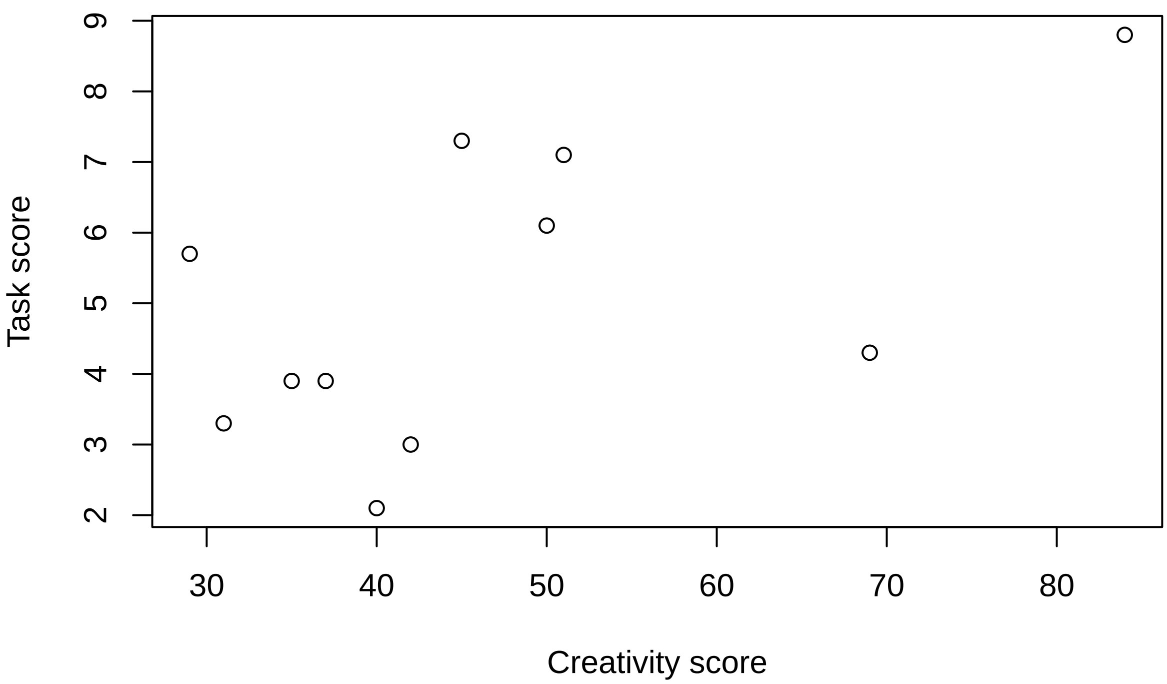
STAT3340 Ass't 1 Solutions, Fall 2024

(out of 35 points)

1. A random sample of 11 elementary school students is selected, and each student is measured on a creativity score (x) using a well-defined testing instrument and on a task score (y) using a new instrument. The task score is the mean time taken to perform several hand-eye coordination tasks. The data are as follows.

```
x=c(35,37,50,69,84,40,29,42,51,45,31)
y=c(3.9,3.9,6.1,4.3,8.8,2.1,5.7,3.0,7.1,7.3,3.3)
plot(x,y,xlab="Creativity score",ylab="Task score")
```



Use R to do the following questions. Make sure to include your commands in the R markdown document.

- 1a) Calculate the summaries S_{xx}, S_{xy}, S_{yy} and $\bar{X}$ and $\bar{Y}$. (**3 points: 1 for each of $\bar{y}$, S_{yy}, S_{xy}**)

```
N=length(x)
xbar=sum(x)/N ;  cat("xbar = ", xbar)
```

```
## xbar =  46.63636
```

```
Sxx=sum((x-xbar)^2) ; cat("Sxx = ",Sxx)
```

```
## Sxx =  2778.545
```

```
ybar=sum(y)/N; cat("ybar = ", ybar)
```

```
## ybar =  5.045455
```

```
Syy=sum((y-ybar)^2) ; cat("Syy = ",Syy)
```

```
## Syy =  44.02727
```

```
Sxy=sum((x-xbar)*(y-ybar)) ; cat("Sxy = ",Sxy)
```

```
## Sxy =  201.5818
```

- 1b) Use these summaries to calculate the least squares estimates of the intercept and slope. (**2 points: 1 for each of $\hat{\beta}_1$ and $\hat{\beta}_0$**)

```
 beta1 = Sxy/Sxx ; cat("beta1 = ", beta1)
```

```
## beta1 =  0.0725494
```

```
 beta0 = ybar - beta1 * xbar ; cat("beta0 =", beta0)
```

```
## beta0 = 1.662014
```

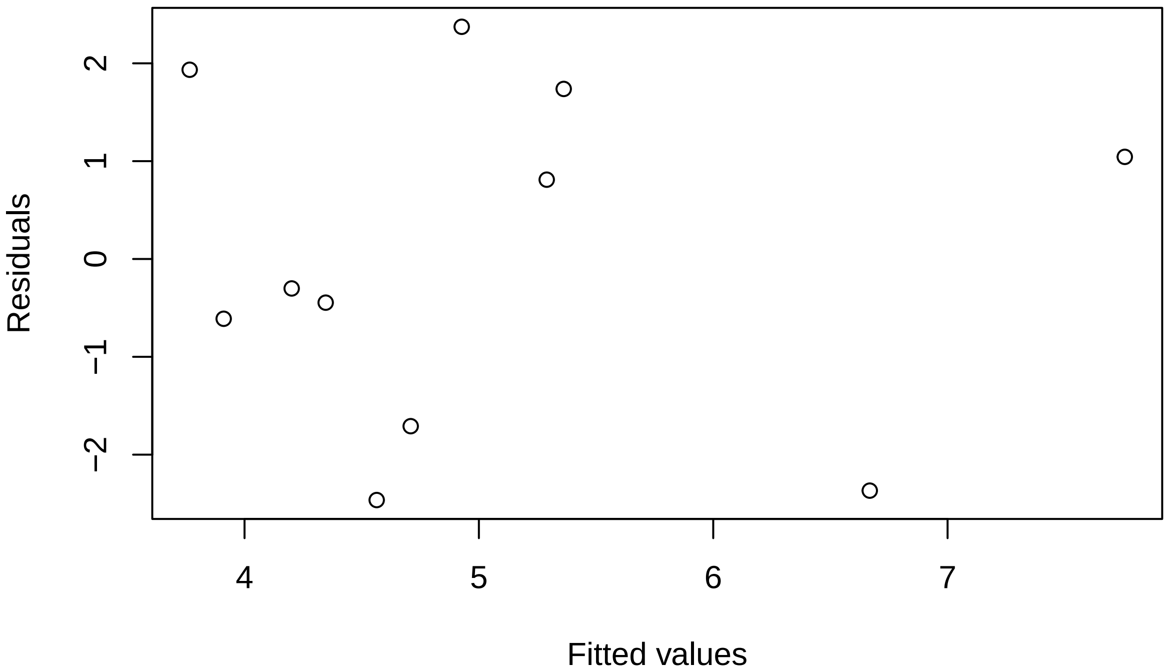
- 1c) Calculate the predicted (fitted) values and the residuals.

```
fit=beta0+beta1*x
resids=y-fit
```

- 1d) Plot the residuals (y axis) vs fitted values (x axis).

(**3 points for plot with reasonably labeled axes. Reduce mark for errors in fits, residuals, or a poorly labeled plot.**)

```
plot(fit,resids,xlab="Fitted values",ylab="Residuals")
```



- 1e) Calculate the mean of the residuals to verify it is zero (to machine precision), and the correlation of the residuals with X to verify it is also zero (to machine precision). The R command to calculate the correlation between u and v is cor(u,v).

(**2 points: 1 for each of mean and correlation. There is no need for rounding.**)

```r
mean(resids)
```

```
## [1] -6.863394e-16
```

```r
cor(resids,x)
```

```
## [1] -1.074939e-16
```

```r
round(mean(resids),5)  #to five decimal digits
```

```
## [1] 0
```

```r
round(cor(resids,x),5) #to five decimal digits
```

```
## [1] 0
```

2. Some data gives the summaries: $n = 10$, $\sum_{i=1}^{10} x_i y_i = 100$, $\sum_{i=1}^{10} x_i = 20$ and $\sum_{i=1}^{10} y_i = 10$. Suppose that the response y is temperature in degrees Celsius.

- $\bar{x} = 20/10 = 2$, $\bar{y} = 10/10 = 1$, so $S_{xy} = \sum_{i=1}^{10} x_i y_i - n\bar{x}\bar{y} = 100 - 10(2)(1) = 80$.
- 2a) (behaviour of sums under linear transformation) If the response was converted to temperature in degrees Fahrenheit ($y' = 32 + 1.8y$), what is $\sum_{i=1}^{10} y_i'$?

$$\begin{aligned}\sum_{i=1}^{10} y_i' &= \sum_{i=1}^{10}(32 + 1.8y_i) \\ &= 32(10) + 1.8\sum_{i=1}^{10} y_i \\ &= 320 + 1.8(10) = 338\end{aligned}$$

(**3 points: The final answer is sufficient for full marks, with no need to show work. If partial marks, subtract 1 point for each error.**)

- 2b) (Behaviour of sums of squares under linear transformation of y.) If the response was converted to temperature in degrees Fahrenheit, what is $S_{xy'}$?

$$\begin{aligned}S_{xy'} &= \sum_{i=1}^{10}(x_i y_i') - (\sum_{i=1}^{10} x_i \sum_{i=1}^{10} y_i')/n \\ &= \sum_{i=1}^{10}(x_i(32 + 1.8y_i)) - (\sum_{i=1}^{10} x_i \sum_{i=1}^{10} y_i')/n \\ &= 32\sum_{i=1}^{10}(x_i) + 1.8\sum_{i=1}^{10}(x_i y_i) - (\sum_{i=1}^{10} x_i \sum_{i=1}^{10} y_i')/n \\ &= 32(20) + 1.8(100) - (20)(338)/10 \\ &= 144\end{aligned}$$

(**4 points: The final answer is sufficient for full marks, with no need to show work. If partial marks, subtract 1 point for each error.**)

3. Find the equation of the line which passes through the points (1,1) and (4,5).

Slope is (5-1)/(4-1) = 4/3. To find intercept, solve $y - y_0 = (4/3)(x - 0)$, where y_0 is the intercept, and (x,y) is either of the given points.

$$(1 - y_0) = (4/3) \to y_0 = 1 - 4/3 = -\frac{1}{3}$$

.

$$y = -\frac{1}{3} + \frac{4}{3}x$$

(**3 points: 2 for slope, 1 for intercept. No work need be shown.**)

4. Derive the partial derivative of $SSE = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$ with respect to β_1.

(**3 points for final answer, subtract 1 point for each error**)

$$\frac{\partial}{\partial \beta_1} SSE = -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i)$$

5. Where $f(x) = log(x) + e^{x^2}$, derive $f'(x)$, the derivative of $f(x)$.

(Note: throughout this course $log(x) = log_e(x) = ln(x)$, the natural logarithm.)

(**3 points for final answer, subtract 1 point for each error**)

$$f'(x) = \frac{1}{x} + 2xe^{x^2}$$

6. Where $f(x) = arcsin(\sqrt{x})$, derive $f'(x)$.

(Note: *arcsin* is the inverse *sin* function. It will be clear later why this derivative is of interest.)

(**3 points for final answer, subtract 1 point for each error**)

$$f'(x) = \frac{1}{2}x^{-1/2}\frac{1}{\sqrt{1-x}}$$

7.(Inverse of a diagonal matrix.)

Where

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

calculate M^{-1}, the inverse of M.

(**3 points for final answer, subtract 1 point for each error**)

$$M^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

8. (Inverse of a 2×2 matrix.)

Where

$$M = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

calculate M^{-1}, the inverse of M.

(**3 points for final answer, subtract 1 point for each error**)

$$M^{-1} = \frac{1}{4-6}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -2 & 1 \\ 1.5 & -.5 \end{bmatrix}$$