Homework
Biostatistics (STAT:3510; Bognar)

1. A sociologist collected a random sample of 13 statistics majors and 14 sociology majors. The students were asked about how many hours per week they spend socializing. The results are summarized in the following table. Assume that the amount of socialization for statistics majors X_1 follows a normal distribution with mean μ_1 and standard deviation σ_1 (i.e. $X_1 \sim N(\mu_1, \sigma_1)$), while the amount of socialization for sociology majors X_2 follows a normal distribution with mean μ_2 and standard deviation σ_2 (i.e. $X_2 \sim N(\mu_2, \sigma_2)$). Because the sample standard deviations s_1 and s_2 are quite similar, lets make the reasonable assumption that $\sigma_1 = \sigma_2$.

Statistics:	$n_1 = 13$	$\bar{x}_1 = 32.1$	$s_1 = 10$
Sociology:	$n_2 = 14$	$\bar{x}_2 = 23.0$	$s_2 = 12$

   (a) Find a 95% confidence interval for $\mu_1 - \mu_2$.
   (b) Based upon your answer in (1a), is there a significant difference in the mean time spent socializing between statistics and sociology majors? Why?
   (c) Suppose we wish to test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ at the $\alpha = 0.05$ significance level. Based upon your answer in (1a), will H_0 be rejected? Why?
   (d) Based upon your answer in (1c), will the $p-$value be less than 0.05 or greater than 0.05? Why?
   (e) Test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ at the $\alpha = 0.05$ significance level. *Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.*
   (f) Based upon your answer in (1e), is there a significant difference between the average socialization times? Why?
   (g) Approximate the $p-$value for the test in (1e) using the $t-$table.
   (h) Use the $t-$Probability Applet at

   `http://www.stat.uiowa.edu/~mbognar/applets/t.html`

   to precisely determine the $p-$value for the test in (1e).
   (i) Consider the test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ at the $\alpha = 0.01$ significance level. Based upon your answer in (1g) and (1h), do you reject H_0? Why?
   (j) Could we do the above analysis if the study times were *not* normally distributed? Why?

2. The level of iodine in Company A's table salt follows a $N(\mu_1, \sigma_1)$ distribution, while the level in Company B's salt follows a $N(\mu_2, \sigma_2)$ distribution. A random sample of each companies' product yielded

Company A:	$n_1 = 16$	$\bar{x}_1 = 22.4$	$s_1 = 1.0$
Company B:	$n_2 = 9$	$\bar{x}_2 = 27.5$	$s_2 = 1.4$

   Assume that $\sigma_1 = \sigma_2$.

   (a) Test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 < \mu_2$ at the $\alpha = 0.01$ significance level. *Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.*
   (b) Based upon your answer in (2a), is the mean iodine level for Company A significantly lower than Company B? Why?
   (c) Approximate the $p-$value for the test in (2a) using the $t-$table.
   (d) Based upon your answer in (2c), is the mean iodine level for Company A significantly lower than Company B? Why?

3. The height of Iowa corn stalks X_1 (in cm) have a $X_1 \sim N(\mu_1, \sigma_1)$ distribution, while the height of Nebraska corn stalks X_2 (in cm) have a $X_2 \sim N(\mu_2, \sigma_2)$ distribution. A random sample of corn stalks from Iowa and Nebraska yielded the following summary statistics.

Iowa:	$n_1 = 15$	$\bar{x}_1 = 155$	$s_1 = 16$
Nebraska:	$n_2 = 20$	$\bar{x}_2 = 145$	$s_2 = 10$

   It is known that $\sigma_1 \neq \sigma_2$.

(a) Suppose we wish to test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ at the $\alpha = 0.10$ significance level. Find the $p-$value for this test using the $t-$table.

(b) Based upon your answer in (3a), is there a significant difference in the mean heights? Why?

(c) Use the $t-$Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/t.html`

to precisely determine the $p-$value for the test in (3a).

(d) Find a 90% confidence interval for $\mu_1 - \mu_2$.

(e) Based upon your answer in (3d), is there a significant difference in the mean heights? Why?

(f) Could we do the above analysis if the heights were *not* normally distributed? Why?

4. The mathematics SAT scores of students at *public* universities have a $N(\mu_1, \sigma_1)$ distribution, while the mathematics SAT scores of students at *private* universities have a $N(\mu_2, \sigma_2)$ distribution. A random sample of students from public and private universities found

Public:	$n_1 = 16$	$\bar{x}_1 = 520$	$s_1 = 100$
Private:	$n_2 = 15$	$\bar{x}_2 = 505$	$s_2 = 85$

It is known that $\sigma_1 \neq \sigma_2$.

(a) Test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 > \mu_2$ at the $\alpha = 0.05$ significance level. *Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.*

(b) Find the $p-$value for the test in (4a) using the $t-$table.

(c) Based upon your answer in (4b), do students at public universities have a significantly higher mean SAT score? Why?

(d) Use the $t-$Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/t.html`

to precisely determine the $p-$value for the test in (4b).

5. The time (in seconds) it took four individuals to run 1/4 of a mile was recorded. The same individuals went through an exercise program and were re-timed. The times before the exercise program, x_i^b, and after the exercise program, x_i^a, are shown below.

Subject:	1	2	3	4
Before x_i^b:	125	160	185	152
After x_i^a:	127	135	169	115
Difference $x_i^d = x_i^b - x_i^a$:	-2	25	16	37

Assume the difference in running times, X_d, is normally distributed, i.e. $X_d \sim N(\mu_d, \sigma_d)$ (because $n < 30$ this assumption is critical). Because subjects appear in both groups, then groups are *not* independent.

(a) Find the sample mean difference, $\bar{x}_d$.

(b) Verify that the sample standard deviation of the differences $s_d = 16.432$.

(c) We would like to determine if the exercise program significantly *decreases* the mean running time. Thus, test $H_0 : \mu_d = 0$ vs. $H_a : \mu_d > 0$ at the $\alpha = 0.05$ significance level using a paired $t-$test. *Find the test statistic and critical value, plot the rejection region, and state your decision and final conclusion.*

(d) Approximate the $p-$value for the test in (5c) using the $t-$table.

(e) Based upon your answer in (5d), does the exercise program significantly decrease the mean running time? Why?

(f) Use the $t-$Probability Applet at

`http://www.stat.uiowa.edu/~mbognar/applets/t.html`

to precisely determine the $p-$value for the test in (5d).

(g) Suppose the significance level was $\alpha = 0.10$ (instead of 0.05). Based upon your answer in (5f), does exercise program significantly decrease the mean running time? Why?