Practice Final Math 15 Spring 2011 ~~Fall 2010~~ Jones Name Key

You must show all work to recieve credit!

Find the indicated probability.

X 1) (5 Points) If you pick a card from a standard 52 card deck, what is the probability that you get a seven or a heart? ch 4

$P(7 \text{ or } \heartsuit) = P(7) + P(\heartsuit) - P(7 \text{ and } \heartsuit)$

$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \boxed{.3077}$

Find the necessary sample size.

✓ 2) (5 Points) Weights of women in one group are normally distributed with a standard deviation of 17 lb. A researcher wishes to estimate the mean weight of all women in this group. Find how large a sample must be drawn in order to be 90% confident that the sample mean will not differ from the population mean by more than 3.2 lb. ch 7.2

Sample Size n for Mean

$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left(\frac{1.645 \cdot 17}{3.2}\right)^2 = 76.37$

Always Round up $\rightarrow$ $\boxed{77 \text{ women}}$

$z_{\alpha/2} = \text{invn}(1 - \alpha/2)$

$z_{\alpha/2} = \text{invnorm}(.95) = 1.645$

Solve the problem.

X 3) (10 Points) My son has a 1950's era electric train. It has 4 unique passenger cars and 5 unique freight cars. Three cars are selected by Trevor at random and he arranges them behind the engine.

Chap. 4

In how many ways can 3 cars be selected from this group of 9 cars? ${}_9C_3$ MATH $\boxed{84}$

In how many ways can 3 cars be selected and arranged from this group of 9 cars? ${}_nP_r$ $\boxed{504}$

In how many ways can 3 of the 4 passenger cars be selected and arranged? ${}_nP_r$ $\boxed{24}$

If 3 cars are randomly selected without replacement from the 9 cars, find the probability that the selected cars will consist of all passenger cars.

$P(\text{All passenger}) = \frac{\#\text{ Set of all passenges}}{\#\text{ of groups of 3 cars}} = \frac{{}_4C_3}{{}_9C_3} = \frac{4}{84}$

$P(\text{All passange}) = \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{2}{7} =$ $\boxed{.0476}$

X 4) (5 Points) Suppose you pay $1.00 to roll a fair die with the understanding that you will get back $3.00 for rolling a 1 or a 4, nothing otherwise. What is your expected value? ch 5.2

E(x)

One Var Stat L_1, L_2

Net $: x		p(x)
1 or 4	2	2/6
other	−1	4/6
		1

$E(x) = \sum x \cdot p(x)$

$= 2 \cdot \frac{2}{6} + -1\left(\frac{4}{6}\right)$

$\boxed{E(x) = 0}$

Insurence

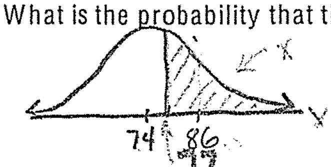
Chap. 6

§ 6.3

5) (10 Points) The amount of snowfall falling in a certain mountain range is normally distributed with a mean of 74 inches, and a standard deviation of 12 inches. Show graphs of the normal distribution with a labeled x-axes for each of the parts below. $\mu = 74$ $\sigma = 12$

a) What is the probability that the amount of snow fall in any given year will exceed 77 inches?

①



Normalcdf(LB, UB, μ, σ)

$P(X > 77) = \text{normalcdf}(77, 999, 74, 12)$

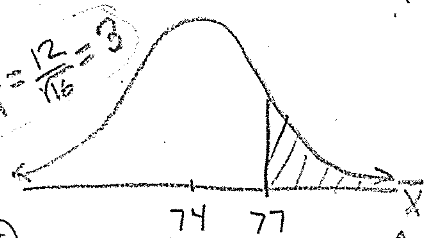
$= \boxed{.4013}$

b) What is the probability that the mean annual snowfall during 16 randomly picked years will exceed 77 inches?

$n = 16$
$\mu_{\bar{x}} = 74$
$\sigma_{\bar{x}} = \sigma/\sqrt{n} = 12/\sqrt{16} = 3$

CLT

§ 6.5



$P(\bar{x} > 77) = \text{normalcdf}(77, 999, 74, 3)$

$= \boxed{.1586}$

$\sigma/\sqrt{n}$

Dist of Sample Means for Samples of size 16

Construct the indicated confidence interval for the difference between the two population means. Assume that the assumptions and conditions for inference have been met.

§ 9.3

6) (10 Points) The table below gives information concerning the gasoline mileage for random samples of trucks of two different types. Find a 95% confidence interval for the difference in the means $\mu_X - \mu_Y$.

2-Sample T Interval

Stats

Not Pooled

	Brand X	Brand Y
Number of Trucks	50	50
Mean mileage	$20.5 = \bar{x}_1$	$24.3 = \bar{x}_2$
Standard Deviation	$2.3 = s_1$	$1.8 = s_2$

a) What is the point estimate for the difference in the milage? $\bar{x}_1 - \bar{x}_2 = 20.5 - 24.3 = -3.8$

b) Use the ~~formula~~ Skip and the table to find the margin of error for this confidence interval.

$E = \frac{UB - LB}{2} = \frac{(-2.98 - (-4.62))}{2} = .82$

No formula for chapter 9, $E = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2} = .82$

use out put of TI to find E

c) Find the confidence interval using both your calculation above and your calculator.

TI $\boxed{(-4.62, -2.98)}$

2 SampT Interval →

$(2.98, 4.62)$ → Switch X & Y is o.k.

d) Interpret the meaning of this confidence interval.

Since zero isn't in this interval there is a significant difference between brand X and Y. We are 95% Confident that brand Y gets between 2.98 and 4.62 mpg better gas milage than brand X.

Find the indicated probability.

7) An airline estimates that 98% of people booked on their flights actually show up. If the airline books 67 people on a flight for which the maximum number is 65, what is the probability that the number of people who show up will exceed the capacity of the plane?

a) (5 points) How many people does the airline expect to show up? What is the mean and standard deviation of this binomial proability distribution? §6.3

$\mu = np = 67 \cdot .98 = 65.66$ = Expected Value = E(x)

$\boxed{\mu = 65.66}$

$\sigma = \sqrt{npq} = \sqrt{67 \cdot .98 \cdot .02} = \boxed{1.15}$

b) (5 points) Use the binomial distribution to fill in the table BinomialPDF(n,p). Draw the right tail of the binomial probability distribution. Chap. 6

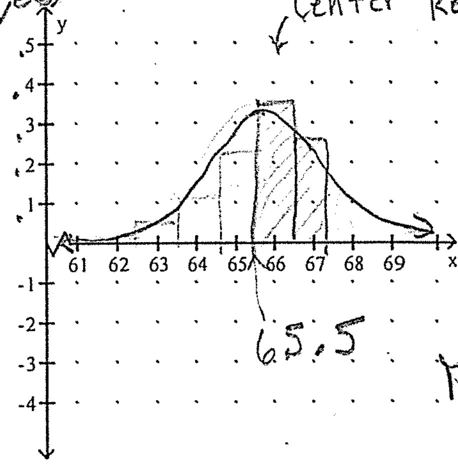
X	P(X)
61	.002
62	.009
63	.034
64	.105
65	.238
66	.353
67	.258

→ Histogram

c) $P(X=66) = \text{binpdf}(67, .98, 66) = .353$

d) $P(X=67) = .258 = L_2(168) = \text{binpdf}(67, .98, 67)$

Center Rectangles on integers



a) (5 points) Find the probability that the number of people who show up will exceed the capacity of the plane? §6.4

$P(X \geq 66) = 1 - \text{Binedf}(\underset{n}{67}, \underset{p}{.98}, \underset{X}{\overline{65}}) = .611$

$= 1 - P(X \leq 65)$

Skip d) (5 points) Use the normal distribution with a continuity correction to find the probability that the number of people who show up will exceed the capacity of the plane? Draw and label the normal distribution needed. §6.6

$P_N(X \geq 65.5) = \text{ncdf}(\underset{LB}{65.5}, \underset{UB}{999}, \underset{\mu}{65.66}, \underset{\sigma}{1.15})$

$= .5553$

e) (5 points) Is is appropriate to use the normal approximation for this problem?

NO, $nq < 5$ $\quad nq = 67 \cdot .02 = 1.34$

Find the P-value for the indicated hypothesis test.

8) (20 Points) A manufacturer claims that at most 6% of its fax machines are defective. In a random sample of 125 such fax machines, 8% are defective. Do all five steps for the hypothosis test and then find the P-value for a test of the manufacturer's claim. Use $\alpha = .05$

(5 Points) State the null and alternate hypothosis. Graph and shade the critical region. Find the critical value.

68.2
8.3

Claim: $p \le .06$

$H_0: p = .06$

$H_1: p > .06$

1-Prop Z Test $X = n \cdot p$

$n = 125$

$x = .08 \cdot 125 = 10$ ← has to be an integer

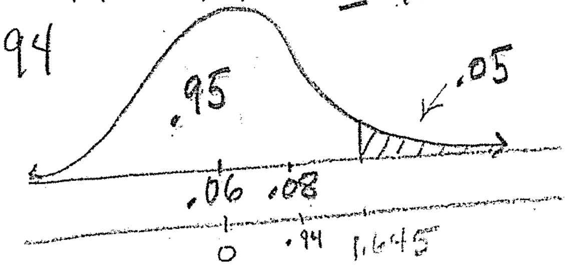
Use $\alpha = .05$, Rt tail CV: $Z = \text{invnorm}(.95) = 1.645$

(5 Points) Find the test statistic, and the point estimate of the population proportion.

$$TS: Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{.08 - .06}{\sqrt{\frac{.06 \cdot .94}{125}}} = .94$$

$X = n\hat{p} = 125 \cdot .08$

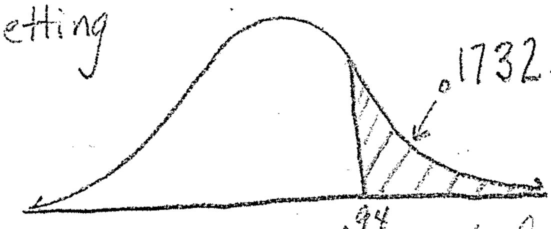
$X = 10$



(5 Points) Find the P-value and explain the meaning of the P-value. Shade a graph showing the area equal to the p-value.

P-Value = normalcdf(.94, 999) = .1732

there is a 17% chance of getting a Sample of 125 with an 8% defect rate when the overall defect rate is at most 6%.



Since Such a Sample is not unusual when the defect rate is l

(5 Points) Clearly state your conclusion. We fail to Reject H_0.

there is Not enough evidence to reject the claim that the defect rate is at most 6%.

(5 Points) If the test gave an incorrect conclusion, what type of error have you made Type I or Type II? Explain what this means in the context of this question. Explain how you could reduce the chances of making this type of error.

Type II. We failed to show that the defect rate is above 6% when it really is.

Increase the Sample Size will reduce the Probability of a Error.

Perform the required hypothesis test for two population means. Assume that the conditions and assumptions for inference are satisfied.

9) (25 Points) A coach uses a new technique in training middle distance runners. The times for 8 different athletes to run 800 meters before and after this training are shown below. TTest on L_3 Chapt 3

Matched Pairs

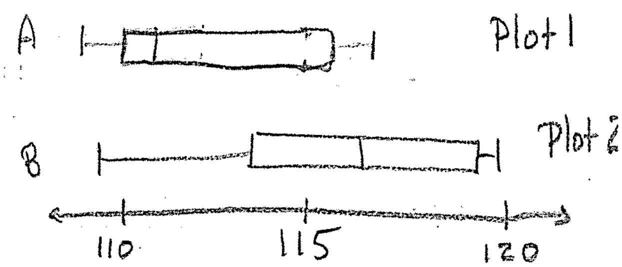
Athlete	A	B	C	D	E	F	G	H	
Time beforetraining (seconds)	119.6	114.9	119.9	108.6	119.7	118.7	112.6	116.3	L_1
Time after training (seconds)	110.2	113.6	117.5	109.4	117.9	108.3	109.0	112.4	L_2

Good or bad 9.4 >0 ⇒ faster ⇒ improved ⇒ $H_1: \mu_d > 0$ $L_3 = L_1 - L_2 = d$

(7 Points) Fill in the summary statistics.

	Before	After
Mean	116.3	112.3
Minimum	108.6	108.3
Q1	113.75	109.2
Median	117.5	111.3
Q3	119.65	115.55
Maximum	119.9	117.9
Standard deviation	4.07	3.77

Draw a side by side box plot of the data. Zoom Stat Stat Plot



(3 Points) After Reviewing the descriptive statistics above, write a sentence comparing their center and spread.

the new technique appears to improve times, the spread of both data appear close to the same.

Test the claim that the training helps to $\boxed{\text{improve}}$ the athletes' times for the 800 meters? $9.4
Perform a test at the 5% significance level.

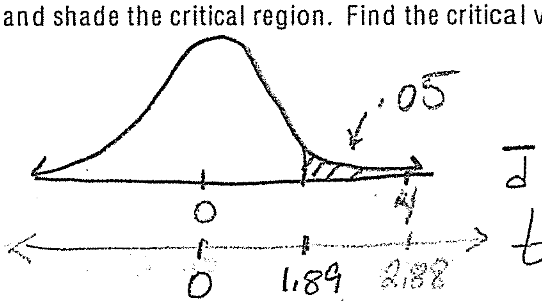
(5 Points) State the null and alternate hypothosis. Graph and shade the critical region. Find the critical value.

Claim: Athletes improved $\mu_d > 0$

$H_0: \mu_d = 0$ CV: $t = \text{invt}(.95, 7)$

$H_1: \mu_d > 0$ $df = n - 1 = 7$

$t^* = 1.89$



(5 Points) Find the test statistic. Give the initial conclusion to your hypothosis test.

TTest on L_3 with $\mu_0 = 0$ $\bar{d} = 4$ sec = point estimate of difference

TS: $t = 2.88$ PV: P-value = .0118 < .05

(5 Points) Clearly state your final conclusion.

TS: $t = \dfrac{\bar{d} - 0}{S_d/\sqrt{n}} = \dfrac{4 - 0}{3.928/\sqrt{8}} = 2.88$

→ Reject H_0, Support H_1

there is sufficent evidence to say that the training helped improve their times.

Perform the appropriate chi-square test and state your conclusion.

✓ 10) (30 Points) Use a 0.01 significance level to test the claim that the proportion of men who plan to vote in the next election is the same as the proportion of women who plan to vote. 300 men and 300 women were randomly selected and asked whether they planned to vote in the next election. The results relating the observed frequencies for intension to vote by gender are shown below. χ^2 Test

§ 11.3

	Men	Women
Plan to vote	170	185
Do not plan to vote	130	115
	300	300

Rt Tail

$\alpha = .01$

$df = (\text{row}-1)(\text{Col}-1) = (2-1)(2-1) = 1$

(5 Points) State the null and alternate hypothosis. Graph and shade the critical region. Find the critical value.

$H_0: O = E$ Proportion who plan to vote is Independent of Gender

$H_1: O \neq E$ Proportion who plan to vote is dependant on Gender

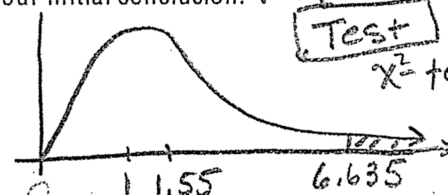
CV: $\chi^2_R = \text{inv}\chi 2(\alpha = .01, df = 1, \text{Rt tail}) = 6.635$

(5 Points) Find the matrix of expected values and the test statistic. State your initial conclusion. Use $\boxed{\text{Stat}}$ $\boxed{\text{Test}}$ χ^2-test

TS: $\chi^2 = 1.552$ $[\text{Exp}] = [B] = \begin{bmatrix} 177.5 & 177.5 \\ 122.5 & 122.5 \end{bmatrix}$

P-Value = .2128

df = 1



(5 Points) Clearly state your final conclusion.

fail to Reject H_0 there is Not evidence to suggest that the proportion who plan to vote is dependent on gender.

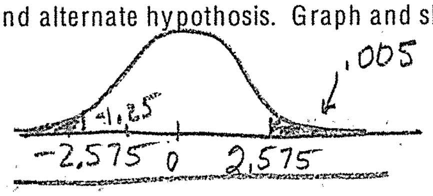
9.2 **Perform an appropriate hypothesis test of the claim that the proportion of men who plan to vote in the next election is the same as the proportion of women who plan to vote. Perform the test at the .01 significance level.**

2-Prop Z Test

(5 Points) State the null and alternate hypothosis. Graph and shade the critical region. Find the critical value.

$H_0: P_1 = P_2$

$H_1: P_1 \neq P_2$



CV: $Z = \text{invnorm}(.995) = \pm 2.575$

(5 Points) Find the test statistic, and a point estimate for the difference between the proportion of men who plan to vote in the next election and the proportion of women who plan to vote in the next election .

TS: $Z = -1.25$ P-Value = .2128 > .05

2 prop Z Test

$x_1 = 170$ $n_1 = 300$ $x_2 = 185$ $n_2 = 300 = 185 + 115$

(5 Points) Clearly state your final conclusion.

fail to Reject H_0. there is Not Sufficent evidence show the proportions are different

or to Reject that they are the Same.

15

Solve the problem.

11) (33 Points) Periodically during the last two and a half years my husband has gone out to do capacity runs to determine the range of his Electric Mustang. The paired data below consist of battery age in years and the range of my husband's EV in miles.

AGE (years)	0.2	0.3	0.8	1.0	1.2	1.5	1.7	1.8	2.0	2.5
Range (miles)	37	35	35	34	10	29	31	33	32	29

Lin Reg T-Test

§10.2 a) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between the age of the batteries and the range of the car? r = $-.244$ r* = $.632$ ← Table A-6 $|r| < r^*$

Is there a significant linear corrolation? Yes **No**

b) (3 Points) Make a scatter plot of this data.

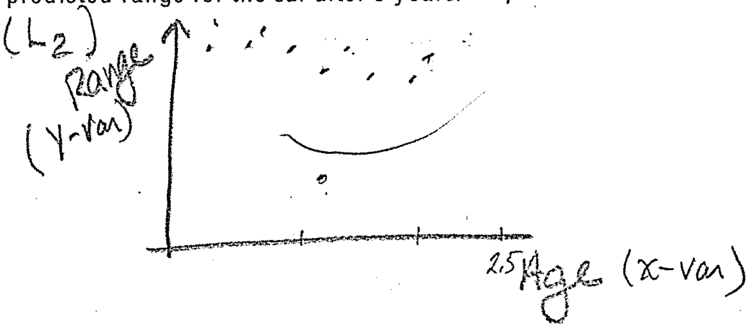
c) (3 Points) Find the equation for and graph the regression line. Don't use

$y = 33.79 - 2.53x$

§10.3 d) (3 Points) Based on the above data what is the best predicted range for the car after 3 years. ________

- OneVarStat (L_2)

$\bar{y} = 32.7$ = mean of Range values



One of the batteries died and had to be exchanged. One of the capacity runs was done right before this happened. The data value that corresponds to this run is an outlier. Remove it and redo the test for correlation.

L_1 Years	0.2	0.3	0.8	1.0		1.5	1.7	1.8	2.0	2.5
L_2 Range	37	35	35	34		29	31	33	32	29

Lin Reg T Test $n = 9$

§10.2 e) (3 Points) At the 5% level of significance, do the data provide sufficient evidence of an association between age and range. r = $-.848$ r* = $.666$ ← Table A-6 $|r| = |-.848| = .848 > .666 = r^*$

Is there a significant linear corrolation? **Yes** No

§10.3 f) (3 Points) Find the equation for the regression line. $y = a + bx = 36.7 - 3x$

g) (3 Points) Find the best predicted range for the car after 3 years. $\hat{y} = 36.7 - 3(3) = 27.7$

§10.4 X h) (3 Points) Interpret the slope of this regression line. $\hat{y} = 36.71 - 2.998x$

If each additional year the range of the car decreases by 3 miles per year

X i) (3 Points) Find and interpret the meaning of r^2. $= .7198$

72% of the Variation in range can be explaned by the Linear correlation with the batteries

§10.4 X j) (6 Points) Find a 95% prediction interval for the range of the car after 3 years. (Se=1.57). Should the battery company replaced my husbands batteries because they did not live up to their guarenteed capacity of 80% of new after 3 years. X

$$E = t_{\alpha/2} \cdot S_e \cdot \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n\sum x^2 - (\sum x)^2}} = 2.365 \cdot 1.57 \cdot \sqrt{1 + \frac{1}{9} + \frac{9(3 - 1.31)^2}{(9 \cdot 20.4 - (11.8)^2)}}$$

$E = 4.83$ (22.9, 32.5)

$t_{\alpha/2}$
$df = n - 2 = 7$
$\alpha = .05$
Two Tail

$\hat{y} \pm E = 27.7 \pm 4.83$ / 80% of 37 = 29.6 ← New range is in PI

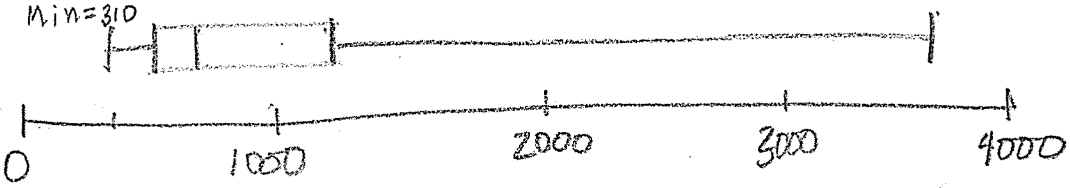
So they They don't have to Replace (but they did!)

Construct a boxplot for the given data. Include values of the 5-number summary in all boxplots.

11) (22) The weekly salaries (in dollars) of 24 randomly selected employees of a company are shown below. Construct a boxplot for the data set.

310 320 450 460 470 500 | 520 540
580 600 650 700 | 710 840 870 900
1000 1200 | 1250 1300 1400 1720 2500 3700

Min = 310 $Q_1 = 510$ Med = 705 $Q_3 = 1225$
Max = 3700



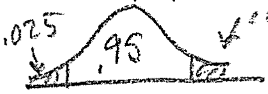
Solve the problem.

12) (13) A researcher wishes to estimate the proportion of fish in a certain lake that are inedible due to pollution of the lake. How large a sample should be tested in order to be 95 percent confident that the true proportion of inedible fish is estimated to within 0.08?

§ 7.2
For Prop
$\hat{p}$ unknown

$$n = \frac{z_{\alpha/2}^2 \cdot .25}{E^2} = \frac{(1.96)^2 \cdot .25}{.08^2} = 150.06$$

Round up $\boxed{151}$



$z_{\alpha/2}$ = Invnorm(.975)
= 1.96

A two-sample z-test for two population proportions is to be performed using the P-value approach. The null hypothesis is $H_0: p_1 = p_2$ and the alternative is $H_a: p_1 \neq p_2$. Use the given sample data to find the P-value for the hypothesis test. Give an interpretation of the p-value.

13) (14) A poll reported that 41 of 100 men surveyed were in favor of increased security at airports, while 35 of 140 women were in favor of increased security. Is there a difference between the proportion of men and women who support an increase in apport security?

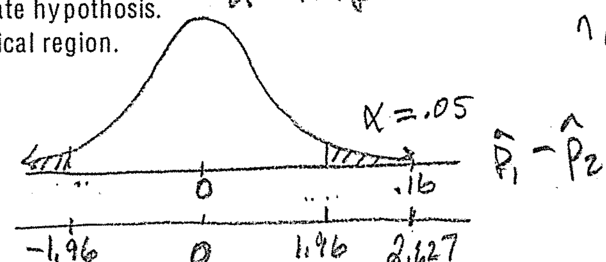
§ 9.2

a) (2 Points) State the null and alternate hypothosis.

b) (3 Points) Graph and shade the critical region.

2-Prop Z Test $X_1 = 41$ $X_2 = 35$
$n_1 = 100$ $n_2 = 140$

$H_0: p_1 = p_2$
$H_1: p_1 \neq p_2$



c) (5 Points) Find a point estimate for the difference in the population proportions, the critical value, and test statistic. Label these on your graph and shade the critical region.

$\hat{p}_1 - \hat{p}_2 = \frac{41}{100} - \frac{35}{140} = .16$ TS: $z = 2.627$
$= .41 - .25 = .16$

d) (5 points) Find the p-value, draw a new graph and label this area. Explain the meaning of this p-value.

P-value = .008014 < .05

If the prop of men and women who support increase security is the same in the populatio the p-value of .008 indicates that it wou be very unusual to get a sample difference of .16. Therefore,

e) (5 Points) Clearly state your conclusion.

Reject H_0 TISE to Reject that the proportion of Men and women who support increased security is the same.

There is sufficient evidence to support that the 41% of men who favor increased airport security is significantly greater than the 25% of women who favor increased security

Estimate the indicated probability by using the normal distribution as an approximation to the binomial distribution.

14) 16) A multiple choice test consists of 40 questions. Each question has 4 possible answers of which one is correct. If all answers are random guesses, estimate the probability of getting at least 20% correct.

X

a) What is the mean and standard deviation of the binomial distribution used for this problem.

$\mu = np = 40 \cdot .25 = 10 \qquad \sigma = \sqrt{npq} = \sqrt{40 \cdot .25 \cdot .75} = 2.739$

b) What proportion do we expect her to get right and what proportioin did she get right in this sample?

$E(x) = \mu = 10$

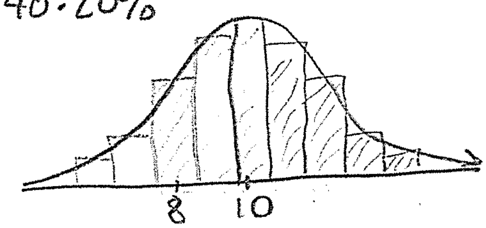
c) What is the proability that we see a sample with 20% or more correct guesses out of 20? 40?
Use the binomial Distribution.

$8 = np = 40 \cdot 20\%$

$P(X \geq 8) = 1 - P(X \leq 7)$

$= 1 - \text{Bincdf}(40, .25, 7)$

$= .8180$



Use the Normal Distribution with a continuity correction.

$P_N(X \geq 7.5) = \text{ncdf}(7.5, 999, 10, 2.739) = .8193$

Perform the appropriate chi-square test and state your conclusion.

15) 16) Decide whether or not the conditions and assumptions for inference with a chi-square test are satisfied. Explain your answer. E > 5 for all cells

✓

(1) SRS (2) (3) Independent

Use the sample data below to test whether car color affects the likelihood of being in an accident. Use a significance level of 0.01. Show your matrix of expected values and clearly state how the result of this hypothosis test applies to this problem.

obs = [A] =	Red	Blue	White
Car has been in accident	28	33	36
Car has not been in accident	23	22	30

$\boxed{\text{MATRX}}$ $\boxed{\text{EDIT}}$ $\boxed{\text{STAT}}$ $\boxed{\text{TEST}}$ χ^2-Test $\boxed{\text{DISTR}}$

(1) TS: $\chi^2 = .429$

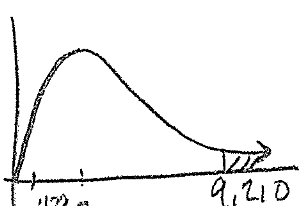
P-value $= .80 = \chi^2\text{cdf}(.429, 999, df$

$df = (r-1)(c-1) = (2-1)(3-1) = 2$

$\alpha = .01$, Rt Tail

(2) CV: $\chi^2_R = 9.210$

fail to Reject H_0



b) $\text{Exp} = [B] = \begin{bmatrix} 28.8 & 31.0 & 37.2 \\ 22.2 & 23.9 & 28.8 \end{bmatrix}$

c) $P(A|R) = \frac{28}{51} = .549$

d) $P(A|W) = \frac{36}{66} = .545$

e) $P(A|B) = \frac{33}{66} = .50$

Provide an appropriate response. a) TINSE to Support that Accidents are dependent on Car Color.

17) Explain what is meant by the coefficient of determination, r^2. Give an example to support your result.

$r^2 = \frac{\text{explained Variation}}{\text{Total Variation}}$ is the proportion of the variation in y that is explained by the regression equation with x.

16) The violent crime rate (number of violent crimes per 100,000 residents) is investigated for nine U.S. cities for the years 1990 and 2000 to see if there has been a change. Use a significance level of 0.05.

T Test on L_3

City	A	B	C	D	E	F	G	H	I	
Violent crime rate in 1990	325	250	199	785	645	259	855	679	301	L_1
Violent crime rate in 2000	379	355	175	925	750	405	1005	902	455	L_2
	-54									

← increase ⇒ difference is Neg

Is there evidence that the violent crime rate has incresed? (Clearly write out each of the 5 steps of your hypothosis test and state your conclusion in terms of the question asked.)

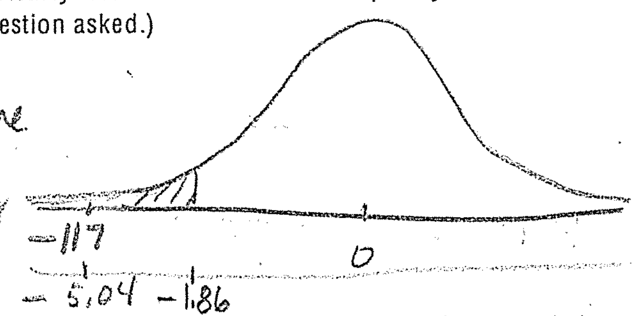
a) (2 Points) State the null and alternate hypothosis.

b) (3 Points) Graph and shade the critical region.

$H_0: \mu_d = 0 \Rightarrow$ No change in crime

$H_1: \mu_d < 0 \Rightarrow$ increase in crime

CV: $t = \text{invT}(.05, 8) = -1.86$



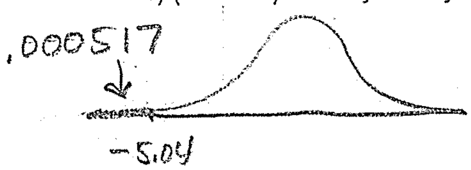
c) (5 Points) Find a point estimate for the population mean of the difference, the critical value, and test statistic. Label these on your graph and shade the critical region.

T Test(L_3). $\bar{d} = -117$ TS: $t = -5.04$

d) (5 points) Find the p-value, draw a new graph and label this area. Explain the meaning of this p-value.

e) (5 Points) Clearly state your conclusion.

P-Value = .000517 = tcdf(-999, -5.04, 8)



there is only a .05% chance of a sample difference this large if in reality there had been No change in homicide rate.

Reject H_0 Accept H_1 there has been an increase in crime.

If you had mistakenly treated these data as two independent samples instead of matched pairs. The significance test would have found no significant differece? Explain why the results are so different.

there is too much variation between cities to show a difference between the means of the Two years.

Find the mean and standard deviation of the given probability distribution.

17) The random variable x is the number of houses sold by a realtor in a single month at the Sendsom's Real Estate office. Its probability distribution is as follows.

mean 3.6 Standard deviation 2.62

Find the probability that a realator sells 5 or more homes. $P(X \geq 5) = .14 + .11 + .21 = .56$

Is it unusual for a realtor to sell 5 or more houses in a month? No, $P(X \geq 5) > .05$

1-VarStat L_1, L_2

Houses Sold (x)	Probability P(x)
0	0.24
1	0.01
2	0.12
3	0.16
4	0.01
5	0.14
6	0.11
7	0.21

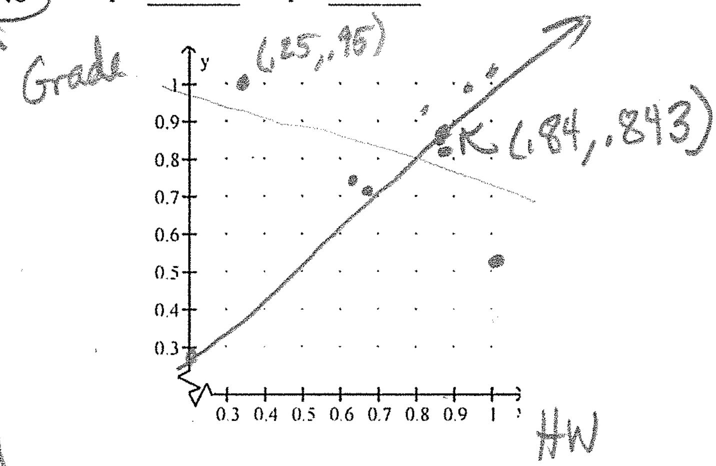
21) (27 Points) The sample data below give the homework grades and final class grades as percentages for 10 statistics students.

a) (3 Points) At the 5% level of significance, do the data provide sufficient evidence that homework score is a good predictor of course grade? abs(−.269) = .269 < .666

Is there a significant linear correlation? Yes No r = −.269 r* = .666

b) (3 Points) Make a scatter plot of this data.

Zoom: zoomStat



Homework Score	Final Grade
.68	.71
.89	.81
.95	.96
.25	.95
.65	.75
.89	.83
.99	.52
.91	.92
.84	.90

Don't use

1-VarStat L2

c) (3 Points) What is the regression equation? $\hat{y} = .946 - .165x$

d) (3 Points) Find the best predicted course grade for student with a homework grade of 0.85. $\bar{y} = .817$

e) (3 Points) **Remove the two outliers.** Without these two students test at the 5% level of significance, if the data provide sufficient evidence that homework score is a good predictor of course grade? n=7

Is there a significant linear correlation? Yes No r = .856 r* = .754

f) (3 Points) Find the equation for the regression line and graph it above. $\hat{y} = .293 + .659x$

g) (3 Points) Find the best predicted course grade for student with a homework grade of 0.85. .853

h) (3 points) Place this point on your graph and label it.

i) (3 points) Find the slope of the regression line and what it means.

j)(3 points) Discuss the significance of the outliers and whether the data should be analyzed with or without the outliers.

i) m = .659 means that for each percent increase in HW grade the % grade in the class will increase by .659 percent.

j) Both or without the Outliers have a significant impact on Regression Results.

Perform the indicated goodness-of-fit test.

18) Among the four northwestern states, Washington has 51% of the total population, Oregon has 30%, Idaho has 11%, and Montana has 8%. A market researcher selects a sample of 1000 subjects, with 450 in Washington, 340 in Oregon, 150 in Idaho, and 60 in Montana. At the 0.05 significance level, test the claim that the sample of 1000 subjects has a distribution that agrees with the distribution of state populations. What kind of sampling method is this researcher using?

State	W	O	I	M
Obs.	450	340	150	60
Exp.	510	300	110	80

Exp = %·1000

$H_0: O=E$
$H_1: O \neq E$
χ^2 Gof test
df = 3
TS: $\chi^2 = 31.938$
P-value $= 5.39 \times 10^{-7}$

CV: $\chi^2_R = \text{INV}\chi^2(\text{Rt}, df=3, \text{Sig}=.05)$
$\chi^2_R = 7.815$

TISE to Reject that the proportion of subjects in each state agrees with the dist of the states populations.

Provide an appropriate response.

19) Four independent samples of 100 values each are randomly drawn from populations that are normally distributed with equal variances. You wish to test the claim that $\mu_1 = \mu_2 = \mu_3 = \mu_4$.

i) If you test the individual claims $\mu_1 = \mu_2$, $\mu_1 = \mu_3$, $\mu_1 = \mu_4$, . . . , $\mu_3 = \mu_4$, how many ways can you pair off the 4 means? ${}_4C_2 = 6$ $\quad \mu_2 = \mu_3, \mu_2 = \mu_4$

ii) Assume that the tests are independent and that for each test of equality between two means, there is a 0.99 probability of not making a type I error. If all possible pairs of means are tested for equality, what is the probability of making at least one type I errors? P(Mistake) = .01 P(right) = .99

iii) If you use analysis of variance to test the claim that $\mu_1 = \mu_2 = \mu_3 = \mu_4$ at the 0.01 level of significance, what is the probability of not making a type I error?

ii) P(At least one) = 1 − P(No Mistakes in 4 trials)
$= 1 - (.99)^4 = .0585$

iii) P(Type I) = .01 So P(Not making a type I) = .99

20) Of the 23 first-year male students at State U. admitted from Jim Thorpe High School, 8 were offered baseball scholarships and 7 were offered football scholarships. The University admissions committee looked at the students' composite ACT scores (shown in the table), wondering if the University was lowering their standards for athletes. Assuming that this group of students is representative of all admitted students, what do you think? Test an appropriate hypothesis and state your conclusion.

Composite ACT Score		
Baseball	Non-athletes	Football
25	21	22
22	27	21
19	29	24
25	26	27
24	30	19
25	27	23
24	26	17
23	23	

$H_0: \mu_B = \mu_N = \mu_F$ Mean incoming ACT is the Same for Football, baseball and Non athletes

H_1: At least one group has a different mean

Input Data into Stat Edit L_1, L_2, L_3 then Run $\boxed{\text{Stat}}$ $\boxed{\text{Test}}$ ANOVA

20) TS: 4.563 = F
P Value = .0233
dfN = 2
dfD = 20
CV: $F^* = 3.493$

TISE to support the claim that the Mean ACT Score of Atletes and Nonathletes are different.

Perform the required t–test for the slope of the regression line and state your conclusion.

17) (~~15~~ Points)The sample data below give the homework grades and final class grades as percentages for ~~10~~ 9 statistics students.

Homework Score	Final Grade
.68	.71
.89	.81
.95	.96
.25	.95
.65	.75
.89	.83
.99	.52
.91	.92
.84	.90

a) (3 Points) At the 5% level of significance, do the data provide sufficient evidence that homework score is a good predictor of course grade? 3

No, p-value = .758 > .05

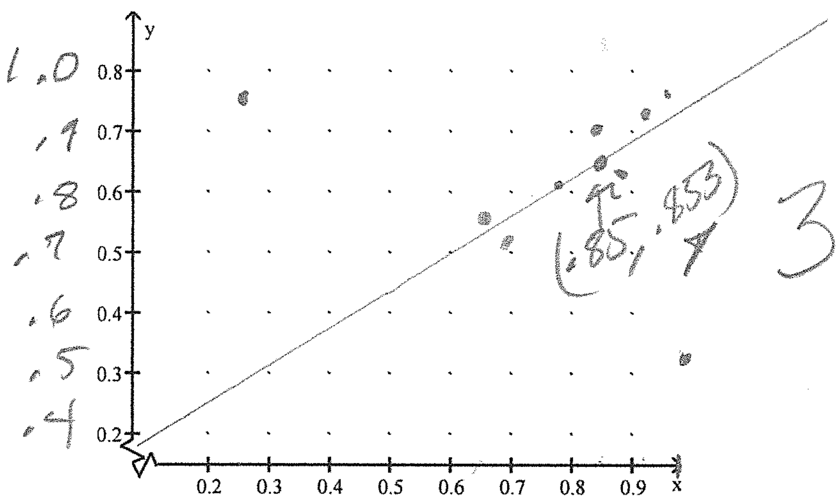
b) (~~3 points~~) Is there a significant linear correlation? Yes No r = -.267 r* = .666

c) (3 Points) Find the best predicted course grade for student with a homework grade of ~~0.85~~. $\hat{y}$ = .817 3

d) (3 Points) Make a scatter plot of this data.

Y = .946 − .165X

3 .70



e) (3 Points) Remove the two outliers. Without these two students test at the 5% level of significance, if the data provide sufficient evidence that homework score is a good predictor of course grade? 3

Is there a significant linear correlation? Yes No r = .846 r* = .754

f) (3 Points) Find the equation for the regression line and graph it. Y = .293 + .659X 3

g) (3 Points) Find the best predicted course grade for student with a homework grade of 0.85. .70

$\hat{p}$ = .293 + .659(.85) = .853 3

h) (3 points) Place this point on your graph and label it. (.85, .853) 3

The outliers are Not representative of most students the data should be analyzed without them. 3

27