EXPLORATIONS

Chapter 12

Tests of Significance

This chapter includes information on how tests of significance can be performed with confidence intervals. Topic 26 clarifies the meaning of a hypothesis test using large sample tests for proportions using a simulation. Topic 27 discusses the use of large sample tests for a mean. Topic 28 introduces large sample tests for the difference between two proportions. The difference between two means is covered in Topic 29, while Topic 30 discusses tests for fit, homogeneity of proportions, and independence.

Topic 26—Large Sample Test for a Proportion and a Simulation to Help Clarify the Meaning of a Hypothesis Test

Example: When a public policy was introduced several years ago, 63% of the people voted for it. Test the claim that a larger percentage of voters would be in favor of the policy today with a significance level of 0.05. A simple random sample of 265 voters has 69.4% agreeing with the policy. For this topic, change to folder **BLDTALL**.

H_0: $p = 0.63$

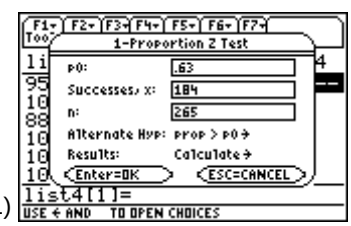
H_a: $p > 0.63$

with significance level $\alpha = 0.05$

Because $n * p = 265 * .694 = 184 > 10$ and $n(1 - p) = 265 * .306 = 265 - 184 = 81 > 10$, you can use a normal distribution to approximate the binomial.

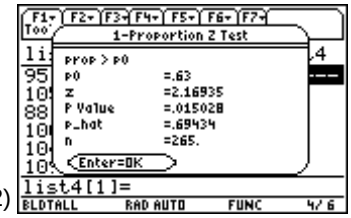
1. In the Stats/List Editor, press 2nd [F6] **Tests**, **5:1-PropZTest**, and enter p_0: **0.63**, Successes, x: **184**, n: **265**, Alternate Hyp: **prop > p_0**, and Results: **Calculate** (screen 1).

(1)

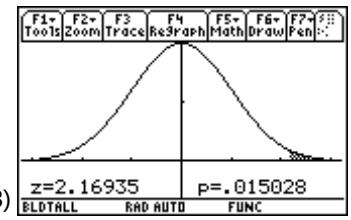


2. Press ENTER to view screen 2 with *p*-value = **0.015028**. If in step 1 you selected Results: **Draw** (be sure all Plots and Functions are turned off under F2 **Plots**), screen 3 would result with the same *p*-value of **0.015028**.

(2)



(3)



A *p*-value as small as 0.015 (0.015 < 0.05) is strong evidence that the new proportion is larger than 63%, so you reject the null hypothesis and conclude that the proportion favoring the policy has (statistically) significantly increased. There are only 15 chances in 1000 that you would get a proportion as large as .694 if you were taking samples from a population with about 63% successes. The 0.63 would be the mean and middle value in screen 3, your *z* calculation of 2.169 puts you way in the right tail. That makes you feel that your sample is probably from another population, with a mean to the right of 0.63. Of 1000 tests that find a *p*-value of .015, you expect 15, on average, to lead you to the wrong conclusion, that is rejecting the hypothesis when you should accept the hypothesis. If so, that would be a type I error.

Note: *If* $\alpha = 0.01$ *you would fail to reject* H_0 *and conclude that there was no significant change in the proportion.*

Home screen calculation:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{184}{265} - .63}{\sqrt{\frac{.63(1-.63)}{265}}} = 2.16935$$

To verify that a *z* value of 2.16935 corresponds to $p = .015$, type **normcdf(2.169,∞)** from the Home screen.

Note: *In Topic 22, screen 6, the above sample information was randomly generated from a population with* p = **.67** *using* **tistat.randbin(265,.67)/265 = .69434**.

Testing a Hypothesis with a Confidence Interval

From Topic 22 screen 3, the 95% confidence interval for this same data was 63.89% to 74.98%, so you were 95% confident that the sample was from a population with a proportion of successes that was greater than 63%, indeed greater than 63.89%. The 95% confidence interval is comparable to a one-tailed test with significance level $\alpha = 0.025$ (in the upper tail), because $0.015 < 0.025$, the fact that the confidence interval gives the same results should not surprise you.

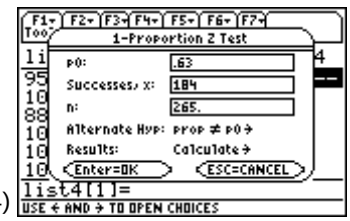
Note: *The 90% confidence interval which is comparable to* $\alpha = 0.05$ *in the upper tail, gives an interval of 64.8% to 74.1%, which is above the null hypothesis value of 63%, so that hypothesis could be rejected at the* $\alpha = 0.05$ *level.*

Two-Tailed Test

If you have no idea or indication that the popularity of the policy has increased or decreased, a two-tailed test would be appropriate.

1. In the Stats/List Editor, press 2nd [F6] **Tests**, **5:1-PropZTest**, and enter p_0: **0.63**, Successes, x: **184**, n: **265**, Alternate Hyp: **prop ≠ p_0**, and Results: **Calculate** (screen 4).

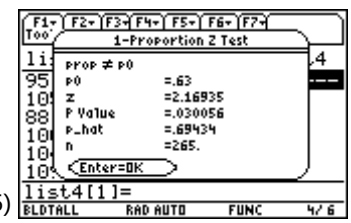
(4)



2. Press ENTER to view screen 5. The only difference in the output is now the *p*-value = $2 * 0.015 = .03$.

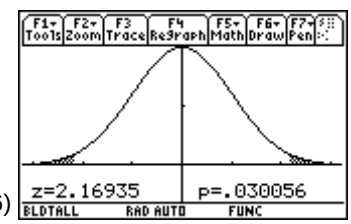
   Your conclusions are the same as with $\alpha = 0.05$, since $0.03 < 0.05$.

(5)



3. Repeat step 1, but with this change:
   Results: **Draw** and press ENTER to display screen 6.

(6)



Note: *If* $\alpha = 0.02$*, you would conclude that there was no significant change in the policy's approval.*

Simulation to Help Clarify the Meaning of Test of Significance

In Topic 22 screens 7 and 8, you simulated 80% confidence intervals of samples of size 50 from a population with p = **.67**. To show the parallels here, test the hypothesis:

H_0: $p_0 = 0.67$

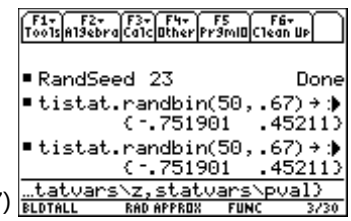
H_a: $p \neq p_0 = 0.67$

with significance level $\alpha = 0.20$.

From the Home screen:

1. Set **RandSeed 23** as in the top line of screen 7.

(7)



2. Take a sample size of 50 from a population with p = **.67** success, perform the hypothesis test, and display the resulting z statistic and p-value with the following: **tistat.randbin(50,.67)→x:tistat.ztest_1P(.67,x,50): {statvars\z,statvars\pval}**. (Press CATALOG F3 **Flash Apps**.)

3. Press ENTER twice for the results **{-.751901 .45211}**, which by coincidence also occurs twice in succession in screen 7. The z value of **-.751901** indicates you are less than 1 standard deviation to the left of the hypothesized population mean (and the mean of all sample proportions, 0.67), and you would expect to be at least |-.7519| deviations from the mean 45.21% of the time if the null hypothesis (p_0 = .67) were true.

4. Press ENTER eight more times for all 10 sample results given in the table.

Note: *All instructions should be on one line. Only part of the instructions will be displayed (screen 7).*

Sample	*Z* statistic	*p*-value
1	-.7519	.4521
2	-.7519	.4521
3	.4511	.6519
4	-1.6542	.0981
5	1.0527	.2925
6	1.6542	.0981
7	-1.3534	.1759
8	-.4511	.6519
9	-.1504	.8805
10	-.7519	.4521

Notice that the fourth simulated sample is far enough from the hypothesized mean that you could only expect it to be this different, or more different 9.8% of the time. Since 0.098 < 0.20, you would reject the null hypothesis, and since the z value is negative (-1.6542) you conclude that the proportion was significantly less than 0.67.

In Topic 22 screen 8, this case gave a confidence interval of 0.47 to 0.65, indicating a value significantly less than 0.67. There are three of 10 simulated cases in which you would reject H_0 (sample 4, 6, and 7). In the long run, you would expect two of 10, or 20% such cases ($\alpha = 0.20$).

Note: *If you were doing a one-tailed hypothesis test with* $\alpha = 0.10$*, only 1 of the 10 (or the 10% you would expect in the long run) would lead to rejecting the null hypothesis:*

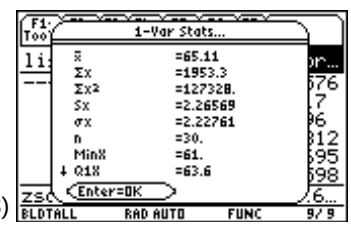
Sample 4 if H_0: $p < p_0$ *as z is negative* (-1.6542) *with p*-value = 0.098 < 0.10*, or*

Sample 6 if H_0: $p > p_0$ *as z is positive.*

Topic 27—Large Sample Test for a Mean

Example: In the past, a population in a certain age group had a mean height of $\mu = 5$ ft 4 in, or 64 inches with a standard deviation $\sigma = 2.4$ inches. Test to see if the mean height has increased by taking a random sample from the population of size n = **30**, with the results $\bar{x}$ = **65.11** and s_x = **2.26569** (screen 8).

(8)



This is the same data used in Topic 23, screen 9.

Since n = **30**, you could replace σ by s_x = **2.27** (as in Topic 23). Assume σ is known to be 2.4 from the past because often populations of measurements might have a mean change over time, but no change in the variation on σ.

Since $n \geq 30$, you can assume that sample means are normally distributed (according to the Central Limit Theorem), so a z test is appropriate.

H_0: $\mu = 64$

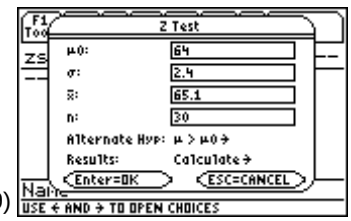
H_a: $\mu > 64$

with significance level $\alpha = 0.05$

1. In the Stats/List Editor, press [2nd] [F6] **Tests**, **1:ZTest**, and use Data Input Method: **Stats**.

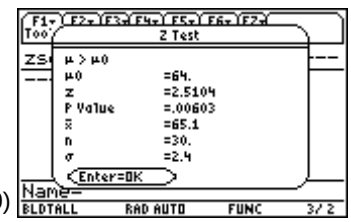
2. Press [ENTER] to display screen 9 and type the values μ_0: **64**, σ: **2.4**, $\bar{x}$: **65.1**, n: **30**, and Alternate Hyp: **$\mu > \mu_0$**.

(9)



3. Press [ENTER] to display screen 10 with p-value = **.00603** < .05 . You reject H_0 and conclude that the mean is significantly greater than 64 inches.

(10)



Notice: $z = \dfrac{65.1 - 64}{\dfrac{2.4}{\sqrt{30}}} = 2.5104$,

p-value = normcdf(2.5104,∞) = 0.00603.

If you used σ **= 2.27 =** s_x, the p-value would be **0.003975 ≈ 0.004**. (See Topic 31 for the t test procedure.)

From Topic 23, screen 12 you could come to the same conclusion with a 90% confidence interval (5% in the upper tail), with σ = **2.27** or (64.4 to 65.8) which has all values greater than 64.

Topic 28—Large Sample Test for the Difference Between Two Proportions

Example: Test whether there is a difference in the proportion of men or women in a population who agree with a certain public policy. Base your conclusion on the results of the simple random samples given below where ***x*** is the number of people in the sample of size ***n*** that agree.

Note: n_1 = 936 = 694 + 242 *(both values > 5).*

n_2 = 941 = 645 + 296 *(both values > 5), therefore the normal approximation to the binomial can be used.*

Women	$x_1 = 694$	$n_1 = 936$	or 694/936 = 74.15%
Men	$x_2 = 645$	$n_2 = 941$	or 645/941 = 68.54%

H_0: $p_1 = p_2$ or $(p_1 - p_2 = 0)$

H_a: $p_1 \neq p_2$

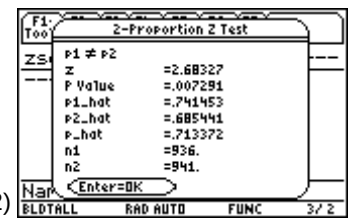
with significance level $\alpha = 0.05$

1. In the Stats/List Editor, press 2nd [F6] **Tests**, **6:2-PropZTest**, and enter Successes, x1: **694**, n1: **936**, Successes, x2: **645**, n2: **941**, and Alternate Hyp: $\mathbf{p_1 \neq p_2}$ (screen 11).

(11)

2. Press ENTER to display screen 12 with p-value = **0.007** < .05. There is very strong evidence that a statistically significantly larger percentage of women than men agree with the public policy.

(12)



Home screen calculation:

The pooled proportion, or
p-hat = **(694 + 645)/(936 + 941) = 0.713**,
z = **(0.741 - 0.685)/√(0.713 ∗ (1 - .713) ∗ (1/936 + 1/941)) = 2.682**,
and p-value = **2 ∗ normcdf(2.682,∞) = 0.0073**.

Note: *The difference was represented in Topic 24, screen 23 with a 95% confidence interval of* 5.6% ± 4.1%.

Topic 29—Large Sample Test for the Difference Between Two Means (Unpaired and Paired)

Unpaired or Independent Samples

Example: Test the claim that teaching Method B results in higher test scores than Method A. Base your conclusion on the following statistics from two random samples of students. The first sample received Method A and the second sample received Method B.

Method	Mean	s_x	n
A	75.2	8.42	32
B	78.4	8.13	30

H_0: $\mu_1 = \mu_2$ or $(\mu_1 - \mu_2 = 0)$

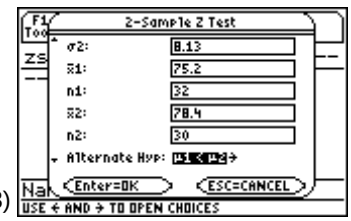
H_a: $\mu_1 < \mu_2$

with significance level $\alpha = 0.05$

Since n_1 and n_2 are both ≥ 30, you will use s_{xA} and s_{xB} to replace σ_1 and σ_2, as justified in Topic 23. By the Central Limit Theorem, the sample means are normally distributed so a 2-Sample *z* test will be used. (See Topic 32 for the 2-Sample *t* test procedure.)

1. In the Stats/List Editor, press 2nd [F6] **Tests**, **3:2-SampZTest**, and use Data Input Method: **Stats**, with σ_1: **8.42**, σ_2: **8.13**, $\bar{x}1$: **75.2**, n1: **32**, $\bar{x}2$: **78.4**, n2: **30**, and Alternate Hyp: $\mu_1 < \mu_2$ (screen 13).

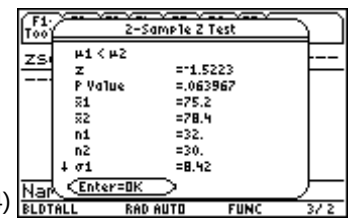
(13)



2. Press ENTER to display screen 14 with *p*-value = **0.063967** and z = **-1.5223**.

   Since 0.063967 > .05, these results are not so unusual if the null hypothesis were true. So you will not reject the null hypothesis H_0 and conclude that Method B does not result in a (statistically) significant higher mean score. The same conclusions result from a confidence interval as in Topic 25, screen 26 with a 90% confidence interval (5% in both tails) which gives $-6.7 < \mu_1 - \mu_2 < 0.3$. This contains the possibility of zero, or no difference in the mean scores for the two methods.

(14)



Note: $z = (75.2 - 78.4) / \sqrt{(8.422 / 32 + 8.132 / 30)} = -1.5223$, *p*-value = normcdf(-∞, -1.5223) = 0.064.

Matched Pairs or Dependent Samples

Example: To test the claim that a blood pressure medication reduces the diastolic blood pressure, a random sample of 30 people with high blood pressure had their pressures recorded. After a few weeks on the medication, their pressure was recorded again. This data was given in Topic 25, screen 28 with the pressures before taking the medication stored in **list1** and the pressures after taking the medication stored in **list2**. **List3** contained the differences in pressures (positive values if the pressure dropped). Screen 15 shows some of this data plus the mean $\bar{d} = 10$ and the stdDev = s_d = 7.55212.

(15)

H_0: $\mu_d = 0$ or H_0: $\mu_B - \mu_A = 0$

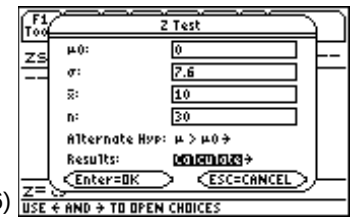
H_a: $\mu_d > 0$

with significance level $\alpha = 0.05$

With n = **30** ≥ 30, you can replace $\sigma_d = s_d = 7.6$, as justified in Topic 23. From the Central Limit Theorem, you can assume the sampling distribution of the mean of the differences $\bar{d}$ is normally distributed, so you will use a z test. (See Topic 32 for the t test procedure.)

1. In the Stats/List Editor, press 2nd [F6] **Tests**, **1:ZTest**, and use Data Input Method: **Stats**, with μ_0: **0**, σ: **7.6**, $\bar{x}$: **10**, n: **30**, and Alternate Hyp: **$\mu > \mu_0$** (screen 16).

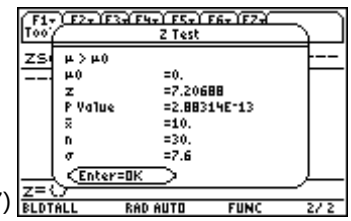
(16)



2. Press ENTER for the results (screen 17).

   With p-value = **$2.88 \times 10^{-13} \approx 0.000$**, there is very strong evidence that there is a statistically significant drop in the blood pressure. Topic 25, screen 33 gave a 90% confidence interval (5% in each tail) $7.7 < \mu_d < 12.3$ with an estimate well above the null hypothesis of zero.

(17)



Note: $z = \dfrac{\bar{d} - 0}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{10}{\dfrac{7.6}{\sqrt{30}}} = 7.21$,

p-value=normCdf (7.6, ∞) = 2.8E-13

Topic 30—Chi-Square Tests for Goodness-of-Fit, Homogeneity of Proportions, and Independence (One- and Two-Way Tables)

Goodness-of-Fit Test

Example: Binomial Distribution (check on randBin generator). For this example, change to folder **MAIN**.

In Topic 16, screen 22, you generated the number of successes in five trials (n = **5**) with the probability of success of 1/3 for each trial. This was repeated 100 times and a histogram (Topic 16, screen 24) was used to make a frequency table. The table is repeated below and in screen 18, with **list1** and **list2** keyed in and **list3** generated with **tistat.binompdf(5,1/3)** and recorded with four decimal places.

X	Freq	p(X)
0	10	.1317
1	32	.3292
2	36	.3292
3	15	.1646
4	7	.0412
5	0	.0041
Total	100	1.0000

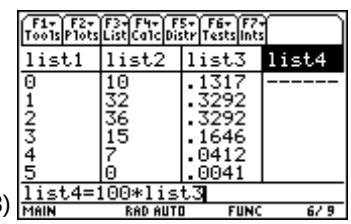
X (list1)	Obs. Freq (list2)	Exp. Freq (list4)
0	10	13.17
1	32	32.92
2	36	32.92
3	15	16.46
4 or more	7	4.53
Total	100	100

Test H_0: the distribution generated by randBin is binomial with n = **5** and p = **1/3**.

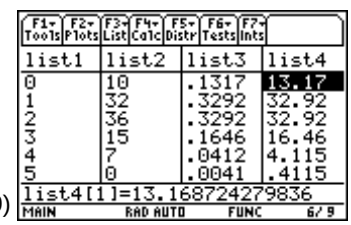
H_a: the data come from some other distribution

1. With the observed values in **list2** adding to 100, and the probability or proportion of the number of success in **list1** given in **list3**, multiply **list3** by 100 (screen 18), with **list4** highlighted for the expected values in **list4** (screen 19).

(18)



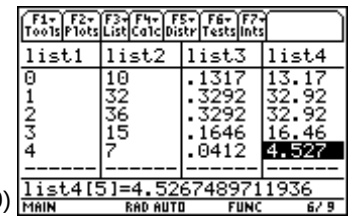
(19)



2. Consolidate the last two rows across lists 1 through 4 so that the **4** in **list1** now indicates getting four or five successes (screen 20).

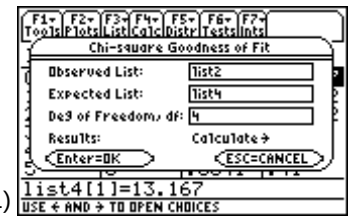
Without doing this, the expected value of five successes (.41) would be too small. All expected frequencies should be at least one, and no more than one in five of the expected frequencies should be less than five for this test to be valid. You meet these criteria, as only one cell of 4.527 is less than 5.

(20)



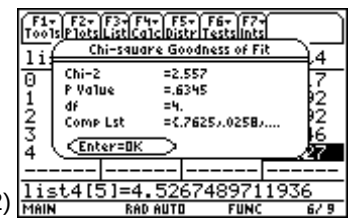
3. Press 2nd [F6] **Tests**, **7:Chi2 GOF**, enter Observed List: **list2**, Expected List: **list4**, Deg of Freedom, df: **4**, and Results: **Calculate** (screen 21).

(21)

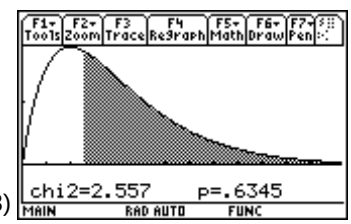


4. Press ENTER for the calculated output in screen 22 or the **Draw** output in screen 23. With *p*-value = **0.6345** > .05, you conclude that there is no significant difference between the observed and expected values (Chi-square close to zero), so these results are reasonable for a binomial distribution with n = **5** and p = **1/3**.

(22)

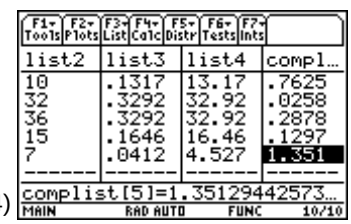


(23)



The components of Chi-square List started in the bottom line of screen 22 with **{.7625, .0258, . . .}** are stored at the end in the Stats/List Editor (screen 24). You see the largest component of 1.351 (with the sum of the components equal to the Chi-square statistic of 2.557) is the last, with Observed List: **7** and Expected List: **4.5267**.

(24)



Chi-Square Test for Two-Way Tables

Example: Assume that a sample of 277 students is selected. The following table shows the results of classifying each student by major and by the science elective they selected. The null hypothesis is: choice of major is independent of choice of science elective, and the alternative is: choice of major is not independent of choice of science elective. Is there a relationship between the major and the elective that is selected?

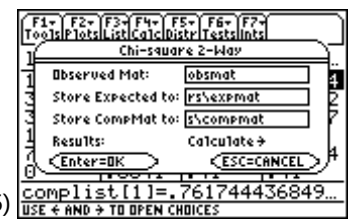
	Liberal Arts	Business	Elementary Education	Theater	Total
Gen Biology	88	46	11	9	154
Gen Physics	43	13	6	4	66
Basic Astronomy	26	13	10	8	57
Total	157	72	27	21	277

From the Home screen:

1. Enter the observed values in a matrix, with **[88,46,11,9;43,13,6,4;26,13,10,8]→obsmat**.

2. In the Stats/List Editor, press 2nd [F6] **Tests**, **8:Chi2 2way**, with the Observed Mat: **obsmat** from above, leaving Store Expected to: and Store CompMat to: with the default names (**expmat** and **compmat**, respectively) in **statvars (statvars\expmat and statvars\compmat)** (screen 25).

(25)

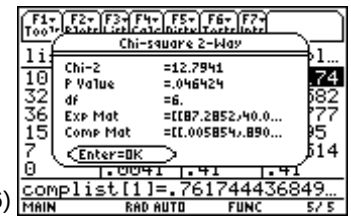


Note: *You could enter the matrix* [88,46,11,9;43,13,6,4;26,13,10,8] *in the* Observed Mat: *box instead of the named matrix. It will then be stored in* **statvars\obsmat**.

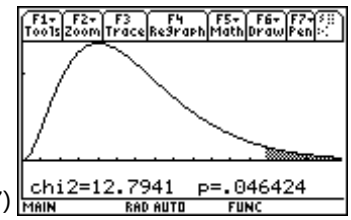
3. Press ENTER for Results: **Calculate** and the **Draw** option (screens 26 and 27). With *p*-value = **.046424** < .05, you reject the null hypothesis of no relationship and conclude the different majors select different science electives. You can explore these differences.

4. Press MODE and choose **FLOAT 3** from Display Digits. Press ENTER.

(26)

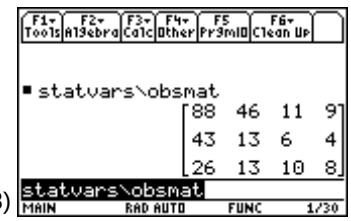


(27)



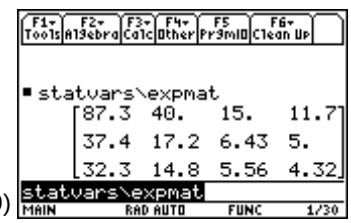
5. Display **statvars\obsmat**, **expmat**, and **compmat** by entering them on the status line on the Home screen (screens 28, 29, and 30).

(28)



Only 1 of 12, or less than 20% of the expected values are less than 5 and all are greater than 1. Notice that the largest relative difference in the observed and expected values and thus the largest components to the Chi-square statistic are the last two values in the last row. More Elementary Education and Theater majors are taking Astronomy than you would expect if there were no relationship. You observe frequencies of 10 and 8, while expecting frequencies more like 5.6 and 4.3 for Chi-square components of $3.55 + 3.13 = 6.68$, more than half the total value of 12.79.

(29)



(30)