1. The president of a university claimed that the entering class this year appeared to be larger than the entering class from previous years but their mean SAT score is lower than previous years. He took a sample of 20 of this year's entering students and found that their mean SAT score is 1,501 with a standard deviation of 53. The university's record indicates that the mean SAT score for entering students from previous years is 1,520. He wants to find out if his claim is supported by the evidence at a 5% level of significance. The population the president is interested in is which of the following?
   A) all entering students to his university this year.
   B) all entering students to all universities in the U.S this year.
   C) all SAT test centers in the U.S. this year.
   D) the SAT scores of all students entering universities in the U.S. this year.

#2 – 4: An entrepreneur is considering the purchase of a coin-operated laundry. The current owner claims that over the past 5 years, the mean daily revenue was \$675 with a population standard deviation of \$75. A sample of 30 days reveals a daily mean revenue of \$625.

2. If you were to test the null hypothesis that the daily mean revenue was \$675, which test would you use?
   A) Z test of a population mean
   B) t test of a population proportion
   C) t test of population mean
   D) Z test of a population proportion

3. If you were to test the null hypothesis that the daily mean revenue was \$675 and decide not to reject the null hypothesis, what can you conclude?
   A) There is not enough evidence to conclude that the daily mean revenue was \$675.
   B) There is enough evidence to conclude that the daily mean revenue was \$675. The null hypothesis would be rejected at a significance level of α=0.05.
   C) There is enough evidence to conclude that the daily mean revenue was not \$675.
   D) There is not enough evidence to conclude that the daily mean revenue was not \$675.

4. TRUE or FALSE: If, rather than the population standard deviation, you had the sample standard deviation there would be no change in the test used.
   A) True B) False

5. A bank tests the null hypothesis that the mean age of the bank's mortgage holders is less than or equal to 45 years, versus an alternative that the mean age is greater than 45 years. They take a sample and calculate a p-value of 0.0202. Which of the following statements is true?
   A) The null hypothesis would be rejected at a significance level of α=0.05.
   B) The null hypothesis would be accepted at a significance level of α=0.05.
   C) The null hypothesis would be rejected at a significance level of α=0.01.
   D) The null hypothesis would be accepted at a significance level of α=0.01.

6. If you are testing the difference between the means of 2 related/dependent populations with samples of n_1 = 20 and n_2 = 20, the number of degrees of freedom is equal to:
   A) 39 B) 38 C) 19 D) 18

7. TRUE or FALSE: In examining the difference between related/dependent samples, we are essentially sampling from an underlying population of differences.
   A) True B) False

#8 – 10: A researcher randomly sampled 30 graduates of an MBA program and recorded data concerning their starting salaries. Of primary interest to the researcher was the effect of gender on starting salaries. The result of the pooled variance *t*-test of the mean salaries of females (Population 1) and males (Population 2) in the sample is given at the right. Assume a level of significance of α=0.05

Starting Salaries by Gender:	
Hypothesized Difference	0
Level of Significance (α)	0.05
Population 1 Sample (female)	
Sample Size	18
Sample Mean	99210
Sample Standard Deviation	13577
Population 2 Sample (male)	
Sample Size	12
Sample Mean	105820
Sample Standard Deviation	11741
Difference in Sample Means	-6610
t Test Statistic	-1.37631
Lower-Tail Test	
Lower Critical Value	-1.70113
P-value	0.089816

8. What type of variable is the underlying variable of interest?
   A) Categorical B) Quantitative

9. The researcher was attempting to show statistically that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. Which of the following is an appropriate alternative hypothesis?
   A) H_A: $\mu_{females} > \mu_{males}$
   B) H_A: $\mu_{females} < \mu_{males}$
   C) H_A: $\mu_{females} \neq \mu_{males}$
   D) H_A: $\mu_{females} = \mu_{males}$

10. The proper conclusion for this test is:
   A) Because P-value = 0.089816 > 0.05 =α, the researcher will FAIL TO REJECT H_0 in favor of H_A. That is, there is not sufficient evidence to indicate that females have a lower mean starting salary than male MBA graduates.
   B) Because −1.70113 < −1.37631, the researcher will REJECT H_0 in favor of H_A. That is, there is sufficient evidence to indicate that females have a lower mean starting salary than male MBA graduates.
   C) Because P-value = 0.089816 > 0.05 =α, the researcher will REJECT H_0 in favor of H_A. That is, there is sufficient evidence to indicate that females have a lower mean starting salary than male MBA graduates.
   D) Because −1.70113 < −1.37631, the researcher will FAIL TO REJECT H_0 in favor of H_A. That is, the researcher will accept H_0 that females do not have a lower mean starting salary than male MBA graduates.

11. Suppose you wish to test H_0: $\mu \le 47$ versus H_A: $\mu > 47$. What will result if you reject H_0 and conclude that the mean is greater than 47 when its true value is really 52?
   A) You have made a correct decision.
   B) You have made a Type I error.
   C) You have made a Type II error.
   D) None of the above is correct.

12. Suppose you wish to test H_0: $\mu \le 47$ versus H_A: $\mu > 47$. What will result if you fail to reject H_0 and conclude that there is not sufficient evidence that the mean is greater than 47 when its true value is really 52?
   A) You have made a correct decision.
   B) You have made a Type I error.
   C) You have made a Type II error.
   D) None of the above is correct.

13. We evaluate an appropriately calculated 95% confidence interval with the following knowledge:
   A) If all possible samples of size n are taken and confidence intervals are developed, 95% of them will cover the true population mean, but 5% of them will not.
   B) We are exactly 95% confident that the sample mean, $\bar{X}$, is contained in the interval.
   C) The probability that the calculated interval contains the true population mean is 0.95.
   D) We are sure that the confidence interval contains the true population mean.

#14 – 15: To test the effectiveness of a business school preparation course, 8 students took a general business test BEFORE and AFTER the course. The results and some calculations are shown in the table below.

Business Preparation Course, BEFORE and AFTER					
Student	Exam Score BEFORE	Exam Score AFTER	Difference: (AFTER - BEFORE)	Difference - mean	(Difference - mean)^2
1	530	670	140	90	8100
2	690	770	80	30	900
3	910	1000	90	40	1600
4	700	710	10	-40	1600
5	450	550	100	50	2500
6	820	870	50	0	0
7	820	770	-50	-100	10000
8	630	610	-20	-70	4900
		sums=	400	0	29600
	n=8	mean=	50	sample std dev=	65.03
				std error=	22.99

14. What assumptions are necessary to solve this problem?
    A) The populations of scores BEFORE and scores AFTER are INDEPENDENT.
    B) The populations of scores BEFORE and scores AFTER are RELATED/DEPENDENT.
    C) The population of difference scores is normally distributed.
    D) Both (A) and (C)
    E) Both (B) and (C)

15. Using the information above, select the correct 95% confidence interval estimate of the differences in scores, AFTER – BEFORE.
    A) $CI = \bar{X} \pm z(std\ dev) = 50 \pm 1.96(65.03) = (-75.4588\ , 177.4588)$
    B) $CI = \bar{X} \pm z(std\ error) = 50 \pm 1.96(22.99) = (4.9306\ , 95.0604)$
    C) $CI = \bar{X} \pm t(std\ dev) = 50 \pm 2.3060(65.03) = (-99.9592\ , 199.9592)$
    D) $CI = \bar{X} \pm t(std\ error) = 50 \pm 2.3060(22.99) = (-3.01494\ , 103.0149)$

16. It is desired to estimate the mean (total) compensation of CEOs in the Service industry. Data were randomly collected from 18 CEOs and the 95% confidence interval was calculated to be ($2,181,260, $5,836,180). Which of the following statements is a correct interpretation?
    A) We are 95% confident that the mean (total) compensation of the sampled CEOs falls in the interval $2,181,260 to $5,836,180.
    B) In the population of Service industry CEOs, 95% of them will have (total) compensations that fall in the interval $2,181,260 to $5,836,180.
    C) We are 95% confident that the true mean (total) compensation of all CEOs in the Service industry is between $2,181,260 and $5,836,180. That is, we can say with 95% confidence that the true mean (total) compensation is greater than $2,000,000.
    D) 95% of the sampled total compensation values fell between $2,181,260 and $5,836,180.

17. The possible responses to the question "How many people in your household are unemployed currently?" result in
    A) categorical ordinal variable.
    B) discrete numerical variable.
    C) categorical nominal variable.
    D) continuous numerical variable.

18. Since a _______ is not a randomly selected probability sample, there is no way to know how well it represents the overall population.
    A) Stratified sample
    B) Convenience sample
    C) Cluster sample
    D) Simple random sample

19. True or False: The question "Is your household income last year somewhere between $50,000 and $100,000?" will most likely result in coverage error.
    A) True
    B) False

#20 – 21: The employees of a company were surveyed on questions regarding their educational background (college degree or no college degree) and marital status (single or married). Of the 600 employees, 400 had college degrees, 100 were single, and 60 were single college graduates.

marital/ education	single	married	Totals
degree	60	340	400
no degree	40	160	200
Totals	100	500	600

20. The probability that an employee of the company is married and has a college degree is:
    A) 0.0667   B) 0.567   C) 0.667   D) 0.833

21. The probability that an employee of the company is married given that the employee has a college degree is:
    B) 0.567   B) 0.667   C) 0.68   D) 0.85

#22 – 24: Use the graphic at the right to answer the questions.
(source: http://www.slideshare.net/deeneshgoundory1/quality-tools-54066272 **)**

22. This data set shown in the histogram is best described as:
    A) Symmetric
    B) Skewed left
    C) Skewed right
    D) Bimodal

23. The mean of this data shown in the histogram is likely:
    A) Greater than the median
    B) Less than the median
    C) Approximately equal to the median
    D) Can't tell from the picture

24. The sampling distribution of the means calculated from samples of size n=50 drawn from this distribution would be:
    A) Right skewed like the underlying data
    B) Left-skewed like the underlying data
    C) Approximately normal
    D) Cannot be determined using the information provided

#25 – 26: A hypothetical book club polled its members to ask how many books each member read in the last 12 months. The table at the right shows the results.

# books read/year	
x	P(x)
6	0.05
12	0.12
24	0.65
36	0.15
48	0.03

25. What is the mean or expected value for the number of books read?
    - A) 25.2
    - B) 4.836
    - C) 24.18
    - D) 1.0

26. What is the probability that a randomly selected member read at least 12 books?
    - A) 0.12
    - B) 0.17
    - C) 0.83
    - D) 0.95

#27 – 28: The average room prices (in US$) paid in 2012 by people of various nationalities while traveling away from their home country are shown below:

135	141	150	158	158	164	169	171

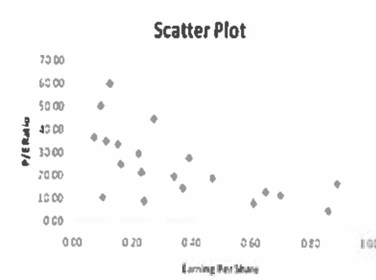
27. The median of the room prices is:
    A) 155.75 B) 158 C) 135 D) 171

28. The mean of the room prices is:
    A) 155.75 B) 158 C) 135 D) 171

29. If you wish to determine whether there is evidence that the proportion of items of interest is the same in population 1 as in population 2, the appropriate test to use is:
    - A) the Z test
    - B) χ^2. test
    - C) Either (A) or (B)
    - D) Neither of (A) nor (B)

30. The Central Limit Theorem is important in statistics because :
    - A) For any data population distribution, it says the sampling distribution of the sample mean is approximately normal, no matter what the sample size.
    - B) For any sample size n, it says that the sampling distribution of the sample mean is approximately normal.
    - C) For a large sample size n, it says the data population distribution is approximately normal.
    - D) For a large sample size n, it says the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population distribution.

31. True or False: Given at right is the scatter plot of the price/earnings ratio versus earnings per share of 20 U.S. companies. There appear to be a positive relationship between price/earnings ratio and earnings per share.
    A) True B) False

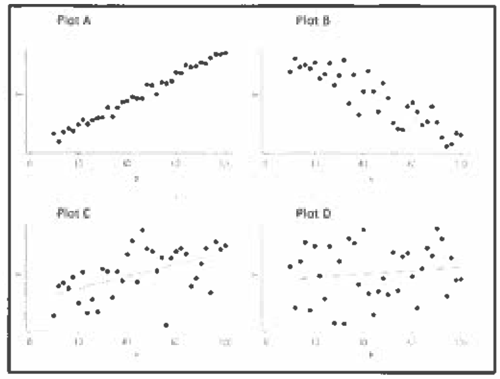


32. The correlation coefficients for the plots at the right are:

    −0.89   0.14   0.52   0.99.

    Which plot has a correlation of 0.14?

    A) Plot A
    B) Plot B
    C) Plot C
    D) Plot D



33. According to the Chebyshev rule, at least 75% of all observations in any data set are contained with a distance of how many standard deviations around the mean?

    A) 1   B) 2   C) 3   D) 4

#34-36: The director of transportation of a large company is interested in the usage of her van pool. She considers her routes to be divided into local and non-local. She is particularly interested in learning if there is a difference in the proportion of males and females who use the local routes. She takes a sample of a day's riders and finds the observed values at the right. Additionally, she calculates the expected values and the χ^2: statistic, as shown. With a level of significance of 0.05, she will use this information to perform a chi-square hypothesis test:

H_0: $p_{male} = p_{female}$ versus H_A: $p_{male} \neq p_{female}$

Evaluate Proportions Using Local Routes:			
Observed:	**Male**	**Female**	**Totals**
Local	27	44	71
Non-Local	33	25	58
Totals	60	69	129
Expected:	**Male**	**Female**	**Totals**
Local	33.02326	37.97674	71
Non-Local	26.97674	31.02326	58
Totals	60	69	129

χ2 calculations:		
Expected:	Male	Female
Local	1.098608	0.955311
Non-Local	1.344848	1.169433
	χ2 =	4.5682

34. The test will involve ________ degree(s) of freedom.

    A) 1   B) 2   C) 3   D) 4

35. What is the overall or mean proportion ($\bar{p}$) of local riders?

    A) 0.4496   B) 0.4691   C) 0.5349   D) 0.5504

36. If the critical value of the hypothesis test is 3.841, what is the conclusion of the test? State your conclusion in terms of the problem.

    A) With $\chi^2_{stat} = 3.841 < \chi^2_{crit} = 4.5682$ we fail to reject the null hypothesis. Thus we do not have sufficient evidence that the proportions of male and female riders taking local routes are different.
    B) With $\chi^2_{stat} = 3.841 < \chi^2_{crit} = 4.5682$ we must accept the null hypothesis that the proportions of male and female riders taking local routes are the same.
    C) With $\chi^2_{stat} = 4.5682 > \chi^2_{crit} = 3.841$ we reject the null hypothesis in favor of the alternative hypothesis. That is, we have evidence at the α=0.05 level of significance that the proportions of male and female riders taking local routes are different.
    D) With $\chi^2_{stat} = 4.5682 > \chi^2_{crit} = 3.841$ we fail to reject the null hypothesis in favor of the alternative hypothesis. Thus we do not have sufficient evidence that the proportions of male and female riders taking local routes are different.

37. Four freshmen are to be assigned to eleven empty rooms in a student dormitory. All the rooms are considered to be alike so that it does not matter who is being assigned to which room. How many different ways can those 4 freshmen be assigned?

Rule 4: permutation. $_{11}P_4 = \frac{11!}{(11-4)!} = 11 \times 10 \times 9 \times 8$
$= 7920$

38. According to a survey of American households, the probability that the residents own 2 cars given an annual household income over $50,000 is 0.80. Of the households surveyed, 0.60 had incomes over $50,000 and 0.70 had 2 cars. The probability that the residents of a household own 2 cars and have an income over $50,000 a year is:

A = A household own 2 cars
B = A household income over $50,000.

$P(A) = 0.7 \quad P(B) = 0.6 \quad P(A \mid B) = 0.8.$

$P(A \text{ and } B) = P(A \mid B) \cdot P(B) = 0.6 \times 0.8 = 0.48.$

39. A Medicare contractor is interested in finding an estimate for the true proportion of fraudulent claims being submitted by a questionable Medicare supplier. A pilot study resulted in a proportion of 0.35 of this supplier's claims being judged fraudulent. What sample size of claims must the Medicare contractor evaluate to determine an estimate of the true proportion of fraudulent claims to within ±0.05 with 90% confidence?

$n_e = \frac{(z)^2 \, p(1-p)}{(MoE)^2} = \frac{(1.645)^2 \cdot 0.35 \times (1-0.35)}{0.05^2}$

$= 246.2$

So. the sample size of claims must be at least 247.

40. A campus program enrolls undergraduate and graduate students (with equal probability). If a random sample of 4 students is selected from the program to be interviewed about the introduction of a new fast food outlet on the ground floor of the campus building, what is the probability that exactly 3 of the students selected are undergraduate students? (Indicate the type of probability distribution by showing the appropriate formula and fill in the information necessary to solve from the problem statement.)

Y = # of undergraduate students.

then $Y \sim \text{binomian}(4, 0.5)$

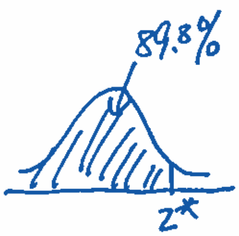
$P(Y=3) = {}_4C_3 \; p^3(1-p^4)$

$= 4 \times (\frac{1}{2})^4$

$= \frac{1}{4}$

the probability that exactly 3 of students selected are undergraduate students is 0.25.

41. The owner of a fish market determined that the mean (μ) weight for a catfish is 3.2 pounds with a standard deviation (σ) of 0.8 pound. Assuming the weights of catfish are normally distributed, below what weight (in pounds) will 89.80% (0.8980) of catfish weights occur?



X = weight of catfish. $X \sim N(3.2, 0.8)$

$P(X < x^*) = 89.8\%$

$P(Z < z^*) = 89.8\%$

$z^* = 1.27$.

$x^* = \mu + 1.27\sigma$

$= 3.2 + 1.27 \times 0.8$

$= 4.216$.

89.8% of catfish weights below 4.216 Lbs.

42. You are given a data set with mean (μ) = 57 and standard deviation (σ) = 14.2. If you can assume the data distribution is normally distributed, what is the probability that a randomly selected observation will have a value between (14 , 85)?

$X \sim N(57, 14.2)$

$z_1 = \frac{14-57}{14.2} = -3.03$

$z_2 = \frac{85-57}{14.2} = 1.97$

$P(14 < X < 85) = P(-3.03 < Z < 1.97)$

$= P(Z < 1.97) - P(Z < -3.03)$

$= 0.9756 - 0.0012$

$= 0.9744$

#43 – 44: One of the biggest issues facing e-retailers is the ability to turn browsers into buyers. This is measured by the conversion rate, the percentage of browsers who buy something in their visit to a site. The proportion of browsers who were converted to buyers for a company's website was 0.101 (10.1%) The website at the company was redesigned in an attempt to increase its conversion rates. A sample of 200 browsers at the redesigned site was selected. Suppose that 24 browsers made a purchase. The company officials would like to know if there is evidence of an increase proportion of browsers who were converted to buyers for the company's newly designed website.

43. Construct a 95% confidence interval estimate for proportion of browsers who were converted to buyers for the company's newly designed website.

CI: $\hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$\hat{p} = \frac{24}{200} = 0.12$

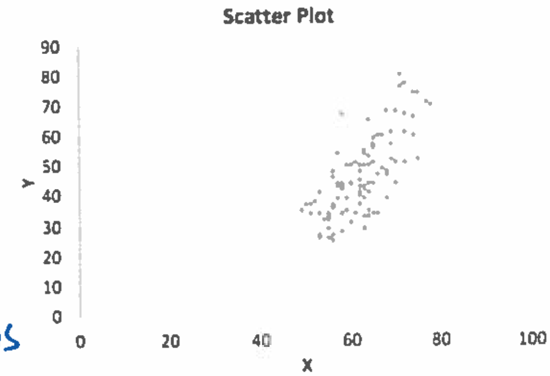
CI: $0.12 \pm 1.96 \cdot \sqrt{\frac{0.12(1-0.12)}{200}}$

$[0.075, 0.165]$

44. Interpret the confidence interval in terms of the problem.

We are 95% confident that the true proportion of browers who were converted to buyers for a company's website is between [0.075, 0.165]. and since 0.101 falls in the confidence interval.. we do not have sufficient evidence to say there is an increase proportion of browsers who were converted to buyers.

#45– 48: Zagat's publishes restaurant ratings for various locations in the United States. The Excel output shows scatterplot and the regression model to predict the cost per person, based on a variable that represents the sum of the ratings for food, décor and service.



45. Interpret the meaning of the slope in terms of the problem.

when x (sum of the ratings for food, décor and service) increased by 1 unit, the cost per person will be increased by 1.4963 units

Excel output:

Regression Statistics	
Multiple R	0.7388
R Square	0.5458
Adjusted R Square	0.5412
Standard Error	9.0940
Observations	100

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	9740.0629	9740.0629	117.7746	0.0000
Residual	98	8104.6871	82.7009		
Total	99	17844.7500			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-46.7718	8.6746	-5.3918	0.0000	-63.9863	-29.5573
Summated Rating	1.4963	0.1379	10.8524	0.0000	1.2227	1.7699

b_1

46. Provide the equation for the least squares regression line based on the output.

$$\hat{y} = -46.7718 + 1.4963 \cdot x$$

47. Predict the mean cost per person for a restaurant with a summated rating of 50.

$x = 50$

$$\hat{y} = -46.7718 + 1.4963 \times 50$$
$$= 28.0432$$

48. Identify the Coefficient of Determination, and interpret its meaning in terms of the problem.

$R^2 = 0.5458$.

54.58 % of the variation in Y. can be explained by ~~using~~ variations in sum of ratings for food. décor and service.