Lecture Notes problem 7.4/2

(a), (b), (d), (f), (i) = 4 pts each,  (c), (e), (g), (h) = 3 pts each,  (j) = 8 pts

Note: For (j), run the R code below (or download 7.4/2(j) and open in R directly), and include the output for each of the parts (e.g., via copy and paste); clearly indicate where the output agrees with your manual computations.

# PART (a)

# INPUT X AND Y DATA VALUES
X = c(0, 3, 6, 9, 12)
Y = c(500, 490, 470, 430, 350)
n = length(X)

# CORRELATION COEFFICIENT
r = cor(X, Y)
print("Correlation Coefficient")
print(r)

# SCATTERPLOT OF Y VERSUS X
plot.new()
plot(X, Y, xlim=c(0,12), ylim=c(350,550), xlab="X = Months", ylab="Y = Assay (mg)", pch=19)

# PART (b)

# LEAST SQUARES REGRESSION LINE OF Y VERSUS X
regline = lm(Y ~ X)
regline

# 95% CONFIDENCE INTERVALS FOR COEFFICIENTS
beta.hats = confint(regline)
print("95% CIs for Coefficients")
print(beta.hats)
# (Ignore the "2.5%" and "97.5%" in the output.)

# PART (c)

# FITTED VALUES
Y.hat = fitted(regline)
print("Fitted Values")
print(Y.hat)

# RESIDUALS
resids = round(resid(regline), 2)
print("Residuals")
print(resids)

# LEAST SQUARES REGRESSION LINE PLOT
abline(regline, lwd=2, col = "blue")
points(X, Y.hat, pch=19, col="blue")
for (i in 1:n)  lines(c(X[i],X[i]), c(Y[i],Y.hat[i]), lty=2)
# PART (d), (e), (f)

**ANOVA TABLE**

```r
anova(regline)
summary(regline)
```

**PLEASE ANSWER:** In the preceding output, why does the p-value of 0.02049 appear twice?

# PART (g) - Estimate Mean Potency at 6 Months

```r
new = data.frame(X = 6)
predict(regline, new, interval = "confidence")
```

# PART (h) - Estimate Months till Expiration

```r
(0.9*500 - coef(regline)[1]) / coef(regline)[2]
```

# PART (i) - Residual Plot

```r
plot.new()
plot(regline, which = 1, id.n = 5, labels.id = resids, pch=19)
```

# PART (j) - Log-Transformed Linear Regression

**FIRST DEFINE RESCALED VARIABLES:** \( \bar{Y} = 510 - Y \) AND \( \bar{X} = X / 3 \)

```r
Xtilde = X / 3
Ytilde = 510 - Y
print(Xtilde)
print(Ytilde)

# SCATTERPLOT OF LN(\( \bar{Y} \)) VERSUS \( \bar{X} \)

```r
ln.Ytilde = log(Ytilde)
plot.new()
plot(Xtilde, ln.Ytilde, xlab = "Xtilde", ylab = "ln(Ytilde)", lwd=2, pch=19)
```

**LEAST SQUARES REGRESSION LINE PLOT OF LN(\( \bar{Y} \)) VERSUS \( \bar{X} \)**

```r
regline.transf = lm(ln.Ytilde ~ Xtilde)
summary(regline.transf)
# Note the Warning message below the summary.
abline(regline.transf, lwd=2)
```

**COMPUTE "SCALE PARAMETER" \( \alpha \) AND "SHAPE PARAMETER" \( \beta \)**

```r
coeffs = coef(regline.transf)
alpha = exp(coeffs[1])
beta = coeffs[2]
```

**PLOT TRANSFORMED MODEL \( \bar{Y} = \alpha \bar{X}^\beta \) ON ORIGINAL (X,Y) SCATTERPLOT**

```r
plot.new()
Y.hat.new = function(X)(510 - alpha * exp(beta * X / 3))
plot(X, Y, xlab = "X = Months", ylab = "Y = Assay (mg)", pch=19)
curve(Y.hat.new, col = "blue", lwd=2, add=TRUE)
```