101B_hw01_Daren_Sathasivam

Daren Sathasivam

2024-04-08

1 In class activity: Making Meringues

Group's experiment design approach:

Group A(control): Using the ingredients - eggwhites, sugar, cream, and vanilla extract (general google recipe)
Group B: Substituting eggwhites with aquafaba(liquid from canned chickpeas) - same portion as group A
Group C: Substituting eggwhites with meringue powder - same portion as group A

- Have several randomly assigned people make meringues based on listed recipe and repeat it several times.
- Measure weight/diameter of meringue
- Find the average size/weight for each group
- Compare results between each group

a. What kind of approach do you use to design the experiment, best guess approach, one factor at a time approach, or factorial design approach?

- For the experiment, a factorial design approach is most suitable as it allows to test multiple different factors to understand how certain different recipes can change the size of the meringues.

b What factors do you consider to affect the outcome?

- Type and quantity of ingredients
- Mixing technique or method
- Baking temperature and duration
- Use of additional additives

c. How will you measure the outcome variable?

- Weighing each meringue after baking to determine its weight
- Measuring the diameter of each meringue

d. Which principle of experiment is used in your design? How do you implement the principles in your experiment?

- Randomization: Assigning participants to the different groups randomly helps ensure that any observed differences is not due to systematic biases(skilled vs unskilled bakers)
- Replication: Each recipe variation should be tested multiple times to obtain reliable and consistent results.

e. Update the design of experiment based on what you hvae learned from Chapter 1.

- Specifying the number or replications for each recipe to improve the precision of the estimates.
- Control other confounding variables such as oven temperature, mixing technique, and no additives.
- Blinding technique: participants do not know which substitute they are using to avoid any unconcsious bias when baking the meringue.

2: Problem 2.25

```r
burn <- c(65, 82, 81, 67, 57, 59, 66, 75, 82, 70, 64, 56, 71, 69, 83, 74, 59, 82, 65, 79)
type <- c(rep(1, 10), rep(2, 10))
data2 <- data.frame(burn = burn, type = type)
# data2
```

a. Test for $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_a : \sigma_1^2 \neq \sigma_2^2$

```r
var.test(burn~type)
```

```
##
##  F test to compare two variances
##
## data:  burn by type
## F = 0.97822, num df = 9, denom df = 9, p-value = 0.9744
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2429752 3.9382952
## sample estimates:
## ratio of variances
##          0.9782168
```

- After testing the hypothesis that the two variances are equal using $\alpha = 0.05$, the results yield a p-value of 0.9744, meaning we fail to reject the null hypothesis. This indicates that the variability in burn times does not significantly differ between the two types.

b. Test for $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$

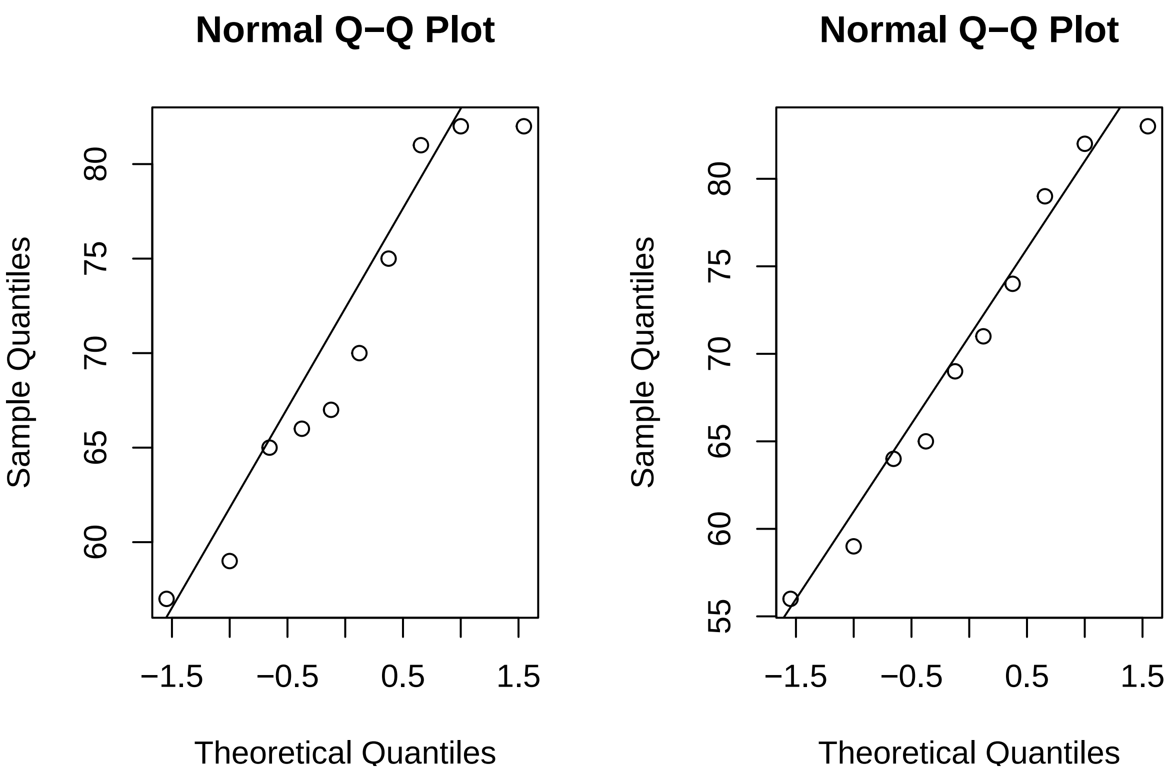
```r
t.test(burn~type, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  burn by type
## t = 0.048008, df = 18, p-value = 0.9622
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -8.552441  8.952441
## sample estimates:
## mean in group 1 mean in group 2
##            70.4            70.2
```

- The p-value for the t-test is 0.9622, meaning we fail to reject the null hypothesis. This indicates that the mean in burn times does not significantly differ between the two types.

c. Discuss the role of normality assumption in this problem. Check the assumption of normality for both types of flares.

```r
par(mfrow = c(1, 2))
qqnorm(burn[type==1]); qqline(burn[type==1])
qqnorm(burn[type==2]); qqline(burn[type==2])
```



- Both results seem to maintain the normality assumption. However, type1 flare deviates from the normal line relatively more in comparison to the type2 flare and also has a slight "S-shaped" curve which can indicate it containing heavier tails than a normal distribution.

3: Problem 2.27 - skip(f) and (g)

```r
thickness <- c(11.176, 7.089, 8.097, 11.739, 11.291, 10.759, 6.467, 8.315, 5.263, 6.748, 7.461, 7.015, 8
temp <- c(rep(95, 8), rep(100, 8))
data3 <- data.frame(thickness = thickness, temp = temp)
# data3
```

a. Is there evidence to support the claim that the higher baking temperature results in wafers with a lower mean photoresist thickness? Use $\alpha = 0.05$

```r
var.test(thickness~temp)
```

```
##
##  F test to compare two variances
##
## data:  thickness by temp
## F = 1.6381, num df = 7, denom df = 7, p-value = 0.5306
## alternative hypothesis: true ratio of variances is not equal to 1
```

```
## 95 percent confidence interval:
##  0.3279572 8.1822436
## sample estimates:
## ratio of variances
##           1.638117
```

```
t.test(thickness~temp, var.equal = TRUE, alternative = "greater")
```

```
##
##  Two Sample t-test
##
## data:  thickness by temp
## t = 2.6751, df = 14, p-value = 0.009059
## alternative hypothesis: true difference in means between group 95 and group 100 is greater than 0
## 95 percent confidence interval:
##  0.8608158       Inf
## sample estimates:
##  mean in group 95 mean in group 100
##          9.366625          6.846625
```

- The F-test's p-value is 0.5306 which is greater than $\alpha = 0.05$. Thus, we fail to reject the null hypothesis of equal variances. Indicating that there is not enough evidence to conclude a significant difference in variances between the two temperature groups. The t-test with an alternative of "greater than" difference for the 95C group has a p-value of 0.009059. Thus we reject the null hypothesis indicating that there is statistically significant evidence to support the claim that higher baking temperatures results in wafers with a lower mean photoresist thickness.

b. What is the p-value for the test conducted in part (a)?

- F-test p-value: 0.5306
- t-test p-value: 0.009059

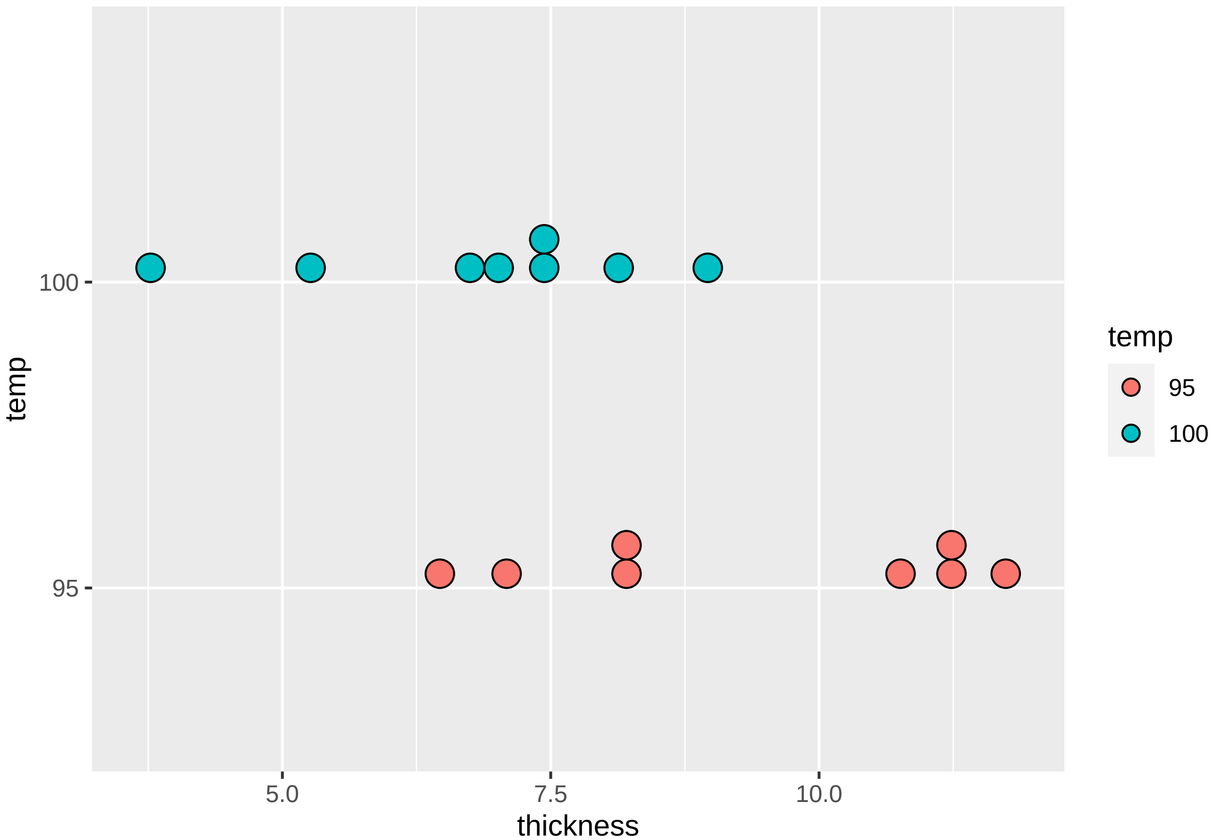
c. Find a 95% confidence interval on the difference in means to provide a practical interpretation of this interval.

- Confidence Interval: $[0.8608158, \infty)$
- Interpretation: The confidence interval suggests that on average, wafers baked at 95C have 0.86 units greater than those baked at 100C. However, since the upper bound is ∞ and unboounded, it produces uncertainty and cannot definitively conclude how much higher the photoresist thickness may be in the 95C group.

d. Draw dot diagrams to assist in interpreting the results from this experiment.

```
data3$temp <- as.factor(data3$temp)
library(ggplot2)
p <- ggplot(data3, aes(x = temp, y = thickness, fill = temp)) +
  geom_dotplot(binaxis = 'y')
p + coord_flip()
```

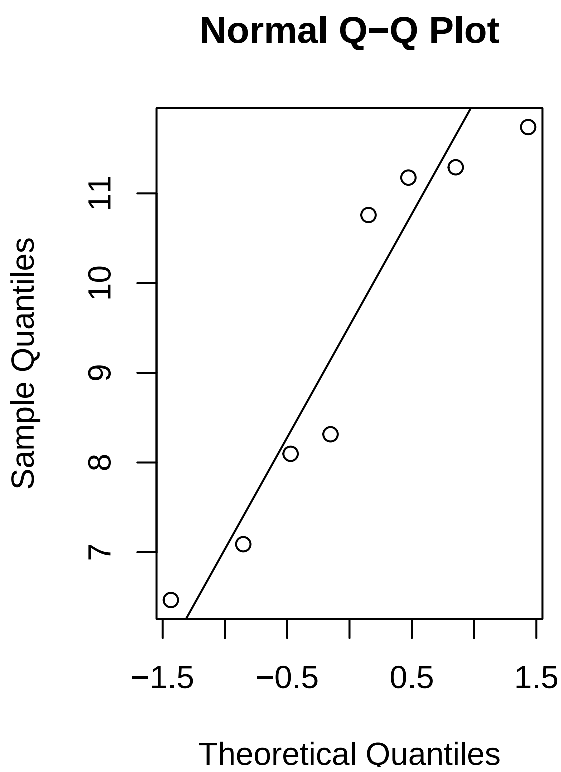
```
## Bin width defaults to 1/30 of the range of the data. Pick better value with
## `binwidth`.
```

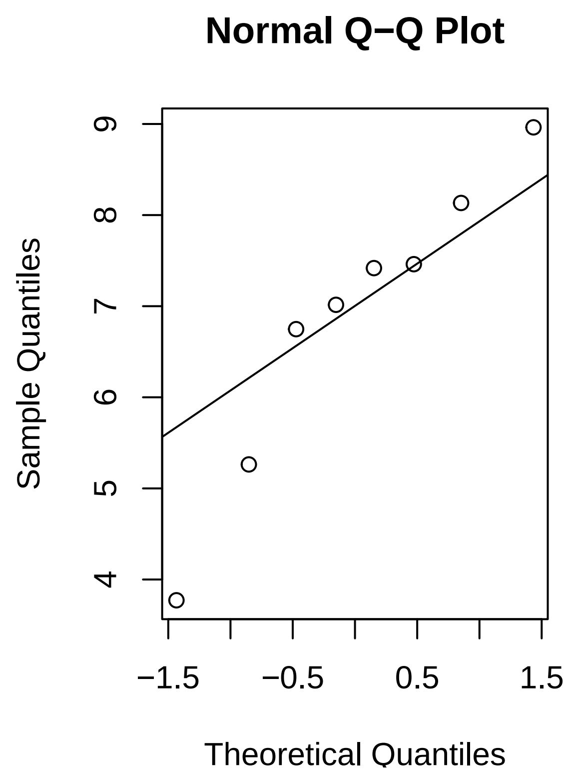


Based off of the dot diagrams, we cannot conclude for certain if the experiment follows a normal distribution. However, we can observe that the 100C group tends to result in a thinner photoresist layer relative to the 95C group. The model and observations can be improved by obtaining a larger sample size.

e. Check the assumption of normality of the photoresist thickness.

```
par(mfrow = c(1, 2))
qqnorm(thickness[temp==95]); qqline(thickness[temp==95])
qqnorm(thickness[temp==100]); qqline(thickness[temp==100])
```





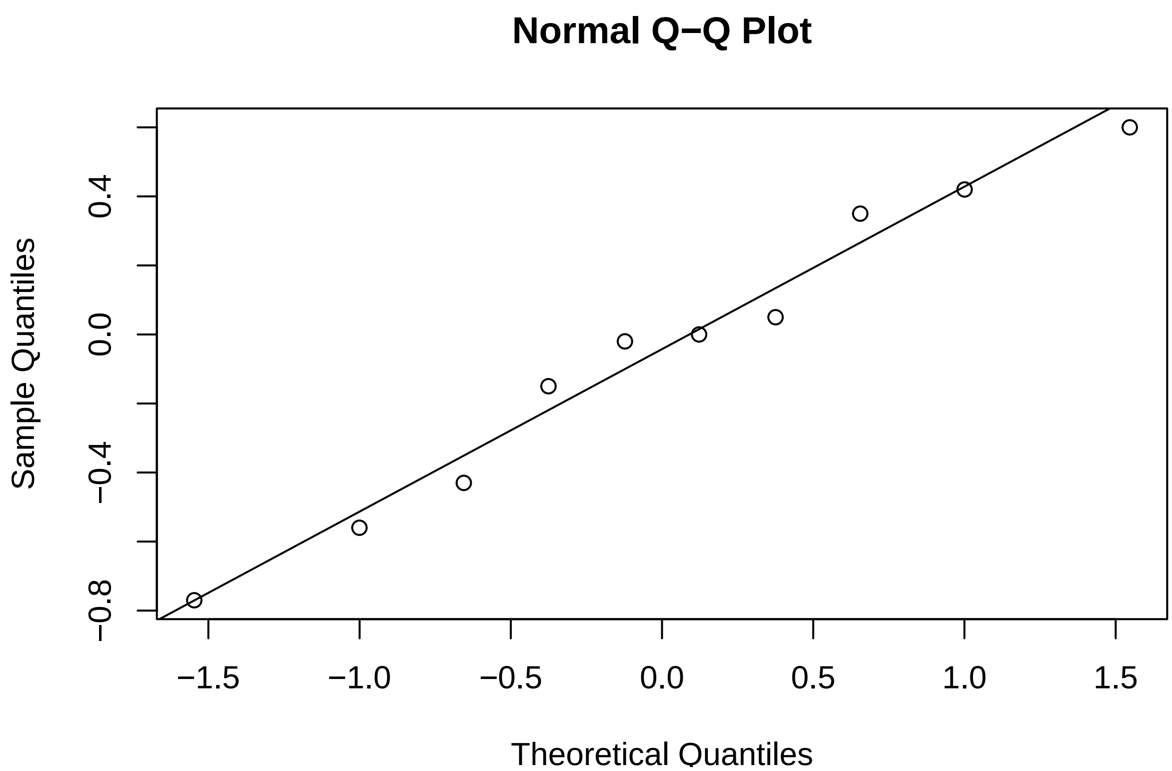
- Both temperature groups result in an "S-shaped" curve which indicates heavier tails than an expected normal distribution. A larger sample size may clarify this observation.

4: Problem 2.30

```
IQ <- c(6.08, 6.22, 7.99, 7.44, 6.48, 7.99, 6.32, 7.60, 6.03, 7.52, 5.73, 5.80, 8.42, 6.84, 6.43, 8.76,
order <- c(rep(1, 10), rep(2, 10))
data4 <- data.frame(IQ = IQ, order = order)
# data4
```

a. Is the assumption that the difference in score is normally distributed reasonable?

```
diff <- IQ[1:10]-IQ[11:20]
qqnorm(diff); qqline(diff)
```



- The normality assumption holds as the observed values on the QQ plot remain relatively near the line, suggesting that the difference in scores is approximately normally distributed.

b. Find a 95% confidence interval on the difference in mean score. Is there any evidence that mean score depends on birth order?

```r
t.test(IQ ~ order, paired = T)
```

```
##
##  Paired t-test
##
## data:  IQ by order
## t = -0.36577, df = 9, p-value = 0.723
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.3664148  0.2644148
## sample estimates:
## mean difference
##          -0.051
```

- Confidence Interval: $[-0.366148, 0.2644148]$
- Since the confidence interval contains a 0, it indicates that there is no statistically significant difference in mean IQ scores between the two birth orders.

c. Test an appropriate set of hypotheses indicating that the mean score does not depend on birth order.

```r
t.test(IQ ~ order, paired = T)
```

```
##
```

```
##  Paired t-test
##
## data:  IQ by order
## t = -0.36577, df = 9, p-value = 0.723
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -0.3664148  0.2644148
## sample estimates:
## mean difference
##          -0.051
```

- Null Hypothesis: $H_0 : \mu_1 = \mu_2$
- Alternative Hypothesis: $H_1 : \mu_1 \neq \mu_2$
- Based on the p-value of 0.723, we fail to reject the null hypothesis. This indicates that there is no statistically significant evidence to conclude a dependence of mean IQ scores on birth order in this experiment.