1. On the normal curve find the area
a) between -.88 and 1.62
b) to the right of .44

Find z if the area
c) to the right is .62
d) between .0 and z is .2088
2. On the t-curve with 26 degrees of freedom

Find the area
a) to the right of -1.058
b) between 1.706 and 2.479

## Find $t$ if the area

c) to the left of $t$ is .0025
d) between -1.315 and $t$ is .895
3. The Wellness Center Poll: WeLikeStudents $U$ built a new Wellness and Exercise Center one year ago on Campus. The university administration wants to know whether the students at their school like the new Wellness Center. The administration sends a letter to all students who live on campus to get their opinion. The letter starts with the following questions: Have you been to our brand new beautiful Wellness Center?

To reply to the poll students are asked to send the letter back to the President of WeLikeStudents U.

Name any biases in this poll. Give both the statistical names and an English description.
4. A statistics professor is interested in the price of statistics textbooks. In a sample of 20 statistics books the average price was $\$ 145$ with a standard deviation of $\$ 20$. Find an $80 \% \mathrm{t}-$ confidence interval for $\mu$ the true average price of a statistics textbook.
4.a) Is $\$ 155$ too high for the true average price of a statistics textbook? Explain.

5
omer wants to know how many French fries are in a box of Super Sized fries to within 2 fries with $99 \%$ probability. If the SD for the number of fries in Super Sized boxes is 9 , how many French fry boxes must be checked?
6. A tomato juice manufacturer wants to add spicy tomato juice to the products it sells. To make the juice spicy, it must add precisely one half teaspoon of pepper to a can of tomato juice. If it adds more it will be too spicy and if adds less than one half teaspoon of pepper it will not be spicy enough. The manufacturer has a machine which is known to have a standard deviation of .1 teaspoons for putting in pepper. A sample of 20 cans averages .55 teaspoons of pepper to it. At level of significance .01 has the correct amount of spice been added? Do a One Sample Ztest.
6.a) What is your conclusion in English?
7. The US Chamber of Commerce believes that $30 \%$ of Americans like to go to Starbucks. In a random sample of 100 Americans 25 of them like to go to Starbucks. With $=.10$ is the Chamber of Commerce correct?

Is this a One Sample Z-test, a One Sample t-test, or a One Sample Test for Proportion (= a One Sample Binomial Test? (You do not have to carry out the test.)
10. In the northern hemisphere is the correlation between temperature and latitude negative, zero or positive?
11. Find the correlation, the regression line and either the root mean square error or the standard error of estimate for the following pairs of data.

| $X$ | $Y$ |
| ---: | :---: |
| 2 | 33 |

12. Multiple regression - see the Excel Printout below:

The following regression was run on consumption of heating oil (in gallons) as a function of average daily temperature and amount of attic installation (in inches).

REGRESSION STATISTICS<br>Multiple R . 86466<br>R Square . 74764<br>Adjusted R Square . 70558<br>Standard Error 70.469<br>Observations 15

ANOVA

|  | df | SS | MS | F | Signif F |
| :--- | :---: | :---: | :---: | :--- | :---: |
| Regression | 2 | 176543 | 88271 | 17.77 | .0003 |
| Residual(Error) | 12 | 59592 | 4966 |  |  |


|  | Coefficient | Std-Err | t Stat | Pvalue |
| :--- | :---: | :---: | :---: | :---: |
| Intercept | 563.11 | 9.11 | 61.83 | .0000 |
| TEMP | -4.51 | .89 | -5.03 | .0003 |
| INSUL | -29.32 | 6.59 | -4.45 | .0008 |

a) What is the regression equation?
b) What is the numerical value of $R$

1. On the normal curve find the area
a) between -.88 and 1.62
Ans. $=.9474-.1894=.7580$
b) to the right of .44
Ans. $=1-.6700=.3300$

Find z if the area
c) to the right is .62

Ans. $\mathrm{z}=-.31$, large area to the right means z is negative, finding . 6200 in the body of the z table gives $\mathrm{z}=.31$
d) between 0 and z is .2088

> Ans. $\mathrm{z}=.55$, z could be positive or negative, assume z is positive. If between 0 and z is .2088 adding the area to the left of $0=.5$ gives an area to the left of z of .7088 . From the body of the normal table .7088 gives $\mathrm{z}=.55$

## 2. On the t-curve with 26 degrees of freedom

Find the area
a) to the right of -1.058 Ans. Area=. $85=1-.15$
b) between 1.706 and 2.479 Ans. Area=. $04=$ area to the right of 1.706 minus the area to the right $2.479=.05-.01$

## Find t if the area

c) to the left of $t$ is .0025 Ans. $=-3.067$, small area to the left means $t$ is negative, .0025 and 26 df give $t=3.067$ but final answer is negative.
d) between -1.315 and t is .895

Ans. $\mathrm{t}=2.779$, t must be positive since there is a large area between -1.315 and $t$ From $t$-table the area to the left of -1.315 is .10 , between -1.315 and $t$ is .895 , so to the right of $t$ is $1-.10-.895=.005$ Finding .005 and 27 df gives $\mathrm{t}=2.779$
3. The Wellness Center Poll: WeLikeStudents $U$ built a new Wellness and Exercise Center one year ago on campus. The university administration wants to know whether the students at their school like the new Wellness Center. The administration sends a letter to all students who live on campus to get their opinion. The letter starts with the following questions: Have you been to our brand new beautiful Wellness Center? you love our new Wellness Center. To reply to the poll students are asked to send the letter back to the President of WeLikeStudents U .

Name any biases in this poll. Give both the statistical names and an English description.
There is Selection Bias or Coverage Bias since students not living on Campus are excluded from the poll. Selection Bias occurs when part of the population is never sampled.

There is Non-Response Bias in this poll because some students will not respond to the letter and will not send anything back to the administration. Non-Response bias happens when you have decided to sample someone and you do not get their opinion.

There is Response Bias in this poll because the questions are slanted. Even if the new Wellness Center is really nice, the word beautiful should not be in the question love our new Wellness Center suggests a particular example Response Bias happens when for whatever reason you get an answer from a person which is different from what they really believe.
4. A statistics professor is interested in the price of statistics textbooks. In a sample of 20 statistics books the average price was $\$ 145$ with a standard deviation of $\$ 22$. Find an $80 \%$ tconfidence interval for $\mu$ the true average price of a statistics textbook.

This is a sample standard deviation (=s) since the average and standard deviation are connected to the sample size $\mathrm{n}=20$. Because this is a sample standard deviation this is a t -confidence interval.

## $\bar{X}$

With an $80 \%$ confidence interval there is a .8 probability in the middle of the curve. This gives a .1 probability in the right tail. From the $t$-table with $20-1=19$ degrees of freedom the multiplier is $\mathrm{t}=1.328$

CI is $145 \pm 1.328 * 22$
CI is $145 \pm 6.533$ or 138.467 to 151.533
If $\mu$ is the true average price of a statistics textbook (=population average) then we believe that $\$ 138$. $\quad 151.533$ with $80 \%$ confidence.
4.a) Is $\$ 155$ too high for the true average price of a statistics textbook? Explain.

Since we think $\$ 138.467 \quad 151.533$. We think $\mu=\$ 155$ is too high for the true average price of a statistics textbook, it is outside the confidence interval.
7. The US Chamber of Commerce believes that $30 \%$ of Americans like to go to Starbucks. In a random sample of 100 Americans 45 of them like to go to Starbucks. With $=.10$ is the Chamber of Commerce correct?

Is this a One Sample Z-test, a One Sample t-test, or a One Sample Z-test for Proportion (= a One Sample Binomial Test? (You do not have to carry out the test.)

For each of the 100 Americans we want to know if they are they like to go to Starbucks. Each American will say yes or no. Here 45 Americans said yes and 55 Americans said no. This is Binomial Data and we would do a One Sample Z-test for Proportion. The problem did not ask the test to be performed but if it did this would be the solution.

$$
\begin{aligned}
& \text { Null p=. } 3 \\
& \text { Alt p } 3 \text { If too many or too few like Starbucks we would reject the null. } \\
& \mathrm{Z}(\text { Stat })=(.45-.3) /[(.3)(1-.3) \quad 100)]=3.273 \\
& \mathrm{Z}(\mathrm{R} \quad 10 \text { in half and look up } .05 \text { in } \\
& \text { normal table to get } \mathrm{z}(\mathrm{rej})=1.645 \&-1.645
\end{aligned}
$$

Since the Critical Values are $\pm 1.645$ and the test statistic $Z(S t a t)=3.273$ lies outside the critical values we reject the null. We would conclude that the US Chamber of Commerce is not correct.
8. Past experience indicates that the standard deviation for the number of bathing suits owned by teenagers is 3 in both California and Florida. A movie agent claims that the mean number of bathing suits is higher in California than in Florida. A sample of 20 California teenagers owned on average 7 bathing suites, while 20 Florida teenagers averaged 5 bathing suites owned. At $=.025$ what is your opinion? Do a Two Sample Z-test.

The standard deviations come from past experience so we are willing to believe that they are $\quad 1$ and $\quad 2$ hence the two sample Z-test.
Null $1=2$ On average both are the same.
Alt. $\quad 1>2$ Population 1 Mean (California) is larger than the Population 2 Mean (Florida) $Z($ Stat $\left.)=\left(7 \quad 3^{2} / 20+3^{2} / 20\right)\right]=2.108$
Z(R 25 in half and look up .025 in normal table to get $\mathrm{z}(\mathrm{rej})=1.9 .42$ EMCI48<1 001207.772
9. Santa Claus needs to know who has more toys, boys or girls. A sample of 20 boys had an average of 60 toys with a standard deviation of 10 toys, while a sample of 25 girls averaged 40 toys with an SD of 15 . Carry out a Two Sample t-test at $=.01$.

The averages and standard deviations are connected to the sample sizes $n_{1}=20$ and $n_{2}=25$ so the SDs are sample SDs giving a Two Sample t-test.
Null $1=2$ On average both are the same.
Alt. 12 Population 1 Mean (Boys) is different than the Population 2 Mean (Girls) The problem does not
11. Find the correlation, the regression line and either the root mean square error or the standard error of estimate for the following pairs of data.

| $X$ | $Y$ |
| ---: | :---: |
| 2 | 33 |
| 4 | 17 |
| 5 | 30 |
| 5 | 34 |
| 6 | 21 |
| 8 | 45 |

To compute the correlation:
Column1 Column2 Column3 Column4 Column5 Column6 Column7

|  | $X$ | Y |  | $\bar{X})$ | $\left(\begin{array}{ll}Y & \bar{Y}\end{array}\right)$ | ( $X$ | $\bar{X})(Y$ | $\bar{Y})$ | ( $X$ | $\bar{X})^{2}$ |  | $\bar{Y})^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 2 | 33 | -3 |  | 3 |  | -9 |  |  | 9 |  | 9 |
|  | 4 | 17 | -1 |  | -13 |  | 13 |  |  | 1 |  | 69 |
|  | 5 | 30 | 0 |  | 0 |  | 0 |  |  | 0 |  | 0 |
|  | 5 | 34 | 0 |  | 4 |  | 0 |  |  | 0 |  | 16 |
|  | 6 | 21 | 1 |  | -9 |  | -9 |  |  | 1 |  | 81 |
|  | 8 | 45 | 3 |  | 15 |  | 45 |  |  | 9 |  | 25 |
| $\bar{X}$ | 5 | $\bar{Y}$ |  |  |  |  | XY 40 |  | SSX | 20 | SSY | 500 |

$\operatorname{SSXY} \quad\left(\begin{array}{ccc}X & \bar{X}\end{array}\right)\left(\begin{array}{ll}Y & \bar{Y}\end{array}\right) \quad\left(\begin{array}{ll}X_{i} & \bar{X}\end{array}\right)\left(\begin{array}{ll}Y_{i} & \bar{Y}\end{array}\right)$,
2

$$
\begin{aligned}
& S_{X} \sqrt{\frac{S S X}{n 1}} \sqrt{\frac{20}{61}} \quad 2 \quad S_{Y} \\
& \text { Corr } \quad r \frac{\sqrt{\frac{S S Y}{n 1}}}{\sqrt{\frac{500}{61}}} 10 \\
& S_{X} \square S_{Y}
\end{aligned} \frac{40 /\left(\begin{array}{ll}
6 & 1
\end{array}\right)}{2 \square 0} \quad .410
$$

For the regression equation $Y$ a $b X$ we have:
the slope has the formula
we have from above $\bar{X} \quad 5, \bar{Y} \quad 30, S_{x} \quad 2, S_{y} \quad 10, r \quad .4$ so that the slope $b \quad r \frac{S_{Y}}{S_{X}} \quad$ (.4) $\frac{10}{2} \quad 2$
or $b \frac{S S X Y}{S S X} \quad \frac{40}{20} \quad 2$
12. Multiple regression - see the Excel Printout below:

The following regression was run on consumption of heating oil Y (in gallons) as a function of average daily temperature $\mathrm{X}_{1}$ and amount of attic installation $\mathrm{X}_{2}$ (in inches).

REGRESSION STATISTICS
Multiple R . 86466
R Square . 74764
Adjusted R Square . 70558
Standard Error 70.469
Observations 15

ANOVA

|  | df | SS | MS | F | Signif F |
| :--- | :---: | :---: | :--- | :--- | :--- | :---: |
| Regression | 2 | 176543 | 88271 | 17.77 | .0003 |
| Residual(Error) | 12 | 59592 | 4966 |  |  |


|  | Coefficient | Std-Err | t Stat | Pvalue |
| :--- | :---: | :---: | :---: | :---: |
| Intercept | 563.11 | 9.11 | 61.83 | .0000 |
| TEMP | -4.51 | .89 | -5.03 | .0003 |
| INSUL | -29.32 | 6.59 | -4.45 | .0008 |

This is a two sided test so split $=.05$ in half. Look up .025 in the $t$-table with error degrees of freedom, here 12df. $t($ Reject $)=2.179$. Reject the null since $t(S t a t)=-5.03$ is outside the critical values $\pm 2.179$
f) Test if Insulation has a negative effect on heating oil, take

