MAST20005/MAST90058: Computer Lab Practice Test Solutions

1. ```r
   marks <- read.csv("marks.csv")  # load the data
   ```

   (a) ```r
   cor(marks)[1, 2]

   ## [1] 0.5666219
   ```

   (b) ```r
   m1 <- lm(semester2 ~ semester1, marks)
   coef(m1)   # estimates of the regression coefficients

   ## (Intercept)   semester1
   ## -6.6516667   0.8916667

   sigma(m1) # estimate of the standard deviation

   ## [1] 12.91006
   ```

   (c) ```r
   confint(m1)

   ##                   2.5 %    97.5 %
   ## (Intercept) -47.8288934 34.525560
   ## semester1     0.3323607  1.450973
   ```

   (d) ```r
   newstudent1 <- data.frame(semester1 = 75)
   predict(m1, newdata = newstudent1, interval = "prediction", level = 0.8)

   ##   fit   lwr   upr
   ## 1 60.22333 42.83703 77.60964
   ```

   (e) ```r
   # We need to fit the 'reverse' of the previous model.
   m2 <- lm(semester1 ~ semester2, marks)
   newstudent2 <- data.frame(semester2 = 70)
   predict(m2, newdata = newstudent2, interval = "prediction", level = 0.8)

   ##        fit      lwr      upr
   ## 1 77.16238 65.99842 88.32635
   ```

2. ```r
   yields <- read.delim('yields.txt')  # load the data
   ```

   (a) ```r
   quantile(yields[yields$machine == "A", "yield"])

   ##   0%   25%   50%   75% 100%
   ## 18.90 30.40 34.50 37.35 48.60
   ```

   (b) ```r
   t.test(yield ~ machine, yields, var.equal = TRUE)$p.value

   ## [1] 0.1152238
   ```

   We cannot reject $H_0$.
   ```

(c)
```r
wilcox.test(yield ~ machine, yields)$p.value
```

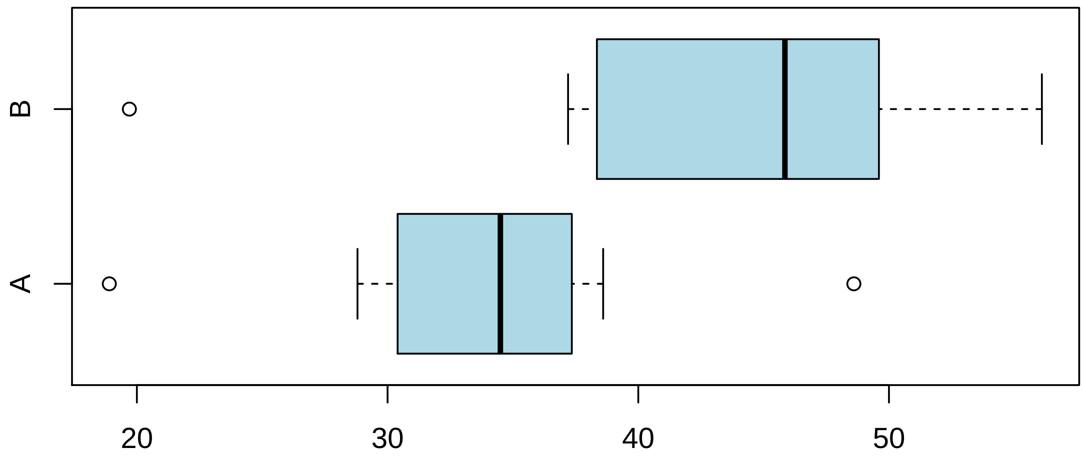
```
## [1] 0.04009324
```

We reject H_0.

(d) No, the results differ. The bulk of the data indicate a difference in mean/median and this is picked up by the Wilcoxon test. However, the presence of outliers in both samples has a substantial influence on the t-test, leading to a non-significant result.

To notice this, you will need to have visualised the data. For example, using a box plot:

```r
boxplot(yield ~ machine, yields, horizontal = TRUE, col = "lightblue")
```



(e)
```r
s <- sd(yields[yields$machine == "A", "yield"])  # sd
k <- qnorm(0.05, 40, s / sqrt(20))  # critical value for x.bar
pnorm(k, 35, s / sqrt(20))  # power
```

```
## [1] 0.7909046
```

The test is for $H_0\colon \mu = 40$ against $H_1\colon \mu < 40$, using $\bar{X} \sim \mathrm{N}(\mu, \sigma_A/\sqrt{n})$ as the test statistic, with σ_A assumed known and equal to s_A.