Charles Liu
304804942
Stats 101B
Disc. 3A
Prof. Shi
HW 1

① (Problem 1.3) Answers Below

i) Statement of the Problem: We would compare the growth of flowers with the different conditions of sunlight, water, fertilizer, and soil conditions.

ii) Selecting Response Variable: The response variable would be growth in height of flowers OR mass of flowers

iii) Choice of Factors, Levels, and Range: (All given below

Non-Controllable Factors:
- Humidity (very humid – not humid)
- Experienced/Inexperiend Gardeners (how much experience)

Control Design Factors:
- Amount of sunlight (0hrs. – 24+hrs.)
- Amount of H_2O (0oz. – 16 oz.)
- Amount of fertilizer (0 oz. – 16oz.)
- Soil conditions
- Initial Flower's Height (1in. – 24in.+)

Blocking Factors:
- Number of flowers
- Growth rate
- stem's height & diameter
- Produces pollen, fruits, or none

② (Problem 1.8) Anwers Below
- Replication is independently repeat run of each factor combinations. It is when the treatment is applied to different (multiple) experiment units.
- We need replication in an experiment because you can obtain the "estimate of experimental error". More importantly, it reduces variability within and between runs.
- Let's say we have Flower 1 (F1) and Flower 2 (F2). Our treatments are A (more water) and B (less water). If we expose say:

F1 to A } 1st run
F2 to B }
F1 to A } 2nd Run
F2 to B }
F1 to A } 3rd run
F2 to B }
F1 to A } 4th run
F2 to B }
⋮
Repeated Measurements

VS.

F1 to A } 1st run
F2 to B }
F1 to B } 2nd run
F2 to A }
F1 to A } 3rd run
F2 to A }
F1 to B } 4th run
F2 to B }
⋮
Replication

③

A	30	29	27
B	32	35	34

$$\left.\begin{aligned} H_0&: \mu_1=\mu_2 \\ H_a&: \mu_1<\mu_2 \end{aligned}\right\} \quad t_0=\frac{\bar{Y}_1-\bar{Y}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2} \text{ under } H_0$$

$$\left.\begin{array}{c}\text{Randomization} \\ \text{Test}\end{array}\right\} D_{obs.}=\bar{Y}_B-\bar{Y}_A$$

$n_A=3$ $\quad$ $n_B=3$

$\bar{Y}_A=\frac{86}{3}$ $\quad$ $\bar{Y}_B=\frac{101}{3}$ $\quad$ $\alpha=0.05$

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

$S_A^2=\frac{7}{3}$ $\quad$ $S_B^2=\frac{7}{3}$

$S_A=1.53$ $\quad$ $S_B=1.53$

$S_p=1.53$

$$S^2=\frac{\Sigma(x_i-\bar{x})^2}{n-1}$$

$$S_A^2=\frac{\Sigma\left[\frac{16}{9}+\frac{1}{9}+\frac{25}{9}\right]}{3-1}=\frac{7}{3}$$

$$S_B^2=\frac{\Sigma\left[\frac{25}{9}+\frac{16}{9}+\frac{1}{9}\right]}{3-1}=\frac{7}{3}$$

$$S_p^2=\frac{(3-1)\frac{7}{3}+(3-1)\frac{7}{3}}{3+3-2}=\frac{7}{3}$$

$S_p=1.53$

$$t_0=\frac{\left(\frac{101}{3}\right)-\left(\frac{86}{3}\right)}{(1.53)\cdot\sqrt{\frac{1}{3}+\frac{1}{3}}}=4.00$$

$P(t\geq|t_0|)$ for $H_a: \mu_1<\mu_2$

$|t_0=4.00| > |t_{0.025,4}=2.776|$

$\boxed{\text{We would Reject the } H_0 \text{ and conclude the mean of the treatments differ.}}$

P-value under t-test:

$P(|t|>|t_0=4.00|)=P(t\geq 4.00 \text{ or } t\leq -4.00)$

$\boxed{\text{P-val.}=0.01601}$

Randomization test:

$D_{obs.}=\frac{101}{3}-\frac{86}{3}=\frac{15}{3}=5$

$\binom{6}{3}=20$ possible combinations

$\text{P-value}=P(D>D_{obs.})+\frac{1}{2}P(D=D_{obs.})$

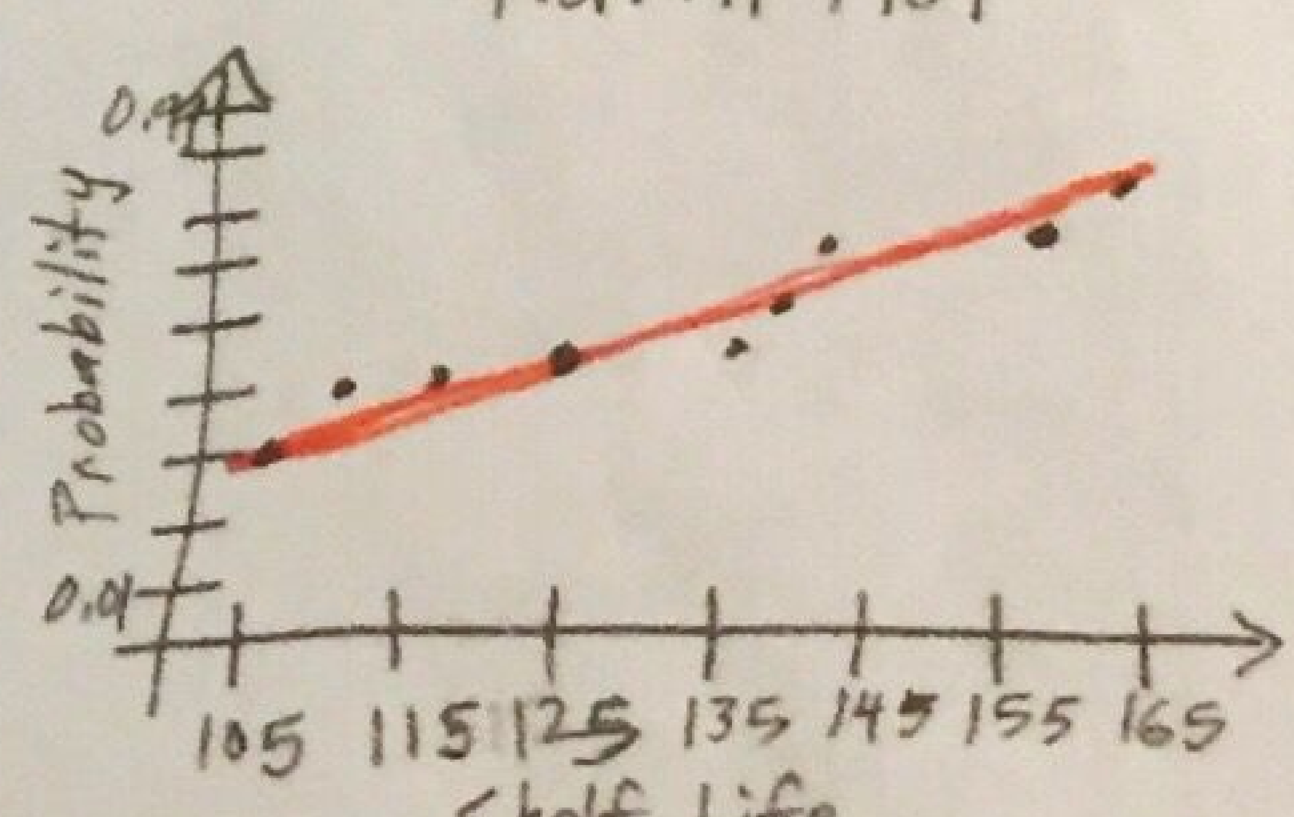
$\boxed{\text{P-val.}=0.01}$

---

④ (Problem 2.23) Answers Below

- We plot it using software shown here. Therefore, we say the Shelf Life can be modeled adequately by Normal Distribution.
- If it is NOT Normally Distributed, then the impact won't be to severe since it uses the t-test rather than Z-scores. However, if it is far from normality, we will see some severe consequences and problems to our model.

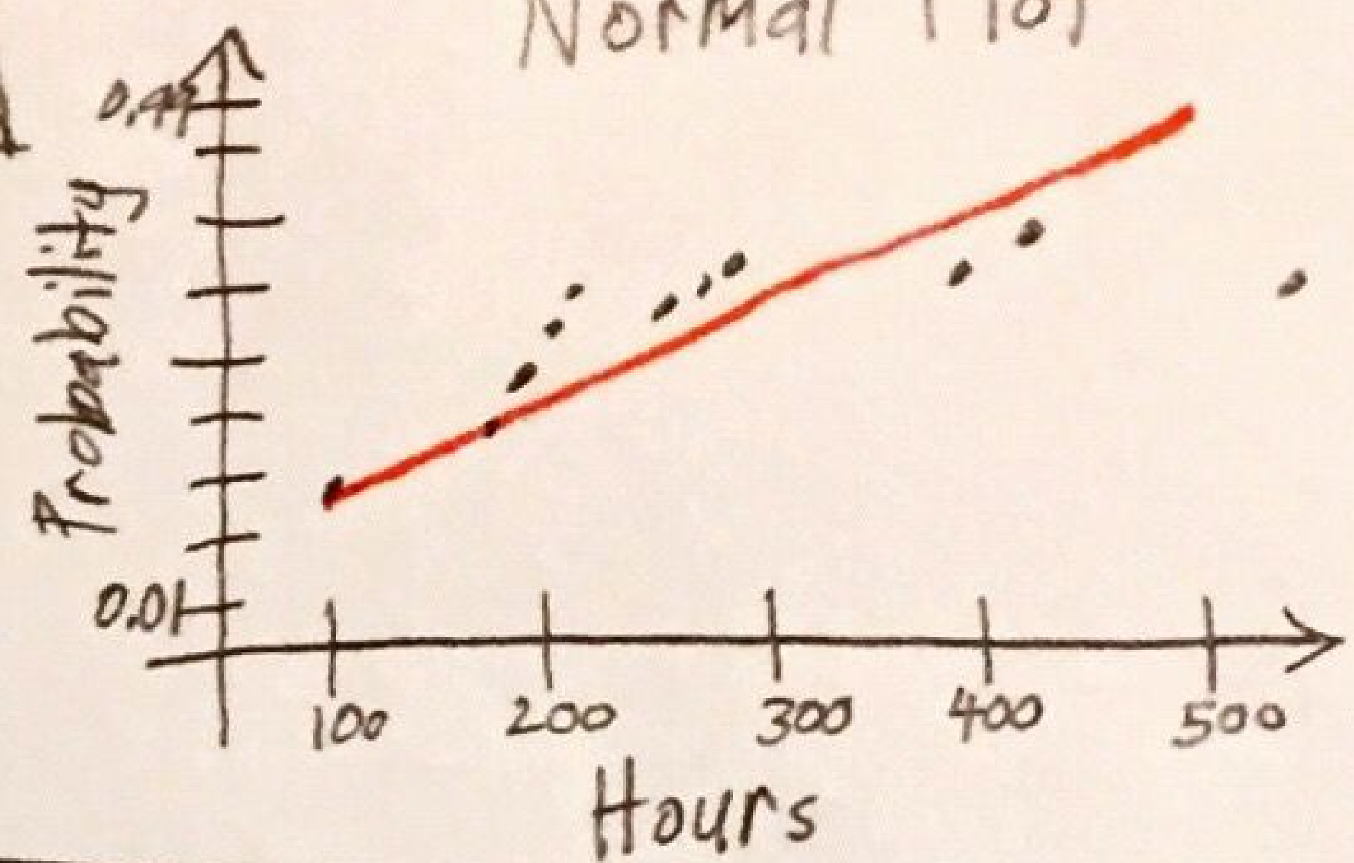
Normal Plot



Charles Liu
304804942
Stats 101B
Disc. 3A
Prof. Shi
HW1 (cont.)

5 (Problem 2.25) Answers Below

- We can plot Hours and be modeled adequately by Normal Distribution.



6 (Problem 2.29)

"Answers done in RStudio & attached to this PDF"

c) $H_0: \sigma_1^2 = \sigma_2^2$

$H_a: \sigma_1^2 \neq \sigma_2^2$

Stats_101B_HW1_Charles_Liu

Charles Liu (304804942)

4/12/2020

Exercise 6 (Problem 2.29)

a)

```r
C2F6_125 <- c(2.7, 4.6, 2.6, 3.0, 3.2, 3.8)
C2F6_200 <- c(4.6, 3.4, 2.9, 3.5, 4.1, 5.1)

t.test(C2F6_125, C2F6_200)
```

```
##
##  Welch Two Sample t-test
##
## data:  C2F6_125 and C2F6_200
## t = -1.3498, df = 9.9404, p-value = 0.207
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.6354454  0.4021121
## sample estimates:
## mean of x mean of y
##  3.316667  3.933333
```

No, C2F6 Flow Rate does not affect average etch uniformity.

b)

```r
t.test(C2F6_125, C2F6_200)$p.value # Our p-value is approximately 0.21
```

```
## [1] 0.2070179
```

c)

```r
var.test(C2F6_125, C2F6_200, alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  C2F6_125 and C2F6_200
## F = 0.85623, num df = 5, denom df = 5, p-value = 0.8689
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1198124 6.1189129
## sample estimates:
```

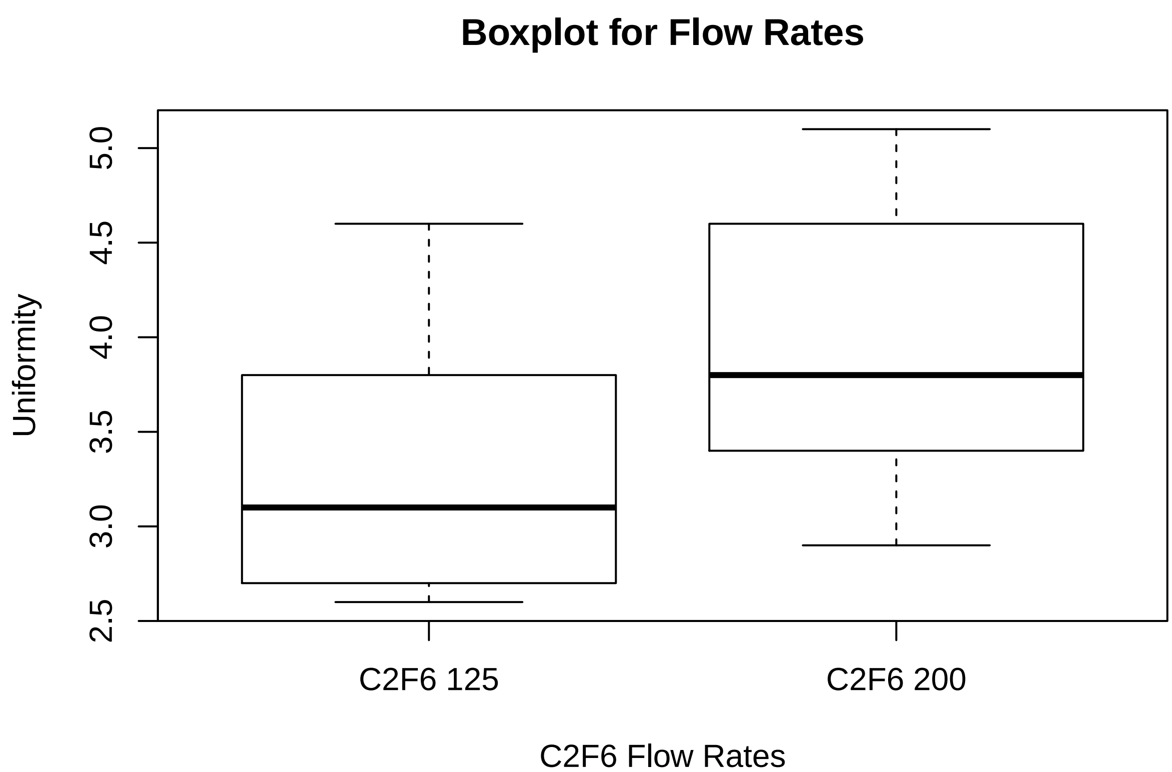
```
## ratio of variances
##          0.8562253
```

```
# F0 = 0.86 and our p-value = 0.87
```

Do NOT reject the Null Hypothesis. The C2F6 Flow Rate does not affect wafer-to-wafer variability.

d)

```
boxplot(C2F6_125, C2F6_200, main = "Boxplot for Flow Rates", names = c("C2F6 125", "C2F6 200"), xlab =
```



The boxplot shown indicates that there is little to no difference in uniformity for the two types of flow rates.