Math 121 **Test Questions** **Spring 2010**

Chapters 13 and 14

1. (10 pts)

(a) Multiply 0.25 by the sample size of 1120 to get 280 for each category.

	Class	Freshman	Sophomore	Junior	Senior
Enrollment	Actual	314	305	261	240
	Expected	(280)	(280)	(280)	(280)

(b) The seven steps:

1. H_0 : The populations are homogeneous.
H_1 : The populations are not homogeneous.

2. $\alpha = 0.05$.

3. $\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}$.

4.

$$\begin{aligned}\chi^2 &= \frac{(314-280)^2}{280} + \frac{(305-280)^2}{280} + \frac{(261-280)^2}{280} + \frac{(240-280)^2}{280}\\ &= 4.1286 + 2.2321 + 1.2893 + 5.7143\\ &= 13.3643.\end{aligned}$$

5. p-value $= \chi^2$`cdf(13.3643,E99,3)` $= 0.0039$.

6. Reject H_0.

7. The populations are not homogeneous.

You could enter the observed counts into list L_1 and the expected counts into list L_2. Then compute $(L_1 - L_2)^2/L_2$. Then use `sum(Ans)` to sum up the values to get 13.36.

2. (10 pts) The seven steps:

1. $H_0 : p_1 = p_2 = \cdots = p_6 = \frac{1}{6}$.
$H_1 : H_0$ is not true.

2. $\alpha = 0.01$.

3. $\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}$.

4. The observed and expected counts are

	1	2	3	4	5	6
Observed	10	14	16	25	23	32
Expected	(20)	(20)	(20)	(20)	(20)	(20)

$$\chi^2 = \frac{(10-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(32-20)^2}{20}$$
$$= 5.0 + 1.8 + 0.8 + 1.25 + 0.45 + 7.2$$
$$= 16.5.$$

5. p-value $= \chi^2$`cdf(16.5,E99,5)` $= 0.00555$.

6. Reject H_0.

7. The numbers do not all have probability $\frac{1}{6}$.

You could enter the observed counts into list $\texttt{L}_1$ and the expected counts into list $\texttt{L}_2$. Then compute $(\texttt{L}_1\texttt{-L}_2)^2\texttt{/L}_2$. Then use `sum(Ans)` to sum up the values to get 16.5.

3. (10 pts) The seven steps:

1. $H_0 : p_1 = p_2 = p_3 = 0.10, p_4 = p_5 = 0.20, p_6 = 0.30.$
$H_1 : H_0$ is not true.

2. $\alpha = 0.01$.

3. $\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}.$

4. The observed and expected counts are

	1	2	3	4	5	6
Observed	8	11	14	29	15	23
Expected	(10)	(10)	(10)	(20)	(20)	(30)

$$\chi^2 = \frac{(8-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(29-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(23-30)^2}{30}$$
$$= 0.4 + 0.1 + 1.6 + 4.05 + 1.25 + 1.6333$$
$$= 9.033.$$

5. p-value $= \chi^2$`cdf(9.033,E99,5)` $= 0.1077$.

6. Accept H_0.

7. The probabilities stated in the null hypothesis are correct.

You could enter the observed counts into list $\texttt{L}_1$ and the expected counts into list $\texttt{L}_2$. Then compute $(\texttt{L}_1\texttt{-L}_2)^2\texttt{/L}_2$. Then use `sum(Ans)` to sum up the values to get 16.5.

4. (12 pts) You need to enter the data into a 2×10 matrix and then run the χ^2 test.

(a) (10 pts) The question says to test "that the two distributions are the same." That means to test whether they are *homogeneous*. The seven steps:

1. H_0 : The populations are homogeneous.
H_1 : The populations are not homogeneous.

2. $\alpha = 0.05$.

3. $\chi^2 = \sum_{\text{all cells}} \frac{(O - E)^2}{E}$.

4. $\chi^2 = 94.91$.

5. p-value $= \chi^2$`cdf(94.91,E99,9)` $= 1.678 \times 10^{-16}$.

6. Reject H_0.

7. The populations are not homogeneous.

(b) (2 pts) The expected number of incidents of crime in Ashland in 2007 is 299.8.

5. (12 pts) You need to enter the data into a 2×3 matrix and then run the χ^2 test.

(a) (10 pts) The seven steps:

1. H_0 : The populations are homogeneous.
H_1 : The populations are not homogeneous.

2. $\alpha = 0.05$.

3. $\chi^2 = \sum_{\text{all cells}} \frac{(O - E)^2}{E}$.

4. $\chi^2 = 0.3333$.

5. p-value $= \chi^2$`cdf(0.3333,E99,2)` $= 0.8465$.

6. Accept H_0.

7. The populations are homogeneous.

(b) (2 pts) The expected values are shown in the following table.

Type of road	Interstate	Primary	Secondary
No. of crashes	765 (764.1)	3505 (3510.9)	1920 (1915.0)
No. of injuries	59 (59.9)	281 (275.1)	145 (150.0)

6. (12 pts)

(a) (8 pts) You need to enter the data into a 2×4 matrix and then run the χ^2 test. The seven steps:

1. H_0 : The populations are homogeneous.
H_1 : The populations are not homogeneous.

2. $\alpha = 0.05$.

3. $\chi^2 = \sum_{\text{all cells}} \frac{(O - E)^2}{E}$.

4. $\chi^2 = 5.378$.

5. p-value $= \chi^2$`cdf(5.378,E99,3)` $= 0.1461$.

6. Accept H_0.

7. The populations are homogeneous.

(b) (2 pts) The first row total is 152, the last column total is 2507, and the grand total is 8333, so the expected count in the first row, last column is

$$\frac{1152 \times 2507}{8333} = 45.7.$$

If you used the TI-83 χ^2`-Test`, you can check the matrix of expected counts and find the value 45.7 in the first row, last column.

7. (12 pts)

(a) (2 pts) One variable is whether the road is straight or curved. The other variable is whether the driver was speeding or not speeding.

(b) (8 pts) Put the data into a 2×2 matrix and use the χ^2 test. The problem was stated in terms of the independence of the variables. Therefore, the hypotheses should also be stated in those terms. The seven steps:

1. H_0 : The variables are independent.
H_1 : The variables are not independent.

2. $\alpha = 0.05$.

3. $\chi^2 = \sum_{\text{all cells}} \frac{(O-E)^2}{E}$.

4. $\chi^2 = 1814.5$.

5. p-value $= \chi^2$`cdf(1814.5,E99,1)` $= 0$.

6. Reject H_0.

7. The variables are not independent. (That is, speed is more likely to be a factor in a traffic death that occurs on a curved road than it is on a straight road.)

(c) (2 pts) The first row total is 13663, the first column total is 31498, and the grand total is 42398. So the expected count in row 1, column 1 is

$$\frac{13663 \times 31498}{42398} = 10150.4.$$

You can check the matrix of expected counts and see the value 10150.4 in row 1, column 1.

8. (25 pts) Enter the SAT-Math values (x) into list $\texttt{L}_1$ and enter the tuition values (y) into list $\texttt{L}_2$.

(a) (5 pts) Use `LinReg(a+bx)` $\texttt{L}_1,\texttt{L}_2,\texttt{Y}_1$. The TI-83 reports that $a = -12.47$ and $b = 0.0556$. So the equation of the regression line is

$$\hat{y} = -12.47 + 0.0556x.$$

(b) (3 pts) In the previous part, the TI-83 also reported that $r = 0.2698$, which is the correlation coefficient.

(c) (3 pts) The value of r is positive and relatively small. That tells us that there is little to no correlation, but what correlation there is, is positive.

(d) (4 pts) Multiply the slope by the change in x to get the change in $\hat{y}$: $0.0556 \times 10 = 0.556$. We would expect the tuition to increase by \$556.

(e) (4 pts) Substitute 573 into the regression equation for x and calculate the value of $\hat{y}$. We get $\hat{y} = 19.388$, or \$19,388. You could also enter and evaluate the expression `Y₁(573)` and get the same answer.

(f) (3 pts) SAT-Math is the explanatory variable and Tuition is the response variable. (We always let x be the explanatory variable and y be the response variable.)

(g) (3 pts) These are *confounding* factors.

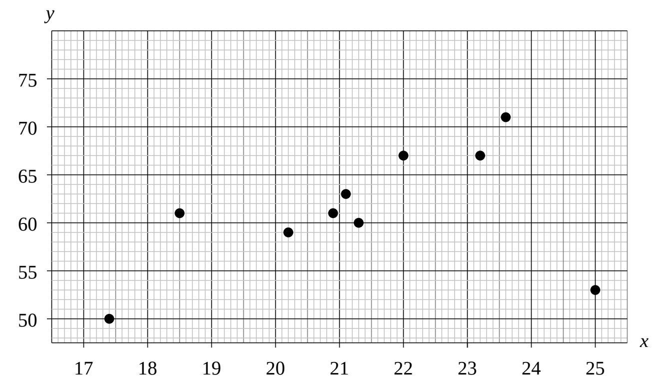
9. (32 pts) The Project on Student Debt recently published a report on student debt after graduation from college in 2007. The report included figures on the average amount of debt by state (average for Virginia is \$18,084) and percentage of students who graduate with debt (percentage for Virginia is 59%). A random sample of 10 states provides the following data.[1]

State	Average Debt (\$1000s) (x)	Percentage with Debt (y)
Alaska	25.0	53
Alabama	20.9	61
Delaware	17.4	50
Indiana	21.3	60
Kansas	18.5	61
Massachusetts	21.1	63
Ohio	22.0	67
Pennsylvania	23.6	71
Rhode Island	23.2	67
South Carolina	20.2	59

Enter the average debt values (x) into list L_1 and enter the percentage values (y) into list L_2.

(a) (4 pts) The scatter plot:

[1] www.projectonstudentdebt.org



(b) (3 pts) Overall, the relationship appears to be positive and strong, with the exception of one value $(25.0, 53)$.

(c) (5 pts) Use `LinReg(a+bx) L₁,L₂,Y₁`. The TI-83 reports that $a = 37.13$ and $b = 1.1289$. So the equation of the regression line is

$$\hat{y} = 37.13 + 1.1289x.$$

(d) (4 pts) The correlation coefficient is $r = 0.4070$, which indicates a moderate to weak positive relationship. (That outlier took its toll on r.)

(e) (3 pts) The coefficient of determination is $r^2 = 0.1657$, which means that 16.57% of the variation in the percentage of students in debt (from state to state) is explained by the variation in average debt (from state to state).

(f) (4 pts) Given that SST = 629.6, find SSR and SSE. Use the formula that $\text{SSR} = r^2 \cdot \text{SST}$. We get $\text{SSR} = 0.1657(629.6) = 104.3$. Then use the formula that $\text{SSE} = \text{SST} - \text{SSR}$ to get $\text{SSE} = 629.6 - 104.3 = 525.3$. (In this problem, I gave the wrong value for SST. It should have been 365.6, in which case SSR = 60.6 and SSE = 305.0.) You could calculate SSR and SSE directly: Compute `Y₁(L₁)` to get the $\hat{y}$ values and store them in `L₃`. Then compute

$$\text{SSR} = \sum(\hat{y} - \overline{y})^2 = 60.6$$
$$\text{SSE} = \sum(y - \hat{y})^2 = 305.0.$$

(g) (4 pts) Substitute 20.921 into the regression equation for x and get $\hat{y} = 60.75$.

(h) (5 pts) Use the TI-83 function `LinRegTTest`. The seven steps are

1. $H_0 : \beta = 0, \rho = 0$.
$H_1 : \beta \neq 0, \rho \neq 0$.

2. $\alpha = 0.05$.

3. $t = \dfrac{b - 0}{\text{SE}(b)}$, where $\text{SE}(b) = \dfrac{s}{\text{SSX}}$ and $s = \sqrt{\dfrac{\text{SSE}}{n-2}}$.

4. (I'll use the *corrected* values from part (f) so that my answer will agree with `LinRegTTest`.) Compute

$$\begin{aligned} s &= \sqrt{\frac{305.0}{10-2}} = 6.1745 \\ \text{SSX} &= \sum (x-\overline{x})^2 = 47.536 \\ \text{SE}(b) &= \frac{6.1745}{\sqrt{47.536}} = 0.8956 \\ t &= \frac{1.1289 - 0}{0.8956} = 1.261. \end{aligned}$$

5. p-value $= 2 \times$ `tcdf(1.261,E99,8)` $= 0.2430$.

6. Accept H_0.

7. The model is not significant. (It's that outlier again!)

Of course, when you use the `LinRegTTest` function, it will do the calculations in steps 4 and 5 for you. It gives the values $t = 1.260$ and p-value $= 0.2430$.