

Data Mining and Machine Learning in Bioinformatics

Group members (Name, Student ID, E-Mail):

- Shinho Kang, 2890169, wis.shinho.kang@gmail.com
- Omar Trinidad Gutierrez Mendez, 2850441, omar.vpa@gmail.com
- Baldomero Valdez, Valenzuela, 2905175, baldmer.w@gmail.com

Exercise Series 5

Task 1:

```
preg.week = c(18,24,30,34,40)
elt.year = c(18,26,30,40,70)

#=====
# Calculate values MANUALLY
#mean(preg.week) = 29.2
preg.week.mean = sum(preg.week) / 5
#mean(elt.year) = 36.8
elt.year.mean = sum(elt.year) / 5

#var(preg.week) = 73.2
preg.week.var = 1/4*sum((preg.week-preg.week.mean)**2)
#var(elt.year) = 407.2
#elt.year.var = 1/4*sum((elt.year-elt.year.mean)**2)

#cov(preg.week, elt.year) = 158.8
preg.elt.cov = 1/4*sum((preg.week-preg.week.mean)*(elt.year-elt.year.mean))
```

a) Determine the coefficients of the linear regression

```
# y_i = Beta_0 + Beta*x_i + Epsilon_i
# => least squares fit
# => Beta = Cov(x,y) / Var(x), Beta_0 = mean(y) - Beta*mean(x)
# here, x is preg.week, y is elt.year.
beta = preg.elt.cov / preg.week.var # 2.169399
beta_0 = elt.year.mean - beta*preg.week.mean # 26.54645
#summary(lm(elt.year~preg.week))
```

b) residual variance

```
# residual variance =  $\text{Eta}^2 = 1/(n-2) * \text{sum}((y_i - \text{Beta}_0 - \text{Beta} * x_i)^2)$   
residual.variance = 1/3 * sum((elt.year - beta_0 - beta * preg.week)**2) # 83.59927
```

c) standard error of the slope coefficient Beta

```
# se(Beta) = sqrt(residual.variance) * sqrt(1/((n-1)*var(x)))  
se.beta = sqrt(residual.variance / (4*preg.week.var) ) # 0.5343376
```

d) 95% confidence interval for the slope coefficient

```
# [Beta - se(Beta)*abs(qt(0.025, n-2)), Beta + se(Beta)*abs(qt(0.025, n-2))]  
interval = beta + se.beta * qt(c(0.025, 0.975), 3) # 0.4688983 3.8698995
```