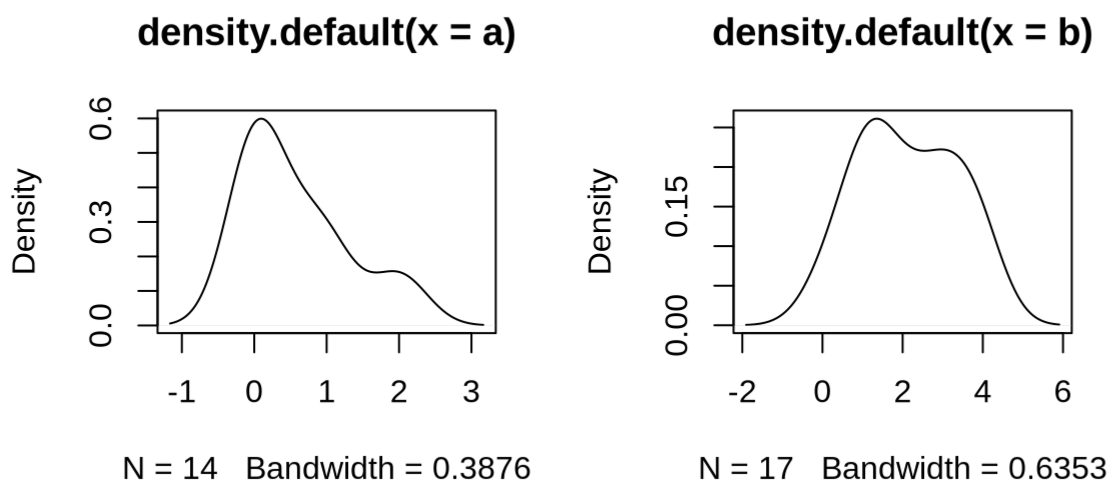
Homework 5

Kevin Ha

October 16, 2021

Problem 7.71

Problem 7.71a



In examining the density of the values, there isn't a clear normal distribution shown.

However since there is no presence of outliers or indication of strong skewness and $15 < n_1 + n_2 < 40$, where $n_1 =$ sample size of neutral group and $n_2 =$ sample size of sad group, the t procedures can still be used.

Problem 7.71b

```
##         Sample Size      Mean  Std. Dev.
## Neutral  14.0000000 0.5714286 0.7300459
## Sad      17.0000000 2.1176471 1.2441037
```

Problem 7.71c

$H_0 : \mu_{Neutral} = \mu_{Sad}$

$H_a : \mu_{Neutral} < \mu_{Sad}$

Problem 7.71d

T-statistic formula:

$$t = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_A^2}{n_B}}} = \frac{0.57 - 2.12}{\sqrt{\frac{0.73^2}{14} + \frac{1.24^2}{17}}} = -4.3030975$$

P-value is calculated using the t-value found above and $df = 13$ since $n_A - 1 < n_B - 1$

```
pt(-4.3030975, 13)
```

```
## [1] 0.0004291661
```

The p-value is extremely low enough to indicate the conclusion that the null hypothesis, H_0 should be rejected and a person's average spending is higher when they are sad.

Problem 7.71e

Confidence interval formula:

$$(\bar{x_1} - \bar{x_2}) \pm t^* \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

First calculate critical T-value, using $df = 13$

```
t_value <- qt(0.025, 13)
t_value
```

```
## [1] -2.160369
```

$$\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = 0.3593269$$

The interval when $df = 13$:

-2.3224971 to -0.7699399

Problem 7.89

Problem 7.89a

$H_0 : \mu_{bf} = \mu_{formula}$

$H_a : \mu_{bf} > \mu_{formula}$

$$t = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_A^2}{n_B}}} = \frac{13.3 - 12.4}{\sqrt{\frac{1.7^2}{14} + \frac{1.8^2}{17}}} = 1.6537344$$

P-value is calculated using the t-value found above and $df = 18$ since $n_A - 1 < n_B - 1$

```
1 - pt(1.6537344, 18)
```

```
## [1] 0.05775748
```

The p-value when using $\alpha = 0.05$ indicates that the data does not provide substantial evidence to reject the null hypothesis, H_0

Problem 7.89b

Confidence interval formula:

$$(\bar{x_1} - \bar{x_2}) \pm t^* \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

First calculate critical T-value, using $df = 18$

```
t_value <- qt(0.05, 18)
t_value
```

```
## [1] -1.734064
```

$$\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}} = 0.6300927$$

The 95% confidence interval when $df = 18$:

-0.1926208 to 1.9926208

Problem 7.89c

The assumptions are that we have two independent SRS from two distinct populations that are both normally distributed.

Problem 7.102

Problem 7.102a

$$F = \frac{\sigma_{larger}}{\sigma_{smaller}} = \frac{s_1^2}{s_2^2} = \frac{9.1}{3.5} = 2.6$$

Problem 7.102b

The appropriate value is given by the $F(15, 10)$ distribution at the 0.05 significance level.

$F(15, 10) = 2.49$

Problem 7.102c

The results are significant at the 5% level since the calculated F test statistic is greater than the F-value found in the 0.05 significance level of the $F(15, 10)$ distribution.

The null hypothesis can also be rejected because of this.

Problem 7.122

Problem 7.122a

```
##          Sample Size      Mean  Variance
## Group 1    10.000000 49.692000  5.372640
## Group 2    10.000000 50.545000  3.703161
```

```
result <- t.test(group1, group2, var.equal=FALSE)
result$parameter
```

```
##       df
## 17.41087
```

T-statistic formula:

$$t = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_A^2}{n_B}}} = -0.8953783$$

P-value (two-tailed test) is calculated using the t-value found above and $df = 17$

```
2 * pt(-0.8953783, 17)
```

```
## [1] 0.383088
```

Problem 7.122b

```
## [1] "Mean of the differences: -0.853"
```

```
## [1] "Variance of the differences: 1.611"
```

```
## [1] "Standard deviation of the differences: 1.269"
```

Paired T-statistic formula:

$$t = \frac{\overline{X}_{diff}}{\sqrt{\frac{S_{diff}^2}{n_{total}}}} = -2.1254263$$

P-value (two-tailed test) is calculated using the t-value found above and $df = 10 - 1 = 9$

```
2 * pt(-2.1254263, 9)
```

```
## [1] 0.06248424
```

Problem 7.122c

The differences in the degrees of freedom were the result of using an unpaired and paired test and if $\alpha = 0.05$ both p-values are not statistically significant and the differences are result of using a two sample vs one sample t test.

Problem 8.71

Problem 8.71a

Proportion for females juveniles:

$$\hat{p}_f = \frac{48}{60} = 0.8$$

Standard error for females juveniles:

$$SE\hat{p}_f = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.8(0.2)}{60}} = 0.0516398$$

Proportion for male juveniles:

$$\hat{p}_m = \frac{52}{132} = 0.3939394$$

Standard error for male juveniles:

$$SE\hat{p}_m = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3939(1-0.3939)}{132}} = 0.0425283$$

Problem 8.71b

Difference in population proportions:

$$D = \hat{p}_f - \hat{p}_m = 0.4061$$

Standard error of D:

$$SE_D = \sqrt{\frac{\hat{p}_f(1-\hat{p}_f)}{n_f} + \frac{\hat{p}_m(1-\hat{p}_m)}{n_m}} = 0.0668979$$

Margin of error for 90% confidence interval:

$$MoE = z * \sqrt{SE_D} = 0.110047$$

90% confidence interval:

$$0.4061 \pm 0.110047$$

0.296053 to 0.516147

Summary: With 90% confidence, the difference in proportion of female references to male references lies in between $(0.296053, 0.516147)$

Problem 8.71c

First, find the pooled sample proportion, p:

$$\hat{p} = \frac{48+52}{60+132} = 0.5208333$$

Next, calculate the pooled standard error:

$$SED_p = \sqrt{\hat{p}(1-\hat{p})\frac{1}{n_1} + \frac{1}{n_2}} = 0.0777823$$

Z statistic formula:

$$z = \frac{\hat{p}_f - \hat{p}_m}{SED_p} = 5.2209822$$

$$P \approx 0$$

```
2*(1-pnorm(5.2209822))
```

```
## [1] 1.779766e-07
```

The near 0 p-value provides strong evidence to indicate that the null hypothesis of the two proportions being equal should be rejected.