ISE 3292/5013 -140

Assignment #3 Guide

Due June 6, 2016

1.

MS 5.54 - Page 215

5.54 Phishing attacks to email accounts. Refer to the *Chance* (Summer, 2007) article on phishing attacks at a company, Exercise 2.24 (p. 38). Recall that *phishing* describes an attempt to extract personal/financial information through fraudulent email. The company set up a publicized email account—called a "fraud box"—which enabled employees to notify them if they suspected an email phishing attack. If there is minimal or no collaboration or collusion from within the company, the interarrival times (i.e., the time between successive email notifications, in seconds) have an approximate **exponential distribution with a mean of 95 seconds**.

a. What is the probability of observing an interarrival time of at least 2 minutes? (The final answer is 0.2826)

b. Data for a sample of 267 interarrival times are saved in the **PHISHING** file. Do the data appear to follow an exponential distribution with β=95? (Compare the sample mean and standard deviation with the mean and standard deviation if data follow an exponential distribution and then draw a conclusion)

Guide:

The Exponential Probability Distribution

An **exponential distribution** is a gamma density function with $\alpha = 1$:

$$f(y) = \frac{e^{-y/\beta}}{\beta} \qquad (0 \le y < \infty)$$

with mean and variance

$$\mu = \beta \qquad \sigma^2 = \beta^2$$

2.

MS 5.56 - Page 215

5.56 Flood level analysis. Researchers have discovered that the maximum flood level (in millions of cubic feet per second) over a 4-year period for the Susquehanna River at Harrisburg, Pennsylvania, follows approximately a **gamma distribution** with **α=3** and **β=0.07** (*Journal of Quality Technology*, Jan.

1986).

a. Find the mean and variance of the maximum flood level over a 4-year period for the Susquehanna River. (Mean is 0.21 and variance is 0.0147)

b. The researchers arrived at their conclusions about the maximum flood level distribution by observing maximum flood levels over 4-year periods, beginning in 1890. Suppose that over the next 4-year period the maximum flood level was observed to be .60 million cubic feet per second. Would you expect to observe a value this high from a gamma distribution with α=3 and β=0.07? What can you infer about the maximum flood level distribution for the 4-year period observed? (Hint: Calculate $\mu \pm 3\sigma$ then make conclusion)

Guide:

The Gamma Probability Distribution

The probability density function for a gamma-type random variable Y is given by

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}e^{-y/\beta}}{\beta^{\alpha}\Gamma(\alpha)} & \text{if } 0 \le y < \infty;\ \alpha > 0;\ \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1}e^{-y}\,dy$$

The mean and variance of a gamma-type random variable are, respectively,

$$\mu = \alpha\beta \qquad \sigma^2 = \alpha\beta^2$$

3.

MS 5.60 - Page 216

5.60 Reaction to tear gas. The length of time Y (in minutes) required to generate a human reaction to tear gas formula **A** has **a gamma distribution** with **α=2** and **β=2**. The distribution for formula **B** is also **gamma**, but with **α=1** and **β=4**.

a. Find the mean length of time required to generate a human reaction to tear gas formula A. Find the mean for formula B.(The final answers: $\mu_A = 4\ and\ \mu_B = 4$)

b. Find the variances for both distributions. (The final answers: $\sigma^2{}_A = 8\ and\ \sigma^2{}_B = 16$)

c. Which tear gas has a higher probability of generating a human reaction in less than 1 minute? (*Hint*: You may use the fact that (Calculate P(Y< 1) for both A and B then compare them)

$$\int ye^{-y/2}\,dy = -2ye^{-y/2} + \int 2e^{-y/2}dy$$

4. MS

5.74 - Page 219

5.74 Washing machine repair time. Based on extensive testing, a manufacturer of washing machines believes that the distribution of the time Y (in years) until a major repair is required has a **Weibull distribution** with **α=2** and **β=4**.

a. If the manufacturer guarantees all machines against a major repair for 2 years, what proportion of all new washers will have to be repaired under the guarantee? (The final answer is 0.63212)

b. Find the mean and standard deviation of the length of time until a major repair is required. (Mean = 1.77246 and standard deviation is 0.9265)

c. Find

$$P(\mu - 2\sigma \le Y \le \mu + 2\sigma).$$

(The final answer is (-0.08054, 3.62546))

d. Is it likely that Y will exceed 6 years?

Guide:

The Weibull Probability Distribution

The probability density function for a Weibull random variable, Y is given by

$$f(y) = \begin{cases} \frac{\alpha}{\beta} y^{\alpha-1} e^{-y^{\alpha}/\beta} & \text{if } 0 \le y < \infty;\ \alpha > 0;\ \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$\mu = \beta^{1/\alpha}\Gamma\left(\frac{\alpha+1}{\alpha}\right)$$

$$\sigma^2 = \beta^{2/\alpha}\left[\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \Gamma^2\left(\frac{\alpha+1}{\alpha}\right)\right]$$

From example 5.15 page 218

The cumulative distribution function for a Weibull distribution is

$$F(y_0) = \int_0^{y_0} f(y)\,dy = \int_0^{y_0} \frac{\alpha}{\beta} y^{\alpha-1} e^{-y^{\alpha}/\beta} dy$$

5.

MS 5.84 - Page 223

5.84 Laser color printer repairs. The proportion *Y* of a data processing company's yearly hardware repair budget allocated to repair its laser color printer has an approximate **beta distribution** with parameters **α=2** and **β=9**.

a. Find the mean and variance of *Y*. (The final answers: mean = 0.18182 and variance= 0.01240)

b. Compute the probability that for any randomly selected year, **at least 40%** of the hardware repair budget is used to repair the laser color printer. (The final answer is 0.0464)

Hint:

From Section 5.9, for cases where α and β are integers,

$P(Y \le p) = F(p) = \sum_{y=\alpha}^{n} p(y)$ where $p(y)$ is a binomial probability distribution with parameters p and $n = (\alpha + \beta - 1)$. Thus, for $n = (\alpha + \beta - 1) = 2 + 9 - 1 = 10,\ p = 0.40$,

$$P(Y > 0.40) = 1 - P(Y \le 0.40)$$

c. What is the probability that **at most 10%** of the yearly repair budget is used for the laser color printer? (The final answer is 0.2639) Hint: For $n = (\alpha + \beta - 1) = 2 + 9 - 1 = 10,\ p = 0.10$ $P(Y \le 0.10) = F(0.10)$

Guide:

The Beta Probability Distribution

The probability density function for a beta-type random variable *Y* is given by

$$f(y) = \begin{cases} \dfrac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} & \text{if } 0 \le y \le 1; \alpha > 0; \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

where

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

The mean and variance of a beta random variable are, respectively,

$$\mu = \frac{\alpha}{\alpha+\beta} \qquad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

6.

MS 5.114 - Page 232

5.114 Lifelengths of memory chips. The lifelength Y (in years) of a memory chip in a laptop computer is a Weibull random variable with probability density

$$f(y) = \begin{cases} \frac{1}{8} y e^{-y^2/16} & \text{if } 0 \le y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

a. What are the values of α and β?

b. Compute the mean and variance of Y. (Mean=3.545 and variance= 3.4335)

c. Find the probability that a new memory chip will not fail before 6 years.(The final answer is 0.1054)

Hint: $P(Y \ge 6) = 1 - P(Y < 6) = 1 - F(6)$

Guide:

The Weibull Probability Distribution

The probability density function for a Weibull random variable, Y is given by

$$f(y) = \begin{cases} \frac{\alpha}{\beta} y^{\alpha-1} e^{-y^\alpha/\beta} & \text{if } 0 \le y < \infty; \alpha > 0; \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$\mu = \beta^{1/\alpha} \Gamma\left(\frac{\alpha + 1}{\alpha}\right)$$

$$\sigma^2 = \beta^{2/\alpha}\left[\Gamma\left(\frac{\alpha + 2}{\alpha}\right) - \Gamma^2\left(\frac{\alpha + 1}{\alpha}\right)\right]$$

The cumulative distribution function for a Weibull distribution is

$$F(y_0) = \int_0^{y_0} f(y)\, dy = \int_0^{y_0} \frac{\alpha}{\beta} y^{\alpha-1} e^{-y^\alpha/\beta} dy$$

7.

MS 6.2 - Page 239

6.2 Tossing dice. Consider the experiment of tossing a pair of dice. Let ***X*** be the outcome (i.e., the number of dots appearing face up) on the **first die** and let ***Y*** be the outcome on the **second die**.

a. Find the joint probability distribution $p(x, y)$.

b. Find the marginal probability distributions $p1(x)$ and $p2(y)$. (The final answer:1/6 and 1/6)

c. Find the conditional probability distributions $P_1(x|y)$ and $P_2(y|x)$.

d. Compare the probability distributions of parts **b** and **c**. What phenomenon have you observed? (Make conclusion that x and y are independent or not)

Guide:

For part a We know $1 \le X \le 6$ and $1 \le Y \le 6$ also 36 outcomes is equally likely.

For part b

$$P(X=1) = p_1(1) = p(1,1) + p(1,2) + p(1,3) + p(1,4) + p(1,5) + p(1,6)$$

For part c

$$\text{The conditional probability distribution of } X \text{ given } Y \text{ is } p_1(x \mid y) = \frac{p(x,y)}{p_2(y)}$$

$$p_2(y \mid x) = \frac{p(x,y)}{p_1(x)}$$

8.

MS 6.4 - Page 240

6.4 Modeling the behavior of granular media. Refer to the *Engineering Computations: International Journal for Computer-Aided Engineering and Software* (Vol. 30, No. 2, 2013) study of the properties of granular media (e.g., sand, rice, ball bearings, and flour), Exercise 3.62 (p. 120). The study assumes there is a system of ***N*** non-interacting granular particles, where the particles are grouped according to energy level, *r*. For this problem (as in Exercise 3.62), assume that N=7 and r=3 then consider the scenario where there is **one** particle (of the total of 7 particles) at energy level 1, **two** particles at energy level 2, and **four** particles at energy level 3. Another feature of the particles studied was the position in time where the particle reached a certain entropy level during compression. All particles reached the desired entropy

level at one of three time periods, 1, 2, or 3. Assume the **7** particles had the characteristics shown in the table. Consider a randomly selected particle and let ***X* represent the energy level** and ***Y* the time period associated with particle.**

Particle ID	Energy Level	Time Period
1	3	1
2	1	1
3	3	3
4	2	1
5	3	2
6	3	2
7	2	1

a. Find the bivariate probability distribution, P(x,y)

b. Find the marginal distribution, $P_1(x)$.

c. Find the marginal distribution, $P_2(y)$.

d. Find the conditional distribution, $P_2(y|x)$.

Guide:

Definition 6.6

Let *f*(*x*, *y*) be the joint density function for *X* and *Y*. Then the **conditional density functions** for *X* and *Y* are

$$f_1(x \mid y) = \frac{f(x, y)}{f_2(y)} \quad \text{and} \quad f_2(y \mid x) = \frac{f(x, y)}{f_1(x)}$$

Guide:

a) The bivariate probability distribution $p\,x\,y\,(\,,\,)$ in table form is:

		x		
		1	2	3
y	1	1/7	2/7	1/7
	2	0	0	2/7
	3	0	0	1/7

B and c) After calculating you will have these tables:

In table form, the marginal probability distribution $p_1(x)$ is:

x	1	2	3
$p_1(x)$	1/7	2/7	4/7

In table form, the marginal distribution $p_2(y)$ is given as:

y	1	2	3
$p_2(y)$	4/7	2/7	1/7

d)

The final answers are: (You must show your work)

The conditional probability distribution of Y given $X=1$ is given in the table:

y	1	2	3
$p_2(y\,|\,1)$	1	0	0

The conditional probability distribution of Y given $X=2$ is given in the table:

y	1	2	3
$p_2(y\,|\,2)$	1	0	0

The conditional probability distribution of Y given $X=3$ is given in the table:

y	1	2	3
$p_2(y\,|\,3)$	1/4	2/4	1/4

9.

MS 6.12 - Page 244

6.12 Distribution of low bids. The Department of Transportation (DOT) monitors sealed bids for new road construction. For new access roads in a certain state, let low bid (thousands of dollars) and let estimate of fair cost of building the road (thousands of dollars). The joint probability density of X and Y is

$$f(x, y) = \frac{e^{-y/10}}{10y}, \qquad 0 < y < x < 2y$$

a. Find $f(y)$, the marginal density function for Y. Do you recognize this distribution?

b. What is the mean DOT estimate, $E(Y)$?

Guide:

Definition 6.5

Let $f(x, y)$ be the joint density function for X and Y. Then the **marginal density functions** *for X and Y are*

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{and} \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

10.

MS 6.14 - Page 245

6.14 Servicing an automobile. The joint density of X, the total time (in minutes) between an automobile's arrival in the service queue and its leaving the system after servicing, and Y, the time (in minutes) the car waits in the queue before being serviced, is

$$f(x, y) = \begin{cases} ce^{-x^2} & \text{if } 0 \le y \le x; 0 \le x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

a. Find the value of c that makes $f(x, y)$ a probability density function. (The final answer is 2)

Hint:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1$$

b. Find the marginal density for X and show that

$$\int_{-\infty}^{\infty} f_1(x)\,dx = 1$$

c. Show that the conditional density for Y given X is a uniform distribution over the interval $0 \le Y \le X$.

Hint:

$$f_2(y|x) = \frac{f(x,y)}{f_1(x)}$$

After calculation the final answer is 1/x.

11.

MS 6.51 - Page 253

6.51 As an illustration of why the converse of Theorem 6.6 is not true, consider the joint distribution of two discrete random variables, X and Y, shown in the accompanying table. Show that COV(X,Y)=0 , but that X and Y are dependent.

		$X = x$		
		-1	0	$+1$
$Y = y$	-1	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{1}{12}$
	0	$\frac{2}{12}$	0	$\frac{2}{12}$
	$+1$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{1}{12}$

Guide:

THEOREM 6.6

If two random variables X and Y are independent, then

$$\text{Cov}(X, Y) = 0$$

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$E(XY)=\sum_x\sum_y xyp(x,y)$$

To find $E(X)$ and $E(Y)$, we must find the marginal distributions of X and Y.

$$E(X)=\sum_x xp_1(x)$$

$$E(Y)=\sum_y yp_1(y)$$

To show that X and Y are not independent, we must show that $p(x,y)\neq p_1(x)p_2(y)$

12.

MS 6.74 - Page 269

6.74 Uranium in the Earth's crust. Refer to the *American Mineralogist* (October 2009) study of the evolution of uranium minerals in the Earth's crust, Exercise 5.17 (p. 199). Recall that researchers estimate that the trace amount of uranium Y in reservoirs follows a **uniform distribution ranging between 1 and 3** parts per million. In a random sample of n=60 reservoirs, let $\overline{Y}$ represent the sample mean amount of uranium.

a. Find $E(\overline{Y})$ and interpret its value (The final answer is 2)

b. Find Var $(\overline{Y})$. (The final answer is 0.00556)

c. Describe the shape of the sampling distribution of $\overline{Y}$. (Apply Central Limit Theorem to answer the question)

d. Find the probability that $\overline{Y}$ is between 1.5 ppm and 2.5 ppm. (The final answer is 1)

e. Find the probability that $\overline{Y}$ exceeds 2.2 ppm. (The final answer is 0.9963)

Guide:

THEOREM 6.9 The Central Limit Theorem

If a random sample of n observations, $Y_1, Y_2, \ldots, Y_n$, is drawn from a population with finite mean μ and variance σ^2, then, when n is sufficiently large, the sampling distribution of the sample mean $\overline{Y}$ can be approximated by a normal density function.

The sampling distribution of $\overline{Y}$, in addition to being approximately normal for large n, has other known characteristics, which are given in Definition 6.15.

13.

MS 6.90 - Page 273

6.90 Mercury contamination of swordfish. *Consumer Reports* found widespread contamination of seafood in New York and Chicago supermarkets. For example, 40% of the swordfish pieces available for sale have a level of mercury above the Food and Drug Administration (FDA) limit. Consider a random sample of 20 swordfish pieces from New York and Chicago supermarkets.

a. Use the normal approximation to the binomial to calculate the probability that fewer than 2 of the 20 swordfish pieces have mercury levels exceeding the FDA limit. (The final answer is 0.0015)

b. Use the normal approximation to the binomial to calculate the probability that more than half of the 20 swordfish pieces have mercury levels exceeding the FDA limit. (The final answer is 0.1271)

c. Use the binomial tables to calculate the exact probabilities in parts **a** and **b**. Does the normal distribution provide a good approximation to the binomial distribution? (0.0005 and 0.1275)

Guide:

Continuity Correction for the Normal Approximation to a Binomial Probability

Let Y be a binomial random variable with parameters n and p, and let Z be a standard random variable. Then,

$$P(Y \le a) \approx P\left(Z < \frac{(a + .5) - np}{\sqrt{npq}}\right)$$

$$P(Y \ge a) \approx P\left(Z > \frac{(a - .5) - np}{\sqrt{npq}}\right)$$

$$P(a \le Y \le b) \approx P\left(\frac{(a - .5) - np}{\sqrt{npq}} < Z < \frac{(b + .5) - np}{\sqrt{npq}}\right)$$

$$Y \sim B(n,p) \sim B(20,0.40);$$

14.

MS 7.108 - Page 362

7.108 Lead and copper in drinking water. Periodically, the Hillsborough County (Florida) Water Department tests the drinking water of homeowners for contaminants such as lead and copper. The lead

and copper levels in water specimens collected for a sample of **10** residents of the Crystal Lake Manors subdivision are shown next.

LEADCOPP

Lead (μg/L)	Copper (mg/L)
1.32	.508
0	.279
13.1	.320
.919	.904
.657	.221
3.0	.283
1.32	.475
4.09	.130
4.45	.220
0	.743

a. Construct a **99%** confidence interval for the **mean lead** level in water specimens from Crystal Lake Manors. (-1.15, 6.92)

b. Construct a **99%** confidence interval for the **mean copper** level in water specimens from Crystal Lake Manors.(0.1519, 0.6647)

c. Interpret the intervals, parts **a** and **b,** in the words of the problem.

d. Discuss the meaning of the phrase, **"99% confident."**

15.

MS 7.114 - Page 364

7.114 Solar irradiation study. The *Journal of Environmental Engineering* (Feb. 1986) reported on a heat transfer model designed to predict winter heat loss in wastewater treatment clarifiers. The analysis involved a comparison of clear-sky solar irradiation for horizontal surfaces at different sites in the Midwest. The day-long solar irradiation levels (in BTU/sq. ft.) at two midwestern locations of different latitudes (St. Joseph, Missouri, and Iowa Great Lakes) were recorded on each of seven clear-sky winter days. The data are given in the table. Find a **95%** confidence interval for **the mean difference** between the day-long clear-sky solar irradiation levels at the two sites. Interpret the results. (156.84, 239.16)

SOLARAD

Date	St. Joseph, Mo.	Iowa Great Lakes
December 21	782	593
January 6	965	672
January 21	948	750
February 6	1,181	988
February 21	1,414	1,226
March 7	1,633	1,462
March 21	1,852	1,698

Guide:

$(1 - \alpha)100\%$ Confidence Interval for $\mu_d = (\mu_1 - \mu_2)$: Matched Pairs

Let $d_1, d_2, \ldots, d_n$ represent the differences between the pairwise observations in a *random sample* of n matched pairs, $\bar{d}$ = mean of the n sample differences, and s_d = standard deviation of the n sample differences.

Large Sample	Small Sample
$\bar{d} \pm z_{\alpha/2}\left(\frac{\sigma_d}{\sqrt{n}}\right)$	$\bar{d} \pm t_{\alpha/2}\left(\frac{s_d}{\sqrt{n}}\right)$
where σ_d is the population deviation of differences.	where $t_{\alpha/2}$ is based on $(n - 1)$ degrees of freedom.
Assumption: $n \geq 30$	*Assumption*: The population of paired differences is normally distributed.

[*Note:* When σ_d is unknown (as is usually the case), use s_d to approximate σ_d.]

In this problem n is small. Therefore, we use this formula:

$$\bar{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

16.

MS 7.116 - Page 364

7.116 Diazinon residue in orchards. Pesticides applied to an extensively grown crop can result in inadvertent areawide air contamination. *Environmental Science & Technology* (Oct. 1993) reported on air deposition residues of the insecticide diazinon used on dormant orchards in the San Joaquin Valley,

California. Ambient air samples were collected and analyzed at an orchard site for each of 11 days during the most intensive period of spraying. The levels of diazinon residue (in mg/m3) during the day and at night are recorded in the table. The researchers want to know whether the mean diazinon residue levels differ from day to night.

DIAZINON

	Diazinon Residue	
Date	Day	Night
Jan. 11	5.4	24.3
12	2.7	16.5
13	34.2	47.2
14	19.9	12.4
15	2.4	24.0
16	7.0	21.6
17	6.1	104.3
18	7.7	96.9
19	18.4	105.3
20	27.1	78.7
21	16.9	44.6

a. Analyze the data using a 90% confidence interval. (-58.894, -18.924)

b. What assumptions are necessary for the validity of the interval estimation procedure of part **a**?

c. Use the interval, part **a**, to answer the researchers' question.

Guide:

Let μ_1 = mean diazinon residue during the day and μ_2 = mean diazinon residue at night. Then $\mu_d = \mu_1 - \mu_2$ is the difference between the mean diazinon residue for day and night.

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

$$\bar{d} = \frac{\sum d_i}{n}$$

$$s_d^2 = \frac{\sum d_i^2 - \frac{\left(\sum d_i\right)^2}{n}}{n-1}$$