

B1. Assuming the dataset is already read in correctly and is called “s”, type

```
> attach(s)
> table(Class, Earprc)
      Earprc
Class  No Yes
FA05   66 155
SP04   69 167
SP05   81 139
```

Thus, the proportions are $155/221 = .701$ for the fall semester and $306/456 = .671$ for the spring semesters. The difference is $.030$. For the hypothesis test, we need to combine the fall with the spring to obtain a single p-hat: $461/677 = .681$. Therefore, the standard error for the difference will be $\sqrt{.681*.319/221 + .681*.319/456} = .0382$. This gives a z-statistic of $(.030-0)/.0382 = 0.79$, leading to a two-sided p-value of $2*.2148 = .4296$. We have no evidence here that the spring proportion is different than the fall proportion.

B2. Assuming the dataset is already read in correctly and is called “s”, type

```
> meanandsd = function(x)
  c(mean(x,na.rm=T), sqrt(var(x,na.rm=T)), sum(!is.na(x)))
> smoke=Cigpacks>0
> meanandsd(GPA[smoke])
[1] 2.8413223 0.6192845 121.00000
> meanandsd(GPA[!smoke])
[1] 3.384054 4.031084 517.00000
```

Notice that the standard deviation for the nonsmoking group (4.03) is extremely large. A bit of exploring will reveal that this is due to a very large GPA: Someone entered 94. We can get rid of this obviously outlying observation and recalculate:

```
> GPA[GPA>5]=NA
> meanandsd(GPA[!smoke])
[1] 3.2084419 0.5532795 516.00000
```

Thus, the t statistic for the test of $H_0: \mu_1 - \mu_2 = 0$ may be calculated as

```
> ((2.841 - 3.208) - 0) / sqrt(.6193^2/121 + .5532^2/516)
[1] -5.982905
```

This t statistic is very extreme. On 120 degrees of freedom, we see a p-value much smaller than $2*.002 = .004$. Thus, we conclude that there is strong evidence that nonsmokers in this population have higher GPAs than smokers. (Be careful! Do not assume that this implies that smoking CAUSES lower GPA, since we don’t know that at all.)

13.12 a. $H_0: \mu_d = 0$

$H_a: \mu_d > 0$ (on average, student height > mom’s height)

$\mu_d = \text{mean “student height–mom’s height” difference for population of college student represented by the sample}$

Note: The null hypothesis could also be written as $H_0: \mu_d \leq 0$

b. $t = \frac{\text{Sample statistic} - \text{Null value}}{\text{Null standard error}} = \frac{1.285 - 0}{0.325} = 3.95$

Sample statistic is observed mean difference in heights, $\bar{d} = 1.285$ inches.
 Null value is $\mu_d = 0$

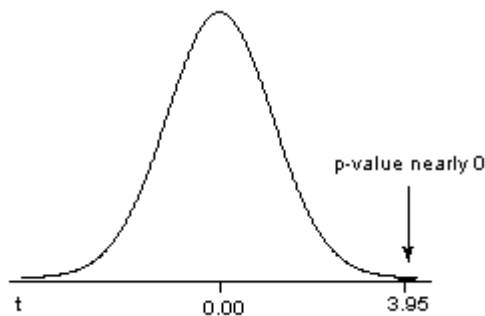
$$\text{Null standard error} = \frac{s_d}{\sqrt{n}} = \frac{3.136}{\sqrt{93}} = 0.325 \text{ (given in output)}$$

c. $df = n - 1 = 93 - 1 = 92$

d. Reported p -value = 0.000. Reject the null hypothesis. For the population of college women represented by the sample, we can conclude that, on average, students' heights are greater than their mother's heights. The observed magnitude of the difference is $\bar{d} = 1.282$ inches.

e. The p -value is the area (probability) to the right of 3.95 in a t -distribution with $df = 92$.

Figure for Exercise 13.12e



13.29 Step 1: $H_0: p_1 - p_2 = 0$, or equivalently, $p_1 = p_2$ (no difference in proportions)

$H_a: p_1 - p_2 > 0$, or equivalently, $p_1 > p_2$ (proportion higher for men with ear pierce)

p_1 = proportion with tattoo in population of college men with an ear pierce

p_2 = proportion with tattoo in population of college men with no ear pierce

Step 2: There are two independent samples and the observed counts in both categories (tattoo or not) are greater than 5 for both groups (ear pierce or not). We must assume the samples represent random samples from the population of all college students.

Test statistic is $z = \frac{\text{Sample statistic} - \text{Null value}}{\text{Null standard error}} = \frac{.1965 - 0}{.03475} = 5.65$. Details are:

Ear pierce, $\hat{p}_1 = \frac{42}{141} = .2979$; no ear pierce, $\hat{p}_2 = \frac{43}{424} = .1014$;

$\hat{p}_1 - \hat{p}_2 = .2979 - .1014 = .1965$

Combined $\hat{p} = \frac{42 + 43}{141 + 424} = \frac{85}{565} = .1504$

Null standard error = $\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} = \sqrt{\frac{.1504(1-.1504)}{141} + \frac{.1504(1-.1504)}{424}} = .03475$

Step 3: p -value ≈ 0.00000001 . It is the area to the right of 5.65 under a standard normal curve. $P(z > 5.65) = P(z \leq -5.65)$. At the bottom of the left page in Table A.1, a cumulative probability is given for $z = -5.61$.

Steps 4 and 5: We can reject the null hypothesis. The conclusion is that in the population(s) represented by the sample(s) the proportion with a tattoo is higher for men with an ear pierce than it is for men with no ear pierce.

- 13.37**
- a.** Both. A confidence interval would be more appropriate to find the magnitude of the mean difference in running times. A hypothesis test could be used to determine if the mean times are significantly different.
 - b.** A confidence interval would be appropriate, to estimate the proportion. A hypothesis test would not be appropriate - there is no obvious null value.
 - c.** Confidence intervals would be appropriate for both parameters. There are no obvious null values to do a hypothesis test.
 - d.** Both. A confidence interval would be appropriate to do find the magnitude of the mean difference in number of visits. A hypothesis test could be used to determine if the mean number of visits for people who fear going to the dentist is lower than for those who do not have the fear.
- 13.54**
- a.** The conclusion was that the results of an IQ test are higher after listening to Mozart than after sitting in silence, but in fact listening to Mozart does not affect IQ test results.
 - b.** The conclusion was that the results of an IQ test are not higher after listening to Mozart than after sitting in silence, but in fact listening to Mozart does increase IQ test results.
 - c.** Listening to Mozart before taking an IQ test has an effect on IQ test results, but this was not found in the study because the sample size was too small.
 - d.** Listening to Mozart before taking an IQ test has a small, but unimportant effect on IQ test results and this small difference was labeled “statistically significant” because the sample size was large.

13.70 Step 1: $H_0: \mu_1 - \mu_2 = 0$, or equivalently $\mu_1 = \mu_2$

$H_a: \mu_1 - \mu_2 > 0$, or equivalently, $\mu_1 > \mu_2$

μ_1 = mean pulse rate in population that does not exercise regularly, and

μ_2 = mean pulse rate in population that does exercise regularly

Step 2: *Necessary conditions*: No outliers are apparent. We assume the sample represents a random sample from a larger population with regard to the effect on resting pulse of exercise (or not).

Step 2 continued and Steps 3,4, and 5: Using the unpooled procedure for the difference between two means, $t = 1.72$ and the associated p -value is slightly larger than 0.057 (see below). We cannot reject the null hypothesis using level of significance $\alpha = 0.05$. Thus, based on the unpooled procedure, we are not able to conclude that the mean pulse rate is lower for those who exercise.

```
> exer = c(62, 72, 60, 63, 75, 64, 60, 52, 64, 80, 68, 64)
> noexer = c(72, 84, 66, 72, 62, 84, 76, 60)
> meanandsd = function(x)
  c(mean(x, na.rm=T), sqrt(var(x, na.rm=T)), sum(!is.na(x)) )
> meanandsd(exer)
```

```

[1] 65.333333  7.487363 12.000000
> meanandsd(noexer)
[1] 72.000000  9.133924  8.000000
> ((72 - 65.333) - 0) / sqrt(7.487^2/12 + 9.134^2/8)
[1] 1.715702

```

The (one-sided) p-value is found by checking 7 degrees of freedom and noting that the p-value of .057 would correspond to a t statistic of 1.80, so our p-value will be slightly larger than .057 because our t statistic of 1.72 is slightly less extreme than 1.80.

13.75 Step 1: $H_0: p \leq .50$ (not a majority)

$H_a: p > .50$ (a majority were dissatisfied)

p = proportion of U.S. adults dissatisfied with K-12 education in August 2000

Step 2: The sample was randomly selected from the population of U.S. adults and the sample size is sufficiently large so that $n\hat{p}$ and $n(1 - \hat{p})$ are both greater than 10.

Sample proportion dissatisfied is $\hat{p} = \frac{622}{1019} = .6104$

Test statistic is

$$z = \frac{\text{Sample statistic} - \text{Null value}}{\text{Null standard error}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.6104 - .50}{\sqrt{\frac{.50(1-.50)}{1019}}} = \frac{.1104}{.01566} = 7.05.$$

Step 3: p -value ≈ 0 . It is the area (probability) to the right of $z = 7.05$. This z -value is beyond the last value given in the “In the Extreme” section of Table A.1, so the area to the right must be nearly 0.

Steps 4 and 5: We can reject the null hypothesis. The conclusion is that in the population of U.S. adults in August 2000, a majority were dissatisfied with the quality of K-12 education.