**MATHEMATICS 201-510-LW**
Business Statistics
Martin Huard
Fall 2008

# Assignment 4

## SOLUTIONS

This assignment is due on **Monday November 24, 2008** at the beginning of the class.

## Question 1 (7.5 points)

Does drinking coffee help with concentration? A random sample of adults was taken where the time to complete a particularly challenging task was measured (in minutes), both before and after drinking a large cup of coffee. Here are the results.

| Before | 23 | 16 | 16 | 22 | 13 | 18 | 24 | 12 | 16 | 11 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| After | 21 | 14 | 15 | 20 | 13 | 19 | 23 | 11 | 10 | 9 | 13 |
| $d = B - A$ | 2 | 2 | 1 | 2 | 0 | -1 | 1 | 1 | 6 | 2 | 1 |

Construct a 95% interval for the mean difference in the time taken to complete the task before and after drinking coffee. Assume that the times to complete the tasks are normally distributed.

Step 1 Assumptions: The populations are normally distributed

Step 2 a) Test statistic: $t$ with $df = n-1 = 10$

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$

Step 3 Point estimate: $\bar{d} = 1.545$ minutes

Step 4 a) $t_{\left(df, \frac{\alpha}{2}\right)} = t_{(10, 0.025)} = 2.228$

b) $E = t_{\left(df, \frac{\alpha}{2}\right)} \frac{s_d}{\sqrt{n}} = 2.228 \frac{1.7529}{\sqrt{11}} = 1.178$

c)
$$\bar{d} - E < \mu_d < \bar{d} + E$$
$$1.545 - 1.178 < \mu_d < 1.545 - 1.178$$
$$0.368 < \mu_d < 2.723$$

Step 5 The 95% confidence interval for the mean difference in the time taken to complete the task before and after drinking coffee is 0.37 to 2.72 minutes.

## Question 2 (7.5 points)

Do car owners living in Quebec do more mileage in a year on their car than people who live in Montreal? To answer this question, two random groups, one from Quebec and the other from Montreal, were taken, where each car owner gave his mileage for the previous year. Here are the results obtained.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Montreal | 22 311 | 25 722 | 12 036 | 12 880 | 25 519 | 39 577 | 23 145 |
| | 19 452 | 18 337 | 10 637 | 16 137 | 28 186 | 18 092 | 19 452 |
| Quebec | 8 536 | 16 272 | 31 066 | 17 469 | 22 782 | 7 351 | 11 644 |
| | 16 868 | 22 199 | 12 842 | 15 138 | 16 720 | 30 490 | 16 868 |

Construct a 90% confidence interval for the difference in the mean mileage between car owners from Quebec and car owners from Montreal. Assume that annual mileages are normally distributed.

Step 1 Assumptions: The populations are normally distributed.
The samples are independent.
The variances are equal

Step 2 a) Test statistic: $t$ with $df = n_M + n_Q - 2 = 14 + 14 - 2 = 26$

b) $1 - \alpha = 0.90$

Step 3 Point Estimate: $\bar{x}_M - \bar{x}_Q = 20820 - 17589 = 3231$ km

Step 4

$$s_p = \sqrt{\frac{(n_M - 1)s_M^2 + (n_Q - 1)s_Q^2}{n_M + n_Q - 2}} = \sqrt{\frac{13 \cdot 7560^2 + 13 \cdot 7074^2}{26}} = 7321$$

$$t_{\left(df, \frac{\alpha}{2}\right)} = t_{(26, 0.05)} = 1.706$$

$$E = t_{\left(df, \frac{\alpha}{2}\right)} s_p \sqrt{\tfrac{1}{n_M} + \tfrac{1}{N_Q}} = 1.706 \cdot 7321 \sqrt{\tfrac{1}{14} + \tfrac{1}{14}} = 4721$$

$$\bar{x}_M - \bar{x}_Q - E < \mu_M - \mu_Q < \bar{x}_M - \bar{x}_Q + E$$

$$3231 - 4721 < \mu_M - \mu_Q < 3231 - 4721$$

$$-1490 < \mu_M - \mu_Q < 7952$$

Step 5 $\therefore$ The 90% confidence interval for the difference in the mean mileage between car owners from Montreal and car owners from Quebec is -1 490 km to 7 952 km.

**Question 3** (7.5 points)

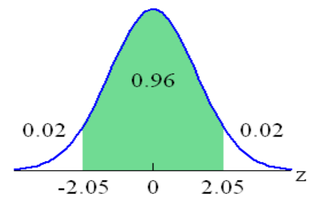
A job recruitment agency is interested in knowing the difference in salary between an engineer and a lawyer. A random sample of 57 engineers was taken, and another of 62 lawyers, where each gave his current yearly salary. It was found that the average salary of an engineer is \$68 230 with a standard deviation of \$14 310 and for a lawyer the average is \$57 510 with a standard deviation of \$29 530. Construct a 96% confidence interval for the difference in the mean annual salary of an engineer and a lawyer.

Step 1 Assumptions: $n_E = 57 \geq 30$, $n_L = 62 \geq 30$

The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.96$ or $\alpha = 0.04$




Step 3 Point estimate: $\bar{x}_E - \bar{x}_L = 68230 - 57510 = 10720$ \$

Step 4 c) $z_{\frac{\alpha}{2}} = z_{0.02} = 2.05$

d) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{s_E^2}{n_E}+\frac{s_L^2}{n_L}} = 2.05\sqrt{\frac{14310^2}{57}+\frac{29530^2}{62}} = 8614.25$

d) $(\bar{x}_E - \bar{x}_L) - E < \mu_E - \mu_L < (\bar{x}_E - \bar{x}_L) + E$

$10720 - 8614 < \mu_E - \mu_L < 10720 + 8614$

$2106 < \mu_E - \mu_L < 19334$

Step 5 The 96% confidence interval for the difference in the mean annual salary of an engineer and a lawyer is \$2 106 to \$19 334.

## Question 4 (7.5 points)

The director of human resources at a large corporation thinks that men and women do not take the same number of sick days during the year. To test this, two random sample, one of men and another of women, were taken, where the number of sick days taken last year was recorded for both. Here are the results obtained.

| Woman | 0 | 5 | 9 | 4 | 5 | 8 | 0 | 1 | 15 | 2 | 3 | 0 | 3 | 2 | 6 | 2 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Man | 1 | 3 | 1 | 4 | 0 | 5 | 2 | 4 | 3 | 2 | 0 | 4 | 3 | 2 | 0 | 1 | 4 | 2 |

Is there sufficient evidence, at the 5% level of significance, to conclude that men and women do not take the same number of sick days during the year? Use the $p$-value approach. Assume the number of sick days a person takes during the year is normally distributed.

Step 1 Assumptions: The populations are normally distributed.
The samples are independent.
The variances are equal

Step 2 $H_O: \mu_W - \mu_M = 0$

$H_A: \mu_W - \mu_M \neq 0$

Step 3 Test statistic: $t$ with $df = n_W + n_M - 2 = 16 + 18 - 2 = 32$

Two-tail test with $\alpha = 0.05$

Step 4

$$s_p = \sqrt{\frac{(n_w - 1)s_w^2 + (n_m - 1)s_m^2}{n_w + n_m - 2}} = \sqrt{\frac{15 \cdot 3.991^2 + 17 \cdot 1.565^2}{32}} = 2.961$$

$$t = \frac{(\bar{x}_W - \bar{x}_M) - (\mu_W - \mu_M)}{s_p \sqrt{\frac{1}{n_W} + \frac{1}{n_M}}} = \frac{4.063 - 1.565}{2.961\sqrt{\frac{1}{16} + \frac{1}{18}}} = 1.754$$

$$2 \cdot 0.040 < p - \text{value} < 2 \cdot 0.050$$

$$0.080 < p - \text{value} < 0.100$$

Step 5 a) $p$-value $> 0.080 > \alpha = 0.05$

b) Fail to reject $H_o$.

$\therefore$ There is insufficient evidence at the 5% level of significance to conclude that women do not take the same number of sick days as men during the year.

**Question 5** (7.5 points)
Laura wants to determine if more women like sushi than men. In a random sample of 320 women, 200 said they liked sushi, and in a random sample of 392 men, 210 said they liked sushi. Construct a 98% confidence interval for the difference in the proportion of men and women who like sushi.

Step 1 Assumptions: $n_W = 320 > 20$ $\quad n_W \hat{p}_W = 200 > 5$ $\quad n_W \hat{q}_W = 120 > 5$

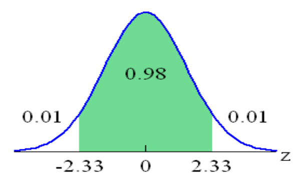
$n_M = 392 > 20$ $\quad n_M \hat{p}_M = 210 > 5$ $\quad n_M \hat{q}_M = 182 > 5$

Samples are independent

Step 2 a) Test statistic: $z$

b) Level of confidence: $1 - \alpha = 0.98$ or $\alpha = 0.02$

Step 3 Point estimate : $\hat{p}_w - \hat{p}_m = \frac{200}{320} - \frac{210}{392} = 0.0893$




Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.01} = 2.33$

b) $E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_w \hat{q}_w}{n_w} + \frac{\hat{p}_m \hat{q}_m}{n_m}} = 2.33 \sqrt{\frac{\frac{200}{320}\frac{120}{320}}{320} + \frac{\frac{210}{392}\frac{182}{392}}{392}} = 0.0861$

c) $(\hat{p}_w - \hat{p}_m) - E < p_w - p_m < (\hat{p}_w - \hat{p}_m) + E$

$0.0893 - 0.0861 < p_w - p_m < 0.0893 + 0.0861$

$0.0031 < p_w - p_m < 0.1754$

Step 5 The 98% confidence interval for the difference in the proportion of men and women who like sushi is 0.3% to 17.5%.

**Question 6** (7.5 points)

Homer thinks that more women watch Oprah then men. To prove his claim, he takes a random sample of 635 women and 956 men, and finds that 123 women and 124 men said the watch Oprah. Is this sufficient evidence, at the 2% level of significance, for Homer to conclude that more women watch Oprah then men? Use the classical approach.

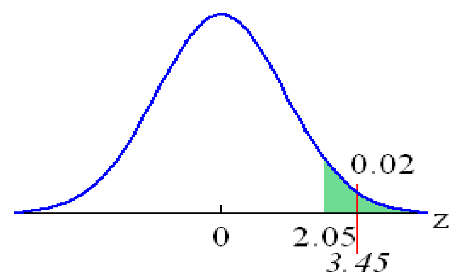
Step 1 Assumptions: $n_W = 635 > 20$ $\quad n_W \hat{p}_W = 123 > 5$ $\quad n_W \hat{q}_W = 512 > 5$

$n_M = 956 > 20$ $\quad n_M \hat{p}_M = 124 > 5$ $\quad n_M \hat{q}_M = 832 > 5$

Samples are independent

Step 2 $H_O: p_W - p_M = 0$

$H_A: p_W - p_M > 0$




Step 3 a) Test statistic: $z$

b) Right-tailed test $\alpha = 0.02$

c) $z_\alpha = z_{0.02} = 2.05$

Step 4 a) $\hat{p}_p = \dfrac{123+124}{635+956} = \dfrac{247}{1591} = 0.15525$

b) $z = \dfrac{\hat{p}_W - \hat{p}_M}{\sqrt{\hat{p}_p \hat{q}_p \left(\frac{1}{n_W} + \frac{1}{n_M}\right)}} = \dfrac{\frac{123}{635} - \frac{124}{956}}{\sqrt{\frac{247}{1591}\frac{1344}{1591}\left(\frac{1}{635} + \frac{1}{956}\right)}} = 3.45$

Step 5 a) $z$ is in the critical region

b) Reject $H_o$.

∴ There is sufficient evidence at the 2% level of significance to conclude that more women watch Oprah then men.

*Questions 7 to 9 are to be done using Excel.*

For these questions, hand-in the printouts of your Excel sheets and copy your Excel work in the Test folder for Business Statistics (W:\Tests\mhuard\Business Statistics\Assignment 4), where your name should be included in the name of the file (for example: Assignment 4 – Your Name). Make sure that your answers are well organized with appropriate labels, and rounded off to an appropriate number of decimal places.

Open the file "Data – Assignment 4" from my web site, and save it under "Assignment 4 – Your Name". Note that you may have to enable macros to be able to generate the data. If the macros are not enabled (that is, if the data does not appear at the click of the button) then go to - EXCEL OPTIONS – TRUST CENTER – TRUST CENTER SETTINGS – MACRO SETTINGS and choose the ENABLE ALL MACROS option. Note that you may need to close your document and open it again.

## Question 7 (5 points)

The manager of a sporting goods store offered a bonus commission to his salespeople when they sold more goods. A new manager dropped the bonus system. To see the effect of this, a random sample of sales people was taken, where the weekly sales (in thousands of dollars) were recorded with and without the bonus.

a) Go to the worksheet "Sheet1", rename it appropriately, make the usual heading in cells A1:A4, then click on the "GENERATE DATA" button to get your data.
b) Test the claim that the average weekly sales dropped when the bonus system was discontinued. Use the classical approach along with a 5% level of significance. Assume the weekly sales are normally distributed.

| Weekly Sales (in $1000) Before Bonus | Weekly Sales (in $1000) After Bonus |
|---|---|
| 4.5 | 6.7 |
| 9.1 | 8.1 |
| 10.2 | 4.6 |
| 10.1 | 10.9 |
| 5.8 | 7 |
| 4.3 | 5.8 |
| 12.8 | 5.6 |
| 11 | -0.9 |
| 7.7 | 5.4 |
| 9.6 | 7.4 |
| 11.6 | 6.4 |
| 10.8 | 9.4 |
| 13.5 | 8.2 |
| 10.5 | 2.2 |
| 13 | 8.1 |

t-Test: Paired Two Sample for Means

| | Weekly Sales (in $1000) Before Bonus | Weekly Sales (in $1000) After Bonus |
|---|---|---|
| Mean | 9.62 | 5.853 |
| Variance | 8.936 | 8.316 |
| Observations | 15 | 15 |
| Pearson Correlation | 0.017879115 | |
| Hypothesized Mean Difference | 0 | |
| df | 14 | |
| t Stat | 3.607 | |
| P(T<=t) one-tail | 0.0014 | |
| t Critical one-tail | 1.761 | |
| P(T<=t) two-tail | 0.0029 | |
| t Critical two-tail | 2.145 | |

| | | | | | |
|---|---|---|---|---|---|
| Step 1 | Assumptions: Population normally distributed | | | | |
| Step 2 | $H_0$: | $\mu_d =$ | 0 | | |
| | $H_a$: | $\mu_d >$ | 0 | | |
| Step 3 | Test statistic: $t$ with $df$ = | | | | 14 |
| | Right-tailed Test | | | $\alpha$ = | 5% |
| | t critical = | 1.761 | | | |
| Step 4 | t = | 3.61 | | | |
| Step 5 | t is in the critical region | | | | |
| | Reject Ho | | | | |
| | There is sufficient evidence, at the 5% level of significance to conclude that the weekly sales dropped when the bonus system was discontinued | | | | |

Question 7 | Question 8 | Question 9

## Question 8 (5 points)

A real estate agent claims that the average value of houses in Sainte-Foy are higher than in Saint-Augustin. To prove his point, he takes random samples of houses in Sainte-Foy and in Saint-Augustin, and notes the value of each.

a) Go to the worksheet "Sheet2", rename it appropriately, make the usual heading in cells A1:A4, then click on the "GENERATE DATA" button to get your data.

b) Test the real estate agent's claim. Use the *p*-value approach with a 5% level of significance. Assume that the value of houses are normally distributed.

| Value (in $) in Ste-Foy | Value (in $) in St. Augustin |
|---|---|
| 257612 | 279055 |
| 286066 | 329460 |
| 379773 | 349049 |
| 324752 | 258978 |
| 354451 | 318024 |
| 332920 | 336218 |
| 400780 | 319649 |
| 292983 | 355881 |
| 385860 | 225203 |
| 386819 | 225971 |
| 344041 | 259752 |
| 410436 | 309900 |
| 369658 | 290326 |
| 333269 | 276962 |
| 312821 | 381241 |
| 365562 | 277463 |
| 386454 | 275369 |

t-Test: Two-Sample Assuming Equal Variances

| | Value (in $) in Ste-Foy | Value (in $) in St. Augustin |
|---|---|---|
| Mean | 350973.56 | 293390.81 |
| Variance | 2028274548 | 2062766829 |
| Observations | 27 | 27 |
| Pooled Variance | 2045520688 | |
| Hypothesized Mean Difference | 0 | |
| df | 52 | |
| t Stat | 4.678 | |
| P(T<=t) one-tail | 0.000011 | |
| t Critical one-tail | 1.675 | |
| P(T<=t) two-tail | 0.000021 | |
| t Critical two-tail | 2.007 | |

| | | | |
|---|---|---|---|
| Step 1 | Assumptions: | Populations are normally distributed | |
| | | Samples are independent | |
| | | Equal Variances | |
| Step 2 | $H_0$: | $\mu_F - \mu_{SA} =$ | 0 |
| | $H_a$: | $\mu_F - \mu_{SA} >$ | 0 |
| Step 3 | Test statistic: t with df = | | 52 |
| | Right-tailed Test with α = | | 5% |
| Step 4 | t = | 4.678 | |
| | p-value = | 0.000011 | |
| Step 5 | p-value < α | | |
| | Reject Ho | | |
| | There is sufficient evidence, at the 5% level of significance to conclude that the value of houses in Ste-Foy is higher than in St. Augustin. | | |

Question 7 | Question 8 | Question 9

## Question 9 (5 points)

Global television is interested in knowing whether men and women watch the same amount of TV. A random sample of men was taken, and another of women, where each person gave the number of hours they watched TV during the previous week.

a) Go to the worksheet "Sheet3", rename it appropriately, make the usual heading in cells A1:A4, then click on the "GENERATE DATA" button to get your data.

b) At the 4% level of significance, can you conclude that men and women do not watch the same amount of TV? Use the classical approach.

| Men (hours of TV) | Women (hours of TV) |
|---|---|
| 13 | 18 |
| 26 | 17 |
| 27 | 16 |
| 18 | 22 |
| 19 | 27 |
| 20 | 19 |
| 26 | 23 |
| 23 | 8 |
| 17 | 3 |
| 9 | 21 |
| 12 | 18 |
| 14 | 19 |
| 13 | 23 |
| 18 | 6 |
| 18 | 22 |
| 21 | 16 |
| 17 | 22 |

Variance (Females): $s_1^2 =$ 21.821

Variance (Males): $s_2^2 =$ 33.758

z-Test: Two Sample for Means

| | Men (hours of TV) | Women (hours of TV) |
|---|---|---|
| Mean | 19.021 | 18.161 |
| Known Variance | 21.821 | 33.758 |
| Observations | 47 | 56 |
| Hypothesized Mean Difference | 0 | |
| z | 0.853 | |
| P(Z<=z) one-tail | 0.202 | |
| z Critical one-tail | 1.751 | |
| P(Z<=z) two-tail | 0.405 | |
| z Critical two-tail | 2.054 | |

| | | | | |
|---|---|---|---|---|
| Step 1 | Assumptions: | Samples are independent | | |
| | | $n_M =$ | 47 | ≥ 30 |
| | | $n_W =$ | 56 | ≥ 31 |
| Step 2 | $H_0$: | $\mu_M - \mu_W =$ | 0 | |
| | $H_a$: | $\mu_M - \mu_W \neq$ | 0 | |
| Step 3 | Test statistic: z | | | |
| | Two-tailed Test with α = | | 4% | |
| | z critical = | 2.054 | | |
| Step 4 | z = | 0.83 | | |
| Step 5 | z is not in the critical region | | | |
| | Fail to reject Ho | | | |
| | There is insufficient evidence, at the 5% level of significance to conclude that men and women do not watch the same amount of TV. | | | |

Question 7 | Question 8 | Question 9