STAT2201, Semester 1 2016

Solution for Quiz #3 (Updated 22/5/2016)

Let M be your month of birth, i.e. $M \in \{1, 2, \dots, 12\}$. $M =$ _______________.

Question 1: You obtain a sample of $n = 15 + M$ (VERSION$_a$) or $n = 5 + M$ (VERSION$_b$) observations. Upon computing the sample mean ($\overline{x}$) and sample standard deviation (s) you find:

$$\text{VERSION}_a : \quad \overline{x} = 163.3, \quad s = 10.2. \qquad \text{VERSION}_b : \quad \overline{x} = 72.4, \quad s = 8.2.$$

(a) Obtain a 90% (in VERSION$_a$) or 95% (in VERSION$_b$) confidence interval for the population mean:

Solution: The formula for the confidence interval to use is: $\overline{x} \pm t_{n-1} \frac{s}{\sqrt{n}}$, where t_{n-1} can be taken from the "t-table" using the $Q(.95)$ column for VERSION$_a$ and the $Q(0.975)$ column for VERSION$_b$. The degrees of freedom (row from the t-table) are $n - 1$. This evaluates to:

	VERSION$_a$			VERSION$_b$	
M	low bound	high bound	M	low bound	high bound
1	158.83	167.77	1	63.7946	81.0054
2	158.981	167.619	2	64.8163	79.9837
3	159.118	167.482	3	65.5446	79.2554
4	159.242	167.358	4	66.0969	78.7031
5	159.356	167.244	5	66.5341	78.2659
6	159.461	167.139	6	66.8912	77.9088
7	159.558	167.042	7	67.19	77.61
8	159.648	166.952	8	67.4448	77.3552
9	159.732	166.868	9	67.6655	77.1345
10	159.81	166.79	10	67.859	76.941
11	159.883	166.717	11	68.0305	76.7695
12	159.952	166.648	12	68.1839	76.6161

(b) What assumptions are required for this confidence interval to be valid?

Solution: This confidence interval assumes a normal distribution for the observations. It also assumes independent observations (not needed for getting credit for the question). Answers based on the CLT with justifiable arguments (indicating the sample size is not too small) are also accepted. Note though, that without normality, the t-distribution result does not exactly hold and the confidence interval is just an approximation.

Question 2: You are comparing two groups of items: 1 and 2. You wish to determine if their population means are the same or not. In the comparison you assume that values are distributed Normally with a mean of μ_i for group i and with the same variance for both groups, denoted by σ^2; but the means and variance are not known.

You obtain two random samples with n_i observations for group i. Upon computing the sample means (denoted by $\overline{x}_i$) and sample standard deviations (denoted by s_i) you find:

$$\overline{x}_1 = 24.3, \qquad \overline{x}_2 = 27.4, \qquad s_1 = 3.2, \qquad s_2 = 2.7.$$

Assume that $n_1 = 5 + M$ (VERSION$_a$) or $n_1 = 12 + M$ (VERSION$_b$) and $n_2 = 10 + M$ (VERSION$_a$) or $n_2 = 8 + M$ (VERSION$_b$).

(a) Use the pooled sample variance to calculate your estimate for the population standard deviation σ:

Solution: Use the formula:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

This evaluates to:

M	VERSION$_a$	VERSION$_b$
1	2.87634	3.00998
2	2.88638	3.00553
3	2.89428	3.00181
4	2.90066	2.99865
5	2.90592	2.99595
6	2.91033	2.9936
7	2.91408	2.99155
8	2.91731	2.98974
9	2.92012	2.98812
10	2.92259	2.98668
11	2.92477	2.98538
12	2.92672	2.9842

(b) Write out the hypotheses associated with this question:

Solution:

$H_0: \ \mu_1 = \mu_2$:

$H_A: \ \mu_1 \neq \mu_2$:

(c) Calculate the test statistic and draw your conclusion with $\alpha = 0.05$ (VERSION$_a$) and $\alpha = 0.1$ (VERSION$_b$):

Solution: The test statistic,

$$T = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

has a t-distribution with $n_1 + n_2 - 2$ degrees of freedom under H_0. It evaluates to:

M	VERSION$_a$	VERSION$_b$
1	-2.12358	-2.37509
2	-2.25825	-2.49115
3	-2.38357	-2.60156
4	-2.50142	-2.70712
5	-2.61309	-2.80842
6	-2.71953	-2.90598
7	-2.82148	-3.00019
8	-2.91949	-3.09138
9	-3.01401	-3.17984
10	-3.10542	-3.26581
11	-3.19403	-3.34948
12	-3.28008	-3.43105

Note that since the t-distribution is symmetric, taking the absolute values is fine too.

The associated p-values (computed numerically) are:

M	VERSION$_a$	VERSION$_b$
1	0.0507488	0.0276597
2	0.0373669	0.0207644
3	0.0277356	0.015649
4	0.0207146	0.0118322
5	0.0155472	0.00897111
6	0.0117156	0.00681794
7	0.0088577	0.00519223
8	0.00671578	0.00396129
9	0.00510404	0.003027
10	0.00388719	0.00231635
11	0.00296583	0.0017748
12	0.00226649	0.00136142

These values cannot be obtained precisely with the "t-table", but may be approximated (although this is not required to obtain full marks for the question).

As a conclusion: With $\alpha = 0.05$ (VERSION$_a$), H_0 should be rejected for $M \geq 2$. With $\alpha = 0.1$ (VERSION$_b$), H_0 should be rejected for all M.