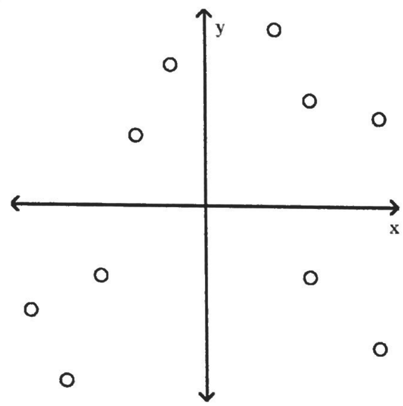
[Solutions]

MULTIPLE-CHOICE : 2 points each

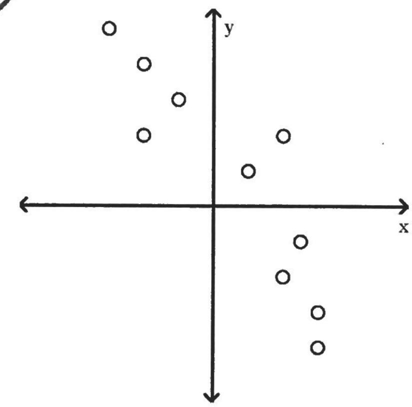
For the following multiple-choice questions, no work is required and no partial credit is possible.
Choose the one alternative that best completes the statement or answers the question and circle that answer choice.

1) Determine which plot shows the strongest linear correlation. - "more linear structure"?

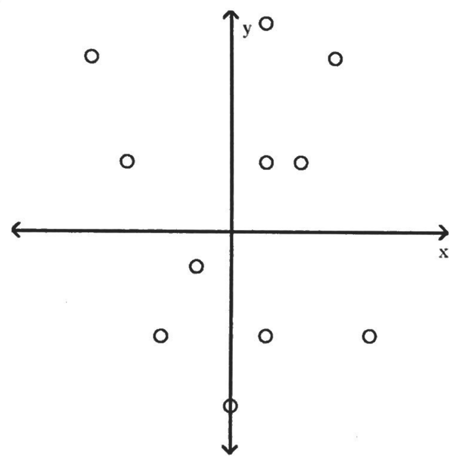
A)



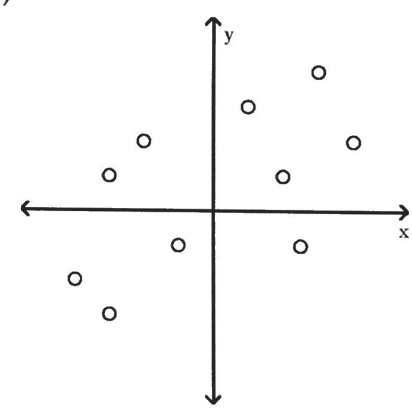
B)



C)



D)



2) Suppose you flip a fair quarter three times in a row. What is the probability the third flip is heads, given that the first two flips were both heads? - Each coin flip is independent.

A) 0.125 B) 0 C) 0.5 D) 0.25

3) A survey showed that 56% of adults are concerned about keeping their private information secure online. Using the binomial distribution, what is the probability that out of 11 adults, more than 9 are concerned about keeping their private information secure online. Binomial $n=11$ $p=0.56$

A) 0.9836 B) 0.0577 C) 0.074 D) 0.0164

$$P(x>9) = 1 - P(x \le 9)$$
$$= 1 - \text{binomcdf}(n=11, p=0.56, x=9)$$

4) A 10 week old german shepherd puppy weighs, on average, 13.8 pounds with a standard deviation of 1.4 pounds. Assuming the weight of german shepherd puppies are normally distributed, what is the probability a 10 week old german shepherd puppy selected at random weighs more than 12 pounds?

A) 0.0993 B) 0.0228 C) 0.9772 D) 0.9007

$P(x>12) = \text{normalcdf}(\underset{\text{lower}}{12}, \underset{\text{upper}}{1e(99)}, \underset{\mu}{13.8}, \underset{\sigma}{1.4})$

5) A random sample of 57 adults are observed for a week and the researchers record x, each subject's total amount of exercise that week (in hours). The researchers also record y, each subject's resting heart rate in beats per minute. If (−3.2, −1.5) is a 85% confidence interval for β, choose the correct interpretation of this interval.

A) For every additional beat per minute of their resting heart rate, we expect the subject's weekly exercise to decrease by 1.5 to 3.2 hours.

B) For every additional hour of weekly exercise, we expect the subject's resting hear rate to increase by 1.5 to 3.2 beats per minute.

C) For every additional beat per minute of their resting heart rate, we expect the subject's weekly exercise to increase by 1.5 to 3.2 hours.

D) For every additional hour of weekly exercise, we expect the subject's resting hear rate to decrease by 1.5 to 3.2 beats per minute.

Interpret slope, x↑1 unit then y changes acording to the slope which is between -3.2 and -1.5 w/ 85% confidence.

6) Suppose your friend computes the coefficient of determination $r^2 = 0.783$ for a data set of (x,y) pairs. Without having access to the scatter plot, or the equation of the least squares line, can you determine if x and y are positively correlated or negatively correlated?

A) Yes, x and y are positively correlated. Since r^2 is positive, there is a positive correlation between x and y.

B) No, we cannot determine the sign of the correlation. The coefficient of determination is not related to the correlation coefficient.

C) No, we cannot determine the sign of the correlation. The coefficient of determination only provides insight into what percentage of the variation in y is explained by the model.

Definition of r^2.

D) Yes, x and y are negatively correlated. If x and y are negatively correlated, then r^2 is positive.

7) What percentage of data is within 1.5 standard deviations of the mean in a normal distribution?

A) 86.6% B) 93.3% C) 13% D) 87%

$P(-1.5 < z < 1.5) = \text{normalcdf}(\underset{\text{lower}}{-1.5}, \underset{\text{upper}}{1.5}, \underset{\mu}{0}, \underset{\sigma}{1})$

8) Suppose a random variable x is distributed with mean μ and random samples of size n are taken. What further assumptions allow you to apply the central limit theorem and conclude the sampling distribution of $\bar{x}$ is normally distributed? Circle all that apply.

A) The distribution of x is normal. {If x is normal, no matter n, $\bar{x}$ will also be normal.

~~B) No additional assumptions need be made. Sampling distributions are always normal.~~ (Not true)

C) The sample sizes should be greater than, or equal to, 30. {If x distribution is unknown, but $n \geq 30$, $\bar{x}$ will be normal.

~~D) The population standard deviation σ must be known.~~ (Does not determine if $\bar{x}$ is normal.)

9) Suppose A and B are mutually exclusive events with P(A)=0.2 and P(B)=0.2. What is P(A and B)?

A) 0.4 B) 0.36 C) 0 D) 0.04

Definition of mutually exclusive.

10) Suppose A and B are independent events with P(A)=0.2 and P(B)=0.4. What is P(A or B)?

A) 0.6 B) 0.52 C) 0 D) 0.08

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
$$= P(A) + P(B) - P(A) \cdot P(B)$$
since A and B are independent
$$= 0.2 + 0.4 - (0.2 \cdot 0.4)$$
$$= 0.52$$

SHORT ANSWER: 5 points each:

For the following short-answer questions, both a correct answer and relevant correct work must be shown (or described) to receive full credit. Partial credit may be earned even if the correct answer is not found. However, an answer alone, even if correct, may result in no credit.

11) To test the effectiveness of a new drug designed to relieve flu symptoms, 261 patients were randomly selected and divided into two groups. One group of 127 patients was given a pill containing the drug while the other group of 134 was given a placebo. Surveys found that 67 of those actually taking the drug felt a beneficial effect, while 53 of the patients taking the placebo felt a beneficial effect.

Group1 Drug

Group2 Placebo

a.) (2 points) What statistical test should the researchers use to determine if the drug was effective at relieving flu symptoms? Why is this test appropriate?

2 PropZTest   Check: $x_1 = 67$ $x_2 = 53$ $n_1 = 127$ $n_2 = 134$ $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$

b.) (1 point) State the null and alternate hypotheses of the test. $n_1\bar{p} > 5$, $n_1\bar{q} > 5$, $n_2\bar{p} > 5$, $n_2\bar{q} > 5$

$H_0: p_1 = p_2$

$H_1: p_1 > p_2$ (Drug was effective i.e. drug prop > placebo prop.)

c.) (2 points) Using a 5% level of significance, what is an appropriate conclusion of the hyposthesis test. Phrase your response in the context of the problem.

P-value $= 0.016 \leq \alpha = 0.05$

At 5% level of sig. we reject H_0. There is statistically sig. evidence in support of the claim that the drug was effective.

12) The weight (in ounces) of 16 large bags of coffee that were randomly selected from a warehouse in Bogotá are listed below. Suppose the weight of the bags of coffee in the warehouse are approximately normally distributed.

L_1

61	85	92	77	83	81	75	78
95	87	69	74	76	84	80	83

a.) (2 points) Which is the more appropriate confidence interval, a t interval or a z interval? Give the reason for your answer.

T-Interval since weights of coffee bags are approx. normal, but we do not know pop. std. dev σ.

b.) (1 point) What is the 99% confidence interval for the true mean weight of all bags of coffee in this warehouse? Round your response to 2 decimal places.

(73.84, 86.16)

c.) (2 point) What is the interpretation of this interval?

With 99% confidence, the true pop. mean μ of coffee bag weight is between 73.84 and 86.16.

13) The number of golf balls ordered by customers of a pro shop has the following probability distribution.

x	P(x)
3	0.18
6	0.28
9	0.40
12	0.10
15	0.04

a.) Find the mean of the probability distribution. Round your answer to two decimal places.

X : L1
P(X) : L2 1 var stats(L1, Freq=L2) mean = 7.62

b.) Find the standard deviation of the probability distribution. Round your answer to two decimal places.

std. dev. = 3.07

14) A researcher wishes to determine whether there is a difference in the average salary of elementary school, high school, and community college teachers. Teachers are randomly selected. Their salaries (in thousands) are recorded below. Assume the salaries of each population is approximately normally distributed of equal standard deviations.

Compare means of 3 Lists => ANOVA

L1 Elementary Teachers	L2 High School Teachers	L3 Community College Teachers
51	51	59
51	52	61
51	53	65
52	51	57
52	53	62
50	52	59
Group 1	Group 2	Group 3

a.) (1 point) What is the null and alternate hypothesis of the researcher?

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1: Atleast one pop. mean is different.

b.) (2 points) Find the value of the test statistic and P-value of the relevant statistical test.

ANOVA(L1, L2, L3) : F = 51.817

P-value = 0.000000184

c.) (2 points) How should the researcher interpret the results of the test at a 1% level of significance? α

At 1% level of sig. reject Ho. There is statistically sig. evidence to say atleast one pop. mean is different.

15) You roll a die 72 times with the following results.

$df = 6 - 1 = 5$

	Number	1	2	3	4	5	6
L_1: Observed →	Frequency	13	9	14	16	13	7
L_2: Expected →		12	12	12	12	12	12

$\Rightarrow \frac{72}{6} = 12$

a.) (2 points) State the null and alternate hypothesis required to test the claim that the die is fair. χ^2 GOF Test

H_0: The distributions are the same. (Die is fair.)

H_1: The distributions are different. (Die not fair.)

b.) (3 points) Use a significance level of 0.01 to test the claim that the die is fair. Be sure to indicate which statistical test was implemented.

P-value $= 0.458 > \alpha = 0.01$

At 1% level of sig. we fail to reject H_0. There is not statistically sufficient evidence to conclude that the die isn't fair.

16) A random sample of 36 packages is collected from all the packages received by a parcel service. The standard deviation of the weight of a package is $\sigma = 3.10$. Suppose the sample mean is 30.2.

a.) (2 points) Which is the more appropriate confidence interval, a t interval or a z interval? Give the reason for your answer.

Z Interval since $n = 36 \geq 30$ and pop. std. dev. σ is known.

b.) (1 point) What is the 95% confidence interval for the true mean weight of all packages received by the parcel service? Round your response to 2 decimal places.

Z Int $\begin{cases} \bar{X} = 30.2 \\ \sigma = 3.10 \\ n = 36 \end{cases}$

$C = 0.95$

$(29.19, 31.21)$

c.) (2 points) What is the intereptation of this interval?

With 95% confidence, the pop. mean weight of the packages recieved is between 29.19 and 31.21.

SHORT ANSWER: 10 points each

For the following short–answer questions, both a correct answer and relevant correct work must be shown (or described) to receive full credit. Partial credit may be earned even if the correct answer is not found. However, an answer alone, even if correct, may result in no credit.For the following short–answer questions, both a correct answer and relevant correct work must be shown (or described) to receive full credit. Partial credit may be earned even if the correct answer is not found. However, an answer alone, even if correct, may result in no credit.

In the case where you use a calculator function to compute the answer, write the function with input as entered into your calculator and the specific output used from that function as the corresponding work.

17) A Boulder City councilwoman wishes to measure the average length of time that a Boulder resident has resided in the city. Previous studies suggest that this length is normally distributed with a population standard deviation of 11.85 months.

σ

a.) (3 points) In order to construct a 99% confidence interval for the true mean time with a margin of error no greater than 3 months, how large of a sample should be gathered?

$$n = \left(\frac{z_c \cdot \sigma}{E}\right)^2$$

$z_c = 2.5758$
$\sigma = 11.85$
$E = 3$

Round up

$n = 104$

b.) (2 points) A sample of 215 Boulder residents has a sample mean of time spent living in Boulder as 85 months. Construct a 99% confidence interval for the true mean time residing in the city.

ZInterval $\begin{cases} \sigma = 11.85 \\ \bar{x} = 85 \\ n = 215 \end{cases}$ $C = 0.99$

$(82.92, 87.08)$

c.) (5 points) Using the same same data as above, test the claim that the mean length of Boulder residency is less than 7.5 years at the 1% significance level. Be sure to indicate your null and alternate hypothesis, as well as phrasing your conclusion in terms of the mean time of residence in Boulder.

7.5 yrs = 90 months

$H_0: \mu = 90$
$H_1: \mu < 90$

ZTest $\begin{cases} \mu_0 = 90 \\ \sigma = 11.85 \\ \bar{x} = 85 \\ n = 215 \\ \mu < \mu_0 \end{cases}$ $\Rightarrow$ P-value

0.0000000003

At 1% level of sig. reject H_0. There is statistically sig. evidence supporting the claim that the mean length of Boulder residency is less than 7.5 years.

this is a population

18) A biologist measures the wingspan of every butterfly (all 119 in total) in the local butterfly house, to the nearest cm. Her measurements are recorded in the table below:

	Wingspan (cm)	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
L_1																	
L_2	Frequency	8	9	15	13	1	2	0	8	16	22	18	2	0	3	1	1

$\underbrace{8}_{8}$ $\underbrace{9, 15, 13, 1, 2, 0}_{40}$ $\underbrace{8, 16, 22, 18, 2, 0}_{66}$ $\underbrace{3, 1, 1}_{5}$

a.) (2 points) What are the mean and standard deviation of the wingspans?

1 var Stats(L_1, Freq=L_2) mean = 11.2521

std. dev. = σ_x = 3.8550

b.) (2 points) Give the 5-number summary for this data set.

min = 5, Q_1 = 7, med = 13, Q_3 = 14, max = 20

The distribution of wingspans of wild butterflies is as follows:

Range of Wingspans (cm)	0 -- 5	6 -- 11	12--17	18--23
Proportion of Butterflies	0.10	0.37	0.48	0.05

Considering the biologist's measurements as a sample of all captive butterflies, conduct a hypothesis test to determine if the wingspan of captive butterflies fits the distribution of wild butterfly wingspans.

c.) (2 points) State the null and alternate hypothesis.

Ho: The captive and wild butterfly wingspan dist. are the same.

H_1: The dist. are different.

d.) (2 points) State the appropriate statistical test and the P-value.

χ^2 GOF Test

P-value = 0.3648

Obs: 8, 40, 66, 5

Exp: 0.10(119), 0.37(119), 0.48(119), 0.05(119)

(Assume α = 5%)

e.) (2 points) What is the conclusion of the hypothesis test. Interpret your conclusion in terms of the distribution of wild and captive butterfly wingspans.

At 5% level of sig. fail to reject Ho. Not sufficient evidence that wingspans of wild vs captive butterflies are distributed differently.

19) Suppose you measure the weight, in grams, of snails found in three locations: the beach, the forest, and your garden. The results are recorded below:

L_1 (1) Beach	L_2 (2) Forest	L_3 (3) Garden
35	29	40
29	26	38
31	29	22
41	27	32
31		39
		39

a.) (2 points) Is this an observational study or an experiment? Explain your choice.

Observational, no control/treatment the measurements are from "nature".

b.) (1 point) Are these measurements qualitative or quantitative?

Quantitative.

c.) (1 point) What level of measurement best describes the measurements being collected?

Ratio. (weight has abs. zero.)

Conduct a hypothesis test to determine if the three populations of snails have different mean weights. Assume the weight of a snail is normally distributed and has equal variance among the locations.

d.) (2 points) State the null and alternate hypothesis.

$H_0: \mu_1 = \mu_2 = \mu_3$

H_1: There is atleast one pop. mean that is different.

e.) (2 points) State the appropriate statistical test and its P value.

ANOVA(L_1, L_2, L_3) P-value = 0.1414

f.) (2 points) What is the conclusion of the hypothesis test at the 5% significance level? Interpret your conclusion in terms of the mean weight of snails in different locations.

At 5% level of significance fail to reject Ho. Not sufficient evidence that atleast one mean snail weight at diff. locations is different.

20) While playing blackjack at a casino, Eugene notices there are very few players getting dealt a blackjack. Of the 300 deals he has seen, only 9 of them were blackjacks. Unbeknownst to Eugene, the probability of being dealt a blackjack is only 4.8% from a standard shoe (the set of cards being dealt from). Assume each deal to a player is independent of every other deal.

a.) (1 point) If the shoe is standard, what is the expected number of blackjacks dealt in 300 deals?

$np = 300(0.048) = 14.4$

b.) (2 points) If the shoe is standard, what is the probability that no more than 9 of 300 deals are blackjacks?

$P(X \leq 9) = \text{binomcdf}(n=300, p=0.048, 9) = 0.3148$

Eugene begins to suspect that the casino's shoe is not standard. He decides to conduct a hypothesis test to determine if the proportion of deals that a blackjack is less than 5%.

c.) (2 points) State the null and alternate hypothesis.

$H_0: p = 0.05$

$H_1: p < 0.05$

d.) (2 points) State the appropriate statistical test and the P-value.

1 Prop Z Test $\quad np = 300(0.05) = 15$ $\quad$ P-value = 0.2134

$nq = 300(0.95) = 285$

e.) (2 points) What is the conclusion of the hypothesis test at the 5% significance level? Interpret your conclusion in terms of the percentage of deals that a blackjack.

At 5% level of sig. fail to reject H_0. There is insufficient evidence that the proportion of deals that are blackjack is less then 5%.

21) The data below are the scores on an exam of 10 randomly selected students from a statistics class and the number of hours that they studied for the exam.

Hours (x)	3	4	2	8	2	3	4	5	6	3
Score (y)	65	80	60	88	66	78	85	90	90	71

a.) (2 points) Find the equation of the least squares line of best fit.

LinReg(a+bx)

$\hat{y} = 57.8 + 4.875x$

($\rho > 0$ pos. correlation)

b.) (4 points) Test the claim that $\rho > 0$ using a significance level of 0.01. Indicate which test was used and explain your results in the context of this problem.

LinRegTTest

$H_0: \rho = 0$

$H_1: \rho > 0$

P-value = 0.0017

At 1% level of sig, reject H_0. There is stat. sig. evidence to support there is positive correlation between hours and score.

c.) (1 point) According to the least squares line, predict the score a student would get on the exam if they studied for 4 hours.

$\hat{y} = 57.8 + 4.875(4) = 77.3$

d.) (3 points) Calculate the 90% confidence interval for the score on the exam of a student that studied for 4 hours.

YINT $\begin{cases} x = 4 \\ df = 8 \\ C = 0.90 \end{cases}$

(64.2, 90.4)