COMP 233 Probability and Statistics for Computer Science

Fall 2017, Assignment 4 Due: December 4, 2017

Question 1 The weights of salmon grown at a commercial hatchery are normally distributed with a standard deviation of 1.2 pounds. The hatchery claims that the mean weight of this year's crop is at least 7.6 pounds. Suppose a random sample of 16 fish yielded an average weight of 7.2 pounds. Is this strong enough evidence to reject the hatchery's claims at the

a) 5 percent level of significance;
b) 1 percent level of significance?
c) What is the p-value?

Solution.

a) From the question, we see that n = 16 and $\sigma = 1.2$. If the null hypothesis that
$$H_0 : \mu \geq 7.6$$
is true, then the distribution of
$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 7.6}{1.2/\sqrt{16}} \approx Z.$$
We notice that
$$-z_{\alpha_1} = -z_{0.05} \approx -1.645 < z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{(7.2 - 7.6)}{1.2/\sqrt{16}} \approx -1.33.$$
Hence, H_0 is accepted at $\alpha_1 = 0.05$.

b) We notice that
$$-z_{\alpha_2} = -z_{0.01} \approx -2.33 < z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{(7.2 - 7.6)}{1.2/\sqrt{16}} \approx -1.33.$$
Hence, H_0 is accepted at $\alpha_2 = 0.01$.

c) Further, the p-value of the test is
$$p - value = P\{Z \leq z_0\} = P\{Z \leq -1.33\} = 1 - P\{Z \leq 1.33\} \approx 1 - 0.9088 = 0.0912.$$
Hence, the p-value is larger than both 0.05 (accept) and 0.01 (accept).

Question 2 The lifetime of special light-bulbs is normally distributed. A sample of 81 bulbs has produced a mean of 738 hours and a standard deviation of 38.2 hours.

Test a hypothesis $H_0 : \mu = 747.5$ versus $H_1 : \mu \neq 747.5$, at the significance levels of $\alpha_1 = 0.05$ and $\alpha_2 = 0.01$. Find the p-value of the test, and use the p-value to verify your answers.

Solution.

We notice that n = 81, so we can assume by the Central Limit Theorem that $\sigma \approx S = 38.2$. If the null hypothesis is true, then the distribution of
$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 747.5}{38.2/\sqrt{81}} \approx Z.$$
We notice that
$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{(738 - 747.5)}{38.2/9} = -2.24 < z_{\alpha_1/2} = -z_{0.025} = -1.96\,.$$

Hence, H_0 is rejected at $\alpha = 0.05$ (since 2.24 > 1.96).
We also notice that

$$z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{(738 - 747.5)}{38.2/9} = -2.24 > z_{\alpha_2/2} = -z_{0.005} = -2.575 .$$

Hence, H_0 is accepted at α = 0.01 (since 2.575 > 2.24).
Further, the p-value of the test is

$$p - value = P\{|Z| > |z_0|\} = P\{|Z| > |-2.24|\} = 2(1 - P\{Z \le 2.24\})$$
$$\approx 2(1 - 0.9875) = 0.025.$$

Hence, the p-value is smaller than 0.05 (reject) and greater than 0.01 (accept).

Question 3 Twenty years ago, entering male high school students of Central High could do an average of 24 pushups in 60 seconds. To see whether this remains true today, a random sample of 36 freshmen was chosen. If their sample mean was 22.5 with a sample standard deviation of 3.1, can we conclude that the mean is no longer equal to 24? Use the 5 percent level of significance.

Solution.

With T_{35} being a *t*-random variable with 35 degrees of freedom, we have the following p-value:

$$p - value = P\left\{|T_{35}| \ge \frac{|\bar{X} - \mu_0|}{S/\sqrt{n}}\right\}$$
$$= P\{|T_{35}| \ge 6|22.5 - 24|/3.1\}$$
$$= 2P\{T_{35} \ge 2.903\} = .0064$$

Hence, the hypothesis that $\mu = 24$ is rejected at the significance level of $\alpha = 0.05$ and we conclude the mean is no longer equal to 24.

Question 4 A gun-like apparatus has recently been designed to replace needles in administering vaccines. The apparatus can be set to inject different amounts of the serum, but because of random fluctuations the actual amount injected is normally distributed with a mean equal to the setting and with an unknown variance σ^2. It has been decided that the apparatus would be too dangerous to use if σ exceeds .10. If a random sample of 50 injections resulted in a sample standard deviation of .08, should use of the new apparatus be discontinued? Suppose the level of significance is $\alpha = .10$. Comment on the appropriate choice of a significance level for this problem, as well as the appropriate choice of the null hypothesis.

Solution.

Test the null hypothesis that:

$$H_0 : \sigma \ge 0.1$$

The value of the test statistic is

$$\chi_0^2 = \frac{(n-1)S^2}{(\sigma_0)^2} = 49(0.08)^2/(0.1)^2 = 31.36$$

So that

$$p - value = P\{\chi_{49}^2 < 31.36\} < 0.023.$$

Hence, the hypothesis that $\sigma \geq 0.1$ is rejected at the significance level of $\alpha = 0.10$ and the apparatus can be utilized. To accept this null hypothesis, the significance level would have to be set at ≤ 0.023 .

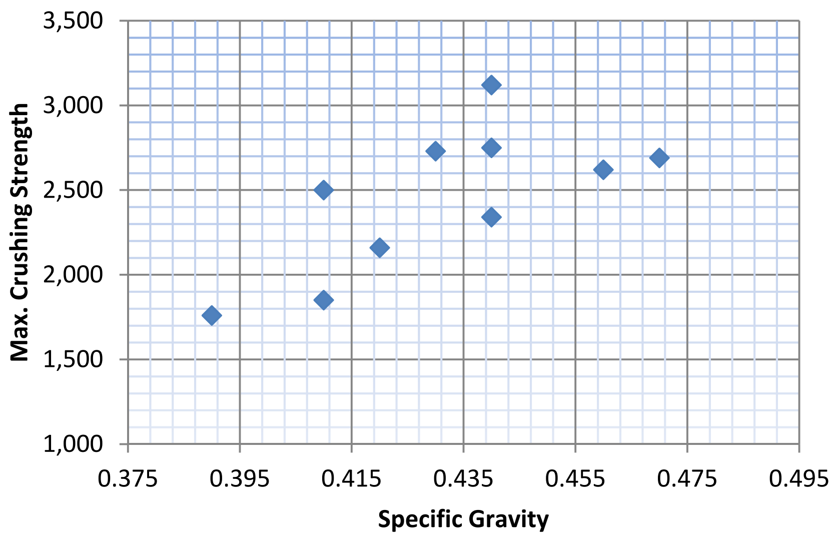
Question 5 The following data indicate the relationship between x, the specific gravity of a wood sample, and Y , its maximum crushing strength in compression parallel to the grain.

x_i	y_i (psi)	x_i	y_i (psi)
.41	1,850	.39	1,760
.46	2,620	.41	2,500
.44	2,340	.44	2,750
.47	2,690	.43	2,730
.42	2,160	.44	3,120

a) Plot a scatter diagram. Does a linear relationship seem reasonable?
b) Estimate the regression coefficients.
c) Predict the maximum crushing strength of a wood sample whose specific gravity is .43.

Solution.

a) A plot would look like:



So a linear relationship seems reasonable.

b) Determining the least squares estimators:

$$n = 10, \quad \bar{x} = 0.431, \quad \sum_{i=1}^{n} Y_i = 24520, \quad \overline{Y} = 2452,$$

$$\sum_{i=1}^{n} x_i Y_i = 10632.9, \quad \sum_{i=1}^{n} x_i^2 = 1.8629.$$

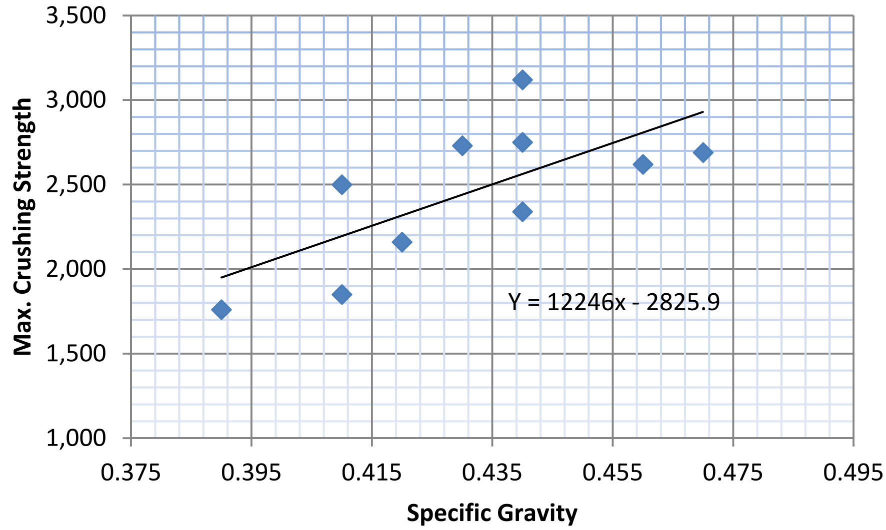
such that

$$B = \frac{\sum_{i=1}^{n} x_i Y_i - \bar{x}\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{10632.9 - 0.431(24520)}{1.8629 - 10(0.431)^2} = 12245.75$$

$$A = \bar{Y} - B\bar{x} = 2452 - 12245.75(0.431) = -2825.92$$

giving the following estimated regression line:

$$Y = -2825.92 + 12245.75x$$



c) The estimated response at $x_0 = .43$ is

$$Y(x_0) = -2825.92 + 12245.75x_0 = 2439.8.$$

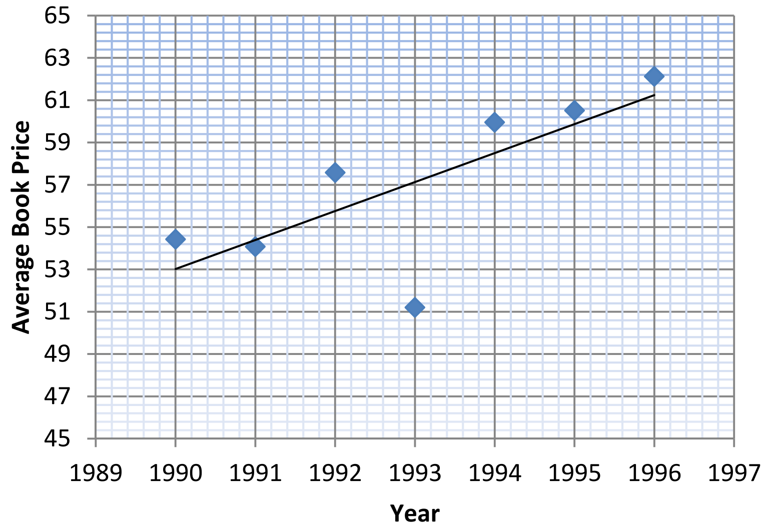
Question 6 The following data give the average price of all books reviewed in the journal *Science* in the years from 1990 to 1996.

Year	1990	1991	1992	1993	1994	1995	1996
Price (Dollars)	54.43	54.08	57.58	51.21	59.96	60.52	62.13

Give an interval that, with 95 percent confidence, will contain the average price of all books reviewed in *Science* in 1997 and in 1998.

Solution.

A scatter diagram looks like:



A 95 percent confidence interval estimate for the expected average prices of all books reviewed in Science in $x_0 = 1997$ and $x_0 = 1998$ requires computing the confidence interval for $Y(x_0)$.
With $100(1 - a)$ percent confidence, $Y(x_0)$ will lie within (the prediction interval)

$$A + Bx_0 \pm \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \sqrt{\frac{SS_R}{n-2}} t_{a/2,n-2}$$

Therefore, we must first compute A and B. Determining the least squares estimators:

$$n = 7, \qquad \bar{x} = 1993, \qquad \sum_{i=1}^{n} Y_i = 399.91, \qquad \bar{Y} = 57.13,$$

$$\sum_{i=1}^{n} x_i Y_i = 797059, \qquad \sum_{i=1}^{n} x_i^2 = 27804371.$$

such that

$$B = \frac{\sum_{i=1}^{n} x_i Y_i - \bar{x} \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{797059 - 1993(399.91)}{27804371 - 7(1993)^2} = 1.370$$

$$A = \bar{Y} - B\bar{x} = 57.13 - 1.370(1993) = -2673.28$$

Also, we need

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \left(\sum_{i=1}^{n} x_i^2 \right) - n\bar{x}^2 = 27804371 - 7(1993)^2 = 28\,.$$

and

$$SS_R = \sum_{i=1}^{n} (Y_i - A - Bx_i)^2 \approx 43.789\,.$$

For a 95 percent confidence interval estimate, a = 0.05, then we have, for $x_0 = 1997$,

$$A + Bx_0 \pm \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \sqrt{\frac{SS_R}{n-2}} t_{a/2,n-2}$$

$$\approx -2673.3 + 1.370(1997) \pm \sqrt{1 + \frac{1}{7} + \frac{(1997 - 1993)^2}{28}} \sqrt{\frac{43.789}{5}} t_{0.025,5}$$
$$\approx 62.61 \pm (3.875)2.571$$
$$= 62.61 \pm 9.962$$
$$= (52.648, 72.572).$$

For a 95 percent confidence interval estimate, a = 0.05, then we have, for $x_0 = 1998$,

$$A + Bx_0 \pm \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \sqrt{\frac{SS_R}{n-2}} t_{a/2,n-2}$$

$$\approx -2673.3 + 1.370(1998) \pm \sqrt{1 + \frac{1}{7} + \frac{(1998 - 1993)^2}{28}} \sqrt{\frac{43.789}{5}} t_{0.025,5}$$
$$\approx 63.98 \pm (3.875)2.571$$
$$= 63.98 \pm 10.856$$
$$= (53.124, 74.836).$$

Question 7 A new drug was tested on mice to determine its effectiveness in reducing cancerous tumors. Tests were run on 10 mice, each having a tumor of size 4 grams, by varying the amount of the drug used and then determining the resulting reduction in the weight of the tumor. The data were as follows:

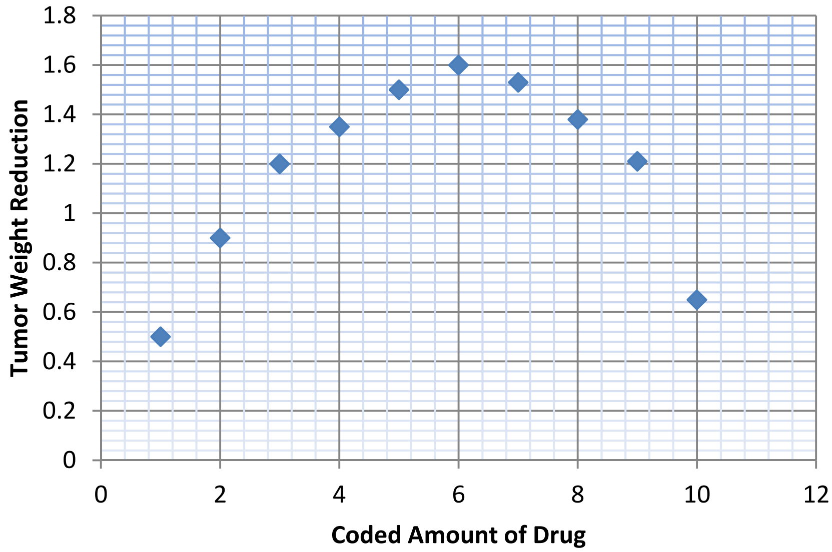
Coded Amount of Drug	Tumor Weight Reduction
1	.50
2	.90
3	1.20
4	1.35
5	1.50
6	1.60
7	1.53
8	1.38
9	1.21
10	.65

Estimate the maximum expected tumor reduction and the amount of the drug that attains it by fitting a quadratic regression equation of the form

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + e$$

Solution.

The scatter diagram is:



To determine the least squares estimators $B_i, i = 0,1,2$, for β_i:

$$\sum_{i=1}^{n} Y_i = B_0 n + B_1 \sum_{i=1}^{n} x_i + B_2 \sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n} x_i Y_i = B_0 \sum_{i=1}^{n} x_i + B_1 \sum_{i=1}^{n} x_i^2 + B_2 \sum_{i=1}^{n} x_i^3$$

$$\sum_{i=1}^{n} x_i^2 Y_i = B_0 \sum_{i=1}^{n} x_i^2 + B_1 \sum_{i=1}^{n} x_i^3 + B_2 \sum_{i=1}^{n} x_i^4$$

where

$$\sum_{i=1}^{n} x_i = 55, \quad \sum_{i=1}^{n} x_i^2 = 385, \quad \sum_{i=1}^{n} x_i^3 = 3025, \quad \sum_{i=1}^{n} x_i^4 = 25{,}333,$$

$$\sum_{i=1}^{n} Y_i = 11.82, \quad \sum_{i=1}^{n} x_i Y_i = 67.54, \quad \sum_{i=1}^{n} x_i^2 Y_i = 457.9$$

giving

$$11.82 = 10B_0 + 55B_1 + 385B_2$$
$$67.54 = 55B_0 + 385B_1 + 3{,}025B_2$$
$$457.9 = 385B_0 + 3{,}025B_1 + 25{,}333B_2$$

Solving this system of equations (using the brute force approach):

The first equation gives

$$B_0 = 1.182 - 5.5B_1 - 38.5B_2$$

Substituting this into the second equation gives

$$67.54 = 55(1.182 - 5.5B_1 - 38.5B_2) + 385B_1 + 3{,}025B_2$$

or

$$2.54 = 82.5B_1 + 907.5B_2$$

Then

$$B_1 \approx 0.03079 - 11B_2$$

and (plugging this into the above equation for B_0)

$B_0 \approx 1.182 - 5.5(0.03079 - 11B_2) - 38.5B_2 \approx 1.01267 + 22B_2$

Now, plugging these two equations into the third equation of the system of equations:

$$457.9 \approx 385(1.01267 + 22B_2) + 3{,}025(0.03079 - 11B_2) + 25{,}333B_2$$
$$\Rightarrow -25.11 \approx 528B_2$$
$$\Rightarrow B_2 \approx -0.0476$$

Then

$$B_1 \approx 0.03079 - 11(-0.0476) \approx 0.554$$

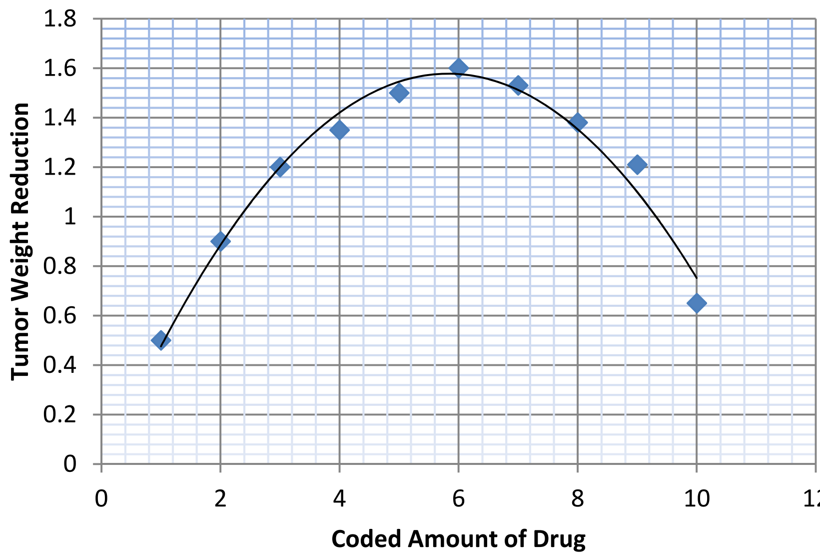
and

$$B_0 \approx 1.01267 + 22(-0.0476) \approx -0.0345$$

So, least squares quadratic regression equation for this data set is (approximately)

$$Y = -0.0345 + 0.554x - 0.0476x^2$$

which looks something like:



The maximum occurs where the tangent of the line is 0, i.e., $dY/dx = 0$. Therefore,

$$\frac{dY}{dx} = 0.554 - (2)0.0476x = 0$$
$$\Rightarrow x = \frac{0.554}{0.0952} \approx 5.82$$

Therefore, the maximum expected tumor reduction occurs when the amount of drug is 5.82.