Topic 26: Goodness of Fit

As an introduction consider a problem that we examined back in Topic 17; it is a hypothesis test for a population proportion. Let us say that we have a strange solid figure that has 8 irregular faces numbered 1 through 8. We roll this solid and it always stops resting on one of those faces. The number on that face is the result of the roll. We believe that the proportion of times that the solid rolls onto a "3" is 0.1. We want to test that hypothesis at the 0.02 level of significance. To do this we will roll the solid over and over and we will record the number of rolls and the number of times that the roll results in a "3". It turns out that we roll the solid 225 times and 13 times it results in a "3". In Topic 17 we learned that we can run the test of the null hypothesis H_0: $p=0.1$ against the alternative H_1: $p\neq0.1$ by using the commands

```
source("../hypo_prop.R")
hypoth_test_prop( 0.1, 13, 225, 0, 0.02)
```

and the console output of that would be

```
> source("../hypo_prop.R")
> hypoth_test_prop( 0.1, 13, 225, 0, 0.02)
                H0_p                   H1                    x
               "0.1"      "prop !=  0.1"                 "13"
                  n           sig level       s.d. of prop
              "225"              "0.02"             "0.02"
            z-score            crit low           crit_high
 "2.32634787404084" "0.0534730425191832"  "0.146526957480817"
          samp prop                    z             attained
"0.0577777777777778"  "-2.11111111111111" "0.0347627626222284"
           decision
      "Do Not Reject"
```

In particular, even though our sample had 13 out of 225 instances of rolling a "3", and this represents a sample proportion of 0.05778, way below the hypothesis value of 0.1, we still do not have enough evidence to reject the null hypothesis, at the 0.02 level of significance, in favor of the alternative that the true proportion is not 0.1.

Now, we want to change the problem and we want to look at a null hypothesis that tells us the proportion of each of the 8 faces. Thus, our H_0 is

H_0: $p_1=0.15$, $p_2=0.09$, $p_3=0.1$, $p_4=0.15$, $p_5=0.135$, $p_6=0.08$, $p_7=0.14$, $p_8=0.155$

and our alternative is just that not all of proportions in the null hypothesis are correct. Although we might be tempted to just run our hypoth_test_prop() for each of the 8 different proportions that is not an appropriate approach. To do that we would be running multiple tests to answer just one question, namely is the entire null hypothesis correct?

This new problem is called a test of the goodness of fit. We have a hypothesis that tells us the proportion for each of the possible outcomes. We have a sample and from that sample we can get the number of instances of each outcome. We want to know if the null hypothesis is true, then how strange would it be to get a distribution of outcomes as strange or stranger than we got in our sample. If it would be really strange to get the sample outcomes then we have evidence to reject the null hypothesis.

To do any work on this problem we need to input the null hypothesis proportions.

```
 9 # for goodness of fit we want the proportions for all
10 #  of the possible outcomes.  These are the null
11 #  hypothesis proportions:
12 #
13 null_props <- c(0.15, 0.09, 0.1, 0.15,
14                  0.135, 0.08, 0.14, 0.155)
```

```
> # for goodness of fit we want the proportions for all
> #  of the possible outcomes.  These are the null
> #  hypothesis proportions:
> #
> null_props <- c(0.15, 0.09, 0.1, 0.15,
+                 0.135, 0.08, 0.14, 0.155)
```

Then we can use those proportions to find the expected number of times each outcome should show up if we roll the solid 225 time (225 being the size of the sample we will take).

```
15 #  Then, knowing that we will take or even have taken a sample
16 #  of size 225, find the expected values for each outcome
17 expected <- null_props * 225
18 expected
```

```
> #  Then, knowing that we will take or even have taken a sample
> #  of size 225, find the expected values for each outcome
> expected <- null_props * 225
> expected
[1] 33.750 20.250 22.500 33.750 30.375 18.000 31.500 34.875
```

It is worth noting that seven of these "expected" values could never happen. For example, we could not possibly get 33.75 instances of a "1". Even though these are impossible values they remain our expected values.

Next we will take our sample and we will find the number of instances of each outcome in that sample.

```
20 source("../gnrnd5.R")
21 gnrnd5(95632022407,985785588)
22 L1
23 table(L1)
```

```
> source("../gnrnd5.R")
> gnrnd5(95632022407,985785588)
style= 7   size= 225    seed= 95632   num digits= 0   alt_sign= 1
8 5 5 8 7 5 8 9
1 1 1 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 6 6
 6 6 6 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8
[1] "DONE "
> L1
  [1] 8 4 2 2 1 8 2 2 7 8 3 3 8 8 5 8 7 1 7 2 1 7 5 4 4 2 1 3 6 5 8
 [32] 5 5 2 8 4 8 8 5 4 1 1 4 2 4 3 8 1 6 1 1 3 6 1 5 1 7 7 2 7 7 4
 [63] 5 4 7 2 7 4 7 5 5 2 2 8 4 2 7 4 4 5 5 6 5 2 7 8 4 1 1 8 4 1 8
 [94] 2 4 7 7 1 6 1 6 4 3 8 3 7 5 8 4 8 4 5 8 7 5 5 8 4 4 3 4 8 4 4
[125] 5 8 5 8 4 4 5 3 7 5 1 5 7 5 2 6 8 2 7 6 2 4 4 1 7 5 7 1 8 8 8
[156] 3 4 7 8 7 7 7 2 3 1 4 8 8 5 4 4 3 5 1 7 8 1 4 7 1 4 4 7 6 4 7
[187] 6 5 7 6 5 8 6 8 6 6 4 7 8 7 6 7 3 8 6 7 8 7 7 3 5 7 8 5 1 6 4
[218] 1 8 6 1 7 8 4 1
> table(L1)
L1
 1  2  3  4  5  6  7  8
27 19 14 39 30 18 39 39
```

Those counts of the frequency of each outcome are our "observed" values. Once we have that we can find the difference between the "observed" and the expected values.

```
#so here are the observed values
observed <- c(27, 19, 14, 39, 30, 18, 39, 39 )
observed
#  then we want to find the   observed - expected values
diff <- observed - expected
diff
```

```
> #so here are the observed values
> observed <- c(27, 19, 14, 39, 30, 18, 39, 39 )
> observed
[1] 27 19 14 39 30 18 39 39
> #  then we want to find the   observed - expected values
> diff <- observed - expected
> diff
[1] -6.750 -1.250 -8.500  5.250 -0.375  0.000  7.500  4.125
```

But those values are both positive and negative and they just cancel each other. Also, we want to really emphasize larger differences. Therefore, let us look at the squares of the differences.

```
#  and we move on from there to get the squares of those
#  differences
diff_sqr <- diff^2
diff_sqr
```

```
> #  and we move on from there to get the squares of those
> #  differences
> diff_sqr <- diff^2
> diff_sqr
[1] 45.562500  1.562500 72.250000 27.562500  0.140625  0.000000
[7] 56.250000 17.015625
```

That took care of the positive and negative values cancelling each other and it gives much more weight to larger differences. However, a difference from a large expected value should not mean as much as the same difference from a lower expected value. To make this adjustment we divide each of our squared differences by the expected value.

```
#  That magnified the values that we big differences and
#  it made everything positive.  Now divide each of those
#  by the respective   "expected"  value so that the same
#  differences from larger expected values carries less
#  weight than do similar differences form lower expected
#  values.
quotients <- diff_sqr / expected
quotients
```

```
> #  That magnified the values that we big differences and
> #  it made everything positive.  Now divide each of those
> #  by the respective   "expected"  value so that the same
> #  differences from larger expected values carries less
> #  weight than do similar differences form lower expected
> #  values.
> quotients <- diff_sqr / expected
> quotients
[1] 1.35000000 0.07716049 3.21111111 0.81666667 0.00462963
[6] 0.00000000 1.78571429 0.48790323
```

Then, the overall strangeness of our sample is measured by the sum of all of those quotients.

```
43 #  Now to find the overall "strangeness" of our observed
44 #  values from the expected values we get the sum of
45 #  all of those quotients.
46 how_strange <- sum( quotients )
47 how_strange
```

```
> #  Now to find the overall "strangeness" of our observed
> #  values from the expected values we get the sum of
> #  all of those quotients.
> how_strange <- sum( quotients )
> how_strange
[1] 7.733185
```

Even if our true population had exactly the proportions given in the null hypothesis, we would not expect a sample of 225 items to have those same proportions. Each sample would have differences between the observed values in that sample compared to the expected values for a sample of size 225. Therefore, each sample would have a value for "how strange" that sample seems to be. The distribution of those "how strange" values will be a χ^2 distribution with the degrees of freedom equal to one less than the number of different outcomes. We have 8 possible outcomes so we have 7 degrees of freedom. Therefore, we can ask, if the null hypothesis is true then how strange is it to get a "how strange" value of 7.733185 or higher?

```
61 pchisq( 7.733185, 7, lower.tail=FALSE)
```

```
> pchisq( 7.733185, 7, lower.tail=FALSE)
[1] 0.3567084
```

With 7 degrees of freedom we could ask for the "critical value", i.e., how large would our "how strange" total have to be so that the probability of getting that value or higher, assuming that the null hypothesis is true, would be, say, 0.02?

```
64 #  Or we could find the critical value for 7 degrees of freedom
65 #  and for a level of significance of 0.02.
66 qchisq( 0.02, 7, lower.tail=FALSE)
```

```
> #  Or we could find the critical value for 7 degrees of freedom
> #  and for a level of significance of 0.02.
> qchisq( 0.02, 7, lower.tail=FALSE)
[1] 16.62242
```

For any goodness of fit problem we would go through all of the same steps.

1) **get the null hypothesis proportions**
2) **determine our desired level of significance**
3) **determine the sample size**
4) **compute the expected values**
5) **take the sample**
6) **find the frequency of the the different categorical values, those frequencies are the observed values**
7) **compute the differences observed - expected**
8) **compute the squares of the differences**
9) **compute the quoitents of the squared differences divided by the expected values**
10) **find the sum of those quotients**
11) **either find how strange it would be to get that sum (pchisqr) or find the critical value (qchisq)**
12) **reject the null hypothesis if the probability of getting that sum is less than our level of significance or if the sum of the quotients is greater than our critical value.**

We could do each step, as outlined and demonstrated above, or we could just use the goodfit() function, telling it the names of the categories, the null hypothesis values, the observed values, and finally the level of significance.

```
68 ###  we can do all of this in one step with the goodfit()
69 ###  function
70 source( "../goodfit.R")
71 goodfit( 1:8, null_props, observed, 0.02)
```

```
> ###  we can do all of this in one step with the goodfit()
> ###  function
> source( "../goodfit.R")
> goodfit( 1:8, null_props, observed, 0.02)
  Items H_null observed expected   diff    diff^2 chi component
1     1  0.150       27   33.750 -6.750 45.562500    1.35000000
2     2  0.090       19   20.250 -1.250  1.562500    0.07716049
3     3  0.100       14   22.500 -8.500 72.250000    3.21111111
4     4  0.150       39   33.750  5.250 27.562500    0.81666667
5     5  0.135       30   30.375 -0.375  0.140625    0.00462963
6     6  0.080       18   18.000  0.000  0.000000    0.00000000
7     7  0.140       39   31.500  7.500 56.250000    1.78571429
8     8  0.155       39   34.875  4.125 17.015625    0.48790323
-----------------------------------
total observations =  225
Number of categories =  8
Number of degrees of freedom =  7
Total of chi component = 7.733185
P(total >=  7.733185 ) =  0.3567083
For  2 % to the right, the
    critical value  is  16.62242
[1] " not enough evidence to reject H0"
```