Test #1 Answers

STAT 494/873

Fall 2025

Complete the problems below. Use R for all calculations and plots. **Make sure to fully explain all answers and show your work to receive full credit.**

1. (24 total points) Researchers would like to estimate a person's body fat percentage (y) by the person's triceps skinfold thickness (x). The researchers take a random sample of 20 people. The corresponding data is stored in a comma delimited file bodyfat.csv on the graded web page of the course website. Below is what part of the data looks like after it is read in.

> set1 <- read.csv(file = "bodyfat.csv")

> head(set1)

person y x

1 1 11.9 19.5

2 2 22.8 24.7

3 3 18.7 30.7

4 4 20.1 29.8

5 5 12.9 19.1

6 6 21.7 25.6

> tail(set1)

person y x

15 15 12.8 14.6

16 16 23.9 29.5

17 17 22.6 27.7

18 18 25.4 30.2

19 19 14.8 22.7

20 20 21.1 25.2

Answer the following questions.

* 1. (10 points) State the estimated regression model by using the data to obtain parameter estimates? Use the lm() function for the computations. Just as a reminder, you need to show all your R code and output with the stated model.

 where y represents body fat and x represents triceps skinfold thickness.

> head(set1)

person y x

1 1 11.9 19.5

2 2 22.8 24.7

3 3 18.7 30.7

4 4 20.1 29.8

5 5 12.9 19.1

6 6 21.7 25.6

> mod.fit <- lm(formula = y ~ x, data = set1)

> summary(object = mod.fit)

Call:

lm(formula = y ~ x, data = set1)

Residuals:

Min 1Q Median 3Q Max

-6.1195 -2.1904 0.6735 1.9383 3.8523

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.4961 3.3192 -0.451 0.658

x 0.8572 0.1288 6.656 3.02e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.82 on 18 degrees of freedom

Multiple R-squared: 0.7111, Adjusted R-squared: 0.695

F-statistic: 44.3 on 1 and 18 DF, p-value: 3.024e-06

* 1. (8 points) Suppose **X** and **Y** are matrices of the standard form in regression analysis:

 and 

Compute ?

> X <- cbind(1, set1$x)

> Y <- set1$y

> solve(t(X)%\*%X)%\*%t(X)%\*%Y

[,1]

[1,] -1.4961046

[2,] 0.8571865

* 1. (6 points) What is the trace of ?

> diag(t(X)%\*%X)

[1] 20.00 13286.29

> sum(diag(t(X)%\*%X))

[1] 13306.29



1. (15 total points) Consider a t-distribution with 30 degrees of freedom.
   1. (7 points) Find the 0.975 quantile (97.5th percentile) from the distribution. In other words, find the t-distribution value such that an area of 0.975 under its curve exists to the left of this value. Use one of the following functions for the calculation: dt(), pt(), qt(), or rt(). Note that we have not used the required function this semester, so you may need to use the help within R.

> qt(p = 0.975, df = 30)

[1] 2.042272

t0.975,30 = 2.04; P(T < 2.04) = 0.975 where T is a random variable with a t-distribution that has 30 degrees of freedom.

* 1. (8 points) Plot the t-distribution with the help of the dt() and curve() functions.

> curve(expr = dt(x = x, df = 30), n = 1000, col = "red", xlim = c(-4,4), main = "t distribution with 30 degrees of freedom", ylab = "f(t)", xlab = "t")

> abline(h = 0)

A diagram of a normal distribution

AI-generated content may be incorrect.

1. (18 total points) Suppose . Answer the following questions.
   1. (10 points) Construct the corresponding contour plot for the probability distribution.

> library(mvtnorm)

> mu <- c(0, 0)

> sigma <- matrix(data = c(1, 0, 0, 1), nrow = 2, ncol = 2, byrow = TRUE)

> x1 <- seq(from = -5, to = 5, by = 0.1)

> x2 <- seq(from = -5, to = 5, by = 0.1)

> all.x <- expand.grid(x1, x2)

> eval.fx <- dmvnorm(x = all.x, mean = mu, sigma = sigma)

> fx <- matrix(data = eval.fx, nrow = length(x1), ncol = length(x2), byrow = FALSE)

> par(pty = "s")

> contour(x = x1, y = x2, z = fx, main = "Multivariate normal contour plot", xlab = expression(x[1]), ylab = expression(x[2]), xlim = c(-5,5), ylim = c(-5, 5), levels = c(0.001, seq(from = 0.02, to = 0.14, by = 0.02)))

A graph with a contour of a contour of a graph

AI-generated content may be incorrect.

* 1. (8 points) What happens to the same contours on the plot in a) as σ12 increases? Provide an intuitive explanation for why the contours change. No plots are needed with your answer.

The contours become more narrow and tilted in a 45-degree angle. This occurs because of the higher, positive correlation between x1 and x2. As x1 increases (decreases), x2 will more likely increase (decrease). Thus, there will be more probability in the corresponding location of the probability distribution.

1. (10 total points) The course notes provide the estimated covariance matrix for p random variables as

Suppose p = 2. By starting with the above equation, show that

Thus, you need to type out the covariance matrix equation in terms of the vectors and multiply out parts of it to form . To help you type out the equations, *some* parts of the answer are included below using the equation editor in Word.

Therefore,



1. (33 total points) Answer the following questions.
   1. (6 points) Why is the last line of an R function important?

The results in the last line are returned as the results for the function. These results can be saved into an object.

* 1. (6 points) What does it mean for two vectors to be orthogonal?

The cross product of two vectors are 0. Thus, .

In a two-dimension case, the vectors are perpendicular.

* 1. (8 points) Explain how one would plot a multivariate normal distribution in three dimensions if p = 2 (bivariate normal distribution). Do not include any R code. Rather, include the process that would need to be used.

Evaluate the multivariate normal density function at a large number of values for x1 and x2 to calculate f(**x**|**μ**,**Σ**). For each (x1, x2) combination, plot the corresponding f(**x**|**μ**,**Σ**) to form the surface of the function.

* 1. (7 points) Why is the covariance matrix equal to the correlation matrix for standardized data?

Because the variances of the random variables are equal to 1. Therefore,



* 1. (6 points) Why is a correlation easier to interpret than a covariance?

A correlation is bounded by -1 and 1, but a covariance is not bounded.