Supplementary exercises 7.91 and 7.93 of IPS7e

Data: Voice onset time (VOT) for 6-year-old children and adults when pronouncing the work "bees". The data include 10 children and 20 adults. Let us denote the VOTs for children and adults by $X_1, \ldots, X_{10}$ and $Y_1, \ldots, X_{20}$, respectively.

Model: The 2 samples are independent and each a simple random sample (i.i.d. sample) from a distribution with unknown mean and standard devation, where μ_1 and σ_1 correspond to the distribution for children and μ_2 and σ_2 to the distribution for adults.

Estimation: The estimates are given in the question (whereas the full data are not available), and are repeated here only for the sake of convenience, with added standard errors for the sample mean in each group (computed as $s_i/\sqrt{n_i}$, $i = 1, 2$).

Statistic	Children	Adults
sample size	$n_1 = 10$	$n_2 = 20$
sample mean	$\bar{X} = -3.67$	$\bar{Y} = -23.17$
sample standard deviation	$s_1 = 33.89$	$s_2 = 50.74$
standard error of mean	$\mathrm{SE}_1 = 10.72$	$\mathrm{SE}_2 = 11.35$

Exercise 7.91

(a) The SEs for children and adults are given above. All calculations, including also the standard error for the difference between VOT means for children and adults, will in this exercise be computed without any assumptions on the variances (i.e., without assuming equal variances); for discussion hereof see Exercise 7.93. By the formula for two-sample inference without assuming equal variances (slide 6L−15) we get

$$\mathrm{SE}(\bar{X} - \bar{Y}) = \sqrt{s_1^2/n_1 + s_2^2/n_2} = \sqrt{10.72^2 + 11.35^2} = \sqrt{243.58} = 15.612.$$

(b) The null hypothesis is $\mathrm{H}_0 : \mu_1 = \mu_2$. There is nothing in the description of the context of the data to suggest a one-sided alternative, so we use $\mathrm{H}_a : \mu_1 \neq \mu_2$. The calculation of the test statistic can be done by hand, as shown below, but the degrees of freedom need to be obtained from software. (*Note:* The lower bound obtained as the smallest of the degrees of freedom in the two samples, i.e. $df = \min(n_1 - 1, n_2 - 1) = 9$, is valid but not recommended.) The subsequent Minitab listing gives $df = 25$, so we use this value.

$$\begin{aligned} t &= (\bar{X} - \bar{Y}) / \mathrm{SE}(\bar{X} - \bar{Y}) = [-3.67 - (-23.17)]/15.612 = 1.25, \\ P &= 2 \cdot \mathrm{P}(t(25) > 1.25) > 2 \cdot \mathrm{P}(t(25) > 1.316) = 2 \cdot 0.10 = 0.20. \end{aligned}$$

The exact P-value from the $t(25)$ distribution is obtained from software as: $P = 0.223$, and from this or the assessment above we conclude the test statistic to be clearly non-significant. There is no (convincing) evidence to indicate that the VOT means for children and adults differ.

(c) The Minitab listing below also includes a 95% confidence interval for the mean difference, but we easily compute it by hand, using $t^* = t_{.975}(25) = 2.060$:

$$\begin{aligned} 95\% \text{ CI for } \mu_1 - \mu_2 : \quad \bar{X} - \bar{Y} \pm t^* \cdot \mathrm{SE}(\bar{X} - \bar{Y}) &= (-3.67 - (-23.17)) \pm 2.060 \cdot 15.612 \\ &= 19.50 \pm 32.16 = (-12.66, 51.66). \end{aligned}$$

As stated in the question, we would know from the non-significant P-value from (b) that the CI contains 0. This is because a test (at a 5% significance level) based on a 95% confidence interval would be significant exactly if 0 was not included in the interval, and we know from (b) the test is non-significant.

Exercise 7.93
Because the pooled variance t-test (based on assuming equal variances) is not part of the VHM 801 course syllabus, we confine ourselves to giving interpretations from the Minitab listing below. It might be possible in this particular situation to justify its use, although with some difference between the sample deviations and no obvious reason why VOT recordings for children and adults should have the same variability, it seems most natural to not make such an assumption.

The listing gives the pooled standard deviation as $s = 46.0$, between the two sample standard deviations and clearly closest to the standard deviation for the largest sample (the largest sample has highest weight). From this value we can (re)compute the SE for the mean differenceas:

$$\mathrm{SE}(\bar{X} - \bar{Y}) = s\sqrt{1/n_1 + 1/n_2} = 46.002\sqrt{1/10 + 1/20} = 17.82.$$

This value differs a bit from the 15.61 we computed in Exercise 7.91. This is because the largest sample standard deviation came from the largest sample, and therefore it has less impact without pooling the standard deviation (it would be the opposite effect if the the largest s came from the smallest sample).

Because of the larger SE, the test statistic drops down to 1.09, but this does not change the conclusion substantially. The pooled df equals 28 (computed as: $df = 10 + 20 - 2$) and is thus a bit larger than the approximate df used in Exercise 7.91, but the biggest difference between the two procedures is the estimated standard error. It seems intuitively clear that the best estimate for the SE comes from the unpooled procedure.

Minitab listings (after all the statistics were entered in the `2-Sample t` menu):

Two-Sample T-Test and CI

Method

μ_1: mean of Sample 1
μ_2: mean of Sample 2
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	10	-3.7	33.9	11
Sample 2	20	-23.2	50.7	11

Estimation for Difference

Difference	95% CI for Difference
19.5	(-12.6, 51.6)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.25	25	0.223

Two-Sample T-Test and CI

Method

μ_1: mean of Sample 1
μ_2: mean of Sample 2
Difference: $\mu_1 - \mu_2$

Equal variances are assumed for this analysis.

Descriptive Statistics

Sample	N	Mean	StDev	SE Mean
Sample 1	10	-3.7	33.9	11
Sample 2	20	-23.2	50.7	11

Estimation for Difference

Difference	Pooled StDev	95% CI for Difference
19.5	46.0	(-17.0, 56.0)

Test

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis H_1: $\mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.09	28	0.283