <- theta ^ y * (1-theta)^(N-y) #dbinom(y, N, theta)
prior <- theta ^ (a-1) * (1-theta)^(b-1) #dbeta(theta, a, b)
posterior1 <- prior * likelihood
posterior2 <- dbeta(theta, y + a, N - y + b)

plot(theta, likelihood, ty='l')
abline(v = y/N, col="red")

par(mfrow=c(1,3))
plot(theta, prior, ty='l')
plot(theta, posterior1, ty='l')
plot(theta, posterior2, ty='l', col="blue")
```

15. Consider the following data:

Catalyst	Temperature	Yield
present	10	15
present	-10	10
absent	10	5
absent	-10	5

(1) Fit a linear model (without an intercept) to the data, relating `Yield` to the other two variables. Write down the fitted model. You need to define all terms used.

Solution: We can define the linear model as

$$y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

where $x_1 = \begin{cases} 1 & \text{if catalyst is present} \\ 0 & \text{otherwise} \end{cases}$, x_2 denotes the temperature, and y the yield.
Therefore we can write it into matrix form:

$$X = \begin{bmatrix} 1 & 10 \\ 1 & -10 \\ 0 & 10 \\ 0 & -10 \end{bmatrix},$$

so

$$X^T X = \begin{bmatrix} 2 & 0 \\ 0 & 400 \end{bmatrix},$$

and

$$X^T y = [2550]$$

thus $\hat{\beta} = [12.50.125]$, the fitted model is

$$\hat{y} = 12.5x_1 + 0.125x_2$$

(2) Estimate the error variance.

Solution: Since

$$SSE = (y - X\hat{\beta})^T (y - X\hat{\beta}) = 56.25$$

therefore,

$$\hat{\sigma}^2 = \frac{SSE}{n-2} = 28.125.$$

(3) Give a point and interval estimate for the mean yield when Catalyst is present and Temperature is 5.

Solution: For $x_0 = [1, 5]^T$, we have

$$\hat{y} = x_0^T \hat{\beta} = 12.5 \times 1 + 0.125 \times 5 = 13.125.$$

and the error should be

$$\sqrt{V(\hat{y}|x_0)} = \sqrt{\hat{\sigma}^2 x_0^T (X^T X)^{-1} x_0} = 075\sigma^2$$

therefore, the 95% confidence interval for $\hat{y}$ given x_0 should be:

$$\hat{y} \pm 1.96 \times 0.75 \times \sqrt{28.125} = 13.125 \pm 7.795 = (5.329, 20.921).$$