STATISTICS
PRACTICE FINAL *[Solutions are at the end]*

(1) Suppose a researcher wishes to test Ho: $\mu \le 100$ milligrams versus H_1: $\mu > 100$ milligrams.
(a) What will result if she concludes that the population mean is less than 100 milligrams when it is actually 104 milligrams? [Type I error, Type II error, no error, impossible to determine].
(b) Suppose she concludes that the population mean is less than 100 milligrams when it is actually 98 milligrams? [Choose one: Type I error, Type II error, no error, impossible to determine].

(2) Ikea claims it takes 60 minutes to assemble the Ikea bunk bed with a population standard deviation of 20 minutes. Numerous random samples consisting of 100 customers who purchased the bed are taken.
(a) The standard deviation of the sample means should be about____. This standard deviation has a name; it is the _____.

(3) A confidence interval was used to estimate the proportion of new Dell laptop computers that are defective. A random sample of 121 laptops produced the following 95% confidence interval as to the proportion of defectives: $2\% \pm .50\%$. (a) Is this statement correct? "We are 95% sure that the true proportion of new Dell laptops that are defective is 2%." (b) What is the best way to state the results?

(4) GE wants to construct a confidence interval for the life of its stoves. Which of the following confidence intervals will be narrower: (a) a 75% confidence interval or a 99% confidence interval? (b) one based on a random sample of 100 stoves or one based on a random sample of 500 stoves?

(5) Intel wants to construct a confidence interval for the life of its computer chips. Which of the following confidence intervals will be narrower: one based on a Z-distribution or one based on a t-distribution?

(6) The 42nd Street Library in NYC wants to estimate how much time a typical patron spends in its library. 25 patrons are randomly sampled with the following results: sample mean = 45 minutes; sample standard deviation = 20 minutes. Construct a 95% confidence interval. You may assume that time spent in a library follows a normal distribution.

(7) Suppose the weight of Perdue chickens is normally distributed with a population mean of 20 pounds and a population standard deviation of 5.5 pounds. A random sample of 121 Perdue chickens is taken. What is the probability of obtaining a sample mean of 20.5 or higher?

(8) Suppose a researcher wishes to test Ho: $\mu \ge 70$ versus H_1: $\mu < 70$. Which of the following possible sample results based on a sample of 25 provides the strongest evidence to reject Ho.
(scenario a) sample mean = 75 ; sample standard deviation = 2 ; or
(scenario b) sample mean = 60 ; sample standard deviation = 5 ; or
(scenario c) sample mean = 66; sample standard deviation = 1

(9) (a) A researcher wants to test whether a bar of Ghirardelli chocolates has no more than 10 grams of fat. What kind of test should be used: one-tail or two tail? Set up the null hypothesis and alternate hypothesis. (b) A company claims that its cereal has at least 30 raisins per serving. What kind of test should be used? Set it up. (c) A company claims that its donut baking machines produces donuts that weigh exactly 6 ounces. What kind of test should be used? Set it up. (d) A college claims that the average SAT score of its freshmen is at least 1000. What are Ho and H_1?

(10) A music company sells "The 50 Best Love Songs of All Time" on CDs using television commercials. It is important for the company to know the average age of its customers. If the average age is more than 40 years old, the company will change its advertising and media plan; if not, everything will remain the same. They are going to take a random sample of customers to test this. What are Ho and H_1?

(11)   A random sample of 81 trucks manufactured by the Gecko Company results in the following: sample mean = 9.80 years, sample standard deviation = 1.80 years. What is the upper limit of a 90% confidence interval? _____

(12)   Molan Springs claims that there are no more than 5 parts per million (ppm) of impurities in its water. A researcher randomly samples 64 bottles of the Molan Springs water and finds that the sample mean is 5.30 ppm and the sample standard deviation is .40 ppm. Should the company's claim be accepted?

(13) A bakery wants to determine whether there is a difference between the male and female bakers when it comes to baking a chocolate pudding pie. They randomly sample 145 bakers with the following results:
Men: average time = 50 minutes, standard deviation = 10 minutes; n = 64
Women: average time = 45 minutes, standard deviation = 12 minutes; n = 81
What is your conclusion? Test at the .05 significance level.

(14) Dell Computers uses two suppliers of motherboards. Dell performs a statistical test to determine whether there is a difference in the average life of motherboards made by two suppliers --one charges more for its motherboards. What is the computed value of the test statistic? Assume a normal distribution. Test at the .05 significance level.
Supplier X: average life= 12.00 years, standard deviation = 0.90 years; n = 14
Supplier Y: average life = 13.50 years, standard deviation = 1.10 years; n = 12

(15) A researcher would like to determine whether the dropout rate for two colleges is equal. At ABC College, 100 out of a sample of 400 students dropped out. At XYZ College, 150 out of 500 dropped out. Test at the .01 level of significance to see whether there is a difference between the two groups.

(16) Given the data below, test the hypothesis that smokers and nonsmokers have the same average life expectancy. Use a significance level of .05 and a two-tail test.

	Smokers	Nonsmokers
Average Lifespan (Yrs)	68	75
Standard Deviation (Yrs)	10	12
Sample Size	100	200

(17) Given the data below, test the hypothesis that computers manufactured by Lenovo and Apple have the same average customer satisfaction score. Use a significance level of .10 and a two-tail test.

	Lenovo	APPLE
Average Score	80	86
Standard Deviation	10	14
Sample Size	15	16

(18) The college wishes to purchase bulbs with an expected life of at least 10,000 hours. A random sample of 25 GE bulbs results in an average life of 9,900 hours with a standard deviation of 3,000 hours. Assume a normally distributed population. (a) Should the bulbs be rejected? Use a .05 significance level. (b) Construct a two-sided 95% confidence interval about the mean, using the above data.

(19) A company sells boxes of disks that are supposed to contain exactly 100 disks. A random sample of 64 boxes is taken and the mean is found to be 102 disks with a standard deviation of 5 disks. (a) Using a .02 significance level , should the disk -packing machinery be adjusted? (b) Construct a two-sided 92% confidence interval about the mean using the above data.

(20) A researcher claims that at least 60% of Americans are in favor of legalizing drugs. A random sample of 100 Americans is taken and 51 are in favor of legalizing drugs. (a) At the .10 significance level, should the researcher's claim be rejected?
(b) Construct a two-sided 75% confidence interval about the proportion using the above data.

(21) The printout below was obtained when a researcher compared 40 college students (variable 1 = males; variable 2 = females) on how much they spend on clothing in a typical month. Explain this printout. How much do male students spend on clothing during a typical month? How much do female students? Is the difference statistically significant or is it possibly sampling error. The test was done at the .05 significance level.

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	54.25	76.90
Variance	447.25	507.99
Observations	20	20
Pooled Variance	477.6197	
Hypothesized Mean Difference	0	
df	38	
t Stat	-3.27738	
P(T<=t) one-tail	0.00112	
t Critical one-tail	1.68595	
P(T<=t) two-tail	0.00224	
t Critical two-tail	2.02439	

(22) A school claims that the average reading score of its students is at least 71. Use the following data and assume a normal distribution. Sample average = 68, standard deviation = 8, n = 16. (a) test this claim at a = .05. (b) Construct a two-sided 95% confidence interval estimate of the true (population) average reading score.

(23) A company claims that its soda machines dispense on average exactly 16 ounces of soda. Use the following data: Sample average = 14.6 ounces, standard deviation = 1.8 ounces, n = 81. (a) test this claim at a = .04. (b) Using the above data, construct a two-sided 90% confidence interval estimate of the true mean.

(24) A company claims that no more than 4% of its computers are defective. In a sample of 100 computers, 9 are found to be defective. (a) Should the computer production line be adjusted? test at a = .01. (b) Using the above data, construct a two-sided 92% confidence interval estimate of the true (population) percentage of defective computers.

(25) A company is interested in determining whether the average salary of men and women is different in their company. A sample of men and women results in the following:

Men: average salary = $69,000 Standard deviation = $10,000 n= 60

Women: average salary = $62,000 Standard deviation = $8,000 n = 30

(a) At a = .03, is there a statistically significant difference between salaries of men and women?

(b) Using the above data, construct a two-sided 95% confidence interval for the true difference between salaries of men and women.

(26) A company is interested in determining whether the average life of HP and Epson printers is different. A sample of HP and Epson printers result in the following:

HP: average life = 7 years Standard deviation = 2 years n= 10

Epson: average life = 5.5 years Standard deviation = 2.5 years n =8

At a = .05, is there a statistically significant difference between the longevity of the two brands of printers?

(27) A company is interested in determining whether the defect rates for computer chips made by two companies is different. A sample of computer chips result in the following:

XYZ: 18 defectives in a sample of 100 computer chips

ADC: 42 defectives in a sample of 200 computer chips

At a = .05, is there a statistically significant difference between the defect rate of the two computer chips?

(28) A researcher wishes to determine whether there is a relationship between the number of job offers one receives and years of college education. A sample of 12 individuals was selected with the following results:

Years of College (X)	No. of Job Offers (Y)
1	0
1	1
2	2
2	3
3	5
3	4
4	7
4	8
5	9
5	8
6	7
6	6

SX = 42   S (X-squared) = 182

SY= 60   S (Y-squared) = 398

SXY = 261

(a) Calculate the regression equation and interpret the meaning of the regression coefficients (a and b).

(b) Compute the correlation coefficient, r, and test it for significance at the a =.05 level.

(c) Compute the coefficient of determination (r-squared) and explain its meaning.

(d) Use the regression equation to predict the number of job offers an individual with 8 years of college could expect.

(29) A researcher wishes to determine whether there is a relationship between IQ and salary. A sample of 10 individuals was selected with the following results:

IQ	Salary (in Thousands)
90	18
95	20
100	22
100	25
105	30
110	30
115	40
120	45
120	50
150	50

$\sum$ IQ=1105
$\sum$ Salary = 330
$\sum$ (IQ) (Salary) = 38,170
$\sum$ IQ squared = 124,775
$\sum$ Salary squared = 12,258

(a) Find the regression coefficients and state the regression equation.
(b) Interpret the meaning of the Y-intercept and the slope in this problem.
(c) Use the regression model to predict the salary for an individual with an IQ of 140.
(d) Compute the correlation coefficient, r. Test r for significance at the .05 level. What does r tell you?
(e) Compute the coefficient of determination, r-squared. What does it tell you?

(30) A researcher studying starting salaries of high school (HS) graduates living in Alabama, correlated the starting salaries with high school average. The results in Excel are shown below. However, we removed some values and you will have to fill them in:

Regression Statistics	
Multiple R	
R Square	
Standard Error	
Observations	25

ANOVA

	df	SS	MS	F	Significance F
Regression	1	818064246	818064246		0.00001
Residual	23	544197810	23660774		
Total	24				

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	2558.423	5721.719	0.447	0.659	-9277.838	14394.683
X Variable 1	402.745	68.494	5.880	0.000	261.055	544.435

(a) Compute the Total Sum of Squares (TSS)
(b) R^2 (coefficient of determination) =
(c) The correlation coefficient =
(d) The sample size consisted of __ students.
(e) The (computed) F-statistic is _____. Is it significant?
(f) What is the regression equation?
(g) Every additional point of high school average should add about how much to one's starting salary (in Alabama)?
(h) A high school student in Alabama with a 0 high school average can expect to earn a starting salary of _____ (Don't worry he went on to become a politician :)
(i) The standard error of estimate is?
(j) A high school graduate in Alabama with a 100 average could expect to earn about ____ (you were not taught about the confidence interval so forget about that).

(31) A researcher wants to see whether there is a relationship between amount spent on clothing per year and IQ. The Excel printout is below:

SUMMARY
OUTPUT

Regression Statistics	
Multiple R	0.891086
R Square	0.794034
Adjusted R Square	0.778190
Standard Error	950.357399
Observations	15

ANOVA

	df	SS	MS	F	Significance F
Regression	1	45264753.9	45264753.9	50.117	0.0000083
Residual	13	11741329.4	903179.2		
Total	14	57006083.3			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	13296.3	1499.81674	8.865296	0.0000007	10056.163	16536.4760
X Variable 1	-99.443	14.04691	-7.079346	0.0000083	-129.789	-69.0964

(a) R^2 (coefficient of determination) =
(b) The correlation coefficient =
(c) The sample size consisted of __ people.
(d) The (computed) F-statistic is _____. Is it significant?
(e) What is the regression equation?
(g) Every point of IQ results in how much spending on clothing?
(h) An individual with a 0 IQ should spend about how much on clothing?
(i) The standard error of estimate is?
(j) An individual with a 100 IQ should spend about how much on clothing?

SOLUTIONS

(1) (a) Type II error (b) no error -- this would be a correct conclusion

(2) 2 minutes -- This is known as the standard error of the mean

(3) No. We are 95% certain that the true (population) proportion of defective Dell laptops is between 1.5% and 2.5%. Or, you might phrase it this way: we are 95% certain that the true population proportion lies somewhere in the interval between 1.5% and 2.5%.

(4) (a) 75% confidence interval will be narrower than a 99% confidence interval. The more confidence you want, the wider the interval.
(b) One based on a sample of 500 will be narrower; the larger the sample, the narrower the interval. In fact, if you take a census (the largest "sample" possible), there will be no interval.

(5) Z distribution will be narrower. Thus, for a 95% confidence interval using Z, the critical values are 1.96; it is more for the t. At infinity degrees of freedom, the t finally becomes a Z.

(6) 45 ± 8.26 minutes (use t-distribution)

(7)   .1587 or 15.87% [$Z = (20.5 - 20)/ (5.5/\sqrt{121})$]

(8) Scenario c this results in a t-value of -20. With scenario a you should not reject the null hypothesis and scenario b produces a t-value of -10. I am assuming a normal distribution so that the researcher can use the t-distribution. If the population is not normally distributed, the researcher might use a different statistical test but the answer will still be the same.

(9) (a) one-tail test: Ho: $\mu \leq 10$ grams versus H_1: $\mu >10$ grams (b) One-tail test: Ho: $\mu \geq 30$ raisins versus H_1: $\mu <30$ raisins (c) Two-tail test: Ho: μ 6 ounces versus H_1: $\neq$ 6 ounces (d) One-tail test: Ho: $\mu \geq 1000$ versus H_1: $\mu <1000$

(10) One-tail test: Ho: $\mu \leq 40$ versus H_1: $\mu >40$
(11) 10.13 years
(12) No. $Z = 6$, $p < .00001$ Claim is very unlikely to be true. Don't drink this water.

(13) $Z = 2.73$ (2.74 if you did not round).   The difference between the male and female bakers is statistically significant at the .05 level. (You compare it with a Z-value of 1.95 if you are testing at the .05 level -- two-tail test).

(14) The difference is statistically significant. Computed t-value = -3.83 (compare this table value of t with 24 degrees of freedom = 2.064). By the way the pooled variance was .99.

(15) Ho: $P_1 = P_2$; Z=-1.66, do not reject Ho, p>.01
(16) Ho: $\mu_1 = \mu_2$; Z= -5.34, reject Ho, p<.05
(17) Ho: $\mu_1 = \mu_2$; t_{29} = -1.36 do not reject Ho, p>.10

(18) (a) Ho: $\mu_1 >=;10,000$ hours t_{24} = -.17, do not reject Ho, p>.05
(b) 8,661.6 hours <- - - - - > 11,138.4 hours

(19) Ho: $\mu_1 = 100$ disks Z = 3.2, reject Ho, p<.02 (b) 100.91 disks <- - - - - - - > 103.09 disks
(20) Ho: P >=.60; Z = -1.84, reject Ho, p<.10 (b) 45.3<--->56.7%

(21) Male students spent $54.25 per month on clothing and female students spent $76.90. These sample means were based on samples of 20 males and 20 females. A t-test was done. If a two-tail test was used, the critical values (from the table) at the .05 level are 2.02439. The calculated value of the t-statistic was -3.27738 (it is negative since men spent less than women). The null hypothesis is that male and female students spend the same amount on clothing per month. The probability of getting the sample evidence if the null hypothesis is true is .00224. Thus, we reject the null hypothesis. Conclusion: Female students spend more on clothing per month than do the male students. The difference is unlikely to be chance variation.

(22) (a) t_{15} = -1.5 p>.05 do not reject H0 (b) 63.7 <--------> 72.3
(23) (a) Z = -7 p<.04 reject H0 (b) 14.27 oz <------> 14.93 oz
(24) (a) Z = 2.56 p<.01 reject H0 (b) 4% <------> 14%
(25) (a) Z = 3.6 p<.03 reject H0 (b) $3,180 <----->$10,820
(26) t_{16} = 1.42 p>.05 do not reject H0
(27) Z = -.61 p>.05 do not reject H0

(28)(a) Y = -.10 + 1.46X every year of college results in an additional 1.46 job offers. 0 years of college results in -.10 offers. (b) r = .87 t10 = 5.6 p <.05 the correlation is significant (c) r-squared = 75.7%. 75.7% of the variation in the number of job offers received is explained by years of college. (d) -.10 + 1.46(8) = 11.58 job offers

(29) (a) Salary = -37.50 + .638(IQ) r =.89 and r-squared = 79.21% The r is significant (t_8 = 5.52, p<.05) Individuals with IQ's of 140 should earn, on average, $51,820

(30) (a) 1,362,262,056 (b) r^2 = .6005 or about 60% (818,064,246/ 1,362,262,056; SS Regression/ SS Total) (c) r = + .7749 (d) 25 students (observations) (e) F-statistic = 34.5747 (MS Regression / MS Residual) and is very, very, very significant. The regression is unlikely to be the result of chance. (f) Starting Salary = $2558.42 + $402.75 (High School Average) -- since this is in dollars, two decimal places are sufficient and we rounded. (g) $402.75 (h) $2558.42 (i) The standard error of estimate is the square root of the Mean Square Residual (23660774) and is 4864.23
(j) $42,832.92 to estimate the confidence interval (which is not covered in this course) you will need to use the standard error of estimate.

(31) (a) R^2 (coefficient of determination) = 79.40% (b) The correlation coefficient = - .891 (it is negative since the slope is negative) (c) The sample size consisted of 15 people. (d) The (computed) F-statistic is 50.117. It is very, very, very, very significant. This is not chance that we are looking at. (e) Amount spent on clothing = $13,296.30 - 99.44 (IQ); Note that the slope term is negative. This indicates an inverse relationship (g) An increase of one point of IQ results in a decrease in amount spent on clothing of $99.44. (h) An individual with a 0 IQ will spend about $13,296.30 on clothing per year. (i) The standard error of estimate is 950.357399. I would round it to 950.36. (j) An individual with a 100 IQ should spend about $3352.30 on clothing.