

343 Proficiency Exam Practice Problems

1. On the normal curve find the area

- a) between -0.88 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? Don't 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.

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% t-confidence interval for μ 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. A loyal McDonald's customer 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 Z-test.

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 $\alpha = .10$ 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.)

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 $\alpha = .025$ what is your opinion? Do a Two Sample Z-test.

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 $\alpha = .01$.

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
4	17
5	30
5	34
6	21
8	45

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

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		

	<u>Coefficient</u>	<u>Std-Err</u>	<u>t Stat</u>	<u>Pvalue</u>
Intercept	563.11	9.11	61.83	.0000
TEMP	-4.51	.89	-5.03	.0003
INSUL	-29.32	6.59	-4.45	.0008

- What is the regression equation?
- What is the numerical value of R^2 ?
- What is the numerical value of SSE?
- What is the numerical value of the typical prediction error?
- Test $b_1 = 0$ vs $b_1 \neq 0$ Temperature, take $\alpha = .05$.
- Test if Insulation has a negative effect on heating oil, take $\alpha = .05$.
- Interpret the results of the tests in question (e) and question (f) in English.

343 Proficiency Exam Practice Problems Solutions

1. On the normal curve find the area

- a) between -0.88 and 1.62 $\text{Ans.} = .9474 - .1894 = .7580$
b) to the right of $.44$ $\text{Ans.} = 1 - .6700 = .3300$

Find z if the area

- c) to the right is $.62$ $\text{Ans. } z = -.31$, large area to the right means z is negative, finding $.6200$ in the body of the z table gives $z = .31$
d) between 0 and z is $.2088$ $\text{Ans. } z = .55$, 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 $z = .55$

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

Find the area

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

Find t if the area

- c) to the left of t is $.0025$ $\text{Ans.} = -3.067$, small area to the left means t is negative, $.0025$ and 26df give $t = 3.067$ but final answer is negative.
d) between -1.315 and t is $.895$ $\text{Ans. } t = 2.779$,
 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 27df gives $t = 2.779$

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? Don't 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. The phrase Don't you 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% t-confidence interval for μ 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 $n=20$. Because this is a sample standard deviation this is a t-confidence interval.

$$\bar{X} \pm t*s/\sqrt{n}$$

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 $t=1.328$

$$\text{CI is } 145 \pm 1.328 * 22 / \sqrt{20}$$

$$\text{CI is } 145 \pm 6.533 \text{ or } 138.467 \text{ to } 151.533$$

If μ is the true average price of a statistics textbook ($=$ population average) then we believe that $\$138.467 \leq \mu \leq \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 \leq \mu \leq \151.533 . We think $\mu = \$155$ is too high for the true average price of a statistics textbook, it is outside the confidence interval.

5. A loyal McDonald's customer 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?

$$n = z^2\sigma^2/E^2 = 2.57^2 9^2 / (2^2) = 133.75 \quad \text{Round up and look at 134 boxes.}$$

To find z put .99 in the middle, which means .01 in both tails and .005 in one tail. Finding .005 in the body of the normal table gives $z = -2.57$

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 for putting in pepper with an SD of .1 teaspoons. 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?

Since the SD is known it is the Population SD = σ . Since we have the Population SD this is a Z-test.

Null $\mu = .5$

Alt $\mu \neq .5$ Do not want too much or too little.

$$Z(\text{Stat}) = (.55 - .5) / [.1 / \sqrt{(20)}] = 2.236$$

$$Z(\text{Reject}) = \pm 2.57$$

Split $\alpha = .01$ in half and look up .005 in normal table to get $z(\text{rej}) = 2.57 \& -2.57$

Since the Critical Values are ± 2.57 and the test statistic $Z(\text{Stat}) = 2.236$ lies inside the critical values we accept the null.

6.a) What is your conclusion in English?

Since we accepted the null we believe the null is true. We believe that $\mu = .5$ and conclude that the amount of pepper in the tomato juice is correct.

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 $\alpha = .10$ 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.

Null $p=.3$

Alt $p \neq .3$ If too many or too few like Starbucks we would reject the null.

$$Z(\text{Stat}) = (.45-.3)/[(.3)(1-.3)/\sqrt{(100)}]=3.273$$

$$Z(\text{Reject})=\pm 2.57$$

Split $\alpha=.10$ in half and look up .05 in normal table to get $z(\text{rej})=1.645 \& -1.645$

Since the Critical Values are ± 1.645 and the test statistic $Z(\text{Stat})=3.273$ 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 $\alpha = .025$ 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 σ_1 and σ_2 hence the two sample Z-test.

Null $\mu_1 = \mu_2$ On average both are the same.

Alt. $\mu_1 > \mu_2$ Population 1 Mean (California) is larger than the Population 2 Mean (Florida)

$$Z(\text{Stat}) = (7-5)/[\sqrt{(3^2/20 + 3^2/20)}]=2.108$$

$$Z(\text{Reject})=\pm 2.57$$

$\alpha=.025$ in half and look up .025 in normal table to get $z(\text{rej})=1.96$

Since the Critical Value is $+1.96$ and the test statistic $Z(\text{Stat})=2.108$ lies to the right of the critical value we reject the null. We conclude that California teenagers have more bathing suits.

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 $\alpha = .01$.

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 $\mu_1 = \mu_2$ On average both are the same.

Alt. $\mu_1 \neq \mu_2$ Population 1 Mean (Boys) is different than the Population 2 Mean (Girls)

The problem does not say who has more so make it a \neq alternative.

$$t(\text{Stat}) = (60-40)/[\sqrt{(10^2/20 + 15^2/25)}]=5.345$$

$$df = \frac{\left[\frac{10^2}{20} + \frac{15^2}{25}\right]^2}{\frac{1}{19}\left[\frac{10^2}{20}\right]^2 + \frac{1}{24}\left[\frac{15^2}{25}\right]^2} = 41.784$$

This calculation for the degrees of freedom can be done with either a TI-83 or TI-84 or TI-Nspire.

Round the df down to 41 degrees of freedom. It is better to round down than to round up.

Split $\alpha=.01$ in half and look up .005 with 41 df in the t-table to get $t(\text{rej})=2.701$

$$t(\text{Reject}) = \pm 2.701$$

Since the Critical Values are ± 2.701 and the test statistic $t(\text{Stat})=5.345$ lies outside the critical values we reject the null. We conclude that Boys and Girls do not have the same number of toys.

If your t-table does not have 41df you can use 40 df or even 30 df if your table does not have 40 df.

10. In the northern hemisphere is the correlation between temperature and latitude negative, zero or positive?

The latitude at the equator is 0° and it is warm there. In the northern hemisphere as latitude increases you are going further north and the temperature gets colder and colder. As latitude goes up the temperature goes down. There is a (strong) negative correlation between temperature and latitude.

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	$(X - \bar{X})$	$(Y - \bar{Y})$	$(X - \bar{X})(Y - \bar{Y})$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
2	33	-3	3	-9	9	9
4	17	-1	-13	13	1	169
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	225
$\bar{X} = 5$	$\bar{Y} = 30$			$SSXY = 40$	$SSX = 20$	$SSY = 500$

$$SSXY = \sum (X - \bar{X})(Y - \bar{Y}) = \sum (X_i - \bar{X})(Y_i - \bar{Y}), \quad SSX = \sum (X_i - \bar{X})^2 = \sum (X_i - \bar{X})(X_i - \bar{X})$$

$$SSY = \sum (Y_i - \bar{Y})^2 = \sum (Y_i - \bar{Y})(Y_i - \bar{Y})$$

$$S_x = \sqrt{\frac{SSX}{n-1}} = \sqrt{\frac{20}{6-1}} = 2 \quad S_y = \sqrt{\frac{SSY}{n-1}} = \sqrt{\frac{500}{6-1}} = 10$$

$$Corr = r = \frac{SSXY / (n-1)}{S_x \cdot S_y} = \frac{40 / (6-1)}{2 \cdot 10} = .4$$

For the regression equation $Y = a + bX$ we have:

the slope has the formula $b = r \frac{S_y}{S_x} = \frac{SSXY}{SSX} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$

and $a = \bar{Y} - b\bar{X}$ where b is the slope.

For the data set $(X, Y) = (2, 33) (4, 17) (5, 30) (5, 34) (6, 21) (8, 45)$

we have from above $\bar{X} = 5$, $\bar{Y} = 30$, $S_x = 2$, $S_y = 10$, $r = .4$ so that the slope

$$b = r \frac{S_y}{S_x} = (.4) \frac{10}{2} = 2$$

or $b = \frac{SSXY}{SSX} = \frac{40}{20} = 2.$

The intercept is $a = \bar{Y} - b\bar{X} = 30 - 2*5 = 20$. So the regression equation is $Y = a + bX = 20 + 2X$

To compute the RMSE you must do the calculation

PredictionError				
X	Y	$\hat{Y} = 20 + 2X$	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
2	33	24	9	81
4	17	28	-11	121
5	30	30	0	0
5	34	30	4	16
6	21	32	-11	121
8	45	36	9	<u>81</u>

$$SSE = 420 = \sum (Y_i - \hat{Y}_i)^2$$

$$RMSE = \sqrt{\frac{SSE}{n}} = \sqrt{\frac{420}{6}} = 8.3666 \quad \text{This is a typical prediction error.}$$

The standard error of estimate is defined as $s = \sqrt{\frac{SSE}{n-2}}$ so in this data set $s = \sqrt{\frac{420}{6-2}} = 10.247.$

There are alternative formulas for correlation and regression. They will give equivalent results.

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 X_1 and amount of attic installation 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

- What is the regression equation?
- What is the numerical value of R^2 ?
- What is the numerical value of SSE?
- What is the numerical value of the typical prediction error?
- Test $b_1 = 0$ vs $b_1 \neq 0$ Temperature, take $\alpha = .05$.
- Test if Insulation has a negative effect on heating oil, take $\alpha = .05$.
- Interpret the results of the tests in question (e) and question (f) in English.

- The regression equation is $Y = 563.11 - 4.51X_1 - 29.32X_2$.
- The numerical value of R^2 is .74764. This is a high R^2 .
- The value of SSE is 59592?
- What is the numerical value of the typical prediction error? The standard error of estimate is 70.469. This is also the approximate value of the RMSE.
- Test $b_1 = 0$ vs $b_1 \neq 0$ Temperature, take $\alpha = .05$.
 Null $b_1 = 0$ (Y = a + 0X) X and Y are not related.
 Alt. $b_1 \neq 0$ X and Y have some relationship.
 t(Stat) = -5.03

This is a two sided test so split $\alpha = .05$ in half. Look up .025 in the t-table with error degrees of freedom, here 12df. $t(\text{Reject}) = 2.179$. Reject the null since $t(\text{Stat}) = -5.03$ is outside the critical values ± 2.179

f) Test if Insulation has a negative effect on heating oil, take $\alpha = .05$.

Null $b_1 = 0$ ($Y = a + 0X$) X and Y are not related.

Alt. $b_1 < 0$ X and Y have a negative or inverse relationship.

$t(\text{Stat}) = -4.45$

This is a one sided test so keep $\alpha = .05$ as it is. Look up .05 in the t-table with error degrees of freedom, here 12df to get $t = 1.782$. $t(\text{Reject}) = -1.782$. Reject the null since $t(\text{Stat}) = -4.45$ is to the left of the critical value -1.782 .

g) Interpret the results of the tests in question (e) and question (f) in English.

It is concluded that Temperature has a relationship with the amount of heating oil.

It is also concluded that Insulation has a negative effect on the amount of heating oil.