Prepared By: Arpan Kumar, Ph. D. in Statistics, Second Year
Role: Teaching Assistant North Carolina State University

Exercise (7.35).

Solution. a. $\bar{x}_1, \bar{x}_2$ are not parameters. They are sample estimates, hence can't be used to form a hypothesis.

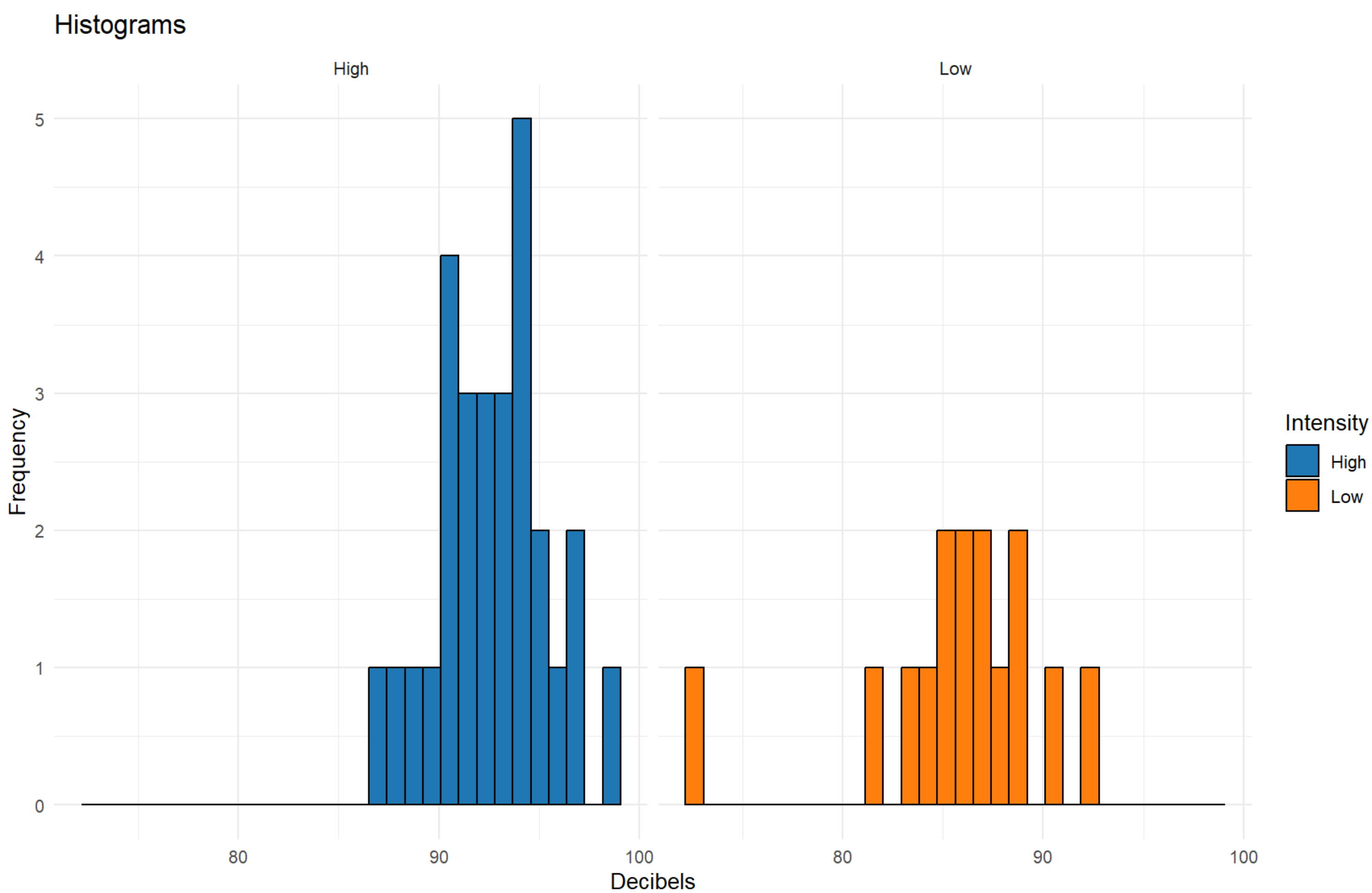
b. The two samples here are not independent.

c. At 90% confidence, to reject H_0, we must have the p-value to be lesser than 0.10, which is clearly not the case here.

d. The researcher claims that the one-sided p-value is 0.018, which is obtained by simply halving the two-sided p-value 0.036. This assumption holds only if the test statistic supports the specified alternative hypothesis.

Exercise (7.39).

Solution. a. The histograms can be found below:



So both the distributions can be assumed to be normal. Just note the fact that there is an outlier for low intensity fitness classes.

b. μ_1 : Mean decibels for high intensity classes
μ_2 : Mean decibels for low intensity classes

The hypotheses are:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Now we have,

$$n_1 = 28, n_2 = 15, \bar{x}_1 = 92.7, \bar{x}_2 = 85.9, s_1 = 2.78, s_2 = 4.50$$

The test statistic is given by:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{92.7 - 85.9}{\sqrt{\frac{2.78^2}{28} + \frac{4.5^2}{15}}} = \frac{6.8}{\sqrt{0.276 + 1.35}} = 5.33$$

and the degrees of freedom is $\min\{28-1, 15-1\} = 14$. Since, $|t| = 5.33 > t_{13,0.025} = 2.145$, we have p-value < 0.05. Hence, we reject the H_0 at 5% significance level.

c. Now, as the low intensity classes appear to have an outlier, t-test is not appropriate in this case.

d. Now we have,

$$n_1 = 28, n_2 = 14, \bar{x}_1 = 92.7, \bar{x}_2 = 86.8, s_1 = 2.78, s_2 = 2.85$$

The test statistic is given by:

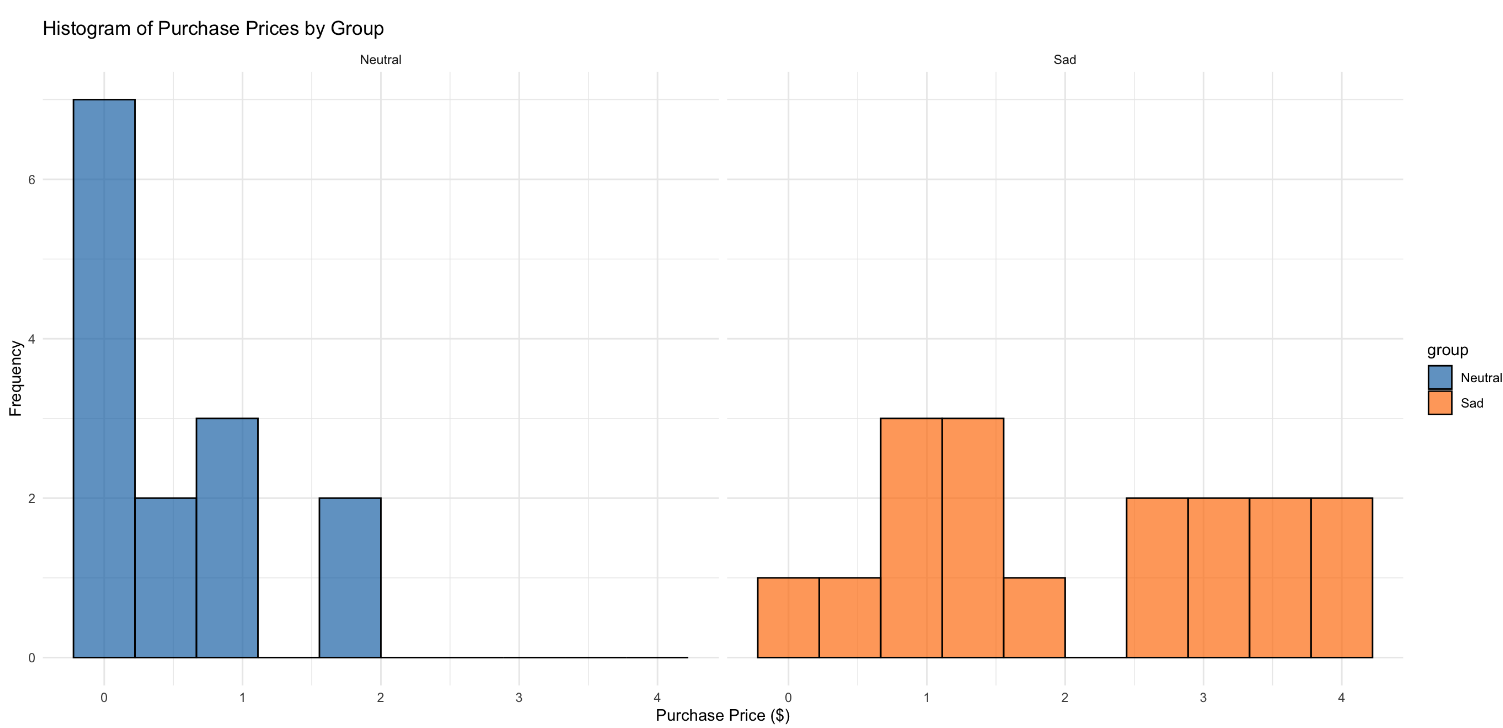
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{92.7 - 86.8}{\sqrt{\frac{2.78^2}{28} + \frac{2.85^2}{14}}} = \frac{5.9}{\sqrt{0.276 + 0.58}} = 6.892$$

and the degrees of freedom is $\min\{28-1, 14-1\} = 13$. Since, $|t| = 6.89 > t_{13,0.025} = 2.1604$, we have p-value < 0.05. Hence, we reject the H_0 at 5% significance level. Thus, we can see that there is no change in result even after removing the outlier.

e. As the outlier is not affecting the result at all, we can report the result.

Exercise (7.47)**.**

Solution. a. The histograms can be found below:



The histograms say that the data are not normally distributed. But, these distributions are not heavily skewed and also don't have outliers. So, we can still use the t-test.

b. Following is the table:

Group	Sample Size	mean	standard deviation
Neutral	14	0.571	0.73
Sad	17	2.117	1.244

c. μ_1 : Mean purchase price for neutral group
μ_2 : Mean purchase price for sad group

The hypotheses are:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

d. The test statistic can be calculated as

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{0.571 - 2.117}{\sqrt{\frac{0.73^2}{14} + \frac{1.244^2}{17}}} = \frac{-1.546}{\sqrt{0.038 + 0.091}} = -4.306$$

and the degrees of freedom is $\min\{14-1, 17-1\} = 13$. Since, $|t| = 4.305 > t_{13,0.025} = 2.1604$, we have p-value < 0.05. Hence, we reject the H_0 at 5% significance level.

e. To calculate the 95% confidence interval we have,

$$\text{estimate} = \bar{x}_1 - \bar{x}_2 = -1.546$$
$$\text{margin of error} = t_{13,0.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.1604 \times 0.359 = 0.776$$

Thus, the confidence interval is given by,

$$\bar{x}_1 - \bar{x}_2 \pm t_{13,0.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = -1.546 \pm 0.776 = (-2.322, -0.77)$$

Exercise (7.48).

Solution. a. μ_1 : Mean TDMS in LC group at 52 week
μ_2 : Mean TDMS in LF group at 52 week

The hypotheses are:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

The test statistic is given by:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{47.3 - 19.3}{\sqrt{\frac{28.3^2}{32} + \frac{25.8^2}{33}}} = 4.1648$$

Since, $|t| > |t_{62,0.025}|$ i.e. $4.1648 > 1.998$, we reject H_0 at 5% level of significance. Hence, we conclude that there is a difference in the TDMS at week 52 between the two dieters group.

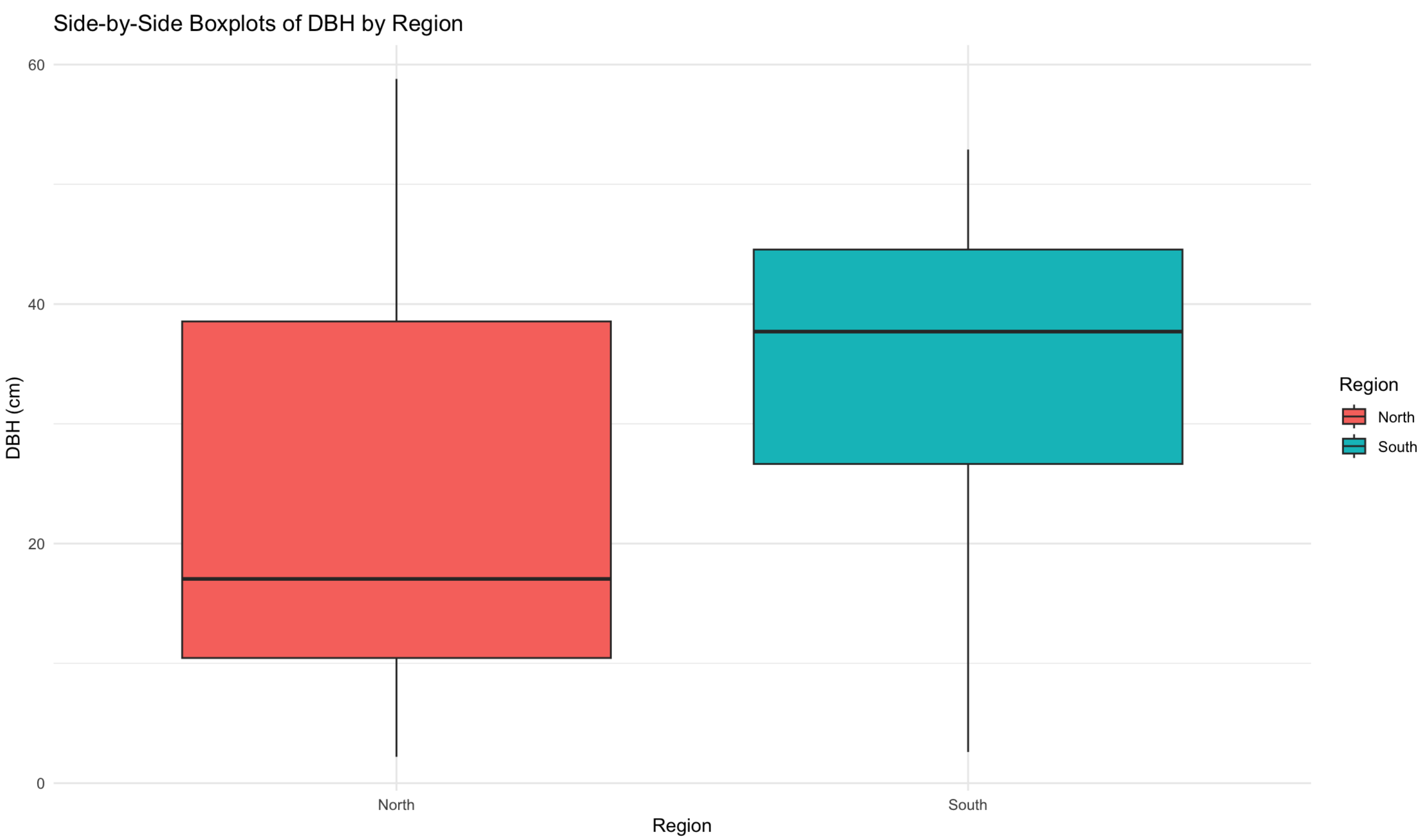
b. Drop out rate is important as sample size is important when it comes to testing of hypothesis. With decreasing sample size, margin of error increases.

Exercise (7.57)**.**

Solution. a. The back-to-back stemplot is given below:

north		south
6554322	0	367
955444410	1	39
8865	2	98632
971	3	267778
7440	4	0004445678
964	5	0123

Also, the side-by-side boxplot is given below:



From these plots we can see that the north distribution is right-skewed, while the south distribution is left-skewed.

b. In this case the two-sample t procedures are still valid as the sample sizes can be considered relatively large and the graphs indicate that there are no outlier in the data.

c. μ_1 : Mean DBH in the north half of the tract
μ_2 : Mean DBH in the north half of the tract

The hypotheses are:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

d. Now we can calculate

$$n_1 = n_2 = 30, \bar{x}_1 = 23.7, \bar{x}_2 = 34.53, s_1 = 17.50, s_2 = 14.26$$

The test statistic can be calculated as

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{23.7 - 34.53}{\sqrt{\frac{17.5^2}{30} + \frac{14.26^2}{30}}} = \frac{-10.83}{\sqrt{10.21 + 6.78}} = -2.629$$

Since, $|t| = 2.629 > t_{29,0.025} = 2.045$, we have p-value < 0.05. Hence, we reject the H_0 at 5% significance level.

e. To calculate the 95% confidence interval we have,

$$\text{estimate} = \bar{x}_1 - \bar{x}_2 = -10.83$$
$$\text{margin of error} = t_{29,0.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.045 \times 4.12 = 8.425$$

Thus, the confidence interval is given by,

$$\bar{x}_1 - \bar{x}_2 \pm t_{29,0.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = -10.83 \pm 8.425 = (-19.255, -2.405)$$

Exercise (7.60)**.**

Solution. a. The hypothesis test framework assumes that hypotheses are pre-specified. Choosing the direction after looking at the data inflates the Type I error rate (false positives) and undermines the validity of statistical inference.

b. As he need to do a two-sided hypothesis test, he should report p-value of $0.046 \times 2 = 0.092$.

!! Thank You !!