CONCORDIA UNIVERSITY
Department of Economics

ECON 221/2 SECTIONS A, B, C and DD
STATISTICAL METHODS I
FALL 2016 – TUTORIAL 4 (SOLUTIONS)
Friday, November 18, 2016

Name: **I.D:** **Section:**

1. The managers of an investment website want to estimate the proportion of users who access their site with cell phones. They take a random sample of 200 investors from among their customers. Suppose that the true proportion of smartphone users is 36 percent.

   a. ***Briefly*** describe the shape that the sampling distribution of the proportion should have.

   Since $np(1-p) = 200 \cdot 0.36 \cdot (1-0.36) = 46.08 > 5$, *the central limit theorem states that the sample size is sufficiently large to assume that the sampling distribution of* $\hat{p}$ *is normal.*

   b. Calculate the mean of the sampling distribution.

   $$E(\hat{p}) = p = 0.36$$

   c. Calculate the standard deviation of the sampling distribution.

   $$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.36 \cdot (1-0.36)}{200}} = 0.0339$$

   d. ***Briefly*** explain how your answers to parts (a) – (c) would change if the sample size increases to 500.

   If the sample size increases to 500, only the answer to part (c) would change to $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.36 \cdot (1-0.36)}{500}} = 0.0214$ *hours.*

   Continue with a sample of 200.

   e. Calculate the probability that the sample proportion of smartphone users is greater than 0.42.

$$\Pr\left(\hat{p} > 0.42\right) = \Pr\left(\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} > \frac{0.42 - 0.36}{\sqrt{\frac{0.36 \cdot (1 - 0.36)}{200}}}\right) = \Pr\left(z > 1.77\right) = 1 - \Pr\left(z < 1.77\right)$$

$$= 1 - 0.9616 = 0.0384$$

f. Calculate the probability that the sample proportion of smartphone users is between 0.30 and 0.40.

$$\Pr\left(0.30 < \hat{p} < 0.40\right) = \Pr\left(\frac{0.30 - 0.36}{\sqrt{\frac{0.36 \cdot (1 - 0.36)}{200}}} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{0.40 - 0.36}{\sqrt{\frac{0.36 \cdot (1 - 0.36)}{200}}}\right)$$

$$= \Pr\left(-1.77 < z < 1.18\right) = \Pr\left(z < 1.18\right) - \Pr\left(z < -1.77\right) = \Pr\left(z < 1.18\right) - \Pr\left(z > 1.77\right)$$

$$= \Pr\left(z < 1.18\right) - \left[1 - \Pr\left(z < 1.77\right)\right] = \Pr\left(z < 1.18\right) + \Pr\left(z < 1.77\right) - 1 = 0.8810 + 0.9616 - 1$$

$$= 0.8426$$

2. Adult heights are approximately normally distributed. Suppose that the population of adult Canadian males has a mean height of 175 centimetres (cm) and a standard deviation of 7cm.

a. Calculate the probability that a randomly chosen adult male is taller than 183cm.

$$\Pr\left(x > 183\right) = \Pr\left(\frac{x - \mu}{\sigma} > \frac{183 - 175}{7}\right) = \Pr\left(z > 1.14\right) = 1 - \Pr\left(z < 1.14\right) = 1 - 0.8729 = 0.1271$$

b. Calculate the probability that the sample mean of two randomly chosen adult males is greater than 183cm.

$$\Pr\left(\bar{x} > 183\right) = \Pr\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{183 - 175}{7/\sqrt{2}}\right) = \Pr\left(z > 1.62\right) = 1 - \Pr\left(z < 1.62\right) = 1 - 0.9474 = 0.0526$$

c. Calculate the probability that the sample mean of five randomly chosen adult males is greater than 183cm.

$$\Pr\left(\bar{x} > 183\right) = \Pr\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{183 - 175}{7/\sqrt{5}}\right) = \Pr\left(z > 2.56\right) = 1 - \Pr\left(z < 2.56\right) = 1 - 0.9948 = 0.0052$$

d. ***Briefly*** explain why the probability of the sample mean being greater than 183cm decreases as *n* increases.

As the sample size increases, we obtain more information about the population; we are more likely to get a few short males in our sample, which will pull the sample mean closer to the population mean.

3. An insurance company checks police records on 582 accidents selected at random and notes that teenagers were at the wheel in 91 of them.

   a. Calculate the margin of error for a 95 percent confidence interval for the true proportion of all auto accidents that involve teenagers.

   $$\hat{p} = \frac{91}{582} = 0.1564 \Rightarrow ME = z_{\alpha/2}\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} = 1.96\sqrt{\frac{0.1564 \cdot \left(1-0.1564\right)}{582}} = 0.0295$$

   b. Construct a 95 percent confidence interval for the true proportion of all auto accidents that involve teenagers.

   $$\hat{p} \pm ME = 0.1564 \pm 0.0295 = \left(0.1268, 0.1859\right)$$

   c. ***<u>Briefly</u>*** interpret the confidence interval.

   There is 95 percent confidence that the population proportion of accidents involving teenagers is between 12.7 and 18.6 percent.

   d. ***<u>Briefly</u>*** explain what "95 percent confidence" means.

   Approximately 95 percent of random samples of size 582 produce intervals that contain the true proportion of accidents involving teenagers.

   e. A politician urging tighter restrictions on drivers' licences issued to teens says, "In one out of every five auto accidents, a teenager is behind the wheel." Briefly explain whether your confidence interval supports or contradicts this statement.

   The confidence interval does not contain the figure that the politician quotes (ie, 0.20), therefore, it contradicts the statement.

   f. Calculate the sample size required to reduce the margin of error for the 95 percent confidence interval by half.

   To reduce the margin of error by half, the sample size must increase by a factor of four. Therefore, the sample size must increase to 2328.

$$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_2}} = ME_2 = \frac{1}{2}ME_1 = \frac{1}{2}z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}} = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{4n_1}} \Rightarrow n_2 = 4n_1$$

4. A city plans to pay for a new downtown public parking garage through parking fees. A random sample of 20 weekdays shows daily fees collected averaged \$126 with a standard deviation of \$15.

   a. ***Briefly*** explain what assumptions are necessary to use these statistics for inference.

   It was necessary to assume that the population is distributed normally so that the estimated sample variance could be used in place of the unknown population variance and the t-distribution could be used to construct the confidence interval.

   b. Construct a 90 percent confidence interval for the true mean daily fees that the parking garage generates.

   $$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 126 \pm 1.729 \cdot \frac{15}{\sqrt{20}} = (120.20, 131.80)$$

   c. ***Briefly*** interpret the confidence interval found in part (b).

   We are 90 percent confident that the true mean daily fees that the parking garage generates are between \$120.20 and \$131.80.

   d. A consultant who advised the city predicts that parking fees collected will average \$128 per day. Based on the confidence interval in part (b), ***briefly*** explain whether the consultant's prediction is realistic.

   Since \$128 lies within the confidence interval, the consultant's prediction is realistic.

   e. Construct a 95 percent confidence interval for the standard deviation of the daily fees that the parking garage generates.

   $$\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}} < \sqrt{\sigma^2} < \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}} \Rightarrow \sqrt{\frac{19\cdot 15^2}{32.852}} < \sigma < \sqrt{\frac{19\cdot 15^2}{8.907}} \Rightarrow 11.4074 < \sigma < 21.9080$$

5. An accounting firm is deciding between IT training conducted in-house or through third-party consultants. Each type of training was implemented at two of its offices with the average cost per employee, the number of employees trained and the (known) population standard deviation of the training cost given in the table below.

	In-House	*Consultants*
Sample Size	210	180
Sample Mean	$490	$500
Population Standard Deviation	$32	$48

a. Construct a 99 percent confidence interval for the true mean difference in IT training costs.

$$\bar{x}-\bar{y}+z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}=490-500\pm 2.575\sqrt{\frac{32^2}{210}+\frac{48^2}{180}}=\left(-20.8261, 0.8261\right)$$

b. Based on the confidence interval in part (a), ***briefly*** explain what can be inferred about the difference in the average IT training costs.

Since the confidence interval contains zero, there is 99 percent confidence that there is no difference in the average training cost.

6. A random sample of 700 Nova Scotia families reports 17 percent are single-parent families, while a random sample of 1000 Alberta families reports 14 percent are. Let p_1 and p_2 be the proportion of single-parent families in Nova Scotia and Alberta, respectively.

a. Calculate the margin of error for a 98 percent confidence interval for $p_1 - p_2$.

$$n_1 = 700,\ \hat{p}_1 = 0.17, n_2 = 1000,\ \hat{p}_2 = 0.14$$

$$ME = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\left(1-\hat{p}_1\right)}{n_1}+\frac{\hat{p}_2\left(1-\hat{p}_2\right)}{n_2}}=2.33\sqrt{\frac{0.17\cdot\left(1-0.17\right)}{700}+\frac{0.14\cdot\left(1-0.14\right)}{1000}}=0.0418$$

b. Construct a 98 percent confidence interval for $p_1 - p_2$.

$$\hat{p}_1-\hat{p}_2 \pm ME = 0.17-0.14\pm 0.0418=\left(-0.0118, 0.0718\right)$$

c. Calculate the width of the 98 percent confidence interval.

$$w = 2\cdot ME = 2\cdot 0.0418 = 0.0836$$

d. Based on the interval calculated in part (b), ***briefly*** explain whether one can infer that the percentage of single-parent families in Alberta is lower than in Nova Scotia.

Since the confidence interval contains zero, there is 98 percent confidence that there is no difference in the percentage of single-parent families in Alberta and Nova Scotia.

7. A city wants to test the effectiveness of its new anti-drinking and driving advertising campaign. To do so, it counts the number of drunk drivers pulled over each day of the week before the campaign starts and each day of the week a month after it starts. The results are shown below.

Day of week	Before	After
M	5	2
T	4	0
W	2	2
J	4	1
F	6	8
S	14	7
D	6	7

a. Calculate the sample mean of the differences.

Day of week	Before	After	Difference
M	5	2	3
T	4	0	4
W	2	2	0
J	4	1	3
F	6	8	-2
S	14	7	7
D	6	7	-1

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{3+4+0+3-2+7-1}{7} = 2$$

b. Calculate the sample standard deviation of the differences.

$$\sqrt{s_d^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} d_i^2 - n\bar{d}^2\right)} = \sqrt{\frac{3^2+4^2+0^2+3^2+(-2)^2+7^2+(-1)^2-7\cdot 2^2}{7-1}} = 3.1623$$

c. Construct a 95 percent confidence interval for the true difference in the mean number of citizens who drive drunk.

$$ME = t_{\alpha/2,n-1}\frac{s_d}{\sqrt{n}} = 2.447\cdot\sqrt{\frac{10}{7}} = 2.9247 \Rightarrow \bar{d} \pm ME = 2 \pm 2.9247 = (-0.9247, 4.9247)$$

d. ***Briefly*** explain whether the advertising campaign was effective in reducing the number of citizens who drive after drinking.

Since the confidence interval contains zero, there is 95 percent confidence that the advertising campaign was ineffective.

8. www.shop.org routinely reports online shopping statistics. Of interest to many online retailers are gender-based differences in shopping preferences and behaviours. Average monthly online expenditures are reported for males and females.

	Male	Female
Sample size	31	31
Mean	\$352	\$310
Sample standard deviation	\$95	\$80

a. Calculate the pooled sample variance.

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(31-1)\cdot 95^2+(31-1)\cdot 80^2}{31+31-2} = 7712.5$$

b. Construct a 95 percent confidence interval for the true difference in the average monthly online expenditures of males and females.

$$ME = t_{\alpha/2,n_1+n_2-2}\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}} = 2.000\cdot\sqrt{\frac{7712.5}{31}+\frac{7712.5}{31}} = 44.6130$$

$$\bar{x}_1-\bar{x}_2 \pm ME = 352-310 \pm 44.6130 = (-2.6130, 86.6130)$$

c. ***<u>Briefly</u>*** interpret the confidence interval you found in part (b).

There is 95 percent confidence that the true difference in the average monthly expenditures of males and females is between -\$2.61 and \$86.61.