Stat 10 Lew FINAL IS TUESDAY 12/14/04 3:00pm-6:00pm

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

1. A study of 150 first year college students, selected at random after their first full year of college, gives the following results for high school ranking (RANK) first year GPA (grade point average) and high school SAT score:

Average GPA =	3.19;	Standard deviation GPA= 0.49
Average RANK=	2.04;	Standard deviation RANK = 1.13
Average SAT =	1291;	Standard deviation SAT = 163

Correlation coefficient for GPA and RANK= -.33
Correlation coefficient for GPA and SAT= .24
Correlation coefficient for RANK and SAT= -.42

Assume the SAT scores had a minimum of 910 and a maximum of 1580 and are normally distributed. RANK has a minimum of 0 and a maximum of 6 and is not normally distributed. GPA has a minimum of 1.56 and a maximum of 3.98 and it is normally distributed.

(a) Of the 3 correlation coefficients given to you above, please identify which pair has the strongest correlation and which has the weakest correlation.

Strongest = RANK + SAT = -.42
Weakest = GPA + SAT = .24

(b) A student is interested in regressing first year college GPA on SAT. Using the information at the top of the page, please find the regression equation. Clear identify the slope, intercept, x and y variables.

Y on X

slope = $r \cdot \frac{SD_y}{SD_x} = (.24)\left(\frac{.49}{163}\right) = .0007$ Y = GPA X = SAT

intercept = $\bar{y} - (b\bar{x}) = 3.19 - (.0007 \cdot 1291) \approx 2.26$

equation is $\underset{(y)}{GPA} = \underset{(\text{intercept})}{2.26} + \underset{(\text{slope})}{.0007}\ \underset{x}{(SAT)}$

(c) Please interpret the values of slope and intercept you calculated in part (b) in plain English.

The intercept suggest that if SAT = 0 GPA = 2.26
The slope suggests that for a one-unit ~~change~~ increase in SAT there is a .0007 increase in predicted GPA

(d) Peter and Paul are USC undergraduates who were selected in this study. Peter had an SAT score of 1040 and Paul had an SAT score of 820, what are their predicted GPAs?

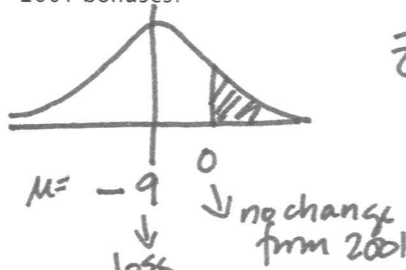
Peter = 2.26 + (.0007 * 1040) ≈ 3.00

Paul - you should not predict b/c that's extrapolation an SAT of 820 is below the range of 910 - 1580

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

2. Lawyers frequently receive a "year-end bonus" because law firms are partnerships and the money earned is shared among partners. There are approximately 1,000,000 lawyers in the U.S. and year 2002's "year-end bonus", when calculated as a percentage change of the 2001 "year-end bonus", is normally distributed with a mean year-end bonus of -9% (a decrease, 2002 was a bad year compared to 2001) and a standard deviation of 16%. SHOW YOUR WORK FOR FULL CREDIT.

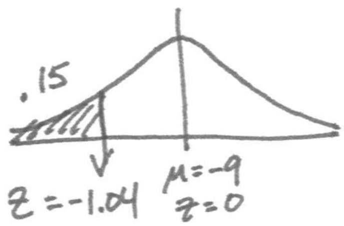
a) What proportion of lawyers received year-end 2002 bonuses that were as large as or larger than their year-end 2001 bonuses?



$$z = \frac{\bar{Y}-\mu}{\sigma} = \frac{0-(-9)}{16} = \frac{+9}{16} \cong .56$$

$1 - .7123 = .2877$ or 28.77%

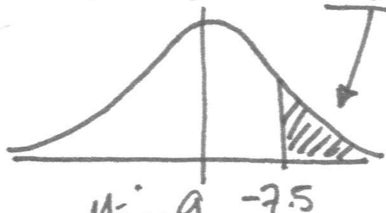
b) A simple random sample of 100 lawyers has an average year-end bonus at the 15th percentile, what is the actual value of that average? $n = 100$



$$z = \frac{Y-\mu}{\sigma/\sqrt{n}} \text{ so } -1.04 = \frac{Y-(-9)}{16/\sqrt{100}}$$

solve for Y $\cong -10.66\%$

c) A simple random sample of 100 lawyers has an average year-end bonus of -7.5%, what is the chance of getting a sample average of -7.5% or higher? $n=100$



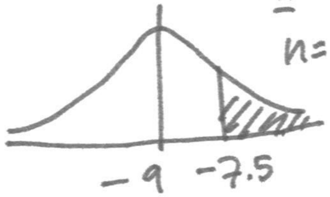
$$z = \frac{Y-\mu}{\sigma/\sqrt{n}} = \frac{-7.5-(-9)}{16/\sqrt{100}} \cong .94$$

there is .1736 in the shaded area or 17.36%

d) What is the chance that a lawyer, selected at random, will have a year-end bonus of -7.5% or higher?

Ch. 6

$n=1$



$$z = \frac{Y-\mu}{\sigma} = \frac{-7.5-(-9)}{16} = .09$$

area is .4641 or 46.41%

e) What percentage of lawyers have year end bonuses between -1% and +5%?

Ch. 6

$$z_{+5} = \frac{Y-\mu}{\sigma} = \frac{+5-(-9)}{16} \cong .88$$

$$z_{-1} = \frac{-1-(-9)}{16} \cong .50$$

subtract areas so $.3085 - .1894 = .1191$ or 11.91%

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

3. The IQ scores of adult humans (age 18 and over) are approximately normally distributed with a mean of 100 and a standard deviation of 15. The highest IQ of a currently living adult as reported by the Guiness Book of World Records belongs to Marilyn vos Savant who scored a 186 (nearly six standard deviations above average). The maximum IQ score is 200, values estimated above it are deemed unreliable. The lowest score on record is 40.

Ch. 6

(a) How low is the lowest 5% of all IQ scores (that is, at or below what IQ score is the lowest 5%) How high is the highest 10% of IQ scores (that is, at or above what IQ Score is the highest 10%)?

so $\underset{(z)}{-1.64} = \frac{y-\mu}{\sigma} = \frac{y-100}{15}$

solve $y = 75.4$

so $z = \frac{y-\mu}{\sigma}$

$+1.28 = \frac{y-100}{15}$

$y = 119.2$

(b) A simple random sample of size 256 is drawn from the adult human population. What is the chance that the sample average will exceed 101? $n=256$

Ch. 18

$z = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} = \frac{101-100}{15/\sqrt{256}} \approx 1.07$ shaded area is .1423 or 14.23%

(c) How large of a sample would a researcher need to select to insure that he or she is within plus or minus 1 IQ point of the population mean IQ with 99% confidence?

so they want $\bar{Y}$ to be w/i ± 1 IQ point so it's like a rearranged CI. i.e. $\bar{Y} \pm \underbrace{z\left(\frac{\sigma_Y}{\sqrt{n}}\right)}_{\text{set this}}$ part equal to 1.0 (margin of error)

$z=2.57$ or 2.58 will give .5% in each tail (so a total of 1% ∴ $100-99=1\%$)

$1 = 2.57\left(\frac{15}{\sqrt{n}}\right)$ solve for n $\approx 1,486$

or use formula $n = z^2 \cdot (\sigma^2)/(MOE)^2$

(d) A simple random sample of 256 USC students is drawn from the adult human population. The sample average is 97 and the sample standard deviation is 30. Please test the hypothesis that USC students have lower IQ scores than the average human. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule.

$H_0: \mu = 100$ $\alpha = .05$

$H_1: \mu < 100$

d.f. $\approx$ 250 (closest lower)

$z = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ or could use $t = \frac{\bar{Y}-\mu}{s/\sqrt{n}} = \frac{97-100}{30/\sqrt{256}} = -1.60$

(← easier b/c σ = known)

do not reject

.10 > p-value > .05

⇓

$\frac{97-100}{15/\sqrt{256}} = -3.2$

p-value is .0007 < .05 so reject NULL

4. Investors ask about the relationship between returns on investments (the money you make by investing your money) in the United States and on investments overseas. Below is a table of total returns on investments on U.S. and overseas stocks over a 10 year period. Assume overseas returns and US returns are both normally distributed.

	Year	Overseas % Return	U.S. % Return
Average	1991.5000	9.8100	16.0300
Standard Deviation	2.7386	15.6493	12.6810

(a) Suppose the correlation, r, of the U.S. and overseas returns is -.3239. Please describe the relationship between U.S. and overseas returns in words, using r to make your description more precise.

If r = -.3239 the relationship is negative (so as one goes up the other one goes down) and WEAK an r = -.3239 is closer to zero than to -1.0

(b) Find the regression line of overseas returns on U.S. returns. Please interpret the values of the slope and of the intercept of this line.

Y on X

$$\text{slope} = r \cdot \frac{SD_Y}{SD_X} = -.3239\left(\frac{15.6493}{12.6810}\right) \approx -.3997$$

$$\text{intercept} = \bar{y} - (b\bar{x}) = 9.8100 - (-.3997 * 16.03) = 16.2172$$

OVERSEAS RET. (Y) = 16.2172 − .3997 (US RETURNS) X

→ interpret.
SLOPE - for every one unit ~~change~~ increase in US Returns there is a -.3997 ~~decrease~~ change in overseas
INTERCEPT If US returns were = 0 the overseas ≅ 16.22%

(c) In 1993, the return on U.S. stocks was 10.1%, what was the predicted return on overseas stocks for that year? Suppose I told you that the actual return on Overseas Stocks that year was 32.9%? Why are they so different?

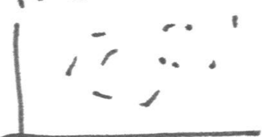
Assuming 10.1 is not extrapolation then

$$16.2172 - .3997(10.1) = 12.18$$

over seas return

If the actual was 32.9%

It's just the case that the line is not going exactly through all of the points and we underpredicted overseas for that particular year. The correlation -.3239 is pretty weak



so we don't expect great predictive power

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

5. Does salt cause high blood pressure? One large study was done at 52 centers in 32 counties. Each center recruited 200 subjects in 8 age- and sex- groups. Salt intake was measured, as well as blood pressure and several possible confounding variables. After adjusting for age, sex, and the confounding variables, 25 of the centers found a positive association between diastolic pressure and salt intake; 27 found a negative association. Do the data support the theory that salt causes high blood pressure? Answer yes or no, and explain briefly.

No, the data support association, that is, the variables appear to be related but the direction (~~direction~~ positive vs negative) is not clear because of conflicting results.

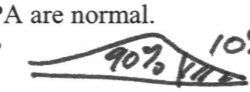
6. A study on pre-meds, selected at random, gives the following results for the medical college admissions test (MCAT) and undergraduate GPA (grade point average):

Average GPA: 3.3;	Standard deviation	= 0.4	Min. GPA = 2.5 Max. GPA = 4.3
Average MCAT: 10;	Standard deviation	= 1.1	Min MCAT=7.0 Max MCAT=13

Correlation coefficient = 0.65

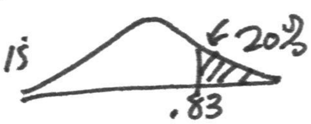
a) Suppose the percentile of one student's GPA is at the 90th percentile. Predict the student's percentile on the MCAT. The scatter diagram is football shaped and the MCAT and the GPA are normal.

If 90th percentile AND normal then $z = +1.28$



so $z_{MCAT} = r * z_{GPA}$ which is $z_{MCAT} = .83 = (.65)(1.28)$
(Y) (X)

a $z = .83$ is



so the student is at the 80th percentile of MCAT scores

b) Please construct the regression equation for MCAT on GPA. Please interpret the values of the slope and intercept in plain English.

Y X

slope $= r \cdot \frac{SDy}{SDx} = (.65)\left(\frac{1.1}{0.4}\right) = 1.7875$ Intercept $= \bar{y} - bx = (10)-(1.7875 * 3.3)$
$= 4.1013$

equation is $\underset{(Y)}{MCAT} = \underset{(\text{intercept})}{4.1013} + \underset{(\text{slope})}{1.7875}(GPA)$

slope ⇒ so for a one point increase in GPA see a 1.7875 increase in MCAT

if GPA=0, MCAT would be 4.1

c) Your wealthy cousin, a graduate of USC, is really, really dumb. He told you his GPA is a whopping 1.9. He's now thinking of going to medical school. What is his predicted MCAT score?

Should not calculate this, it's extrapolation. There is no data to support it b/c 1.9 is less than the lowest GPA used to calculate the regression equation

Stat 10 Lew FINAL IS TUESDAY 12/14/04 3:00pm-6:00pm

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

$\sigma = 16$

7. The pregnancy duration of human females (age 18 and over) is approximately normal with a mean of 266 days and a standard deviation of 16 days. It is believed that older pregnant women have longer pregnancy durations. A simple random sample of 121 older pregnant women is drawn from the population of all pregnant women. The average pregnancy duration for the sample is 267 days and the sample standard deviation is 35.

(a) Please test the hypothesis that older women have longer pregnancy durations than the average woman. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule. $\alpha = .05$

$H_0: \mu = 266$ can use z test because σ = known = 16

$H_1: \mu > 266$

$z = \frac{267-266}{16/\sqrt{121}} \doteq +.68$ so p-value is $\approx .25$

p-value > α (.25 > .05)

Do not reject. Not stat sig. No evidence of longer durations.

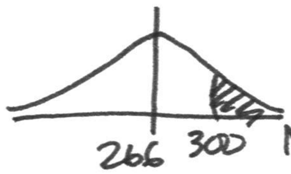
If use t-test then $t = \frac{\bar{Y}-\mu}{s/\sqrt{n}} = \frac{267-266}{35/\sqrt{121}} = +.31$

df = 120 p-value > .10

SO NOT STAT SIGNIF. DO NOT REJECT NULL (same conclusion as z test)

(b) What proportion of pregnancies have durations as long as or longer than 300 days?

Ch. 6



$z = \frac{Y-\mu}{\sigma} = \frac{300-266}{16} = +2.13$

area is about .0166 or 1.66%

(c) Suppose a researcher is only interested in studying the proportion of pregnancies that have durations as long as or longer than 300 days. How large of a sample would he or she have to select in order to properly invoke the Central Limit Theorem to create confidence intervals or test hypotheses about these 300+ day pregnancies?

if 1.66% are longer than 300 days (from part B above) and if this is p then 1-p = 100 - 1.66 = 98.34 to make np > 10 then n should be at least 603.

8. Here are two statistics on all persons who consider themselves computer programmers in 1999:

$820,000 dollars per year
$141,000 dollars per year

Which one of these numbers is the mean salary from computer programming and which one is the median salary from computer programming in 1999? Assume the samples were of good quality.

The mean is 820,000

The median is 141,000

Explain your choice in the space below. Be brief. This is not a long answer.

Because salaries cannot be < 0 but they can be extremely high, outlier will probably be on the right and the distribution will be right skewed so mean > median

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

9. High Bias and High Variance are both considered undesirable features of certain sample statistics (such as a sample mean for example). You are working with a team on a marketing study, a sample of size 100 is drawn. One of the variables you are interested in is the average time spent on the internet on any day. You plan to construct confidence intervals and perform some unspecified hypothesis tests. Studies always have problems, and today you have your choice: High Bias or High Variance. Which one would you rather deal with and why?

HIGH VARIANCE. High variance can be addressed by gathering a larger sample recall $s/\sqrt{n}$ or $\sqrt{\frac{pq}{n}}$ the larger n is, the smaller the variability.
Bias – no matter how large your sample gets bias will give you incorrect estimates.

10. Los Angeles International Airport handles an average of 6,000 international passengers an hour. Suppose 80% can pass through primary security, but the rest are detained for interrogation by the FBI. And suppose the FBI can handle 1,500 passengers an hour without unreasonable delays for travelers and extra costs to the airlines (due to missed flights and connections).

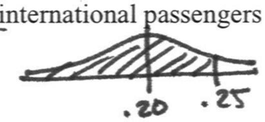
a. Over break, it is expected that as many as 8,000 international passengers will arrive per hour. When that occurs, what is the expected proportion of passengers who will be detained?

.20 is the expected proportion (20%)

b. Referring to part a, find the approximate chance that less than 2,000 out of 8000 international passengers will be detained? $2000/8000 = .25 = \hat{p}$

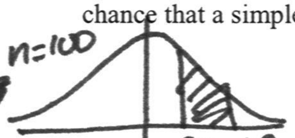
$n = 8000$

$z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}} = \frac{.25-.20}{\sqrt{\frac{(.2)(.8)}{8000}}} \approx +11.18$



virtually 100%

c. Suppose the FBI decides to randomly sample passengers in order to speed up the screening process. What is the chance that a simple random sample of 100 will have between 22 and 28 passengers detained by the FBI?



$\frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$

$z_{.28} = \frac{.28-.20}{\sqrt{\frac{(.20)(.80)}{100}}} = +2.0$ area is .9722

shaded area is .9722 − .6915 $\boxed{= .2857}$

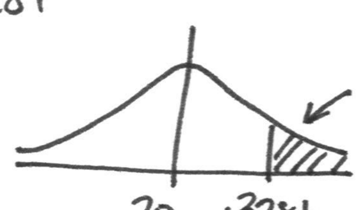
$z_{.22} = \frac{.22-.20}{\sqrt{(.20)(.80)/100}} = +.50$ area is .6915

d. Certain ethnic/racial groups appear to be detained at much higher rates than others. Suppose a human rights organization sends 64 persons who appear to be of middle eastern origin through the airport and 21 are detained for interrogation. Please test the hypothesis that persons of middle eastern origin are detained in higher proportions than the typical traveler. State a null and an alternative hypothesis, perform a test, state a p-value and explain your result (do you reject or not reject the null and why). Use a 5% level of significance as your decision rule. You may treat the 64 as if it were a simple random sample and it is of reasonable size. $\alpha = .05$

$H_0: p = .20$ $\hat{p} = 21/64 = .3281$

$H_1: p > .21$

$z = \frac{.3281-.20}{\sqrt{\frac{(.20)(.80)}{64}}} = \frac{.1281}{.05} = +2.56$



area is .0052

.0052 < .05 so reject the null

the results are statistically significant. Persons of M.E. origin are detained at higher rates.

11. A marketing survey interviewed 1000 adults selected at random from the population of all U.S. adults. Of the adults, 529 said they currently own a personal computer. When asked about the manufacturer of their computer, 144 of them said "Dull", 115 of them said "Compact", 175 of them said "some other company" and the rest of them said "I don't know". The mean time of ownership (in months) for the 529 was 12.9 with a standard deviation of 8.7.

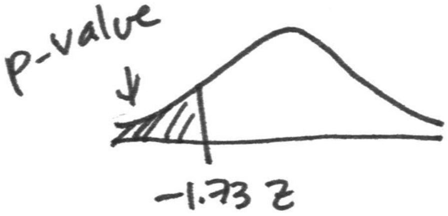
(a) A Compact executive saw the survey and is now upset, he believes that the survey was poorly done and argues that Compact's true market share is 25% (i.e. he thinks that 25% of all adults who own computers own a Compact) and cannot be nearly as low as the survey suggests.

Let's help the executive out. Please test the hypothesis that Compact's market share is actually 25%. Use a 5% level of significance as your decision rule. State the null hypothesis, the alternative hypothesis, perform a test, give a p-value, and state your conclusion in plain English: would you reject the null and on the basis of your test result do you also think the survey was poorly done?

$H_0: p = .25$ $\qquad \hat{p} = \frac{115}{529} = .2174$ $\qquad \alpha = .05$

$H_1: p < .25$

$$\text{Test } z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{.2174 - .25}{\sqrt{\frac{(.25)(.75)}{529}}} = \frac{-.0326}{.0188} = -1.73 = z$$



p-value is .0418

.0418 < .05 so REJECT THE NULL

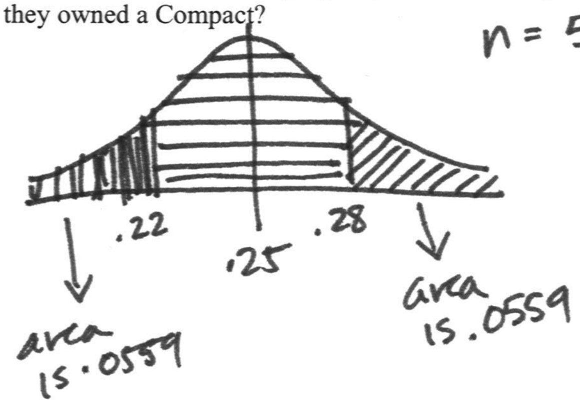
THIS RESULT IS STATISTICALLY SIGNIF.

The executive is wrong, Compact does have less than 25%.

(b) Suppose Compact's true market share is REALLY 25%. What is the chance that among 529 computer owners you would get less than 22% of them saying they owned a Compact? What is the chance that you would get between 22% and 28% saying they owned a Compact? What is the chance that you would get at least 28% saying they owned a Compact?

$n = 529$, true $p = .25$ samp. dist normal

$$SD = \sqrt{\frac{(.25)(.75)}{529}}$$



$$z_{.22} = \frac{.22 - .25}{\sqrt{\frac{(.25)(.75)}{529}}} = -1.59$$

$$z_{.28} = \frac{.28 - .25}{\sqrt{\frac{(.25)(.75)}{529}}} = +1.59$$

So area is $1.0 - (.0559 + .0559) = .8882$

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

12. A student issued the following command for analysis variable FAMILYSEI

```
. summarize familysei, detail

                         familysei
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%         22.5              0
10%         28.2              0       Obs                1428
25%         37.3              0       Sum of Wgt.        1428

50%         63.5                      Mean            68.9631
                        Largest       Std. Dev.      39.68948
75%         92.3            175
90%        129.5          180.6       Variance       1575.255
95%        145.7          194.4       Skewness       .6213121
99%        166.9          194.4       Kurtosis       2.792152
```

Please answer the following questions based on the Stata results for variable FAMILYSEI. You may round the numbers given above to one or two decimal places. For example, 39.68948 can be rounded to 39.7 or 39.69

A. Using the Stata results above, please calculate the range, the interquartile range, and list the values of the quartiles (i.e. Q1, Q2 and Q3) (10 points)

range = 194.4 − 0 = 194.4

IQR = ~~[illegible]~~ 92.3 − 37.3 = 55.0

$Q_1 = 37.3$, $Q_2 = 63.5$, $Q_3 = 92.3$

B. Is the distribution for this variable skewed? (circle one): YES NO (1 point)

Please justify your response in the space below. If you think it is skewed, please indicate the direction (left or right skewed) of the skewness. (4 points)

this is right skewed  where the mean > median

it is also known as positively skewed.

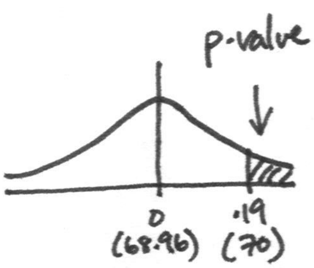
C. Suppose we could treat the 1,428 persons as a population, from it, a simple random sample of size 51 is drawn, assume it is large enough to overcome the non-normality in the population. The mean familysei for the sample is 70. Please test the hypothesis that something is wrong with the sample because it is biased towards higher values. Please state a null and alternative hypothesis, perform the appropriate test and use an alpha=.05 as your decision rule. Please state clearly whether you reject or do not reject the null and give a brief verbal (simple English) conclusion.

$H_0: \mu = 68.96$ $\sigma = 39.68948$ we don't know s (sample SD)

$H_1: \mu > 68.96$ so we could use Z as a test

$\alpha = .05$

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} = \frac{70 - 68.96}{39.6895/\sqrt{51}} = +.19$$



p-value is .4247 $\underset{\text{p-value}}{.4247} > \underset{\alpha}{.05}$

DO NOT REJECT THE NULL, NOT STAT SIG., THIS MEANS THERE IS NOT ENOUGH EVIDENCE TO SUPPORT THE IDEA THAT THIS SAMPLE IS SOMEHOW WEIRD OR BIASED

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

13. There are 20,000 restaurants in the County of Los Angeles, 50% of them received a letter grade of "A" during inspections, 40% received either a B or a C grade and 10% failed their inspections. Restaurant grades are not normally distributed. My financial adviser, the Oracle, has hired you as a temporary personal assistant. Your job is to schedule his next 16 dinners (Oracle never eats at home). Unfortunately, you didn't know about the rating system and you never eat out because you don't have the money. So you listened to your best friend and picked 25 restaurants at random with replacement from an internet database of the 20,000 restaurants in Los Angeles. The Oracle will give you +3 points if you choose "A" restaurants, +1.25 points if you choose "B" or "C" restaurants, and -20 points if you choose a restaurant with a failing grade. Treat your restaurant selections as if they were a simple random sample of restaurants.

A. Please construct the probability distribution for this problem

	A	B or C	Failing	
outcome	+3	+1.25	-20	
probability	.50	.40	.10	= 1.0

B. What is the expected value for the probability distribution? What is the expected value for the mean restaurant scores for a sample of 25 restaurants selected at random with replacement (please assume that the sample is large enough to use Z)?

$\mu = EV = \sum_{i=1}^{n} x_i p_i = (3 * .50) + (1.25 * .40) + (-20 \times .10)$

$\mu = 1.5 \quad .50 \quad -2.0 \quad = 0$

for a sample of size 25 μ is still = 0

C. What is the standard deviation for this distribution?

$\sigma = \sqrt{\sum (x_i - \mu)^2 (p_i)} = \sqrt{(3-0)^2(.5) + (1.25-0)^2(.4) + (-20-0)^2(.10)}$

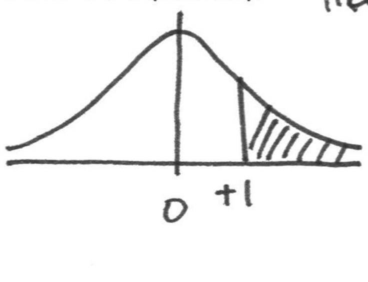
$= 6.7175$

D. What is the standard deviation for the average score of a sample of 25 restaurants?

$\frac{\sigma}{\sqrt{n}} = \frac{6.7175}{\sqrt{25}} = 1.344$

Ch. 18 E. To convert your temporary job into a permanent job, you must have accumulated an average of at least +1 points from the Oracle after picking 25 restaurants for him. What's your chance of getting an average of at least +1 points after picking 25 restaurants? If it is not possible to calculate the chance, please write "not possible" below and explain why.

like a R.S. of size 25 from a sampling dist. w/ mean $\mu = 0$ and $\sigma = 1.344$



$z = \frac{Y - \mu}{\sigma/\sqrt{n}} = \frac{1 - 0}{1.344/\sqrt{25}} = .74$

the chance is about .23 or 23%

Stat 10 Lew FINAL IS TUESDAY 12/14/04 3:00pm-6:00pm

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

14. You know that every UCLA student will definitely get a job after graduation. The only uncertainty is the salary. Suppose this is what you know about the job prospects of UCLA students after graduation:

There is a 35% chance that the salary will be $20,000 per year; a 45% chance that it will be $90,000 per year; and a 20% chance that it will be $40,000 per year.

a. Construct the probability distribution for the salary of UCLA students

OUTCOME	20000	40000	90000	
PROBABILITY	.35	.20	.45	= 1.0

b. Find the expected value and the standard deviation.

Expected Value = $\mu = (20000 \times .35) + (40000 \times .20) + (90000 \times .45) = 55{,}500$

$\sigma = \sqrt{(.35)(20000 - 55{,}500)^2 + (40000 - 55{,}500)^2(.20) + (.45)(90{,}000 - 55{,}500)^2}$

$= 32{,}011.72$

c. Suppose a random sample of 36 UCLA students who have graduated is drawn. What is their expected average salary? What is the standard deviation of the salary for the sample of 36? What is the chance that the sample average will exceed $57,000

Their expected average is the same as the pop. average of 55,500. The standard deviation of the average for a sample of size 36 is $\sigma/\sqrt{n}$ or $32{,}011.72/\sqrt{36}$ $= 5{,}335.29$

the chance of getting $\bar{Y} > 57{,}000$ is

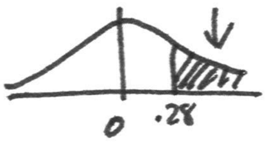
d. Suppose a random sample of 49 UCLA students who have graduated is drawn. What is their expected average salary? What is the standard ~~error~~ of the salary? What is the chance that the sample average will exceed $57,000 ?

Deviation of average

$z = \dfrac{57{,}000 - 55{,}500}{32{,}011.72/\sqrt{36}}$

$= .28$

so the chance is .3897

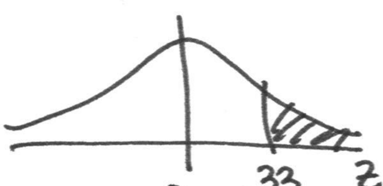


Expected average is still 55,500

SD of the average is $\sigma/\sqrt{n} = 32{,}011.72/\sqrt{49}$
$= 4{,}573.1029$

Chance > 57,000 now is

$z = \dfrac{57{,}000 - 55{,}500}{32{,}011.72/\sqrt{49}} = .33$ about .3707%



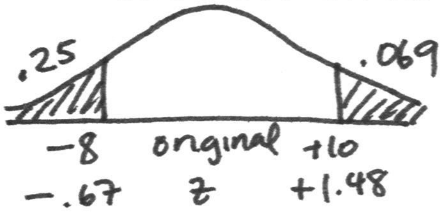
$(1 - .6293) = .3707$

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

15. An investment advisor advertises that he is "da very best" and proclaims that his clients will "retire in style" if they hand over all of their savings to him. However, after an audit, the Securities and Exchange Commission (SEC) learned that only 6.9% of his clients had percentage returns greater than or equal to 10%. Additionally, 25% of his clients reported percentage returns as low as or lower than -8%. One client had lost all of her savings (a total loss of 100%) but another client was at the maximum with returns of +91%. Noting the variation in returns, the SEC assumes that his clients' percentage returns are normally distributed (and you should too) (Note: outliers can occur in normal distributions).

Using the information above please answer the following:

A. Please calculate the mean (μ) and standard deviation (σ) of all of his client's percentage returns.



original distance = 10 - -8 = 18%

z score distance = 1.48 - -.67 = 2.15 z

if 2.15z = 18% then

2.15σ = 18% so

σ = 18/2.15 = 8.3721 to find μ use z

$-.67 = \dfrac{-8-\mu}{8.3721}$ or $+1.48 = \dfrac{+10-\mu}{8.3721}$ solve to get -2.39%

B. Please list and clearly identify the values of the five number summary for the percentage returns of all of his clients. Please show the numeric computations when appropriate.

Minimum = -100%

Q1 or 25th percentile = -8%

Q2 or 50th or Median = mean = -2.39%

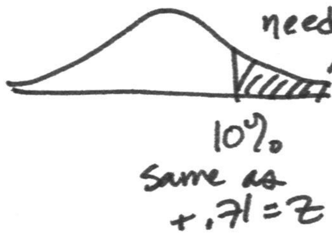
MAX = +91%

Q3 is 75th or

$+.67 = \dfrac{y-(-2.39)}{8.3721}$

Q3 is +3.22%

C. Hoping to maintain his claim of "retiring in style" the advisor believes that by changing his investment decisions for his clients, it will increase the percentage of clients with percentage returns in excess of 10% to 24% instead of the 6.9% reported by the SEC. If the standard deviation (calculated in part A, we will assume it is correct) remains the same, what new mean would he need to achieve his goal?



needs to be 24%
.24 now

so $.71 = \dfrac{10\% - \mu}{8.3722\%}$

new μ should be 4.06%

D. Suppose the SEC performed a poor audit and in fact, the advisor has client percentage returns that are normally distributed with mean 2.5% and standard deviation 6%. Please calculate the probability that at least one of three randomly selected returns are ≥ -4%. Please assume independence.

← qualifying area, what's the chance a stock comes from here?

$z = \dfrac{-4-2.5}{6} = -1.083$

probability is like .86 or 86%

so to pick 3 and get AT LEAST ONE that is ≥ -4%

that's 1-(chance of getting none) so $1-(.14)^3 = 99.73\%$

virtually 100% chance.

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

16. Here is a correlation table:

	insurance	unemploy	poverty	income	incarceratio	violence
insurance	1.0000					
unemploy	0.4872	1.0000				
poverty	0.1106	0.5222	1.0000			
income	0.4115	-0.1345	-0.7099	1.0000		
incarceratio	0.4473	0.4389	0.4756	-0.1276	1.0000	
violence	0.5284	0.4381	0.3883	-0.0014	0.8144	1.0000

strongest pos = .8144
strongest neg = −.7099
weakest = −.0014

Please identify the strongest positive correlation, the strongest negative correlation and the weakest correlation in the table.

17. The UCLA housing office wants to estimate the mean monthly rent for one bedroom apartments in the 90024 zip code. A random sample of size 36 is selected from the zip code area and the sample mean is found to be $1,100. The sample standard deviation is $200.

$n-1=35$ $t=2.030$ for 95% conf.

a. Please construct a 95% confidence interval for the mean monthly rent of all one bedroom apartments in the zip code.

$$\bar{y} \pm t^*_{n-1}(s/\sqrt{n}) \Rightarrow 1{,}100 \pm 2.03\left(\frac{200}{\sqrt{36}}\right) \Rightarrow 1{,}100 \pm 67.667$$

b. Please construct a 99% confidence interval for the mean monthly rent of all one bedroom apartments in the zip code.

$$\bar{y} \pm t^*_{n-1}\left(\frac{s}{\sqrt{n}}\right) \Rightarrow 1{,}100 \pm 2.725\left(\frac{200}{\sqrt{36}}\right) \Rightarrow 1{,}100 \pm 90.83\bar{3}$$

t is 2.725 for 99% confidence and $n-1=35$

c. What is the minimum sample size needed to generate a 95% confidence interval which has a margin of error of ± $50?

DONT WORRY ABOUT THIS ONE.
No t sample size problem on the final

18. Mark ONE of the columns

True	False	Statement
	X	You can determine whether a distribution is right-skewed if you know its standard deviation and its mean
X		Standard Deviations are either zero or are positive values
X		Selection Bias is a result of mistakes made by the person or persons who design a study which involves sampling
	X	Non-response Bias is caused by researchers who fail to write a survey questionnaire properly
X		Observational studies differ from randomized controlled experiments in that the researchers do not assign the subjects to treatment or control groups
X		Association may point to causation, but it is not the same as causation
	X	An extremely large biased sample (e.g. > 10,000) generates better estimates population parameters than a small random sample (e.g. a little over size 100)
X		Quantitative (or numerical) variables can be discrete or continuous
	X	If X is a continuous random variable which is normally distributed with a mean of 100 and a standard deviation of 15 then the probability that X>115 is 0.5.
X		In order to have a valid probability distribution, the sum of the probabilities must equal to 1 or 100% and the probabilities themselves cannot be negative or greater than 100%.

FINAL WILL BE HELD IN THE LECTURE HALL Boelter 3400

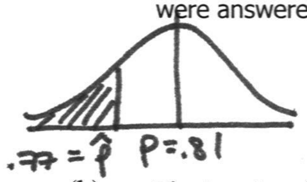
19. The Los Angeles Department of Transportation fines Taxicab companies that take too long to respond to customers. Each year, the Department of Transportation tests each taxicab company by making 100 independent calls to that company (you can treat this as a random sample of size 100) and uses this sample to check the proportion (or percentage) of calls that are answered in a timely manner where within 15 minutes is defined as timely.

$p = .81$ true pop. prop.

Suppose that a taxi company's true proportion of calls answered in a timely manner is .81 (or 81%). That is, in the course of thousands and thousands of calls, the taxicab company would be able to successfully answer 81% of the calls within fifteen minutes.

TYPO

(a) What is the chance that a random sample of 100 calls made to this taxi company will only show 77% OR LESS (sorry) were answered in a timely manner? $n=100$



$$z = \frac{.77 - .81}{\sqrt{\frac{(.81)(.19)}{100}}} = -1.02$$

chance is .1539

(b) The Los Angeles Department of Transportation doesn't know that this company has a true proportion of calls answered in a time manner of .81 or 81%. Instead, it makes 100 independent calls to the company (treat it as a random sample) and finds that only 75% were answered in a timely manner. Please construct a 98% confidence interval for the proportion of calls answered in a timely manner.

$np > 10$
$nq > 10$

$$\hat{p} \pm z^* \left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \Rightarrow .75 \pm 2.33\left(\sqrt{\frac{(.75)(.25)}{100}}\right)$$

$$\Rightarrow .75 \pm .101$$

20. A bank wonders whether omitting the annual credit card fee for customers who charge at least $2400 in a year would increase the amount charged on the credit card. The bank makes this offer to a simple random sample of 25 of its existing credit card customers. It then compares how much these customers charge this year with the amount they charged last year. The mean percentage change is +12% with a standard deviation of 6%. The amounts are nearly normally distributed.

(a) Is there a statistically significant increase in the mean amount charged? Please state a null and alternative hypothesis, perform the appropriate test, state a p-value and use an alpha=.05 as your decision rule. Give you conclusion in plain English please – do you reject or not reject the null and what does this mean?

$H_0: \mu = 0$ (no change)

$H_1: \mu > 0$ (positive change - an increase)

test is t b/c $\sigma = ?$

$df = 25 - 1 = 24$

$$t = \frac{12 - 0}{6/\sqrt{25}} = 10$$ do a lookup, 10 is "off the chart" so

$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

we know that the p-value is much less than .05 so Reject the null, this is STAT. SIG. which means the new ~~method~~ offer seems to make people charge more.

(b) Please construct a 99% confidence interval for the mean percentage change.

$$\bar{y} \pm t^*_{n-1}\, s/\sqrt{n} \Rightarrow 12 \pm 2.797\left(\frac{6}{\sqrt{25}}\right) \Rightarrow 12 \pm 3.36$$

$n - 1 = 24$