A N S W E R K E Y

STAT 1051: QUIZ 8 Take Home

Name

1. Marketing strategists want to characterize purchasers and non-purchasers of Crest toothpaste according to household income. The average income of sample #1 consisting of 20 purchasers of Crest is $47,200 with a sample standard deviation of $13,600 while the average income of sample #2 consisting of 20 non-purchasers of Crest is $39,800 with a sample standard deviation of $10,039. Test the hypothesis: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 > 0$ at the .05 level of significance.

   (a) **Question 1** (5 pts) $\alpha = .05$ $\nu = n_1 + n_2 - 2 = 20 + 20 - 2 = 38$ Is this a one-sided test or two-sided test? **One-sided test.**

   (b) **Question 2** (5 pts) $t_{38,.05} = \mathbf{1.685954}$

   (c) **Question 3** (5 pts) $S_p^2 = \frac{19(13600^2)+19(10039^2)}{38} = 142870760$ $S_p = \mathbf{11952.86}$

   (d) **Question 4** (5 pts) $T = \frac{\mu_1 - \mu_2 - d_0}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{\mathbf{47200} - 39800 - \mathbf{0}}{11952.86(.3162278)} = 1.957762$

   (e) **Question 5** (5 pts) Should we reject the null hypothesis? **Yes** Why? Is $T = 1.957762 > 1.685954$? Yes, therefore, reject the null hypothesis Note that the p-value=0.02881552

2. A paired difference experiment in which eight women who enrolled in an aerobics class were weighed before and after a five week class. The weights expressed in units of pounds are tabulated below. Test the hypothesis: $H_0 : \mu_D = 0$ vs $H_1 : \mu_D > 0$ where $\mu_D = (\mu_1 - \mu_2)$. Use $\alpha = .10$

	Before #1	After #2	d
1	198	194	4
2	154	151	3
3	124	126	-2
4	110	104	6
5	127	123	4
6	162	155	7
7	141	129	12
8	180	165	15

(a) **Question 6** (5 pts) $\bar{d} = \mathbf{6.125}$ $s_d^2 = \frac{\sum(x_i - \bar{d})^2}{n-1} = \mathbf{28.41071}$ $s_d = \mathbf{5.33017}$

(b) **Question 7** (5 pts) $\alpha = .10$ $n - 1 = \mathbf{7}$

(c) **Question 8** (5 pts) $t_{7,.10} = \mathbf{1.41492}$

(d) **Question 9** (5 pts) $T = \frac{\bar{d}-0}{\frac{s_d}{\sqrt{8}}} = \mathbf{3.250199}$

(e) **Question 10** (5 pts) Should we reject the null hypothesis? **Yes** Why? Is $T = 3.250199 > 1.89$? Yes, therefore, reject the null hypothesis.
Note that p-value=0.007026709.

3. A specialist in hypertension conducted an experiment with 50 volunteers who suffer from high blood pressure. Half of them, group #1, underwent medication; the other half, group #2, exercised. The percent decrease in blood pressure was recorded and the aggregate statistics are given below. Test the hypothesis that aerobic exercise is better than medication i.e.: $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 < \mu_2$ at $\alpha = .05$ (Hint: Write the hypothesis like: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 < 0$)

Group	n	Sample Mean	Sample Variance
1	25	9.92	13.16
2	25	13.52	5.76

(a) **Question 11** (5 pts) $\alpha = \mathbf{.05}$ $\nu = n_1 + n_2 - 2 = 25 + 25 - 2 = \mathbf{48}$ One-sided test.

(b) **Question 12** (5 pts) $t_{48,.05} = \mathbf{1.677224}$

(c) **Question 13** (5 pts) $S_p^2 = \frac{24(13.16)+24(5.76)}{48} = \frac{24(13.16)+24(\mathbf{5.76})}{48} = \mathbf{9.46} \rightarrow S_p = \mathbf{3.0757}$

(d) **Question 14** (5 pts) $T = \frac{\mu_1 - \mu_2 - d_0}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{9.92 - 13.52 - 0}{3.0757(.2828427)} = \mathbf{-4.1382}$

(e) **Question 15** (5 pts) Should we reject the null hypothesis and thereby conclude that percent reduction in high blood pressure is greater due to aerobic exercising than taking medication? **Yes** Why? Because $T = -4.1382 < -1.677224$; therefore, we reject the null hypothesis. For all practical purposes, the effects of medication and of aerobic exercising are different. Note that p-value=7.0188e-5

4. A sample of 100 electric light bulbs produced by manufacturer A showed a mean lifetime of 1190 hours with a standard deviation of 90 hours. A sample of 75 bulbs produced by manufacturer B showed a mean lifetime of 1250 hours with a standard deviation of 120 hours. Is there a difference in mean lifetimes by manufacturer at alpha=.05?

I would make a stronger assertion in that manufacturer A produces light bulbs which have a mean lifetime less than that of manufacturer B. Consequently, I would conduct a one-sided test. The author, on the other hand, is forcing us to conduct a two-sided test.

Accordingly, let us test the hypothesis: $H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 \neq 0$ at $\alpha = .05$.

Let us denote manufacturer A by 1 and manufacturer B by 2.

Question 16 (5 pts) Gather the given information and put it into a table.

Descriptive Statistic	Population 1	Population 2
n	**100**	75
$\bar{x}$	1190	**1250**
s	**90**	120

(a) $\alpha = .05$ $\frac{\alpha}{2}$ $\nu = 100 + 75 - 2 = 173$ Two-sided test.

(b) **Question 17** (5 pts) $t_{173,\mathbf{.025}} = 1.973771$

(c) **Question 18** (5 pts) $S_p^2 = \frac{99(90^2)+74(120^2)}{173} = 10794.8 \rightarrow S_p = \mathbf{103.898}$

(d) **Question 19** (5 pts) $T = \frac{\mu_1 - \mu_2 - d_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{\mathbf{1190} - 1250 - \mathbf{0}}{103.898\sqrt{\frac{1}{100}+\frac{1}{75}}} = -3.78$

(e) **Question 20** (5 pts) Should we reject the null hypothesis, that is, are the lifetimes of the light bulbs different according to manufacturer? **Yes** Why? Because $|T| = 3.78 > 1.973771$? Note that the p-value=.0002152

5. The Institute for Health Metrics and Evaluation which is run by the University of Washington has published predictions of required hospital resources, number of deaths per day, and the cumulative number of deaths which are attributed to the coronavirus. See http://www.healthdata.org/. This organization published on 7 April 2020 a prediction of the cumulative deaths in the United States. It appears in Figure 1

Question 21 (2 pts) What is the predicted number of deaths in the United States due to the coronavirus as of 1 May 2020?

Answer 1 *About 53,000=53k*

Question 22 (2 pts) What is its 95% confidence interval?

Answer 2 *(30k,100k)*

Question 23 **BONUS** (4 pts) Interpret the meaning of this confidence interval to your favorite American politician.

Answer 3 *We are 95% confident that the actual number of deaths in the U.S which can be attributed to the coronavirus will be in the confidence interval.*

Note: Although we do not know the formula which the researchers used to obtained the confidence interval, let us assume that

$$\begin{aligned} a &= \bar{x} - \frac{s}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} \to \bar{x} - \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}} \\ b &= \bar{x} + \frac{s}{\sqrt{n}} z_{\frac{\alpha}{2}} \end{aligned}$$

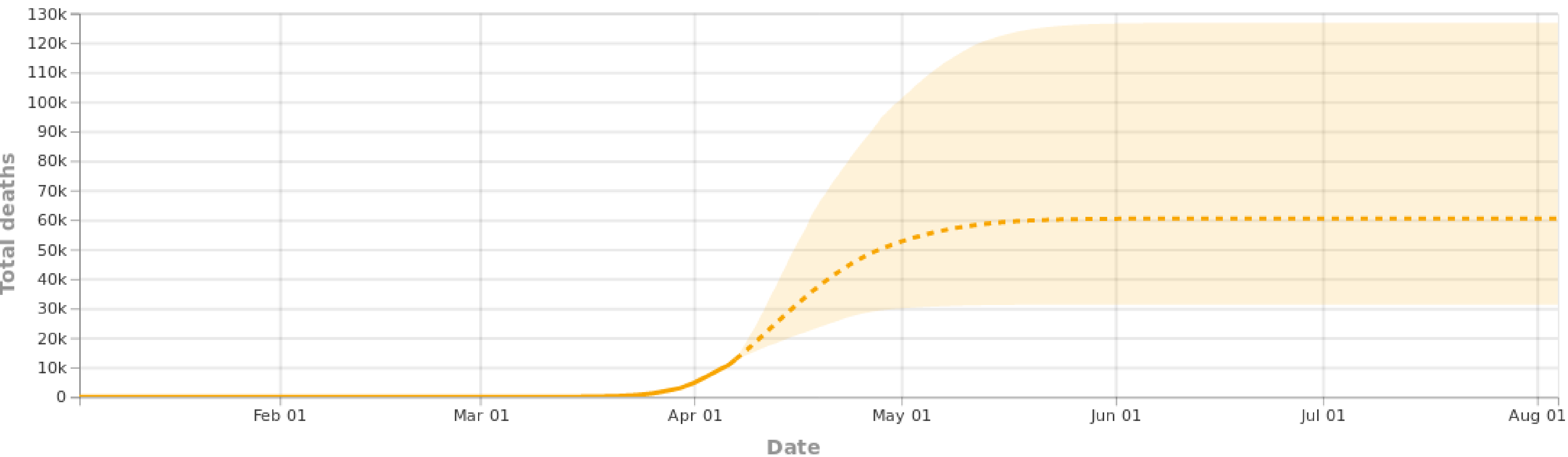
So that, $\frac{s}{\sqrt{n}} = \frac{b-a}{2z_{\frac{\alpha}{2}}} = 17.5$ *when* $\alpha = .05$ *a=30 and b=100.*

The coefficient of variations, CV, of the estimate is: $CV = \frac{\frac{s}{\sqrt{n}}}{est} = \frac{17.5}{53} = 33\%$ *A CV of 2% is considered a good value for a government survey. A CV of 33% is very high. The Institute for Health Metrics and Evaluation estimate and confidence interval are good enough for these troubled times to give policy makers some sense of direction in how to combat potential consequences of the coronavirus epidemic.*

Question 24 (2 pts) Is this an example of interpolation or of extrapolation?

Answer 4 *Extrapolation*

Figure 1:


Total deaths
130k
120k
110k
100k
90k
80k
70k
60k
50k
40k
30k
20k
10k
0
Feb 01
Mar 01
Apr 01
May 01
Jun 01
Jul 01
Aug 01
Date