

Chapter 13

**Special Cases of Normally Distributed Data**

Topics 31 and 32 discuss handling special cases of normally distributed data. Single and two-sample *t* procedures are discussed. Topic 33 covers inference for least-square regression line, including confidence and prediction intervals for the response.

**Topic 31—Single-Sample *t* Procedures for the Mean (Includes Simulation)**

**The T-Distribution**

1. Press  $\square$  [Y=]. Highlight **y1**, **y2**, and **y3** and press **CLEAR** on each to clear the previous data.
2. Press **CATALOG**, **F3** **Flash Apps** to paste the following functions into **y1**, **y2**, and **y3**:
  - **y1 = tistat.normpdf(x)**
  - **y2 = tistat.tpdf(x, 2)**
  - **y3 = tistat.tpdf(x, 5)**

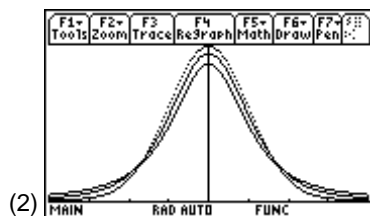
(screen 1).



3. Set up the window using  $\blacklozenge$  [WINDOW] with:

- $xmin = -4$
- $xmax = 4$
- $ymin = 1$
- $ymax = 1$
- $xsc1 = 1$
- $yscl = 0$
- $xres = 1$

4. Press  $\boxed{F2}$  Zoom, A:ZoomFit (screen 2).



Notice that in the centers of the distributions, the standard normal distribution is the highest, followed by  $t$  distribution of 5 and 2 degrees of freedom — the greater the degrees of freedom, the closer the  $t$  distribution is to the normal distribution — with  $n \geq 30$  they are very close. Notice that the  $t$  distributions have larger tails than the normal distribution. Because the area under a probability density function (pdf) must equal one, if the peak of a density is lower than the normal curve's peak, then its tails must be higher than the normal curve's tails so that the area under the curve remains one. (If the peak is lower, the tails have to be higher to have the same area.)

*Example:* Store a random sample of size 10 from a population of heights in **list1**:

**{63.0, 63.6, 66.3, 67.9, 69.3, 66.0, 68.7, 64.2, 66.9, 66.7}**  $\rightarrow$  **list1**.

**Note:** The tails of each of these distributions extend to  $\pm \infty$ , so they cannot be "longer."

**Note:** These are the first 10 scores from Topic 23. Use this data to estimate the population mean with a 90% confidence interval and to test the hypothesis:

$$H_0: \mu = 64$$

$$H_a: \mu > 64.$$

## Requirement of Sampling from a Normal Distributed Population

The  $t$  procedures are quite robust and give good answers as long as the population is mound-shaped with short tails like a normal distribution. Outliers are an indication that the population is not normal. You will analyze the distribution of the sample to gain information about the distribution of the population.

1. From the Stats/List Editor, turn off all functions and plots with **[F2] Plots, 4:FncOff** and **[F2] Plots, 3:Plots Off**.
2. From the Stats/List Editor, press **[F2] Plots, 2:Norm Prob Plot**.
3. Enter Plot Number: **Plot 6**, List: **List1**, Data Axis: **X**, Mark: **Box**, and Store Zscores to: **statvars\zscores**, and press **[ENTER]**.
4. Press **[♦] [GRAPH]** to display the normal probability plot in screen 3.

The  $t$  procedures are often called small sample procedures but they work quite well for large samples also, as the Central Limit Theorem assures  $\bar{x}$ 's are normally distributed, and for large  $n \geq 30$  the  $t$  distribution and the normal distribution have, approximately, the same shape. (See Topic 23.)

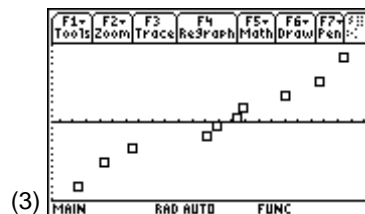
The large sample  $z$  procedures of Topics 23 and 27 can also be used for small samples if  $\sigma$  is known and you can assume the populations sampled are normally distributed. Samples of any size from a normal distribution have means that are normally distributed, with mean =  $\mu$  and standard

$$\text{deviation} = \frac{\sigma}{\sqrt{n}}.$$

## Critical Value

For this topic, change the decimals back to six.

1. Press **[MODE]** and choose **FLOAT 6** from Display Digits. Press **[ENTER]**.
2. Press **[F5] Distr, 2:Inverse, 2:Inverse t** with Area: **0.95** (5% in the upper tail) and Deg of Freedom,  $df = 9$ .



3. Press **[ENTER]** to display screen 4, with the critical  $t = 1.83311$  (screen 4). (You can also obtain this from the Home screen by typing `tistat.inv_t(.95,9)`.)



### Confidence Interval ( $t$ Interval):

What is the 90% confidence interval for the population mean?

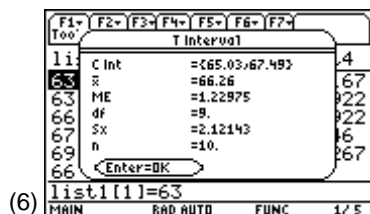
1. Press **[2nd]** **[F7]** **Ints**, **2: TInterval** and use Data Input Method: **Data**, with List: **list1**, Freq: **1**, and C Level: **0.90** (screen 5).



2. Press **[ENTER]** to display screen 6 with the 90% confidence interval of  $66.26 \pm 1.22975$  inches, or  $65.03 \leq \mu \leq 67.49$ .

### Home screen calculation:

$$ME = t_{\alpha/2} \frac{s_x}{\sqrt{n}} = 1.83311 * \frac{2.12143}{\sqrt{10}} = 1.22975.$$



### Test of Significance ( $t$ test):

Use the above data to test the hypothesis:

$$H_0: \mu = 64$$

$$H_a: \mu > 64$$

with significance level  $\alpha = 0.05$

From the Stats/List Editor:

1. Press **[2nd]** **[F6]** **Tests**, **2:T Test** and use Data Input Method: **Data**, with  $\mu_0$ : **64**, List: **list1**, Freq: **1**, Alternate Hyp:  $\mu > \mu_0$ , and Results: **Calculate** (screen 7).



2. Press **[ENTER]** to display screen 8.

With  $p$ -value = **.004136** < .05, you reject  $H_0$  and conclude that the mean of the population is significantly greater than 64, which agrees with the results of the confidence interval from 65.03 to 67.49, all greater than 64 (screen 6).

(8)

F1	F2	F3	F4	F5	F6	F7
Tool	13	Calc	1:1	2:2	3:3	4:4
li	$\mu > \mu_0$	=64				.4
63	t	=3.36884				67
66	P Value	=.004136				922
67	df	=9				922
69	$\bar{x}$	=66.26				6
66	sx	=2.12143				67
66	n	=10				6
	<b>[Enter]=OK</b>					
MAIN	RAD	AUTO	FUNC	1/5		

### Home screen calculation:

$$t = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}} = \frac{66.26 - 64}{\frac{2.12143}{\sqrt{10}}} = 3.36883 > 1.8331$$

(critical  $t$  value).

To verify the calculated  $p$ -values:

- Press **[F5] Distr, 6:t Cdf**, with Low Val = **3.36883**, Up Val =  $\infty$ , and df = **9**.
- Press **[ENTER]** to display screen 9 with Cdf(3.36883, $\infty$ ,9) = **.004136**.

(9)

F1	F2	F3	F4	F5	F6	F7
Tool	13	Calc	1:1	2:2	3:3	4:4
li	t Cdf...					.4
63	Cdf	=.004136				67
66	Low Val	=3.36883				922
67	Up Val	= $\infty$				922
69	df	=9				6
66	<b>[Enter]=OK</b>					
	list1[1]=63					67
MAIN	RAD	AUTO	FUNC	1/5		

**Note:**  $\infty$  is found above the **[CATALOG]** key.

### Simulation

The sample used in this topic example consists of the first 10 values used in Topics 23 and 27 and can be generated with **RandSeed 789**, and then **tistat.randNorm(65,2.5,10)**. To generate a sample, calculate the 90% confidence interval, then display the results:

- Set **RandSeed 789** (screen 10).
- Enter the following on one line:  
**tistat.randnorm(65,2.5,10)**→**x1:tistat.tint(x1,1,.90)**:  
{**statvars\lower,statvars\upper,statvars\upper-statvars\lower**}.
- Press **[ENTER]** twice to display the rest of screen 10.

(10)

F1	F2	F3	F4	F5	F6
Tool	13	Calc	1:1	2:2	3:3
	RandSeed 789	Done			
	tistat.randnorm(65,2.5,10)	→	x1:tistat.tint(x1,1,.90)		
	{statvars\lower,statvars\upper,statvars\upper-statvars\lower}				
	→	{65.029 67.5094 2.4803}			
	tistat.randnorm(65,2.5,10)	→	x1:tistat.tint(x1,1,.90)		
	{63.3507 65.8985 2.547}				
	→	{statvars\upper-statvars\lower}			
MAIN	RAD	AUTO	FUNC	3/30	

**Note:** The 1 in **tint(x1,1,.90)** is for Freq: 1 for list x1.

**Note:** If **[ENTER]** is pressed eight more times, two of the 10 intervals do not contain 65.

The first interval of 65.029 to 67.5094 was the interval in the previous *Confidence Interval* section. It does not contain the population  $\mu = 65$ . The second interval, 63.3507 to 65.8985, does contain 65.

In the long run, you would expect only one in 10 (90% CI) intervals not to contain the population value. The first two interval widths are  $67.5094 - 65.029 = 2.48$  and

$65.8985 - 63.3507 = 2.55$ . These values equal  $2 * t_{\alpha/2} \frac{s_x}{\sqrt{n}}$ . The values vary from sample to sample since  $s_x$  varies from sample to sample.

## Topic 32—Two-Sample (Independent and Matched Pairs) $t$ Procedures

### Independent Samples

*Example:* From the Home screen, store a random sample of size 10 from a population of male heights in **list1**, and a random sample of size 9 from a population of female heights in **list2** in the **MAIN** folder.

Males: {**63.0, 63.6, 66.3, 67.9, 69.3, 66.0, 68.7, 64.2, 66.9, 66.7**} → **list1**

Females: {**63.3, 60.5, 65.7, 61.9, 63.6, 60, 59, 63.5, 63.8**} → **list2**.

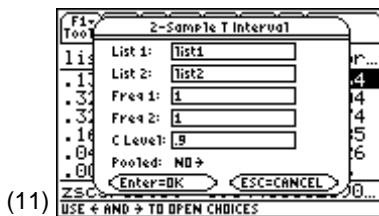
*Confidence Interval (t Interval):*

Find the 90% confidence interval for the difference in the mean heights of males and females in the sampled population.

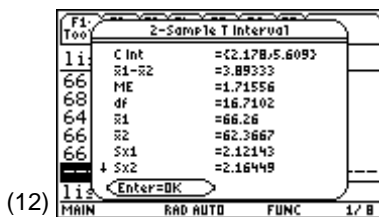
- From the Stats/List Editor, press  $\boxed{2nd}$   $\boxed{[F7]}$  **Ints**, **4:2-SampTInt** and use Data Input Method: **Data**, with List 1: **list1**, List 2: **list2**, Freq 1: **1**, Freq 2: **1**, C Level: **0.90**, and Pooled: **NO** (screen 11).

**Note:** When using a  $Z$  Interval,  $2 * z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$  is a constant from sample to sample.

**Note:** The normal probability plots for **list1** and **list2** are linear enough to assume that these are samples from normal populations.



- Press  $\boxed{ENTER}$  to display the results (screen 12).



3. Repeat steps 1 and 2, and change Pooled: **YES** (screen 13). The results are basically the same for the 90% confidence interval  $2.2 < \mu_M - \mu_F < 5.6$  inches, or  $(3.9 \pm 1.7)$ . You are 90% confident that the mean height of males in the population is between 2.2 to 5.6 inches taller than the mean height of females.

(13)

F1	2-Sample T Interval	
li	C Int	=2.181,5.605
66	x1-x2	=3.89333
68	ME	=1.71193
64	df	=17
66	s1	=2.12
66	s2	=2.1667
66	sx1	=2.12143
66	sx2	=2.16449
li	Enter=OK	
MAIN	RAD AUTO	FUNC 1/B

The pooled and nonpooled results are so close because their degrees of freedom (df) are close, which is correct because  $s_x^1$  is close to  $s_x^2$  (2.12 compared to 2.16) suggesting, even with these small samples, that  $\sigma_1 = \sigma_2$ , which satisfies the assumption needed to pool the standard deviation. If these values are not close, it may suggest that the assumption is not true and that the nonpooled procedure is the correct procedure. This nonpooled procedure gives the most conservative answer — the wider interval ( $\pm 1.72$  compared to  $\pm 1.71$ ), so some authors suggest it is best to use it all the time.

*To pool or not to pool:*

**Not pooled (use  $s_1$ ,  $s_2$ ,  $n_1$ , and  $n_2$ ):**

$$df = \left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right) / \left[ \frac{1}{n_1 - 1} \left( \frac{s_1^2}{n_1} \right) + \frac{1}{n_2 - 1} \left( \frac{s_2^2}{n_2} \right) \right]$$

$$df = \frac{\left( \frac{2.12^2}{10} + \frac{2.16^2}{9} \right)}{\frac{\left( \frac{2.12^2}{10} \right)}{(10-1)} + \frac{\left( \frac{2.16^2}{9} \right)}{9-1}} = 16.7$$

$$t_{a/2} = \text{inv}_t(\text{Area}, df) = \text{inv}_t(0.95, 16.7) = 1.74141 \approx 1.74$$

$$ME = t_{a/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$ME = 1.74 \sqrt{\frac{2.12^2}{10} + \frac{2.16^2}{9}} = 1.74 * 0.984 = 1.7$$

**Pooled:**

$$df = n_1 + n_2 - 2 = 10 + 9 - 2 = 17, t_{\alpha/2} = \text{inv}_t(.95, 17) = 1.73961 = 1.74$$

$$ME = t_{\alpha/2} \sqrt{\frac{(n_1 - 1) * s_1^2 + (n_2 - 1) * s_2^2}{(n_1 + n_2 - 2)}} * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$ME = 1.74 * \left( \sqrt{\frac{(10-1) * 2.12^2 + (9-1) * 2.16^2}{(10+9-2)}} \right) * \sqrt{\left( \frac{1}{10} + \frac{1}{9} \right)} = 1.74 * 0.983 = 1.7$$

**Test of Significance (t test):**

Test the claim that males are taller than females on average.

$$H_0: \mu_M = \mu_F \text{ (or } \mu_M - \mu_F = 0)$$

$$H_a: \mu_M > \mu_F, \text{ (or } \mu_M - \mu_F > 0)$$

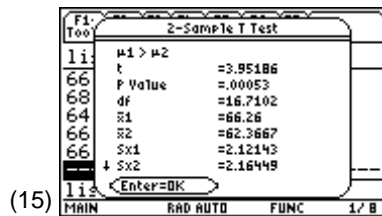
with significance level  $\alpha = 0.05$

1. Press **2nd** **[F6]**, **4:2-SampTTest** and use

Data Input Method: **Data**, with List 1: **list1**, List 2: **list2**,  
Freq 1: **1**, Freq 2: **1**, Alternate Hyp:  $\mu_1 > \mu_2$ , and Pooled: **NO**  
(screen 14).

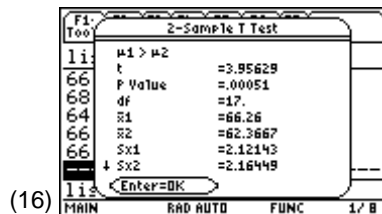


2. Press **ENTER** to display the results for Pooled: **NO**  
(screen 15).



3. Repeat steps 1 and 2, and change Pooled: **YES**  
(screen 16).

In both cases,  $p\text{-value} = .0005 < .05$ , so reject  $H_0$  and conclude the males are significantly taller than the females, which agrees with the results of the confidence interval.

**Not pooled (use  $s_1$ ,  $s_2$ ,  $n_1$ , and  $n_2$ ):**

$$t = \frac{\mu_M - \mu_F}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \frac{(66.3 - 62.4)}{0.984} = \frac{3.89}{0.984} = 3.95$$

$$p\text{-value} = \text{tcdf}(3.95, \infty, 16.7) = 0.00053$$



**Pooled:**

$$t = \frac{3.89}{0.983} = 3.96$$

$$p\text{-value} = tcdf(3.96, \infty, 17) = 0.00051$$

**Matched Pairs**

*Example:* To test a blood pressure medication, diastolic blood pressure readings for a random sample of 10 people with high blood pressure were recorded. After a few weeks on the medication, pressures were recorded again. The data below is stored in **list1** and **list2**, with differences stored in **list3**.

Enter this data in the **MAIN** folder. Clear the lists of previous data first.

Subject	Before (list1)	After (list2)	(list1 - list2) → list3
1	95	99	-4
2	105	86	19
3	88	92	-4
4	100	95	5
5	104	86	18
6	109	89	20
7	89	87	2
8	93	85	8
9	100	87	13
10	97	86	11

Find the 90% confidence interval for the mean change in blood pressure.

Test to determine if there is a significant decrease in blood pressure after taking the medication.

Set up and define **Plot 1** as a normal probability plot for the differences in **list3**.

1. From the Stats/List Editor, turn off all functions and plots with **[F2] Plots, 4:FnoFF** and **[F2] Plots, 3:Plots Off**.

- Press **F2** **Plots, 1:Plot Setup**, highlight **Plot 1**, press **F3** **CLEAR**, and then press **ENTER**.
- Press **F2** **Plots, 2:Norm Prob Plot**, with Plot Number: **Plot 1**, List: **List3**, Data Axis: **X**, Mark: **Box**, and Store Zscores to: **statvars\zscores**.
- Press **ENTER** to return to the Stats/List Editor.
- Press **F2** **Plots, 1:Plot Setup**, and observe that **Plot 1** has been set up with Plot Type: **Scatter**, Mark: **Box**, X List: **npplist**, and Y List: **zscores**.
- Press **F5** **ZoomData** to display screen 17.

There is no indication of long tails or outliers. There is a fairly straight line pattern in the middle, indicative of a sample from a population that is mound-shaped with short tails, so  $t$  procedures should be appropriate.

*Confidence Interval (t Interval):*

What is the 90% confidence interval for the population mean?

- From the Stats/List Editor, press **2nd** **F7** **Ints, 2:TInterval**, and use Data Input Method: **Data**, with List: **list3**, Freq: **1**, and C Level: **0.90** (screen 18).

- Press **ENTER** to display a 90% confidence interval of  $3.6 < \mu_d < 14$ , or  $8.8 \pm 5.2$  (screen 19). You are 90% confident that the blood pressure has been reduced, on average, between 3.6 to 14 points. This is a wide interval, both because of the small sample ( $n = 10$ ) and large  $s_x$  ( $s_x = 8.98$ ).

Note that  $t_{\alpha/2} = \text{Inv}_t(\text{Area}, df) = \text{Inv}_t(0.95, 10-1) = 1.833$

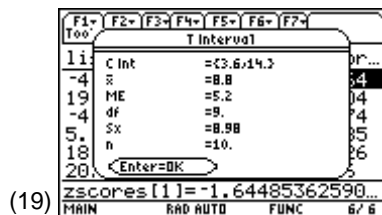
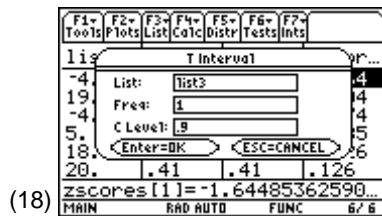
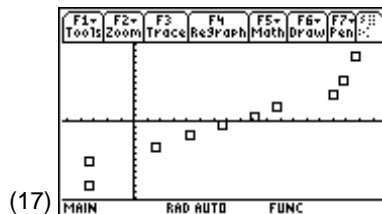
$$ME = \frac{1.833 * 8.98}{\sqrt{10}} = 5.2$$

*Test of Significance (t test):*

$$H_0: \mu_d = 0$$

$$H_a: \mu_d > 0 (\mu_1 > \mu_2)$$

with significance level  $\alpha = 0.05$



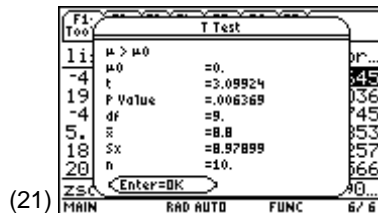
**Note:** Topic 25, screen 33 which used this same data plus 20 more subjects, had an interval of 7.7 to 12.3, or  $10 \pm 2.3$ . This narrower interval was contained in the interval above.

- Press  $\boxed{2\text{nd}} \boxed{F6}$  **Tests, 2:T Test**, and use Data Input Method: **Data**, with  $\mu_0$ : **0**, List: **list3**, Freq: **1**, Alternate Hyp:  $\mu > \mu_0$ , and Results: **Calculate** (screen 20).



(20)

- Press  $\boxed{\text{ENTER}}$  to display  $p$ -value = **.006369** < .05. You would reject  $H_0$  and conclude that there was a significant decrease in the blood pressure after taking the medication. The confidence interval in the first part indicated between a 3.6 to 14 point drop (screen 21).



(21)

**Home screen calculation:**

$$y = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{8.8}{\frac{8.98}{\sqrt{10}}} = 3.09924 > 1.833 \text{ (from } t_{\alpha/2} \text{ in the previous example)}$$

$$p\text{-value} = t\text{cdf}(3.099, \infty, 9) = .0064$$

**Note:** Although the TI-89 has no built-in paired t test as such, this test could have been conducted with  $\boxed{2\text{nd}} \boxed{F6}$  **Tests, D:ANOVA 2-Way**, with Design: **Block** (Topic 40) with two Levels of the Column Factor in **list1** and **list2** with the same results:

$$t = \sqrt{F} = \sqrt{9.60529} = 3.09924 \text{ and } p = \frac{.012738}{2} = .006369$$

**Topic 33—Inference for Least-Square Regression Line (Includes Confidence and Prediction Intervals for the Response)**

*Example:* In Topic 11, you used the folder **CARS** for information on 13 automatic transmission cars, including their city gas mileage in miles per gallon (**mpg**) and their weight (**wt**) in pounds. The data are included here, in the same order, but without the names of the cars.

- Change to the **CARS** folder for this example.
- Verify that the lists **wt** and **mpg** still contain the data given below. If not, create list headings **wt** and **mpg** and enter the data in the following table under the appropriate heading.

<b>Wt (x)</b>	2795	2600	3515	3930	3115	3995	3115	4020	3175	3225	3985	2500	2290
<b>Mpg (y)</b>	23	23	19	17	20	17	22	17	19	19	17	29	28

Is there a significant linear relationship between weight (**wt**) and gas mileage (**mpg**)?

Find the 90% confidence interval for  $\beta$ , the slope of the linear least-square regression line relating **mpg** and **wt**.

Using confidence and prediction intervals, estimate the mean gas mileage of all cars weighing 3200 pounds and the gas mileage of a particular car weighing 3200 pounds.

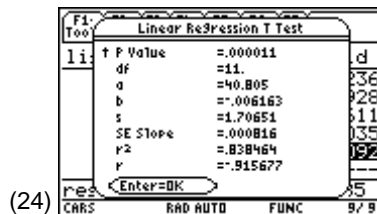
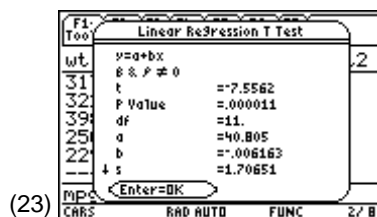
### LinRegT-Test

Test  $H_0: \beta = 0$  or  $p = 0$ , (the slope of the regression line = 0 or the correlation coefficient = 0).

3. From the Stats/List Editor, press  $\boxed{2\text{nd}}$   $\boxed{F6}$  **Tests**,  $\odot$  for the bottom of the menu, and then press **A:LinRegTTest**. Set X List: **wt**, Y List: **mpg**, Freq: 1, Alternate Hyp:  **$\beta \neq 0$** , Store RegEqn to: **none**, and Results: **Calculate** (screen 22).



4. Press  $\boxed{\text{ENTER}}$  to display screen 23, and then press  $\odot$  three times to display the next part shown in screen 24.



5. Press  $\boxed{\text{ENTER}}$  to display list **resid** (screen 25).

If the assumptions for this test are appropriate (the residuals are investigated below), and since the

$p$ -value = **0.000011**  $< 0.05$ , you reject  $H_0: \beta = 0$  and conclude that the slope,  $-0.006163$ , is significantly different from zero and thus  $p \neq 0$  and  $r = -.915677$  is significant negative correlation. The regression equation is

$$y = a + bx = \hat{y}_{\text{mpg}} = 40.805 - .006163wt.$$

Refer to the *Simulation for Test* section to assist in your understanding of this hypothesis test.

(25)

wt	mpg	resid
2795.	23.	-.5783
2600.	23.	-1.78
3515.	19.	-.1407
3930.	17.	.41713
3115.	20.	-1.606
3995.	17.	.81775

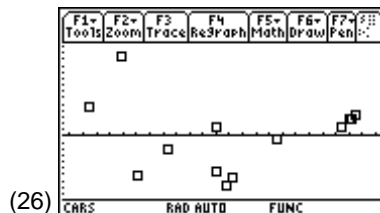
resid = (-.57831068043062, ...)

## Residual Plot

In order that the  $t$  test is valid, the residuals should be randomly distributed about the regression.

1. From the Stats/List Editor, turn off all functions and plots with **[F2] Plots, 4:F<sub>n</sub>Off** and **[F2] Plots, 3:Plots Off**.
2. Set up and define **Plot 1** as Plot Type: **Scatter**, Mark: **Box**, x: **wt**, y: **statsvar\resid**, Use Freq and Categories?: **NO**, and then press **[ENTER]**.
3. From the Plot Setup menu, press **[F5] ZoomData** (screen 26).

In Topic 11, screen 17, you detected curvature in this pattern of residuals. In Topic 12 you used transformations to obtain better fits. You should assume that the pattern is satisfactory for only 13 data points, although a 14<sup>th</sup> point falling in the upper half and about 3/4 of the way over would be better.

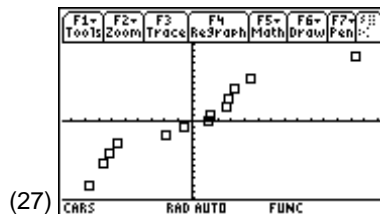


## Normal Probability Plot of Residuals

In order that the  $t$  test is valid, the residuals should come from a population that is normally distributed.

1. From the Stats/List Editor, turn off all plots with **[F2] Plots, 3:Plots Off**.
2. Press **[F2] Plots, 2:Norm Prob Plot** and set up **Plot 2** (or the first available plot) with List: **statvars\resid**, Data Axis: **X**, and Mark: **Box**.
3. Return to the Plot Setup screen and then press **[F5] ZoomData** (screen 27).

These results are less than ideal, particularly the outlier skewed to the right. Continue and assume this pattern is satisfactory for showing that the residuals are approximately normally distributed.



## Finding the 90% Confidence Interval for $\beta$

With the previous assumptions satisfied:

- From the Stats/List Editor, press  $\text{2nd}$   $\text{[F7]}$  **Ints**, **7:LinRegTInt**, with X List: **wt**, Y List: **mpg**, Freq: **1**, Store RegEqn to: **y1(x)**, and Interval: **Slope** (screen 28).



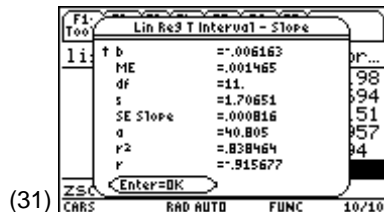
- Press  $\text{⏏}$  to enter C Level: **0.90** (screen 29).



- Press  $\text{ENTER}$  to display screen 30. The 90% confidence interval  $-.008 < \beta < -.005$ , or  $-.0062 \pm .0015$  does not contain zero. This is consistent with the conclusion from the previous test of significance. For every 100 lb. increase in a car's weight, its car mileage is lower, on average, by  $1/2$  to  $4/5$  of a mile per gallon.



- Press  $\text{⏏}$  twice to display the off-screen values (screen 31).



## Estimate Mean

Estimate mean **mpg** for all cars weighing 3200 pounds, and for a particular car that weighs 3200 pounds.

- Press **[2nd] [F7] Ints, 7:LinRegTTest**, with X List: **wt**, Y List: **mpg**, Freq: **1**, Store RegEqn to: **y1(x)**, Interval: **Response**, X Value: **3200**, and C Level: **0.95** (screens 32 and 33).



- Press **[ENTER]** to display screen 34.

95% Confidence Interval:

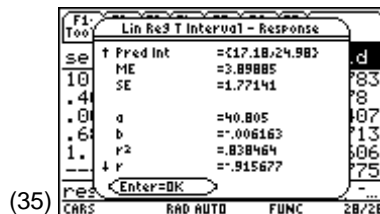
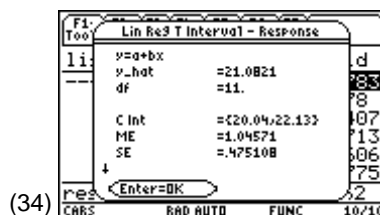
$$\hat{y} \pm ME = 21.1 \pm 1 \text{ or } 20.1 \leq \mu_{mpg} \leq 22.1$$

(interval width = 2.1 mpg)

95% Predicted Interval:

$$\hat{y} \pm ME = 21.1 \pm 3.9 \text{ or } 17.2 < mpg < 25.0$$

(interval width = 7.8 mpg)



**Note:** Two cars in the sample weigh close to 3200 lbs. (3175 and 3225) and both get 19 mpg, which fits in the prediction interval.

## Transformed Data (Doing the Estimation with Transformed Data)

In Topic 12 you found a better fit than the previous linear model with the  $(\text{mpg})^{-2}$  transformation.

From the Home screen:

- Type  $\text{mpg}^{-2} \rightarrow \text{tmpg}$  and press **[ENTER]** to display the transformed mpg.
- From the Stats/List Editor, press **[2nd] [F7] Ints**, **7:LinRegTInt**, with X List: **wt**, Y List: **tmpg**, Freq: **1**, StoreRegEqn to: **y2(x)**, Interval: **Response**, X Value: **3200**, and C Level: **0.95** (screens 36 and 37).



- Press **[ENTER]** twice to display the calculated residuals (screens 38 and 39).

$$95\% \text{ CI: } .002469 \pm .000129 \text{ or } .002340 < (\mu_{\text{mpg}})^{-2} < 0.002598$$

Solving for  $\mu_{\text{mpg}}$ , you get:

$$\frac{1}{\sqrt{0.002598}} < \mu_{\text{mpg}} < \frac{1}{\sqrt{0.002340}}$$

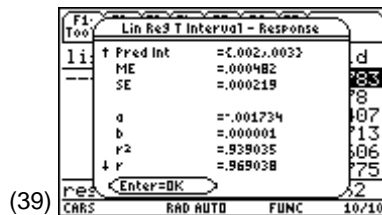
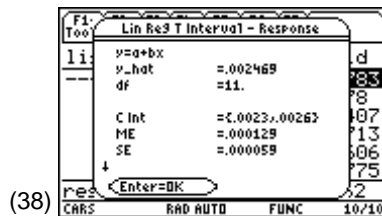
$$19.62 < \mu_{\text{mpg}} < 20.67 \text{ (interval width} = 1.05 \text{ mpg)}$$

$$95\% \text{ PI: } .002469 \pm 0.000482$$

$$.001987 < (x_{\text{mpg}})^{-2} < .002951$$

Solving for  $x_{\text{mpg}}$ , you get:

$$18.4 < x_{\text{mpg}} < 22.4 \text{ (interval width} = 4 \text{ mpg)}$$





Notice that the intervals associated with the model produced from the transformed data are narrower (about half the width of the intervals from the linear model). This is understandable if you plot both equations with the original data (as in Topic 12, screen 36).

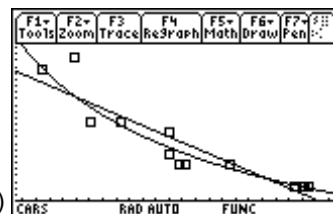
1. Press  $\square$  [Y=] to observe that  $y_1 = 40.805 - 0.006163x$  and  $y_2 = -.001754 + 1.313328E-6x$ .
2. Deselect  $y_2$  by highlighting the  $y_2$  equation and pressing  $\square$  [F4].
3. Type the following function for  $y_3$ :  $y_3 = (y_2(x))^{(-1/2)}$  (screen 41).
4. Make sure all plots are turned off by using  $\square$  [F4] to deselect them while you are in the  $y =$  editor.
5. Set up and define **Plot 1** as Plot Type: **Scatter**, with Mark: **Box**, X List: **wt**, Y List: **mpg**, and Use Freq and Categories?: **NO**. Press  $\square$  [ENTER].
6. From the Plot Setup menu, press  $\square$  [F5] **ZoomData** for the plots in screen 42.

F4→ Tools	F2→ Plots	F3→ List	F4→ Calc	F5→ Distr	F6→ Tests	F7→ Ints	
wt	mpg		tmpg		resid		
2795	23		.00189		-5.E-5		
2600	23		.00189		.00021		
3515	19		.00277		-.0001		
3930	17		.00346		.00003		
3115	20		.0025		.00014		
3995	17		.00346		-5.E-5		
wt [6]=3995							
CARS		RAD AUTO		FUNC		1/11	

(40)

F4→ Tools	F2→ Zoom	F3→ Edit	F4→ All	F5→ Style	F6→ Style	F7→ Style	
PLOT5 1							
Plot 1: <input type="checkbox"/> cars\wt <input type="checkbox"/> cars\mpg							
y1=40.804964 + -.006163·x							
y2=-.001734 + 1.313328E-6·x							
y3=(y2(x)) <sup>-1/2</sup>							
y4=							
y5=							
y5(x)=							
CARS		RAD AUTO		FUNC			

(41)



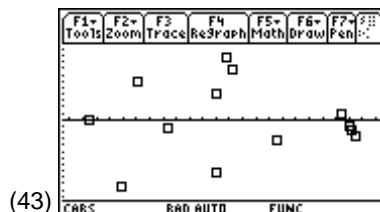
(42)

### Check Residuals Plots for Transformed Data

Check the plots for the transformed data.

1. From the Stats/List Editor, turn off all functions and plots with  $\square$  [F2] **Plots, 4:FnOff** and  $\square$  [F2] **Plots, 3:Plots Off**.
2. Clear plots **Plot 2** and **Plot 3** using  $\square$  [F2] **Plots, 1:Plot Setup** and  $\square$  [F3] **CLEAR**.
3. Set up and define **Plot 2** as Plot Type: **Scatter**, with Mark: **Box**, X List: **wt**, Y List: **statsvar/resid**, and Use Freq and Categories?: **NO**.
4. Press  $\square$  [F5] **ZoomData** (screen 43).

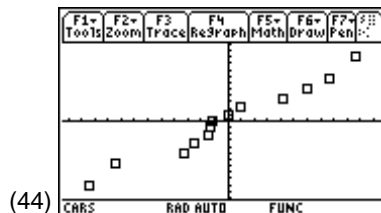
This plot indicates an improvement over screen 26.



(43)

5. Deselect **Plot 2**.
6. From the Stats/List Editor, press **[F2] Plots**,  
**2:Norm Prob Plot** and set up with Plot Number: **Plot 3**,  
List: **statvars\resid**, Data Axis: **X**, Mark: **Box**,  
Store Zscores to: **statvars\zscores**. Press **[ENTER]**.
7. Return to the Plot Setup screen and press **[F5] ZoomData**  
to display screen 44.

This plot indicates an improvement over screen 27.



### Simulation for Test $H_0: \beta = 0$ or $\rho = 0$

Source: "Correlation from Throws of the Dice," *Teaching Statistics*,  
Volume 9, Number 2, May 1987. Used with written permission of the  
Teaching Statistics Trust.

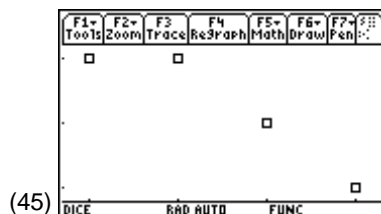
Since  $b = r \frac{S_y}{S_x}$ , if the correlation coefficient  $r = 0$  from a  
sample, then the slope from the sample is zero.

For small samples, the correlation coefficient might be far  
away from the population value. If you toss two dice, one  
marked **x1** and one marked **x2**, you could come up with the  
following samples:

Sample 1:  $(x_1, x_2)$ : **(1, 6), (4, 4), (2, 6), and (3, 5)**.

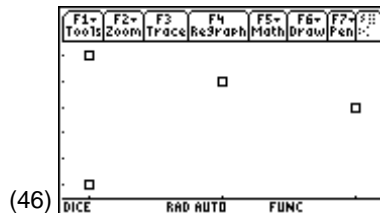
1. From the Stats/List Editor, clear **list1** and **list2**.
2. Enter the  $x$ -coordinates of the ordered pairs 1, 4, 2, and  
3 under the **list1** heading.
3. Enter the  $y$ -coordinates 6, 4, 6, and 5 under the **list2**  
heading.
4. Press **[F2] Plots, 1:Plot Setup** to define **Plot 1** as  
Plot Type: **Scatter**, with Mark: **Box**,  $x$ : **list1**,  $y$ : **list2**, and  
Use Freq and Categories?: **NO** and press **[ENTER]**.
5. From the Plot Setup menu, press **[F5] ZoomData**  
(screen 45).

This plot results in  $r = -0.94$ , which looks fairly linear.  
LinRegTTest gives a  $p$ -value = **.05612**  $> .05$ .  $\rho$  is not  
significantly different from zero — befitting the dice. When  
you are working with small samples, be careful about  
jumping to conclusions if  $|r|$  is large.



Sample 2:  $(x_1, x_2)$ : **(2, 1)**, **(4, 4)**, **(2, 6)**, and **(3, 5)**. Repeat steps 1 through 5 with this sample to observe screen 46.

The first data point is the only difference from sample 1, but the correlation coefficient has been reduced to  $r = -10.16$ .



From the Home screen:

1. Set **RandSeed 123** (screen 47).
2. Print out the correlation coefficient  $r$  and the  $p$ -value with the following instruction:

```
tistat.randInt(1,6,5)>x1: tistat.randInt(1,6,5)>x2:
tistat.LR_tTest(x1,x2):{statvars\r, statvars\pval}
```



3. Press **ENTER** twice for the two simulations in screen 47.

The first simulation displays  $r = -0.866025$  with a  $p$ -value = **0.057669**  $> .05$ . This  $r$  value is not close enough to  $\pm 1$  to reject  $H_0: \beta = 0$  or  $\rho = 0$ , but you would expect to get a value this large or larger only about 6% of the time when  $\rho = 0$ .

The second simulation with  $r = +.46414$  and a  $p$ -value = **.431002** is much more likely to occur when  $\rho = 0$ . Screen 48 repeats screen 47, but with the sample size now 50 instead of 5. Notice that  $r$  is much more likely to be closer to the population correlation coefficient  $\rho = 0$  with these larger sample sizes.

