ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Fall Semester 2015–2016

Statistics for Business

Assignment 3: Suggested Solutions

Confidence Intervals

1. The formula for the CIs is

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

with $\sigma = \$30,000$ and $n = 80$.

(a) Since we want a 90% confidence interval, we use $z_{0.05} = 1.645$, so

$$119,155 \pm 1.645(30,000/\sqrt{80}) = 119,155 \pm 5517 \text{ or } \$113,638 \text{ to } \$124,672$$

(b) Since we want a 95% confidence interval, we use $z_{0.025} = 1.96$, so

$$119,155 \pm 1.96(30,000/\sqrt{80}) = 119,155 \pm 6574 \text{ or } \$112,581 \text{ to } \$125,729$$

(c) Since we want a 99% confidence interval, we use $z_{0.005} = 2.576$, so

$$119,155 \pm 2.576(30,000/\sqrt{80}) = 119,155 \pm 8640 \text{ or } \$110,515 \text{ to } \$127,795$$

2. (a) The point estimate of the population mean is the sample mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}[10+8+12+15+13+11+6+5] = \frac{80}{8} = 10$$

(b) The point estimate of the population standard deviation is the square root of the point estimate for the variance. We have that

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{8-1}[(10-10)^2+(8-10)^2+...+(6-10)^2+(5-10)^2] = \frac{84}{7},$$

so

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sqrt{\frac{84}{7}} = 3.464$$

(c) The margin of error is

$$t_{n-1,0.025}\frac{s}{\sqrt{n}} = t_{7,0.025}\frac{3.464}{\sqrt{8}} = 2.365 \times \frac{3.464}{\sqrt{8}} = 2.896.$$

(d) The 95% confidence interval estimate of the population mean is

$$\bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} = 10 \pm 2.365 \times \frac{3.464}{\sqrt{8}} = 10 \pm 2.896 \quad \text{or} \quad 7.104 \text{ to } 12.896$$

3. To find these probabilities, I just used Excel.

(a) The margin of error is

$$t_{n-1,0.025}\frac{s}{\sqrt{n}} = t_{45-1,0.025}\frac{\$65}{\sqrt{45}} = 2.015 \times \frac{\$65}{\sqrt{45}} = \$19.525.$$

(b) The 95% confidence interval estimate of the population mean is

$$\bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} = \$273 \pm 2.015 \times \frac{\$65}{\sqrt{45}} = \$273 \pm \$19.525 \quad \text{or} \quad \$253.48 \text{ to } \$292.53$$

(c) At 95% confidence, the population mean is between $\$253.48$ and $\$292.53$. The left tail is some $\$24$ above the prior year's $\$229$ level, or the current average spending is well above the average spending two years ago, so average spending is increasing. The point estimate of the increase is $\$273 - \$229 = \$44$ or 19.21% of the price two years ago.

4. (a) The point estimate of the population proportion is

$$\hat{p} = \frac{46}{200} = 0.23.$$

(b) The 95% confidence interval for the population proportion is

$$\begin{aligned}\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.23 \pm 1.96\sqrt{\frac{0.23(1-0.23)}{200}} = 0.23 \pm 1.96 \times 0.0298 \\ &= 0.23 \pm 0.0584 \quad \text{or} \quad 0.1716 \text{ to } 0.2884\end{aligned}$$

5. We know that $\hat{p} = 0.09$ (so that $1400 \times 0.09 = 126$ individuals voted MySpace) and $n = 1400$. So the margin of error (with 95% confidence) is

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.09(1-0.09)}{1400}} = 0.01499.$$

The 95% confidence interval for the population proportion is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.09 \pm 1.96\sqrt{\frac{0.09(1-0.09)}{1400}} = 0.09 \pm 0.01499 \quad \text{or} \quad 0.075 \text{ to } 0.105$$

6. We have $\bar{x}_M = \$135.67$ and $n_M = 40$ with $\sigma_M = \$35$; and $\bar{x}_F = \$68.64$ and $n_F = 30$ with $\sigma_F = \$20$.

   (a) The point estimate of the difference between the population mean expenditure for males and the population mean expenditure for females is

$$\bar{x}_M - \bar{x}_F = \$135.67 - \$68.64 = \$67.03.$$

   (b) With known variances at 99% confidence, the margin of error is

$$z_{\alpha/2}\sqrt{\frac{\sigma_M^2}{n_M} + \frac{\sigma_F^2}{n_F}} = 2.576\sqrt{\frac{(35)^2}{40} + \frac{(20)^2}{30}} = 17.079$$

   (c) The 99% confidence interval for the difference between the two population means is

$$\begin{aligned}(\bar{x}_M - \bar{x}_F) \pm z_{\alpha/2}\sqrt{\frac{\sigma_M^2}{n_M} + \frac{\sigma_F^2}{n_F}} &= 67.03 \pm 2.576\sqrt{\frac{(35)^2}{40} + \frac{(20)^2}{30}} = 67.03 \pm 17.079 \\ &\text{or} \quad \$49.951 \text{ to } \$84.109.\end{aligned}$$

7. The data are

(x) Sample 1	10	7	13	7	9	8
(y) Sample 2	8	7	8	4	6	9

   (a) For the means we have

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{10+7+13+7+9+8}{6} = 9; \\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{8+7+8+4+6+9}{6} = 7.\end{aligned}$$

   Similarly, for the standard deviations, we have

$$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 2.28; \text{ and } s_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1.79.$$

   (b) The point estimate of the difference between the two population means is

$$\bar{x} - \bar{y} = 9 - 7 = 2.$$

(c) The 90% confidence interval estimate of the difference between the two population means is (assuming equal variances)

$$(\bar{x}-\bar{y}) \pm t_{n_x+n_y-2,\alpha/2}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}$$

The 'pooled' variance is estimated as

$$s_p^2 = \frac{(n_x-1)s_x^2+(n_y-1)s_y^2}{n_x+n_y-2} = \frac{(6-1)(2.28)^2+(6-1)(1.79)^2}{6+6-2} = 4.201.$$

The reliability factor $t_{n_x+n_y-2,\alpha/2}$ at 90% confidence is $t_{10,0.05} = 1.812$. So the CI is

$$\begin{aligned}(\bar{x}-\bar{y}) \pm t_{n_x+n_y-2,\alpha/2}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}} &= 2 \pm 1.812\sqrt{\frac{4.201}{6}+\frac{4.201}{6}} = 2 \pm 2.144 \\ &\text{or} \quad -0.144 \text{ to } 4.144\end{aligned}$$

Hypothesis Tests

8. Let

$$\begin{aligned} p_1 &= \text{population proportion of adults planning to travel by airplane for 2003} \\ p_2 &= \text{population proportion of adults planning to travel by airplane for 1993} \end{aligned}$$

(a) The hypothesis of interest is

$$\begin{aligned} H_0 &: p_1 - p_2 = 0 \\ H_1 &: p_1 - p_2 \neq 0 \end{aligned}$$

(b) The relevant sample proportions are

$$\begin{aligned} \hat{p}_1 &= \frac{141}{523} = 0.2696 \\ \hat{p}_2 &= \frac{81}{477} = 0.1698 \end{aligned}$$

(c) With $\alpha = 0.01$ we use $z_{\alpha/2} = z_{0.005} = 2.576$. The estimate for the common, overall proportion is

$$\bar{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2} = \frac{523(141/523)+477(81/477)}{523+477} = \frac{222}{1000} = 0.222$$

The test statistic for $p_1 - p_2 = 0$ is

$$z = \frac{(\hat{p}_1 - \hat{p}_1) - (0)}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n_1} + \frac{\bar{p}(1-\bar{p})}{n_2}}} = \frac{(0.2696 - 0.1698)}{\sqrt{\frac{0.222(1-0.222)}{523} + \frac{0.222(1-0.222)}{477}}} = 3.793.$$

Since $|z| = 3.793 > z_{0.005} = 2.576$ we reject the null in favor of the alternative.

(d) We note that $\hat{p}_1 - \hat{p}_1 = 0.2696 - 0.1698 = 0.0998 > 0$, which may be explained by the fact that airfares became quite cheaper over the period under study.

9. The hypothesis of interest is

$$\begin{aligned} H_0 &: \mu_1 - \mu_2 = 0, \\ H_1 &: \mu_1 - \mu_2 \neq 0. \end{aligned}$$

Assuming that population variances are equal, the test statistic is[1]

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{(s_p^2/n_1) + (s_p^2/n_2)}} \sim t_{n_1+n_2-2}; \text{ where } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

We have

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(35-1)(5.2)^2 + (40-1)(8.5)^2}{35+40-2} = 51.193.$$

Hence ($t_{n_1+n_2-2} = t_{35+40-2} = t_{73}$)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{(s_p^2/n_1) + (s_p^2/n_2)}} = \frac{(13.6 - 10.1)}{\sqrt{\frac{51.193}{35} + \frac{51.193}{40}}} = 2.113$$

The p-value for this test statistic is[2] 0.038. Hence at $\alpha = 0.05$ we reject the null hypothesis in favor of the alternative (note that $t_{73,0.025} = 1.993$).

10. (a) It is necessary to use a paired-difference test, since the two samples are not random and independent. The hypothesis of interest is

$$\begin{aligned} H_0 &: \mu_1 - \mu_2 = 0 \text{ or } H_0 : \mu_d = 0 \\ H_1 &: \mu_1 - \mu_2 \neq 0 \text{ or } H_1 : \mu_d \neq 0 \end{aligned}$$

We have

Population	1	2	3	4	5
X:	1.3	1.6	1.1	1.4	1.7
Y:	1.2	1.5	1.1	1.2	1.8
d_i	0.1	0.1	0	0.2	-0.1
d_i^2	0.01	0.01	0	0.04	0.01

[1] It is also possible to test the hypothesis assuming unequal population variances. I do not follow this route here.

[2] In Excel, just use ‘`=T.DIST.2T(2.113;35+40-2)`’

so

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{0.3}{5} = 0.06; \text{ and}$$
$$s_d^2 = \frac{\sum_{i=1}^{n} d_i^2 - (1/n)\left(\sum_{i=1}^{n} d_i\right)^2}{n-1} = \frac{0.07 - (1/5)(0.3)^2}{5-1} = 0.013.$$

The test statistic is

$$t = \frac{\bar{d} - d_0}{s_d/\sqrt{n}} = \frac{0.06 - 0}{\sqrt{0.013/5}} = 1.177$$

which is distributed as a t with $n - 1 = 4$ degrees of freedom. The rejection region with $\alpha = 0.05$ is $|t| > t_{4,0.025} = 2.776$, and H_0 is not rejected. We cannot conclude that the means are different.

(b) The p-value is

$$\Pr(|t| > 1.177) = 2\Pr(t > 1.177) = 2(0.152) = 0.304$$

(c) A 95% confidence interval for $\mu_1 - \mu_2 = \mu_d$ is

$$\bar{d} \pm t_{4,0.025}\frac{s_d}{\sqrt{n}} = 0.06 \pm 2.776\sqrt{\frac{0.304}{5}} = 0.06 \pm 0.142 \text{ or } -0.082 < (\mu_1 - \mu_2) < 0.202$$

(d) In order to use the paired-difference test, it is necessary that the n paired observations be randomly selected from normally distributed populations. We note that 0 is contained in the CI, which is in line with the hypothesis test above.

11. (a) We first calculate

$$s_x^2 = 15.333 \text{ and } s_y^2 = 10.3.$$

Hence

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} = \frac{(4-1)(15.333) + (5-1)(10.3)}{4+5-2} = 12.457.$$

(b) A 90% confidence interval for $(\mu_1 - \mu_2)$ is given as

$$(\bar{x} - \bar{y}) \pm t_{n_x+n_y-2,\alpha/2}\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}} = (7 - 8.6) \pm t_{7,0.05}\sqrt{\frac{12.457}{4} + \frac{12.457}{5}}$$
$$= -1.6 \pm 1.895 \times 2.368 = -1.6 \pm 4.487 \text{ or } -6.087 < (\mu_1 - \mu_2) < 2.887.$$

(c) The test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{(s_p^2/n_1) + (s_p^2/n_2)}} = \frac{7 - 8.6}{\sqrt{\frac{12.457}{4} + \frac{12.457}{5}}} = -0.676.$$

The rejection region is one-tailed, based on $df = 7$ degrees of freedom. With $\alpha = 0.05$, the rejection region is $t < -t_{7,0.05} = -1.895$. Since the observed value, $t = -0.676$ does not fall in the rejection region, H_0 is not rejected. We do not have sufficient evidence to indicate that $(\mu_1 - \mu_2) < 0$.

12. (a) The hypothesis of interest is

$$\begin{aligned} H_0 &: \mu_1 - \mu_2 = 0 \\ H_1 &: \mu_1 - \mu_2 \neq 0 \end{aligned}$$

(b) The rejection region is two-tailed, based on $df = n_1 + n_2 - 2 = 16 + 13 - 2 = 27$ degrees of freedom. With $\alpha = 0.01$, the rejection region is $|t| > t_{27,0.005} = 2.771$.

(c) The pooled estimator of σ^2 is calculated as

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(16-1)(4.8)^2 + (13-1)(5.9)^2}{(16-1) + (13-1)} = 28.271$$

and the test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{(s_p^2/n_1) + (s_p^2/n_2)}} = \frac{34.6 - 32.2}{\sqrt{28.271\left(\frac{1}{16} + \frac{1}{13}\right)}} = 1.209.$$

(d) The $p-$value is

$$p\text{-value} = \Pr(|t| > 1.209) = 2\Pr(t > 1.209) = 0.237.$$

(e) Comparing the observed $t = 1.209$ to the critical value $t_{27,0.005} = 2.771$ or comparing the p-value (0.237) to $\alpha = 0.01$, we find that H_0 is not rejected and we conclude that $\mu_1 = \mu_2$

13. The hypothesis of interest is

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 \\ H_1 &: \sigma_1^2 > \sigma_2^2 \end{aligned}$$

and the test statistic is

$$F = \frac{s_1^2}{s_2^2} = \frac{114^2}{103^2} = 1.22$$

which is distributed as an $F_{n_1-1,n_2-1} = F_{16-1,15-1} = F_{15,14}$. The critical values of F for various values of α are given below using $df_1 = 15$ and $df_2 = 14$

α	0.10	0.05	0.025	0.01	0.005
$F_{15,14,\alpha}$	2.010	2.463	2.949	3.656	4.247

Moreover since the p-value is 0.358, we conclude that H_0 is not rejected. There is no evidence to indicate that the variances are different.