STAT5002 Lab9 Solution Sheet

Introduction to Statistics

STAT5002

One-sample T-test

In his original paper W.S. Gossett ("Student") demonstrated his new testing technique on various examples. One was where 10 patients tried two different drugs designed to increase sleep time. An image of the table is linked to below.

Additional hours' sleep gained by the use of hyoscyamine hydrobromide.

Patient	1 (Dextro-)	2 (Laevo-)	Difference (2–1)
1.	+ ·7	+ 1·9	+ 1·2
2.	− 1·6	+ ·8	+ 2·4
3.	− ·2	+ 1·1	+ 1·3
4.	− 1·2	+ ·1	+ 1·3
5.	− 1	− ·1	0
6.	+ 3·4	+ 4·4	+ 1·0
7.	+ 3·7	+ 5·5	+ 1·8
8.	+ ·8	+ 1·6	+ ·8
9.	0	+ 4·6	+ 4·6
10.	+ 2·0	+ 3·4	+ 1·4
	Mean + ·75	Mean + 2·33	Mean + 1·58
	S. D. 1·70	S. D. 1·90	S. D. 1·17

from "The Probable Error of a Mean", Student(1908), Biometrika

Note that there is a typographical error: the -1 for patient 5 under "Dextro-" should be -0.1. Also, the SDs presented under the table are actually computed using the population SD, not the sample SD!

Below is the corrected version of the first column of the data (labelled "Dextro-") and the the second column labelled "Laevo-" in R. We will use these two variables in this workshop.

```
dextro = c(.7, -1.6, -.2, -1.2, -.1, 3.4, 3.7, .8, 0, 2)
laevo = c(1.9, .8, 1.1, .1, -.1, 4.4, 5.5, 1.6, 4.6, 3.4)
```

1 The `dextro` data

1.1 Determine the sample mean and SD (note the original paper uses the population SD)

```
# type your code here
x.bar=mean(dextro)
sig.hat=sd(dextro)
cbind(x.bar, sig.hat)
```

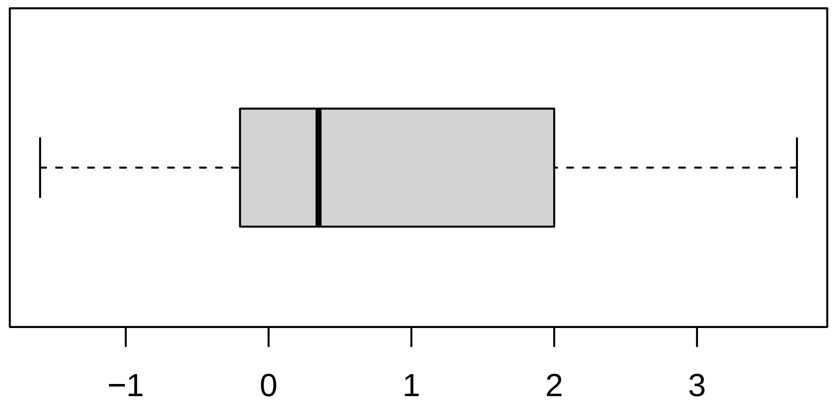
```
     x.bar sig.hat
[1,]  0.75 1.78901
```

```
sqrt(mean(dextro^2)-(mean(dextro)^2)) # the population SD
```

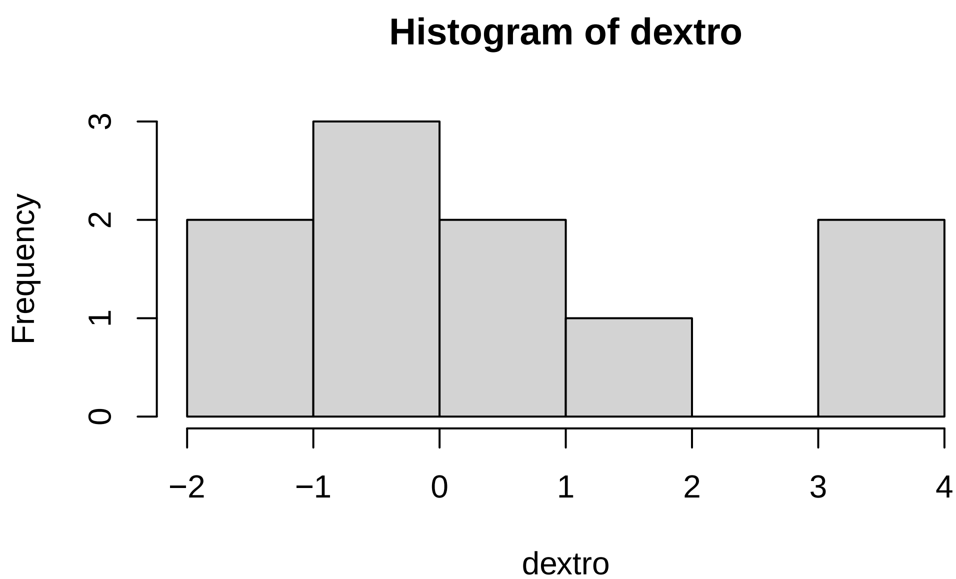
```
[1] 1.697204
```

1.2 Is it reasonable to assume that this is a sample from an "approximately normal box"? Use some plots to check this.

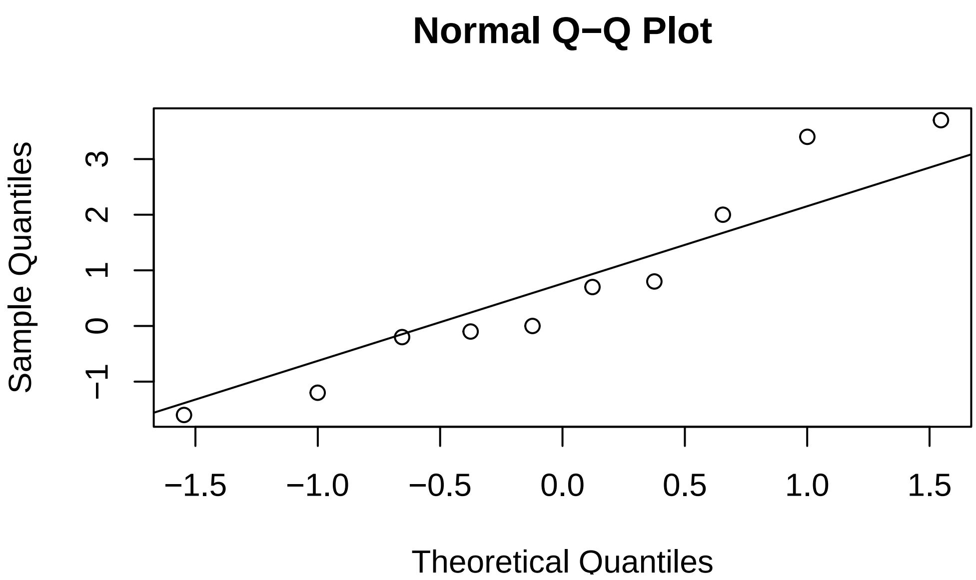
```
# type your code here
boxplot(dextro, horizontal=T)
```




```r
hist(dextro)
```




```r
qqnorm(dextro)
qqline(dextro)
```



Solution: The sample size is rather small (10 observations), so it is rather hard to summarise information from the histogram. However, the boxplot is reasonably symmetric, without outliers. The QQ plot may also suffer from a small sample size, but the quantile points are not far from the QQ line, suggesting the normal assumption may be okay here.

1.3 Provide a 95% confidence interval for the average increase in hours' sleep using Dextro-.

```
# type your code here
n = length(dextro)
q = qt(.975, df=n-1)
x.bar + c(-1,1)*q*sig.hat/sqrt(n)
```

```
[1] -0.5297804  2.0297804
```

1.4 Would we reject a two-sided test that Dextro makes no difference to the duration of sleep at the 5% level of significance? Explain.

Solution: No we would not. The 95% confidence interval includes zero, so the data is consistent with the true increase in sleep being zero. And for any such value, the corresponding two-sided test at the 5% level would not reject.

2 The `laevo` data

2.1 Verify the calculations of the sample mean and SD

```
# type your code here
x.bar = mean(laevo)
sig.hat = sd(laevo)
cbind(x.bar, sig.hat)
```

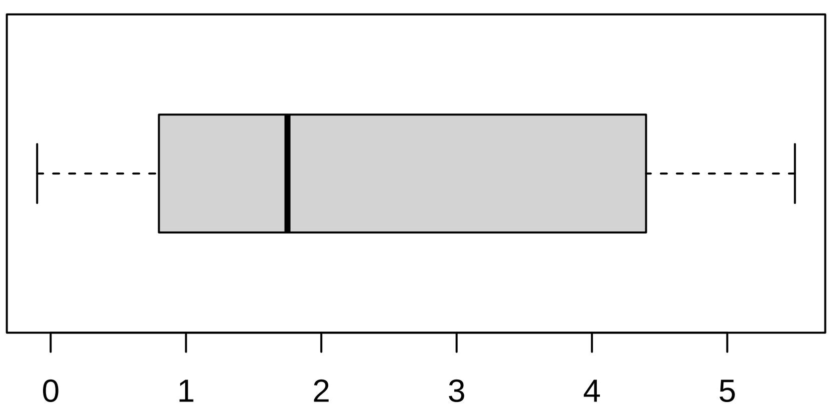
```
     x.bar  sig.hat
[1,]  2.33 2.002249
```

```
sqrt(mean(laevo^2)-(mean(laevo)^2))
```

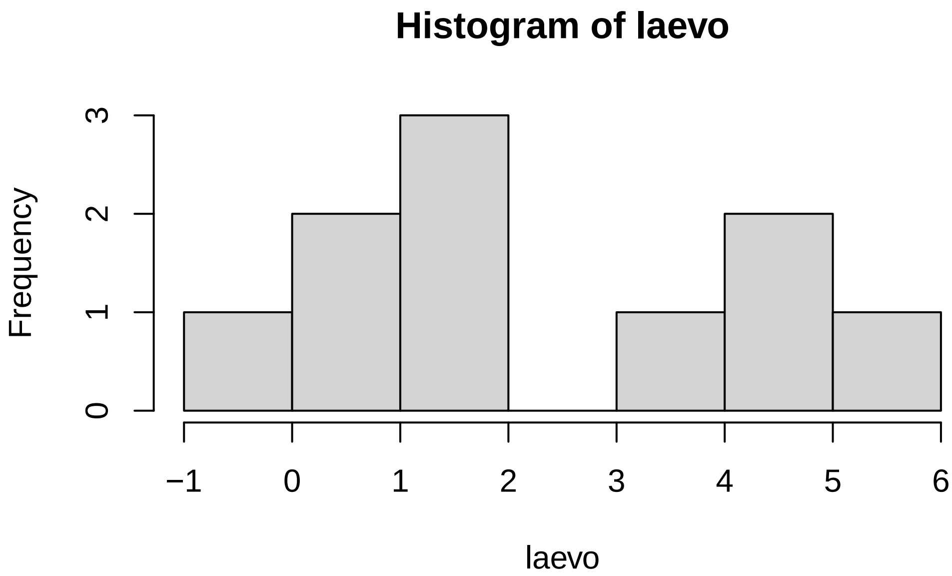
```
[1] 1.8995
```

2.2 Is it reasonable to assume that this is a sample from an "approximately normal box"? Use some plots to check this.

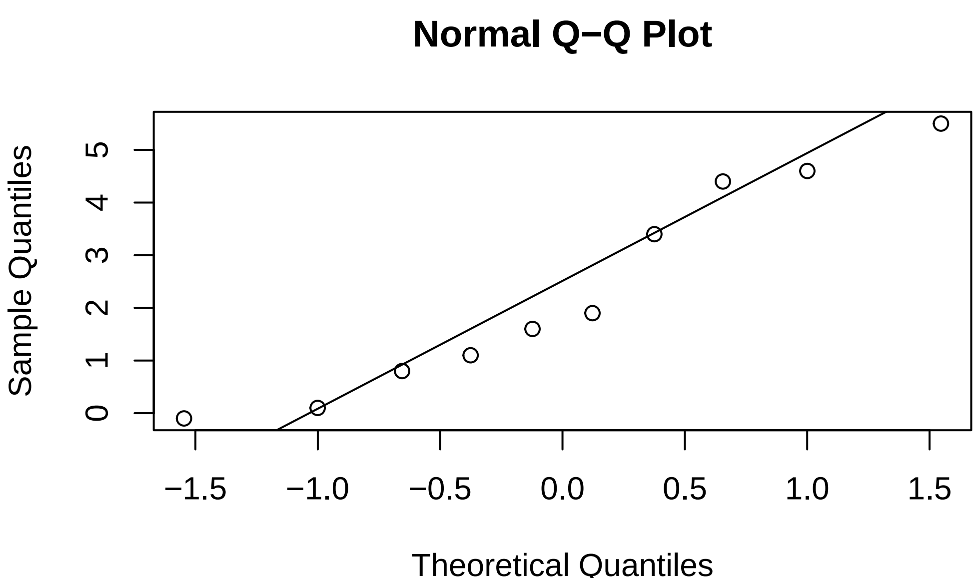
```
# type your code here
boxplot(laevo, horizontal=T)
```




```
hist(laevo)
```




```r
qqnorm(laevo)
qqline(laevo)
```



Solution: The sample size is rather small (10 observations), so it is rather hard to summarise information from the histogram, but there is a suggestion of two "bumps" or "clusters"" in the data. The boxplot is reasonably symmetric, without outliers. The QQ plot may also suffer from a small sample size, but the quantile points are not far from the QQ line, suggesting the normal assumption may be okay here.

Note that, in this case, it is also okay to argue that the normal assumption does not hold (based on the two modes in histogram), and we have to rely on simulation-based estimates of quantiles.

2.3 Provide a 95% confidence interval for the average increase in hours' sleep using Laevo-.

```
# type your code here
x.bar + c(-1,1)*q*sig.hat/sqrt(n)
```

```
[1] 0.8976775 3.7623225
```

2.4 Would we reject a two-side test that Laevo makes no difference to the duration of sleep at the 5% level of significance? Explain.

Solution: The interval does *not* include zero and so the data is not consistent with the true mean increase in sleep being zero. We would thus reject a two-sided test that the true mean is zero at the 5% level of significance.

3 Are the two treatments being equally effective?

Are your results consistent with the two treatments being equally effective? Discuss with reference to **overlap between confidence intervals**.

Solution: Individually, Dextro- is consistent with its true mean being zero, while for Laevo- this is not the case. However, There exist a range of values for the true mean that both samples are consistent with, since the two confidence intervals overlap. So in this sense, it cannot be ruled out that the two true means are the same.

4 Test the difference

Repeat the steps of question 1 again but this time using the third column of "Laevo – Dextro" differences for each patient.

- You can obtain a vector of differences by `d = laevo - dextro`

```
# type your code here
d = laevo - dextro
d
```

```
 [1] 1.2 2.4 1.3 1.3 0.0 1.0 1.8 0.8 4.6 1.4
```

4.1 Verify the calculations of the sample mean and SD

```r
# type your code here
mean(d)
```

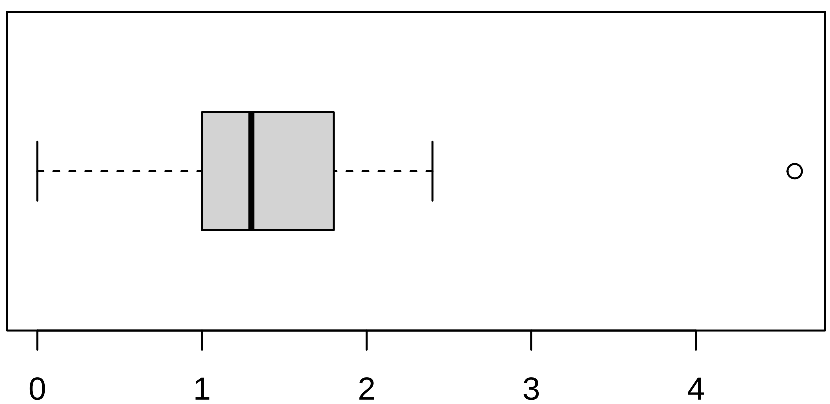
```
[1] 1.58
```

```r
sd(d)
```

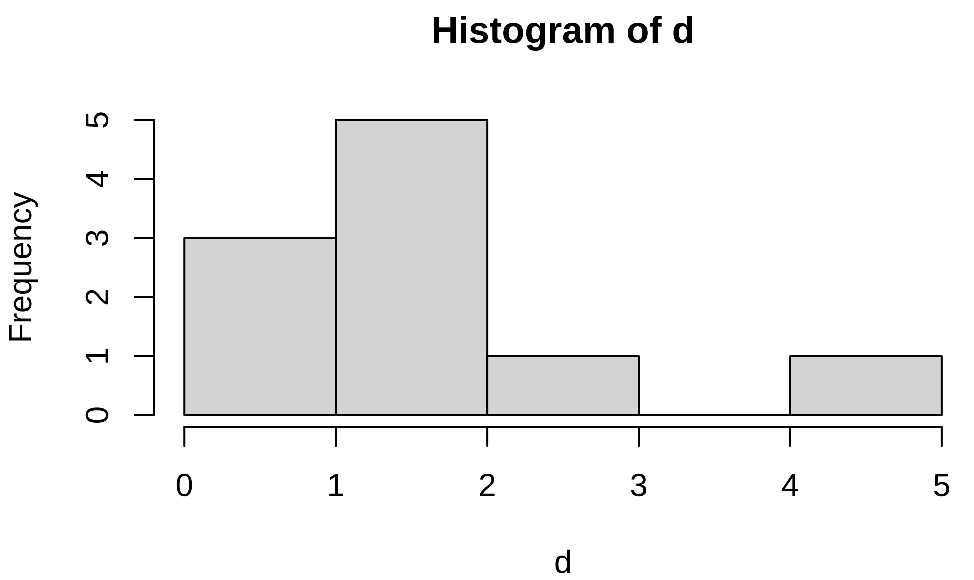
```
[1] 1.229995
```

4.2 Is it reasonable to assume that these differences are like a sample from an "approximately normal box"? Use some plots to check this.

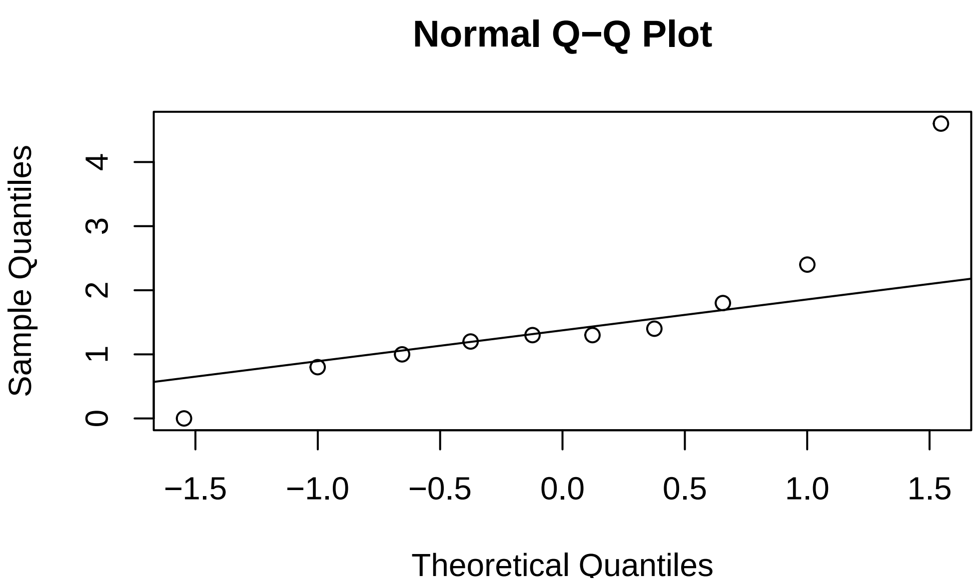
```r
# type your code here
boxplot(d, horizontal=T)
```




```r
hist(d)
```




```r
qqnorm(d)
qqline(d)
```



Solution: The boxplot suggests outliers, whereas the quantile points in the QQ plot depart from the QQ line. The normal assumption may not hold. We may need to compare to a simulation-based bootstrap approach.

4.3 Provide a 95% confidence interval for the average difference in increase in hours' sleep:

```
# type your code here
mean(d) + c(-1,1)*q*sd(d)/sqrt(10)
```

```
[1] 0.7001142 2.4598858
```

4.4 Is this consistent with the two treatments being equally effective? In your answer, refer to your answer to question 3 above.

Solution: This interval does not include 0. Thus this data is not consistent with the true average difference being zero (in this sense), suggesting a difference between the two treatments.

This is the opposite conclusion from question 3. This is because the readings are correlated, reflected in the fact that the SD of the differences is much smaller than the individual SDs of the two drugs considered separately. So this interval is narrower than the individual ones obtained above. As it says in the original paper: "*The low value of the S.D. is probably due to the different drugs reacting similarly on the same patient, so that there is correlation between the results.*"

The "loose" approach taken in question 3 ignores this correlation and so is "less sensitive" at detecting a difference between the treatments, that is the confidence intervals are wider.

5 Perform a formal t-test on the difference.

Perform a formal t-test that the "true" average difference over all (potential) patients is zero. Stating and checking any required assumptions. Will you reject based on a 5% level of significance?

5.1 Write down the testing procedure step-by-step

Let μ_d denote the true average difference over all potential patients.

- **Null hypothesis:** $H_0: \mu_d = 0$.
- **Alternative hypothesis:** $H_0: \mu_d \neq 0$.
- **Assumptions:** The differences $D_1, \dots, D_{10}$ are like a random sample from an approximately normal box.

  - **Comment**: the assumption may not hold – we may need to check our results via simulation

- **Test statistic:** T-statistic following a Student's t distribution with 9 degrees of freedom:

$$T = \frac{\bar{D}}{\hat{\sigma}_d/\sqrt{n}} \sim t_9$$

  if H_0 true. Both small and large values of test statistic argue against H_0.
- **P-value:** If T takes value `t`, P-value would be given by `2*pt(abs(t), df=9, lower.tail=F)`

```
# type your code here
t = sqrt(10)*(mean(d))/sd(d)
t
```

```
[1] 4.062128
```

```
2*pt(t, df=9, lower.tail=F)
```

```
[1] 0.00283289
```

- **Conclusion**: this is a very small P-value, providing strong evidence against the hypothesis of no true difference. Based on a 5% level of significance, we reject H_0.

5.2 Check your answers using `t.test()`.

```
# type your code here
t.test(d)
```

```


    One Sample t-test

data:  d
t = 4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.7001142 2.4598858
sample estimates:
mean of x
     1.58
```

5.3 Using simulation

Now use a bootstrap simulation to construct a 95% confidence interval and perform a test. How do your results here differ from those given by `t.test()`, and why?

Solution

```
# write your code here
sim.T.stats=0
n = length(d)
box.guess = d
for(i in 1:10000) {
    samp = sample(d, size=n, replace=T)
    sim.T.stats[i] = sqrt(n)*(mean(samp)-mean(d))/sd(samp)
    # note we subtract the 'true mean' of box.guess
}
```

```
hist(sim.T.stats, pr=T, n=50)
curve(dt(x, df=9), add=T, lty=2)
```

The outlier in the data means that the simulated T-statistics are not really distributed like a t_9. The heavy lower tail corresponds to simulations where the sample does not include the large positive outlier, thus giving a lower mean, smaller sample SD and thus large, negative T-statistic.

A two-sided P-value uses the proportion of simulated T-statistics that exceed the observed value from the data:

```r
mean(abs(sim.T.stats)>=abs(t))
```

```
[1] 0.0223
```

This is still quite small, still suggesting a difference between the two treatments. But it is not as small as the P-value given by `t.test()`.

To construct a confidence interval, we use the lower and upper 2.5% points of the simulated statistics.

```r
u.l = quantile(sim.T.stats, probs=c(.975, .025))
u.l
```

```
    97.5%      2.5% 
 1.629926 -3.964278 
```

```r
mean(d) - u.l*sd(d)/sqrt(n)
```

```
    97.5%      2.5% 
0.9460261 3.1219405 
```

Note that here the interval is wider than that provided by `t.test()`, and it is not symmetric about the "point estimate" 1.58.