Violating Conditions in Significance Testing
(Teacher Notes)

Julie Hicks, Woodbridge High School, Woodbridge, VA
jjhicks@home.com

John Lieb, The Roxbury Latin School, West Roxbury, MA
jlieb19@yahoo.com

NCSSM Statistics Leadership Institute
July 2001

The goal of the following activities is to provide students with a more concrete idea of what it means for a test of significance, specifically the *t*-test, to be ***robust***. The following materials describe these activities and the necessary ideas that these activities address.

Contents

NOTE: These teacher notes refer to a "Student Materials" handout that includes worksheets for students to complete. It would be helpful to have that packet when reading these teacher notes.

Introduction

Every significance test has conditions that must be satisfied in order for the test to be of value. What happens when these conditions are not satisfied? Well, that depends on which condition is not satisfied and how severely the condition is violated. The activities we will describe examine what happens when conditions for significance tests are violated to varying degrees. Conditions will be violated for two varieties of the *t*-test that are covered in the AP syllabus.

Suppose a test of significance has a significance level of α. When the null hypothesis is true and the conditions for the test are satisfied exactly, then the probability of the test leading to a Type I error equals α. In some situations, violations of the conditions of the test do not cause the probability of a Type I error to differ greatly from α. When this happens, the test is called ***robust***. In other words, if the conditions for a test of significance are not met, does it still behave "as advertised"?

Through simulation on the calculator students will examine how the probability of a Type I error compares to the stated significance level α when conditions for the *t*-test are violated. Before using the calculator, students will complete a worksheet that demonstrates and leads them through the idea behind the simulation. (These worksheets are in the document **"Violating Conditions in Significance Testing (Student Materials)"**)

The two specific tests of significance we will look at are

- **One sample *t*-test for the population mean**
- **Linear Regression *t*-test for the population slope**

Before doing these activities, students should have a firm understanding of the mechanics and details of significance testing, including the idea of a Type I error. Here are two possible ways these materials could be used:

- After teaching each of the two tests, the corresponding activity on violating a condition of that test could be completed.
- After teaching all of the tests of significance that a class will do, take a look at these activities as a whole and see what happens when conditions are violated in these two tests.

Considering the problem of whether α is equal to the probability of a Type I error, when conditions are violated, can also lead to consideration of important topics such as binomial distributions, confidence intervals, and power. These ideas may serve as a good review at the end of the year. (See section titled **EXTENSIONS**)

One-sample *t*-test for the population mean

Overview of necessary conditions

Consider the *t*-test for a population mean. Two necessary conditions for this test are

(1) The data must be a simple random sample from the population of interest.

(2) The population from which the sample is taken must be normally distributed.

We will refer to the second condition listed above as the **normal population** condition. It is unlikely that in reality the population distribution will be exactly normal, and most textbooks tell us that, even if the sample size is small, the *t*-test is robust to departures from population normality. But what does ***robust*** mean? To what does it refer?

We can think about the idea of a ***robust*** test in terms of significance levels and Type I errors.

- ***If*** the population is normally distributed, then the probability that the *t*-test will lead to a Type I error (when the null hypothesis is true) ***is equal*** to the stated significance level. In order to call the *t*-test ***robust***, the probability of the test making a Type I error should be close to the stated significance level ***even if*** the population is ***not*** normally distributed. How robust the *t*-test is will depend primarily on two factors: the sample size and how far away from normal the population is.

- Another way to put it: ***If the condition of normality for the t-test is not met***, the probability of a *t*-test leading to a Type I error generally does not equal the stated significance level α. This probability can be close to the stated significance level when the normality condition is not met; if it is close then we say the *t*-test is ***robust*** against the violation of the normality condition.

(The above line of reasoning also applies to the linear regression *t*-test for the population slope. The major aspect that changes when looking at the *t*-test for the population slope is the different conditions that test requires.)

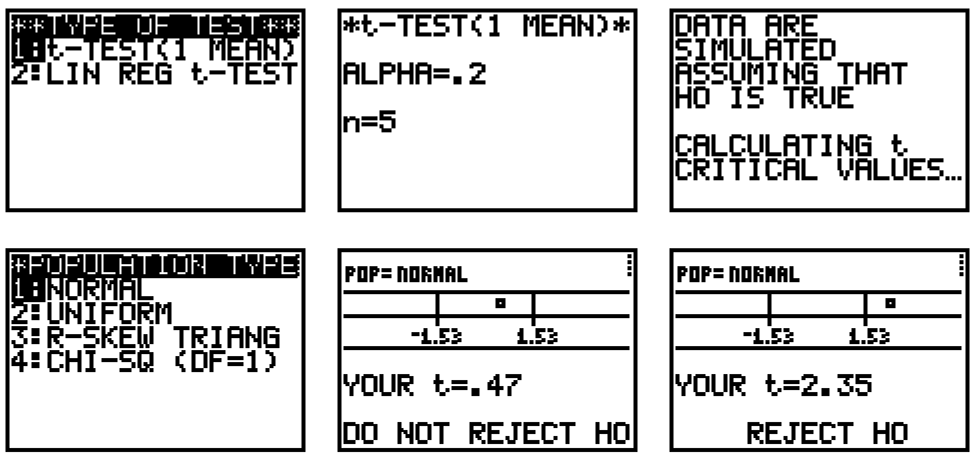
The TI-83 program **ROBUSTT** simulates *t*-tests for one mean in the context of different types of population distributions and for different sample sizes in order to see how the stated significance level α and the probability of a test leading to a Type I error compare. Using the program is integrated into the worksheet **"Violating a condition of the *t*-test for one mean"** which can be found in the **Student Materials** handout.

Program ROBUSTT for the *t*-test for one mean

The calculator generates samples of size *n* from one of four known populations. It then calculates the *t*-score for each sample and makes the decision on whether we should reject the null hypothesis or not. The data are generated by using the value of μ from the null hypothesis, so if we reject the null hypothesis we know we have made a Type I error (rejecting the null hypothesis when it is true.) The user inputs the significance level (alpha) and the sample size.

- **See "Student Materials" for the worksheet "Violating a condition of the *t*-test for one mean" that explains what the program ROBUSTT is doing. It is recommended that you look through this worksheet in conjunction with these Teacher Notes.**

Screen Shots from program ROBUSTT for the *t*-test for the population mean



- ❑ Keep pressing **ENTER** to generate more samples of size 5. Press **ON** to quit the program.
- ❑ Keep track on the appropriate data recording worksheet of whether or not the null hypothesis is rejected. (See **Student Materials** handout.)
- ❑ The table below displays the results of using the calculator program **400** times for each of four population distributions for α **= 0.20** and ***n* = 5**. While simulation results will vary, it is interesting to note that the probability of rejecting the null hypothesis gets further away from the stated significance level of α **= 0.20** as the population deviates further from the normal distribution. This indicates the *t*-test may not be robust when the population is strongly skewed and the sample size is small.

POPULATION	NORMAL		UNIFORM		R-SKEW TRIANGULAR		CHI-SQ (DF=1)	
DECISION	Reject Ho	Fail to Reject Ho	Reject Ho	Fail to Reject Ho	Reject Ho	Fail to Reject Ho	Reject Ho	Fail to Reject Ho
TOTALS	80	320	84	316	88	312	103	297
% REJECTED Ho	20%		21%		22%		25.75%	

Linear Regression *t*-test for the population slope

Overview of necessary conditions

Consider the linear regression *t*-test for the population slope. The necessary conditions for this test are

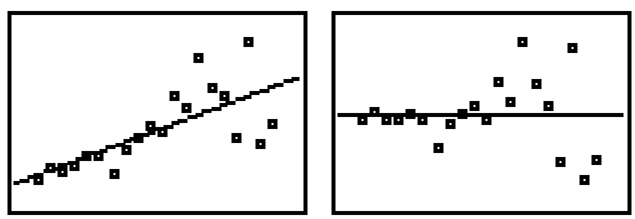
(1) The data must be a simple random sample from the population of interest.

(2) The standard deviation of the errors is constant for all values of *x*.

(3) For a fixed value of *x*, the distribution of the errors around the mean value of *y* is normally distributed with a mean of 0 and a standard deviation of σ.

We will refer to the second condition listed above as **"constant standard deviation of errors"**. We will refer to the third condition listed above as **"normality of errors"**.

- The **constant standard deviation of errors** condition may be violated when we have residual plots that "fan out" as *x* increases. Consider the pictures below. The picture on the left is a scatter plot of bivariate data with the least squares regression line (LSRL) drawn in. The picture on the right is the residual plot for that regression line. Note how the points on the residual plot tend to be further away from zero as *x* increases. This is evidence that the **constant standard deviation of errors** condition might be violated.



Scatter plot of *y* vs. *x* (LSRL included) **Residual Plot**

- The **normality of errors** condition is violated when the population of errors for a fixed value of *x* is not normally distributed. Now, when we collect data we often have only ***one*** value of *y* for a fixed value of *x*. Then how do we check this condition? Well, we usually look at ***all*** the residuals from our sample data and check for skewness and outliers.

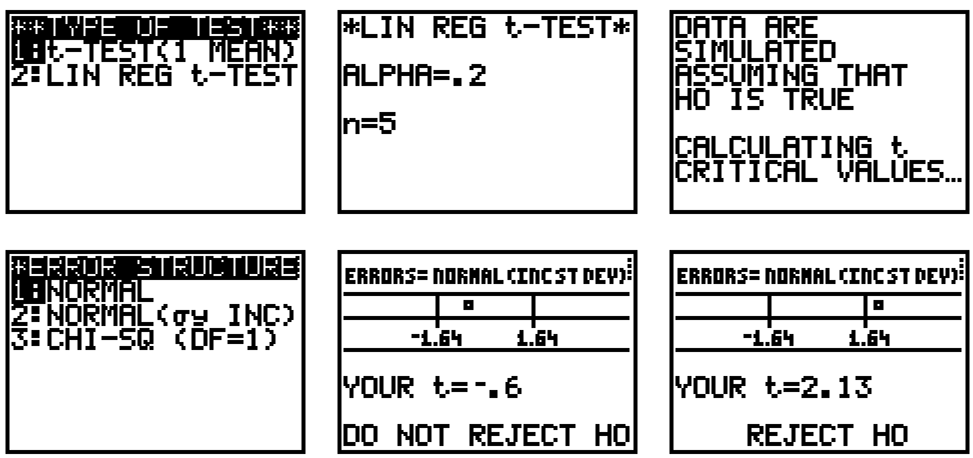
It is unlikely that, in reality, the **constant standard deviation of errors** and **normality of errors** conditions will be met exactly. How ***robust*** is the linear regression *t*-test for the population slope to violations of these conditions? The process of determining this is what the worksheets in the **Student Materials** section will develop. The TI-83 program **ROBUSTT** will speed up the process once the students understand the underlying process. Our criterion for robustness will be the same as it was when we looked at the *t*-test for the population mean. We will violate a condition and compare the stated significance level α to the rate at which we incorrectly reject the null hypothesis in our simulation. If the two numbers are close to the same, then the *t*-test is ***robust*** against violations of this condition.

Program ROBUSTT for the *t*-test for the slope

The calculator generates samples of size *n* from a population on which two quantitative variables are measured. In this activity it will always be given that the slope (b) of the population regression line is 0; that is, these variables have ***no association***. But if we take a sample of size *n* from this population and use the slope of the ***sample*** regression line to test whether or not $b = 0$, we will, on occasion, incorrectly reject the null hypothesis. Ideally, the rate at which we incorrectly reject the null hypothesis should equal the stated significance level α of the test. If we violate conditions of the linear regression *t*-test, will the rate at which we make a Type I error still equal the stated significance level? That is what we will attempt to determine through this activity.

- ➢ **See "Student Materials" for the worksheet "Violating conditions of the *t*-test for the population slope" that explains what the program ROBUSTT is doing. It is recommended that you look through this worksheet in conjunction with these Teacher Notes.**

Screen Shots from program ROBUSTT for the linear regression *t*-test

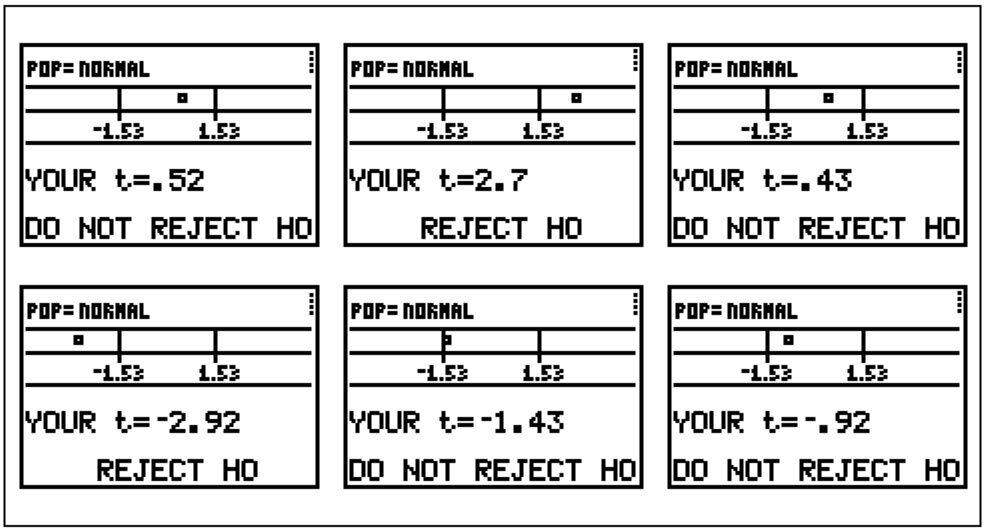


- ❑ Keep pressing **ENTER** to generate more samples of size 5. Press **ON** to quit the program.
- ❑ Keep track on the appropriate data recording worksheet of whether or not the null hypothesis is rejected. (See **Student Materials** handout.)
- ❑ The class can combine results in a table like the one pictured below.

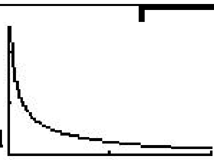
DISTRIBUTION OF ERRORS ®	NORMAL		NORMAL WITH INCREASING ST. DEV.		CHI-SQUARE (DF = 1)	
DECISION	Reject Ho	Fail to Reject Ho	Reject Ho	Fail to Reject Ho	Reject Ho	Fail to Reject Ho
TOTALS						
% REJECTED Ho						

Directions for using data recording worksheet

The data recording worksheet lets students keep track of whether or not an individual sample leads to rejecting or failing to reject the null hypothesis. Space is provided for students to record the values of the significance level α and the sample size n that they are using. This worksheet is in the **Student Materials** handout. Suppose we do a *t*-test for the population mean using a sample size of 5 and a significance level of 0.20. Calculator screen shots are provided below for 6 trials and the appropriate box is marked on the data sheet. It is interesting to record the values of *t* for which we reject the null hypothesis. The non-normal populations tend to give the more extreme *t*-scores. The numbers **–1.53** and **1.53** in the pictures below are the critical values for our *t*-test. *t*-scores outside of this range lead to the null hypothesis being rejected.



Alpha =	**0.20**
Sample Size =	**5**

POPUL 	**NORMAL**		**UNIFORM**		**RIGHT TRIANGULAR**		**CHI-SQUARE (DF=1)**	
TRIAL	**Reject Ho**	**Fail to Reject Ho**	**Reject Ho**	**Fail to Reject Ho**	**Reject Ho**	**Fail to Reject Ho**	**Reject Ho**	**Fail to Reject Ho**
1		**x**						
2	**2.7**							
3		**x**						
4	**-2.92**							
5		**x**						
6		**x**						

–	–	–	–	–	–	–	–	–

Miscellaneous comments about this activity

- Many teachers use the word **conditions** instead of **assumptions** when referring to the criteria that should be checked when performing significance tests. The word **conditions** suggests that there are criteria the students can actually check before performing their test. The word **assumptions** might imply that students can "assume" the criteria are true and not bother to check or investigate. The word **conditions** is used throughout these materials with these facts in mind.

- After doing the activities and simulations, students may get the idea that ***every time*** you reject the null hypothesis that you have made a Type I error. Students should be reminded that the data were simulated by ***using*** the value of the parameter in the null hypothesis. Thus, we ***are certain*** that if we reject the null hypothesis, we have made a Type I error. Point out that when doing tests of significance "for real", we ***do not know*** if the null hypothesis is true. Thus, when we reject the null hypothesis, we could certainly be (and hopefully are) making a correct decision!

- Have students choose a significance level that will lead to the null hypothesis being rejected. Though the examples in the Teacher Notes use an $\alpha = 0.20$, we recommend using an $\alpha = 0.10$ when doing this activity with your students. Using a significance level of 0.05 will often cause no rejections to be made in 25 trials when the population distribution is normal. Point out that a α of 0.10 would not normally be used in practice – we use it here so that we will see some rejections of the null hypothesis when it is true.

  Students could check more standard values of α, such as 0.05 or 0.01, with different sample sizes for homework assignments.

- Students need to know how to perform a linear regression t-test for the population slope before completing the worksheet **"Violating conditions of the *t*-test for the population slope"**.

- When using the program **ROBUSTT**, press ENTER to simulate the t-test repeatedly. The only way to stop the program is by pressing the ON button.

- In the program **ROBUSTT**, the two numbers that appear on the top half of the screen (e.g. –1.53 and 1.53) are the t critical values for the sample size selected by the user. t-scores outside of this range will lead to the null hypothesis being rejected.

- The program **ROBUSTT** puts the calculator in split screen mode (called "HORIZ" on the calculator.) To reset the calculator to full screen mode, go to MODE and go down to the last line and select FULL.

Extensions

1. Consider the case when we take 25 samples of size n from a normal distribution using a significance level α of 0.10. We can think of the number of tests out of 25 in which we reject the null hypothesis as following a binomial distribution with $n = 25$ and $p = 0.10$. After completing 25 trials students can calculate the probability of getting at least as many "Reject Ho" tests as they obtained using the binomial distribution.

2. Consider a specific population distribution (normal, uniform, right-triangular, chi-square). After combining student results (say 40 rejections out of 400 trials), students can make confidence intervals for the probability of a test making a Type I error for each of the four types of population distributions. Then look to see which of the confidence intervals captures the stated significance level α.

3. Suppose that the stated significance level α is 0.20. Let p represent the probability of a t-test (with $n = 5$) leading to a Type I error if the underlying population is a chi-square distribution with 1 degree of freedom. If we simulate n independent t-tests, what is the probability that we will conclude that p is greater than 0.20 if, in fact, it is 0.25?

   We are testing the following hypotheses:
   $$H_O : p = 0.20$$
   $$H_A : p > 0.20$$

   If we let n represent the number of independent t-tests that are generated and use a significance level of $\alpha = 0.05$, then the following tables give the probability of rejecting the null hypothesis for various values of p for which the alternative hypothesis is true – this probability is the **power** of the z-test for one proportion.

   Notice that power increases as n increases (for the fixed alternative $p = 0.25$) and as the alternative value p increases (for fixed n).

n	***p***	**POWER**
25	0.25	**0.173**
50	0.25	**0.241**
100	0.25	**0.358**
200	0.25	**0.545**
400	0.25	**0.785**

n	***p***	**POWER**
200	0.20	**0.05**
200	0.25	**0.545**
200	0.30	**0.951**
200	0.35	**0.9989**
200	0.40	**0.999995**

TI-83 program code for program ROBUSTT

```
ClrHome
Full
Disp "THIS PROGRAM","CLEARS
LISTS","L•,L,,Lƒ,L„,L…","","HIT
ON TO QUIT","","ENTER TO BEGIN"
Pause

ClrHome
ClrDraw
AxesOff
PlotsOff
FnOff
1üQ
Menu("**TYPE OF TEST**","t-TEST(1
MEAN)",A,"LIN REG t-TEST",C)

"T-TEST FOR MEAN"

Lbl A
1üW
{1,7,14,31}üL,
{6,7,17,17}üLƒ
"NORMALUNIFORMR-SKEW
TRIANGULARCHI-SQUARE (DF=1)"üStr1
Output(1,1,"*t-TEST(1 MEAN)*")
Disp ""
Disp ""
Input "ALPHA=",A
Disp ""
Input "n=",N
1-AüC
ClrHome
Disp "DATA
ARE","SIMULATED","ASSUMING
THAT","HO IS
TRUE","","CALCULATING
t","CRITICAL VALUES..."
TInterval 0,ð(N),N,C

Menu("*POPULATION
TYPE*","NORMAL",1,"UNIFORM",2,"R-
SKEW TRIANG",3,"CHI-SQ (DF=1)",4)

Lbl 1
1üK
randNorm(0,1,N)üL•
mean(L•)/(stdDev(L•)/ð(N))üT
Goto 5
Lbl 2
2üK
rand(N)üL•
(mean(L•)-.5)/(stdDev(L•)/ð(N))üT
Goto 5
Lbl 3
3üK
(1-ð(rand(N)))üL•
(mean(L•)-
(1/3))/(stdDev(L•)/ð(N))üT
Goto 5
Lbl 4
4üK
(randNorm(0,1,N))ÜüL•
(mean(L•)-1)/(stdDev(L•)/ð(N))üT
Goto 5

"LINREG T-TEST"

Lbl C
3üW
{1,7,26}üL„
{6,19,17}üL…
"NORMALNORMAL (INC ST DEV)CHI-
SQUARE (DF=1)"üStr1
Output(1,1,"*LIN REG t-TEST*")
Disp ""
Disp ""
Input "ALPHA=",A
Disp ""
Input "n=",N
1-AüC
ClrHome
Disp "DATA
ARE","SIMULATED","ASSUMING
THAT","HO IS
TRUE","","CALCULATING
t","CRITICAL VALUES..."
seq(X,X,1,N,1)üL•
TInterval 0,ð(N-1),N-1,C

Menu("*ERROR
STRUCTURE","NORMAL",21,"NORMAL(Çy
INC)",22,"CHI-SQ (DF=1)",23)

Lbl 21
21üK
randNorm(0,1,N)üL,
LinRegTTest L•,L,,0
tüT
Goto 5

Lbl 22
22üK
randNorm(0,1,N)üL,
L•*L,üLƒ
LinRegTTest L•,Lƒ,0
tüT
Goto 5

Lbl 23
23üK
(randNorm(0,1,N))ÜüL,
LinRegTTest L•,L,,0
tüT
Goto 5

"DRAWING PICTURE"
```

```
Lbl 5
If Q>1
Goto 6
loweruL
upperuU
U-LuB
U+BuXmax
L-BuXmin
0uYmin
1uYmax
Horiz
Line(L,.2,L,.6)
Line(U,.2,U,.6)
Line(Xmin,.6,Xmax,.6)
Line(Xmin,.33,Xmax,.33)

"TOP LINE"

If W=1
Then
Text(3,1,"POP=")
Text(3,18,sub(Str1,L,(K),Lƒ(K))
End
If W=3
Then
Text(3,1,"ERRORS=")
Text(3,30,sub(Str1,L„(K-20),L…(K-
20))
End

Text(24,22,round(L,2))
Text(24,58,round(U,2))
DispGraph
StorePic 9

"EVERYONE GOES HERE"

Lbl 6
Pt-On(T,.45,2)
If T>Xmax
Text(13,80,"-->")
If T<Xmin
Text(13,6,"<--")
Q+1uQ
ClrHome
If W=1 or W=3
Output(2,1,"YOUR t=")
If W=2
Output(2,1,"YOUR z=")
Output(2,8,round(T,2))
If abs(T)>U
Then
Output(4,5,"REJECT HO")
Else
Output(4,1,"DO NOT REJECT HO")
End
Pause
ClrHome
ClrDraw
RecallPic 9

"STARTING OVER"

If K=1
Goto 1
If K=2
Goto 2
If K=3
Goto 3
If K=4
Goto 4
If K=21
Goto 21
If K=22
Goto 22
If K=23
Goto 23
```