HW1, Experimental Design

Luke Beebe

2024-02-06

2.21

A normally distributed random variable has an unknown mean μ and a known variance $\sigma^2 = 9$. Find the sample size required to construct a 95 percent confidence interval on the mean that has a total length of 1.0.

[illegible]

```
n = ceiling((qnorm(1-0.05/2)*3/0.5)^2)
n
```

```
## [1] 139
```

n=139 samples required to construct a 95 percent confidence interval on the mean that has a total length of 1.0.

2.22

The shelf life of a carbonated beverage is of interest. Ten bottles are randomly selected and tested, and the following results are obtained:

```
days <- c(108, 124, 124, 106, 115, 138, 163, 159, 134, 139)
```

a. We would like to demonstrate that the mean shelf life exceeds 120 days. Set up appropriate hypotheses for investigating this claim.
H0: μ <= 120
H1: μ > 120

b. Test these hypotheses using $\alpha = 0.01$. What are your conclusions?

```
t <- t.test(days, alternative="greater", mu=120, conf.level=0.99)
t
```

```
##
##  One Sample t-test
##
## data:  days
## t = [illegible], df = 9, p-value = [illegible]
## alternative hypothesis: true mean is greater than 120
## 99 percent confidence interval:
##  [illegible]  Inf
## sample estimates:
## mean of x
##       131
```

My conclusion is that **we cannot reject the null.**

c. Find the P-value for the test in part (b).

```
t$p.value
```

```
## [1] [illegible]
```

[illegible]

d. Construct a 99 percent confidence interval on the mean shelf life. [illegible]

```
[illegible]
```

```
## [1] [illegible]
## attr(,"conf.level")
## [1] 0.99
```

[illegible]

2.23

Consider the shelf life data in Problem 2.22. Can shelf life be described or modeled adequately by a normal distribution?

No, there are not enough observations at n=10.

What effect would the violation of this assumption have on the test procedure you used in solving Problem 2.22?

[illegible]

2.31

Photoresist is a light-sensitive material applied to semiconductor wafers so that the circuit pattern can be imaged on to the wafer. After application, the coated wafers are baked to remove the solvent in the photoresist mixture and to harden the resist. Here are measurements of photoresist thickness (in kÅ) for eight wafers baked at two different temperatures. Assume that all of the runs were made in random order.

```
celsius95 <- c([illegible])
celsius100 <- c([illegible])
```

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

```
t <- t.test(celsius95, celsius100, var.equal=TRUE, conf.level=0.95)
t
```

```
##
##  Two Sample t-test
##
## data:  celsius95 and celsius100
## t = [illegible], df = 14, p-value = [illegible]
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4884278 4.5515722
## sample estimates:
## mean of x mean of y
## [illegible]
```

Yes, there is evidence to support the claim that higher baking temperature results in wafers with a lower mean photoresist thickness at $\alpha = 0.05$.

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

```
t$p.value
```

```
## [1] 0.01884639
```

0.01884639

c. Find a 95 percent confidence interval on the difference in means. Provide a practical interpretation of this interval.

```
t$conf.int
```

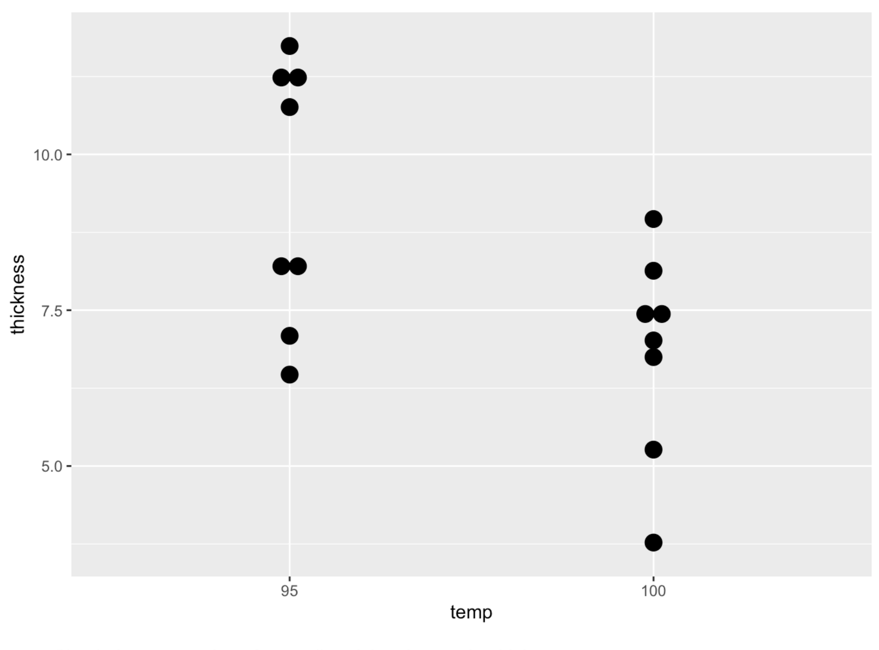
```
## [1] 0.4884278 4.5515722
## attr(,"conf.level")
## [1] 0.95
```

The CI does not contain '0'. This signifies, **at 95% confidence, that the difference in means is not equal to '0'.** Which means we have evidence that the means are different.

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

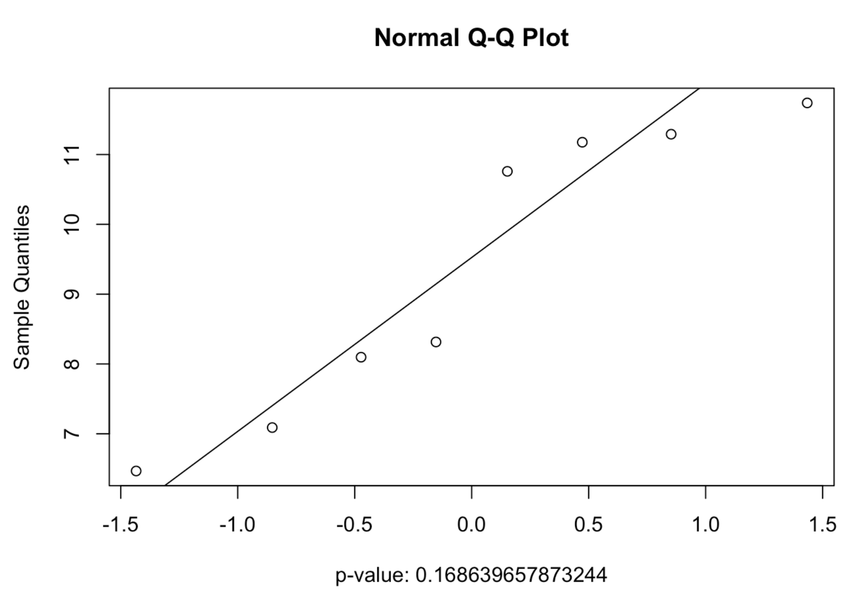
```
celsius_df <- data.frame(append(celsius95,celsius100), 95)
celsius_df[9:16,2]<- 100
colnames(celsius_df) <- c("thickness", "temp")
celsius_df$temp <- as.factor(celsius_df$temp)
ggplot(celsius_df, aes(x=temp,y=thickness)) + geom_dotplot(binaxis = 'y', stackdir = "center")
```

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

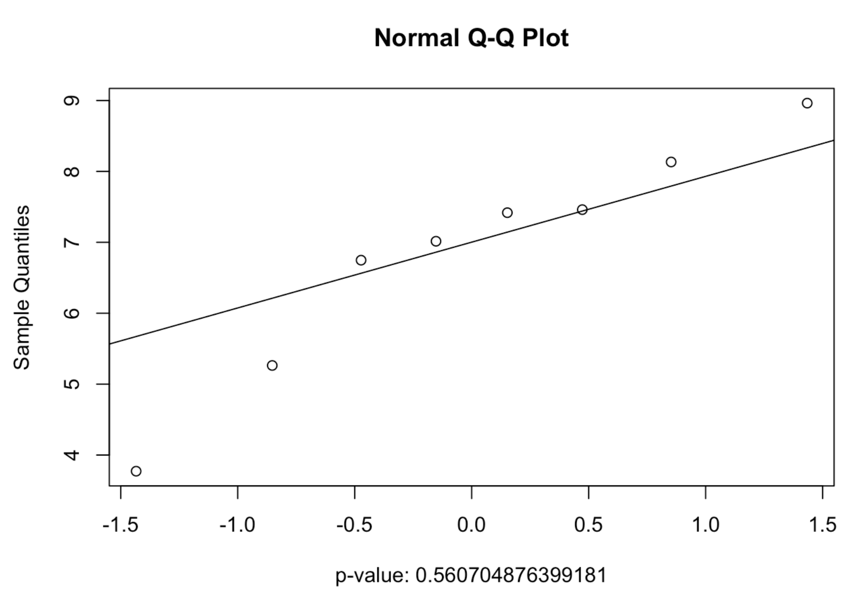


e. Check the assumption of normality of the photoresist thickness.

```
qqnorm(celsius95, xlab = paste("p-value:", shapiro.test(celsius95)$p))
qqline(celsius95)
```




```
qqnorm(celsius100, xlab = paste("p-value:", shapiro.test(celsius100)$p))
qqline(celsius100)
```



We cannot reject the null, therefore we can assume normality.

f. Find the power of this test for detecting an actual difference in means of 2.5 kA.

```
power.t.test(n=8, delta=2.5, sd=sd_pooled(celsius95,celsius100))
```

```
##
##      Two-sample t test power calculation
##
##               n = 8
##           delta = 2.5
##              sd = 1.884034
##       sig.level = 0.05
##           power = 0.6945829
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

The power of this test for detecting an actual difference in means of 2.5 kA is **0.6945829**. This is using the pooled sd of celsius95 and celsius100.

g. What sample size would be necessary to detect an actual difference in means of 1.5 kA with a power of at least 0.9?

```
power.t.test(delta=1.5,power=0.9, sd=sd_pooled(celsius95,celsius100))
```

```
##
##      Two-sample t test power calculation
##
##               n = 34.14244
##           delta = 1.5
##              sd = 1.884034
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

The sample size necessary is **n=35** for *each* group. This is using the pooled sd of celsius95 and celsius100.

2.32

Front housings for cell phones are manufactured in an injection molding process. The time the part is allowed to cool in the mold before removal is thought to influence the occurrence of a particularly troublesome cosmetic defect, flow lines, in the finished housing. After manufacturing, the housings are inspected visually and assigned a score between 1 and 10 based on their appearance, with 10 corresponding to a perfect part and 1 corresponding to a completely defective part. An experiment was conducted using two cool-down times, 10 and 20 seconds, and 20 housings were evaluated at each level of cool-down time. All 40 observations in this experiment were run in random order. The data are as follows.

```
seconds10 <- c(1,2,1,3,5,1,5,2,3,5,3,6,5,3,2,1,6,8,2,3)
seconds20 <- c(7,8,5,9,5,8,6,4,6,7,6,9,5,7,4,6,8,5,8,7)
```

a. Is there evidence to support the claim that the longer cool-down time results in fewer appearance defects? Use $\alpha = 0.05$.

```
t <- t.test(seconds10, seconds20, conf.level = 0.95)
t
```

```
##
##  Welch Two Sample t-test
##
## data:  seconds10 and seconds20
## t = -5.5696, df = 35.601, p-value = 2.706e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.297473 -2.002527
## sample estimates:
## mean of x mean of y
##      3.35      6.50
```

Yes, at $\alpha = 0.05$, there is enough evidence to support the claim that the 20 second cool-down time results in fewer appearance defects.

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

```
t$p.value
```

```
## [1] 2.705576e-06
```

2.705576e-06

c. Find a 95 percent confidence interval on the difference in means. Provide a practical interpretation of this interval.

```
t$conf.int
```

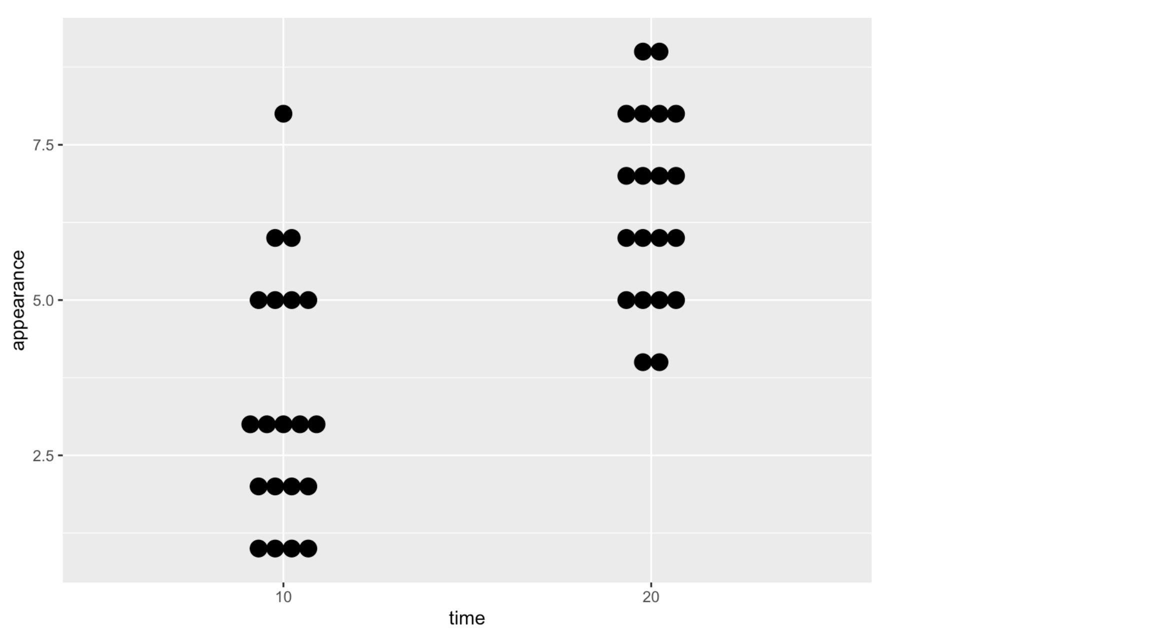
```
## [1] -4.297473 -2.002527
## attr(,"conf.level")
## [1] 0.95
```

I am 95% confident that the true difference of means is within the interval (-4.297473, -2.002527). This signifies that **there is a statistically significant difference in means between the two groups.**

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

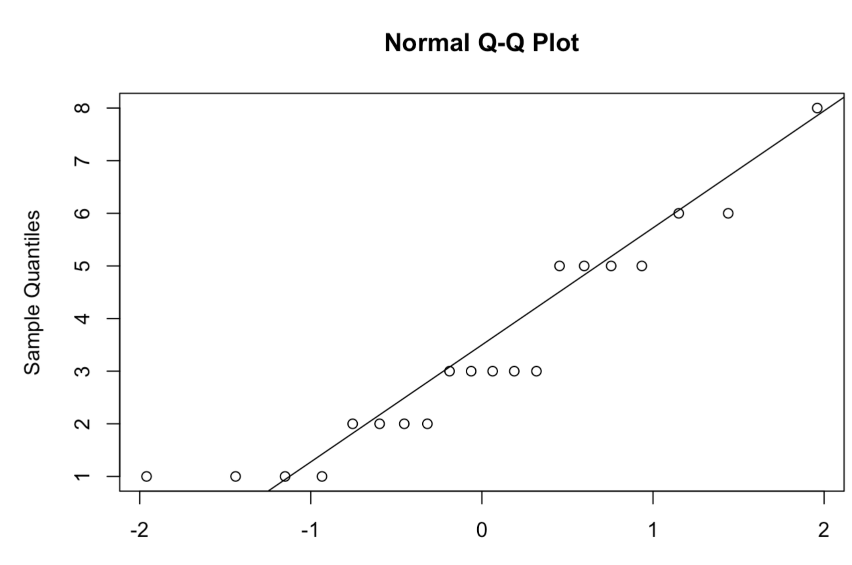
```
seconds_df <- data.frame(append(seconds10, seconds20),10)
seconds_df[21:40,2]<- 20
colnames(seconds_df) <- c("appearance", "time")
seconds_df$time <- as.factor(seconds_df$time)
ggplot(seconds_df, aes(x=time,y=appearance)) + geom_dotplot(binaxis = 'y', stackdir = "center")
```

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

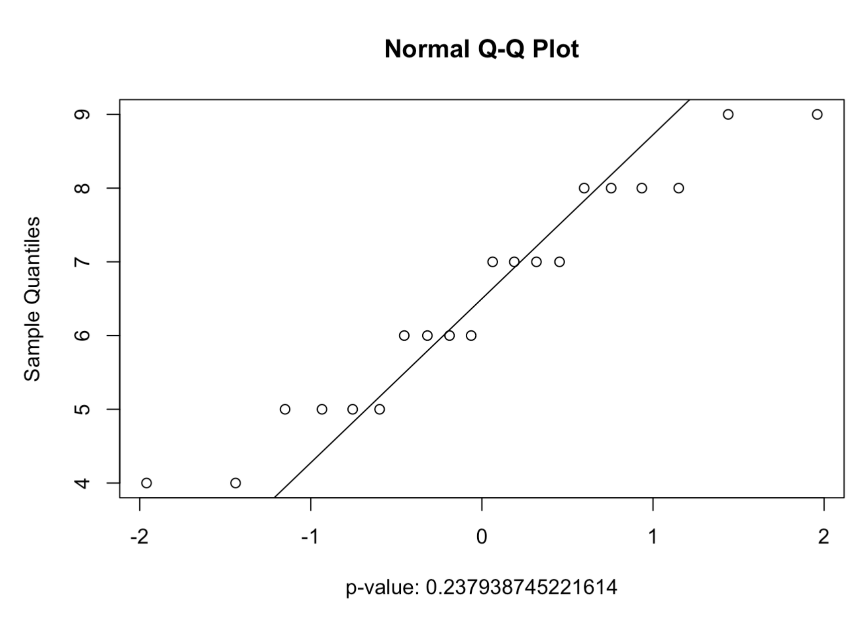


e. Check the assumption of normality for the data from this experiment.

```
qqnorm(seconds10, xlab = paste("p-value:", shapiro.test(seconds10)$p))
qqline(seconds10)
```




```
qqnorm(seconds20, xlab = paste("p-value:", shapiro.test(seconds20)$p))
qqline(seconds20)
```



The p-value of the 10 second cool-down seems to be more askew than normal. We will have to test this further.

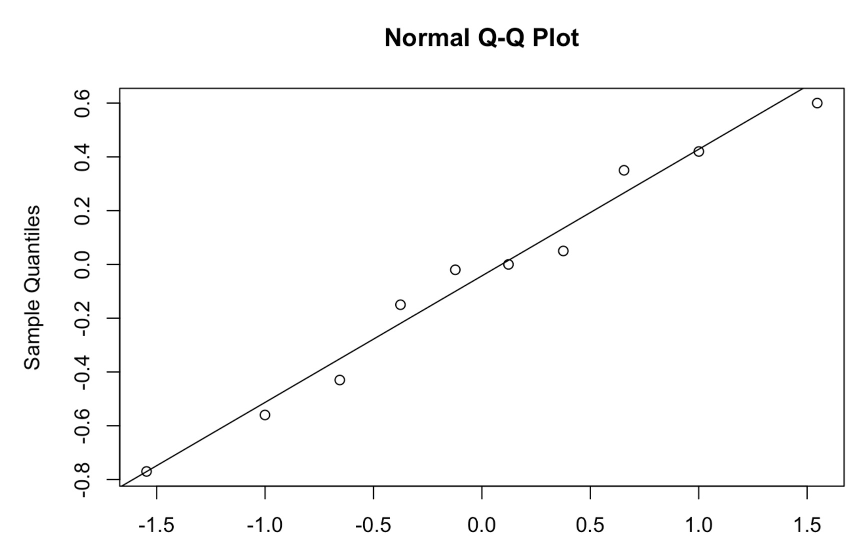
2.35

An article in the journal Neurology (1998, Vol. 50, pp. 1246–1252) observed that monozygotic twins share numerous physical, psychological, and pathological traits. The investigators measured an intelligence score of 10 pairs of twins. The data obtained are as follows:

```
pair <- c(1,2,3,4,5,6,7,8,9,10)
order1 <- c(6.08,6.22,7.99,7.44,6.48,7.99,6.32,7.6,6.03,7.52)
order2 <- c(5.73,5.8,8.42,6.84,6.43,8.76,6.32,7.62,6.59,7.67)
```

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

```
qqnorm(order1-order2, xlab = paste("p-value:", shapiro.test(order1-order2)$p))
qqline(order1-order2)
```



Yes, the assumption is reasonable that the difference in score is normally distributed. Our p-value is very high from the Shapiro-Wilk test and the qqplot trends linear.

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

```
t <- t.test(order1, order2, conf.level = 0.95, paired = T)
t$conf.int
```

```
## [1] -0.3664148  0.2644148
## attr(,"conf.level")
## [1] 0.95
```

From our 95% confidence interval on the difference in mean score **we cannot find any evidence that the mean score depends on birth order** as it contains 0.

c. Test an appropriate set of hypotheses indicating that the mean score does not depend on birth order. H0: The difference in means is equal to 0 H1: The difference in means is not equal to 0

```
t$p.value
```

```
## [1] 0.7229845
```

As 0.7229845 is larger than α = 0.05, we cannot reject the null. This indicates that the mean score does not depend on birth order.

2.47

Construct a data set for which the paired t-test statistic is very large, but for which the usual two-sample or pooled t-test statistic is small. In general, describe how you created the data. Does this give you any insight regarding how the paired t-test works?

```
group1 <- c(99,103,99,103)+1
group2 <- c(99,103,99,102)
t.test(group1, group2, paired=T)$statistic
```

```
## t
## 5
```

```
t.test(group1, group2, paired=F)$statistic
```

```
##         t
## 0.8075729
```

From above you see the data sets for which I found the paired t-test statistic to be large, and for which the usual two-sample t-test statistic small. This is accomplished by using the difference in variances to our advantage.

```
var(group1)
```

```
## [1] 5.333333
```

```
var(group2)
```

```
## [1] 4.25
```

```
var(group1-group2)
```

```
## [1] 0.25
```

From above we see the variances calculated for our two-samples are larger (5.33, 4.25) than that of our paired values (0.25), so that when the pooled variance is calculated it stays noticeably larger than our paired values. **The difference in pooled and paired variances is what shrinks the two-sample t-test statistic and enlarges the paired t-test statistic.**

2.48

Consider the experiment described in Problem 2.28. If the mean burning times of the two flares differ by as much as 2 minutes, find the power of the test. What sample size would be required to detect an actual difference in mean burning time of 1 minute with a power of at least 0.90?

In problem 2.28 there are two groups of observations, each of size n=10.

```
power.t.test(10, 2,sd=sd_pooled(c(65,81,57,66,82,82,67,59,75,70),c(64,71,83,59,65,56,69,74,82,79)))
```

```
##
##      Two-sample t test power calculation
##
##               n = 10
##           delta = 2
##              sd = 9.315459
##       sig.level = 0.05
##           power = 0.0661748
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

The power of the test if the mean burning times of the two flares differ by as much as 2 minutes is **0.0661748**. This is using the pooled sd from the observations of two types of chemical flares.

```
power.t.test(delta=1, power=0.9, sd=sd_pooled(c(65,81,57,66,82,82,67,59,75,70),c(64,71,83,59,65,56,69,74,82,79)))
```

```
##
##      Two-sample t test power calculation
##
##               n = 1824.583
##           delta = 1
##              sd = 9.315459
##       sig.level = 0.05
##           power = 0.9
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

The sample size required to detect an actual difference in the mean burning time of 1 minute with a power of at least 0.90 is **n=1825** in each group.