Challenge Set 5

Deadline: Mar 17 2017 at 5pm

Challenge problems are ***optional*** problems for those interested in testing their abilities. For each correct answer to a challenge question, bonus points of 0.1 are given towards the *final overall grade,* i.e., you can potentially earn up to 1.5 points towards the final grade if you get all questions correct. Proper workings must be shown to get any points, and there is no partial credit. Also, because these are bonus questions, instructors will not provide any help or hints (unlike typical problem or practice set questions where generous assistance will be provided) to be fair to all students. Please submit your solutions via the Turnitin assignment "Challenge Set 5" on TED@UCSD (you can simply take a good resolution photo/scan of your solutions with your student ID number and name clearly labelled and convert it to a PDF for upload) by the deadline.

Q1: The rate of evaporation for a small droplet suspended in air is given by the following equation, where M is the mass flow rate, ρ is the density of the evaporating species, D is the diffusion coefficient, R is the radius of the droplet, w_s is the mass fraction of the evaporating species at the surface of the droplet, and w_e is the mass fraction of the evaporating species in the surrounding environment. All variables are in SI units.

$$M = \frac{\rho D}{R} \ln\left(1 + \frac{w_s - w_e}{1 - w_s}\right)$$

Data was collected in lab and is tabulated below. Estimate w_s and its uncertainty. Show all work by hand for full credit.

M	ρ	D	R	w_e
1.73	791	$1.08 * 10^{-5}$	$6.1 * 10^{-3}$	0.008
1.68	775	$1.25 * 10^{-5}$	$5.2 * 10^{-3}$	0.012
1.81	783	$0.98 * 10^{-5}$	$5.9 * 10^{-3}$	0.011
1.79	802	$1.12 * 10^{-5}$	$7.1 * 10^{-3}$	0.006
1.75	793	$1.20 * 10^{-5}$	$6.6 * 10^{-3}$	0.015
1.61	799	$1.03 * 10^{-5}$	$6.3 * 10^{-3}$	0.009

Solution:

Rearrange equation for w_s

$$\frac{RM}{\rho D} = \ln\left(\frac{1 - w_s}{1 - w_s} + \frac{w_s - w_e}{1 - w_s}\right)$$

$$\ln\left(\frac{1 - w_e}{1 - w_s}\right) = \frac{RM}{\rho D}$$

$$w_s = 1 + (w_e - 1) * \exp\left(-\frac{RM}{\rho D}\right)$$

Calculate the average and standard deviation for each variable. Since they are not given, they need to be estimated from the data with these equations.

$$\bar{X} = \frac{\sum X}{n}$$

$$SS_X = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$s = \sqrt{\frac{SS_X}{n - 1}}$$

$$s_X = \frac{s}{\sqrt{n}}$$

M -

$$\bar{M} = \frac{\sum M}{n} = 1.7283$$

$$SS_M = \sum M^2 - \frac{(\sum M)^2}{n} = 17.9501 - \frac{107.5369}{6} = 0.0273$$

$$s_{\bar{M}} = \sqrt{\frac{SS_M}{n * (n - 1)}} = \sqrt{\frac{0.0273}{6 * (6 - 1)}} = 0.0302$$

ρ -

$$\bar{\rho} = \frac{\sum \rho}{n} = 790.5$$

$$SS_{\rho} = \sum \rho^2 - \frac{(\sum \rho)^2}{n} = 3749849 - \frac{22496049}{6} = 507.5$$

$$s_{\bar{\rho}} = \sqrt{\frac{SS_{\rho}}{n * (n-1)}} = \sqrt{\frac{507.5}{6 * (6-1)}} = 4.1130$$

D –

$$\bar{\mathrm{D}} = \frac{\sum \mathrm{D}}{n} = 1.1100 * 10^{-5}$$

$$SS_{\mathrm{D}} = \sum \mathrm{D}^2 - \frac{(\sum \mathrm{D})^2}{n} = 7.4446 * 10^{-10} - \frac{4.4356 * 10^{-9}}{6} = 5.1933 * 10^{-12}$$

$$s_{\bar{\mathrm{D}}} = \sqrt{\frac{SS_{\mathrm{D}}}{n * (n-1)}} = \sqrt{\frac{5.1933 * 10^{-12}}{6 * (6-1)}} = 4.1607 * 10^{-7}$$

R –

$$\bar{\mathrm{R}} = \frac{\sum \mathrm{R}}{n} = 0.0062$$

$$SS_{\mathrm{R}} = \sum \mathrm{R}^2 - \frac{(\sum \mathrm{R})^2}{n} = 2.3272 * 10^{-4} - \frac{1.3838 * 10^{-3}}{6} = 2.0867 * 10^{-6}$$

$$s_{\bar{\mathrm{R}}} = \sqrt{\frac{SS_{\mathrm{R}}}{n * (n-1)}} = \sqrt{\frac{2.0867 * 10^{-6}}{6 * (6-1)}} = 2.6373 * 10^{-4}$$

w_e –

$$\overline{w_e} = \frac{\sum w_e}{n} = 0.0102$$

$$SS_{w_e} = \sum {w_e}^2 - \frac{(\sum w_e)^2}{n} = 6.7100 * 10^{-4} - \frac{3.7210 * 10^{-3}}{6} = 5.0833 * 10^{-5}$$

$$s_{\overline{w_e}} = \sqrt{\frac{SS_{w_e}}{n * (n-1)}} = \sqrt{\frac{5.0833 * 10^{-5}}{6 * (6-1)}} = 1.3017 * 10^{-3}$$

Compute all the partial derivatives and plug in average values:

$$\frac{\partial w_s}{\partial M} = \frac{-R}{\rho D}(w_e - 1)\exp\left(-\frac{RM}{\rho D}\right) = 0.2062$$

$$\frac{\partial w_s}{\partial \rho} = \frac{RM}{\rho^2 D}(w_e - 1)\exp\left(-\frac{RM}{\rho D}\right) = -4.5090 * 10^{-4}$$

$$\frac{\partial w_s}{\partial D} = \frac{RM}{\rho D^2}(w_e - 1)\exp\left(-\frac{RM}{\rho D}\right) = -3.2112 * 10^{4}$$

$$\frac{\partial w_s}{\partial R} = \frac{-M}{\rho D}(w_e - 1)\exp\left(-\frac{RM}{\rho D}\right) = 57.4904$$

$$\frac{\partial w_s}{\partial w_e} = \exp\left(-\frac{RM}{\rho D}\right) = 0.2949$$

Find average calculated w_s by plugging in average values:

$$w_s = 1 + (w_e - 1) * \exp\left(-\frac{RM}{\rho D}\right) = 0.7081 = w_s$$

Propagation of Error equation:

$$s_{w_s} = \sqrt{\left(\frac{\partial w_s}{\partial M}\right)^2 s_M{}^2 + \left(\frac{\partial w_s}{\partial \rho}\right)^2 s_\rho{}^2 + \left(\frac{\partial w_s}{\partial D}\right)^2 s_D{}^2 + \left(\frac{\partial w_s}{\partial R}\right)^2 s_R{}^2 + \left(\frac{\partial w_s}{\partial w_e}\right)^2 s_{w_e}{}^2}$$

$$s_{w_s} = 0.0212$$

Q2: In this problem we will learn about two possible errors in statistical hypothesis testing and how they are applied in real world problems

Definitions

A **type I error** is the incorrect rejection of a true null hypothesis (a "false positive").

A **type II error** is incorrectly retaining a false null hypothesis (a "false negative").

Familiarizing with the definitions

Joey picks a sample by sampling the uniform distribution U (0, K) where K is a positive number. He decides to test the Null hypothesis: K=5 against Alternate Hypothesis: K! =5 by rejecting Null Hypothesis if K<0.5 or K>4.5

- Compute the probability of **type I** error.
- Compute the probability of **type II** error if true value of K is 6.

Soln

$$\mathrm{P(type\ 1\ error) = P(reject\ Null\ Hyp\ |Null\ Hyp\ is\ true)}$$

$$= \mathrm{P}(k < 0.5 \text{ or } k > 4.5\ |K = 5) = \frac{0.5+0.5}{5} = 0.2$$

$$\mathrm{P(type\ 2\ error) = P(dont\ reject\ Null\ Hyp\ |K = 6)}$$

$$= \mathrm{P}(0.5 < k < 4.5\ |K = 6) = \frac{4}{6} = 0.66$$

Real world application of the definitions

Over speeding is one of the leading cause of accidents In order to punish drivers driving over 40 mph. Police have setup three speed detection sensors on the road. From the physics of the sensor it is know that the measurement error of the sensor can be modelled as a Normal distribution with mean 0 and variance 25. The readings from the three sensors are averaged to give a single estimate(denoted by **K**) of the speed of the vehicle. Considering Null hypothesis as vehicle is not over speeding. Our task is to help the police in determining a threshold for **K** for issuing speeding ticket so that no more than 4 percent of the speeding tickets are given in error.

Soln:

Giving a ticket to a non-speeding vehicle is a type I error (rejecting H0 when it is true). Since we want maximum of 4 percent error and ticket is given if speed is more than 40 it implies we want one-sided rejection region.

Sampling distribution of **K** is N (40, 25/3) we are interested in alpha =0.04 from z table we get it to be for x=1.7507.

So threshold is $40 + z_{0.04} * (\frac{5}{\sqrt{3}})$ = 45.054.

It means that if the averaged output from sensor is more than 45.05 we can give a speeding ticket and still error is bound to be less than 4%.

Q3. An article in the Food Technology Journal (1956, Vol. 10, pp. 39–42) described a study on the protopectin content of tomatoes during storage. Four storage times were selected, and samples from nine lots of tomatoes were analyzed. The protopectin content (expressed as hydrochloric acid soluble fraction mg/kg) is in the table below. The researchers in this study hypothesized that mean protopectin content would be different at different storage times. Perform a test at α = 0.05 to confirm this hypothesis. Which specific storage times are different? Would you agree with the statement that protopectin content decreases as storage time increases?

	Lot								
Time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 days	1694.0	989.0	917.3	346.1	1260.0	965.6	1123.0	1106.0	1116.0
7 days	1802.0	1074.0	278.8	1375.0	544.0	672.2	818.0	406.8	461.6
14 days	1568.0	646.2	1820.0	1150.0	983.7	395.3	422.3	420.0	409.5
21 days	415.5	845.4	377.6	279.4	447.8	272.1	394.1	356.4	351.2

$$H_0: F \approx 1$$
$$H_1: F \neq 1$$

$$\sum X^2 = 30168976\,, \sum \frac{T^2}{n} = 24541327\,, \frac{G^2}{N} = 22568675$$

$$SSTr = \sum \frac{T^2}{n} - \frac{G^2}{N} = 1972651.7\,, dof_{Tr} = I - 1 = 3$$

$$SSE = \sum X^2 - \sum \frac{T^2}{n} = 5627648.6\,, dof_E = N - I = 32$$

$$MSTr = \frac{SSTr}{dof_{Tr}} = 657550.56$$

$$MSE = \frac{SSE}{dof_E} = 175864.02$$

$$F = \frac{MSTr}{MSE} = 3.739$$

For a numerator dof of 3 and a denominator dof of 32 → $F_{crit} = 2.90$

$$F > F_{crit} \rightarrow \ reject\ H_0$$

For four groups and a denominator dof of 32 → $q = 3.832$

$$HSD = q\sqrt{\frac{MSE}{n}} = 535.66$$

$$\overline{X_0} = \frac{T_0}{n} = 1057.44\ , \overline{X_7} = \frac{T_7}{n} = 825.82\ , \overline{X_{14}} = \frac{T_{14}}{n} = 868.33\ , \overline{X_{21}} = \frac{T_{21}}{n} = 415.5$$

$$|\mu_0 - \mu_7| = 231.62 < HSD$$

$$|\mu_0 - \mu_{14}| = 189.11 < HSD$$

$$|\mu_0 - \mu_{21}| = 641.94 > HSD$$

$$|\mu_7 - \mu_{14}| = 42.51 < HSD$$

$$|\mu_7 - \mu_{21}| = 410.32 < HSD$$

$$|\mu_{14} - \mu_{21}| = 452.83 < HSD$$

The protopectin content is significantly different between 0 days and 21 days. It can only be concluded that the protopectin content decreases between 0 and 21 days.