Goodness of Fit HT Exercises

Pages	**Suggested Reading**
465 – 473	Section 11.1, 11.3, 11.4

Pages	**Problems**
487 – 490	(***Section 11.11***) 2, 3, 4, 5 ,7

I mentioned something called "post hoc" testing in class. These tests help us answer the question (if we get a small P – value after a χ^2 GOF test) "which proportion(s) are the ones that are different from what was expected?" (remember that, if we get a big P – value, all the proportions are as expected) . Let's walk you through how to do one after a χ^2 GOF test. To review, you consider doing this post hoc testing once you're gotten a small P – value after a χ^2 GOF test ($\boxed{\text{PRGM}}$ GOF).

E1. "Sean, this post hoc stuff just seems like more work. Why do you make us do it? What's the point?"

Consider this:

National health care is currently a big issue for Americans. The October 21, 2009, *USA Today* article "Poll: Americans Skittish over Health Care Changes" reported the following percentages with respect to "Insurance company requirements you have to meet to get certain treatments covered" if the health care bill passes:

Viewpoint Sept. 11–13	Percentage
Get better	22
Not change	35
Get worse	38
Unknown	5

One month later, during October 16–19, another poll was taken of 1,521 adults. Those viewpoints are categorized in the table below.

Viewpoint Oct. 16–19	Number
Get better	380
Not change	380
Get worse	700
Unknown	61

When I ran a GOF test on this data, I got $\chi^2 \approx 78.5$, **P** ≈ 0. That means that, yep, at least one of the proportions has definitely changed from September. But which ***one***? That's what post hoc helps us discover.

(another reason you need post hoc testing…you can simply compare the "observed" column to the "expected" column, and see if they're off in some way…but what does "off" mean? How far "off" is "off enough"? Good question? Post hoc testing will take care of those gray areas for us)

Now, in order to identify which proportions are the culprit(s), we need to run 4 individual Ztests (or confidence intervals…but we'll run Ztests, because that's what's most often done in the research I've seen). But you know that you have to be careful…if we run 4 1PropZTests, each at α = 5%, we're now down to $(0.95)^4$, or about 81.5% confidence in our answer. That 81.5% is called a "family – wise confidence level", and its complement (19.5%) is called the "family wise error rate". So, you can see, as the number of categories goes up, so does the chance of getting one or more false positives.

There are a couple of ways to deal with this, but the simplest I've found is to individually test each proportion using a 1PropZTest, but test at a smaller significance (α) level. This will offset the increased chance of a Type 1 error (false positive) that you create by running so many tests. The accepted α to use is found by dividing 0.05 by the number of categories that you have; so, in our example,

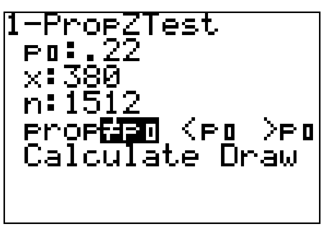
$$\textbf{Our GOF adjusted post hoc } \boldsymbol{\alpha} \textbf{ - level} = \frac{\mathbf{0.05}}{\textbf{number of categories}} = \frac{\mathbf{0.05}}{\mathbf{4}} = \mathbf{0.0125.}$$

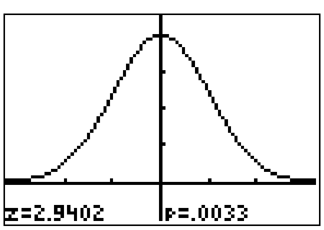
In other words, instead of comparing our post hoc P – values to their usual 5%, we now compare it to 1.25%. If they come in under 1.25%, then the percentage we observed is significantly different. If not, then it's not.

<u>E2</u>. Of course, while testing at a smaller α value reduces the chance of a Type 1 (false positive error), it increases the…

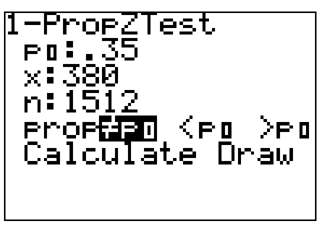
Let's see how to do this…we run a 1PropZTest on each data set with **n** = 1521. Here's what I see in each:

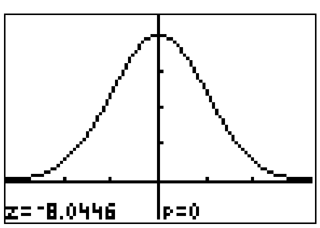
"Get Better"



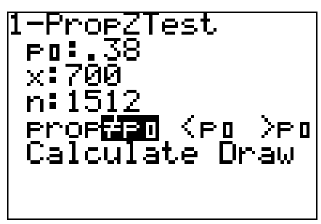


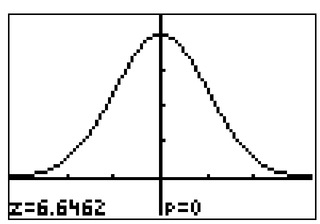
"Not Change"



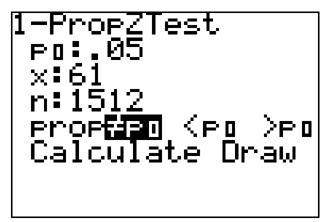


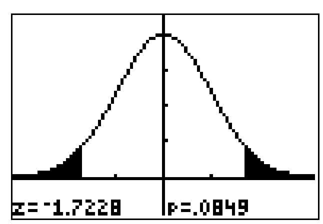
"Get Worse"





"Unknown"





So, here's my synopsis:

<u>Category</u>	<u>Result?</u>	<u>Conclusion?</u>
"Get Better"	**P** < 1.25%	There is a difference in the proportion of "Get Better" in October than in September.
"Not Change"	**P** < 1.25%	There is a difference in the proportion of "Not Change" in October than in September.
"Get Worse"	**P** < 1.25%	There is a difference in the proportion of "Get Worse" in October than in September.
"Unknown"	**P** > 1.25%	There is no difference in the proportion of "unknown" in October than in September*.

Now, I know you were only testing for a difference in these. It is possible, however, to see which way the proportions moved from September to October (i.e., got bigger or smaller).

E3. Which proportions got bigger? Which got smaller? Which, if any, remained unchanged?

E4. Which of the homework questions would require post hoc testing?

Answers.

E1. Remember that a small P – value from a GOF test only tells us that at least one of the proportions we were expecting was wrong, but it doesn't tell us ***which*** one(s). The post hoc tests will identify the culprit(s) that caused the small P – value.

E2. …chance of a Type 2 (false negative) error (β). Remember, all things being equal, α and β are inversely proportional. This is one of the criticisms of this method (this method, BTW, is called a "Bonferroni correction")

E3. "Get Better" and "Get Worse" are higher, "No Change" is lower, and "Unknown" is unchanged (look at the test statistics). And, the reason that you're allowed to say which direction they changed is due to the answer in question E1(e) from your last homework assignment (2 proportion Z – testing).

E4. In leiu of an answer to this, I'll supply you with a flowchart to help you with post hoc testing (link: http://coccweb.cocc.edu/srule/MTH244/homework/PHflowchart.xlsx). It also includes ANOVA post hoc testing (ANOVA is the next topic in class).

Goodness of Fit HT Quizzes

Quiz 1.

An actuary for a certain insurance company wants to ascertain if high performance cars with powerful engines are more susceptible to accidents than other cars. In 50 insurance claims, she classifies the cars as high performance, sub compact, mid size or full size:

Type of Car	High Performance	Sub Compact	Mid Sized	Full Sized
Number of Accidents	20	14	7	9

The distribution of cars (by type) is as follows: 10% are high performance, 40% are subcompacts, 30% are mid-sized and 20% are full-sized. Does the data imply that the distribution of accidents varies significantly from the distribution of car types?

Quiz 2.

1. There's a neat statistical construct called Benford's Law which states that, in many datasets, the leading digit should not follow a uniform distribution (like the songs did on our iPod), but, instead, follow this distribution:

Leading digit…	1	2	3	4	5	6	7	8	9
…should appear in this percentage of the data:	30%	18%	12%	10%	8%	7%	6%	5%	4%

There was some media furor that erupted after the 2009 Iranian elections, claiming that the election was rigged after incumbent Mahmoud Ahmadinejad won with a comfortable margin among alleged shenanigans (I know…how ***unlike*** politics).

One thing we can do, statistically, is see whether or not the distribution of votes cast follows Benford's Law (the logic being, if votes ***were*** being fabricated, it wouldn't occur to normal, non – statistical thinking folks to make sure the leading digits followed some vague law that they were supposed to follow).

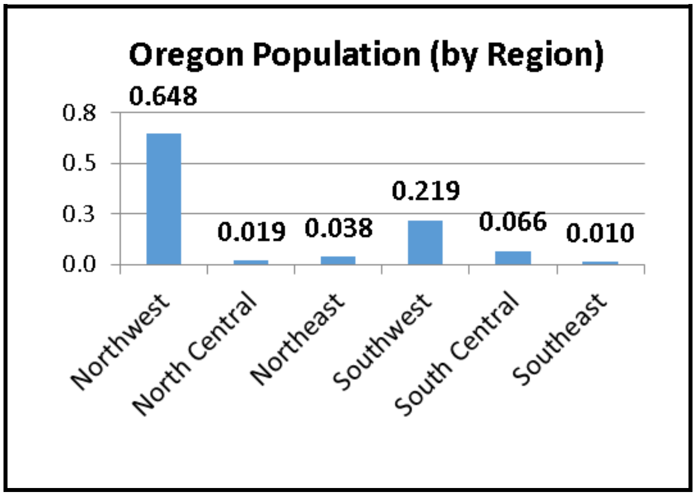
Here are the results from the votes cast across all provinces in Iran among all legitimate candidates:

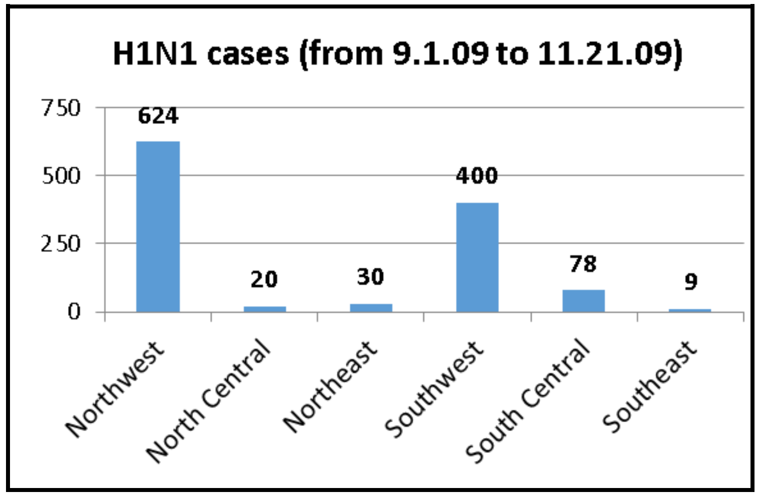
Leading digit…	1	2	3	4	5	6	7	8	9
Number of vote totals starting with that leading digit….	38	21	13	14	11	6	12	4	5

Do these data deviate significantly from what Benford's Law would predict?

Quiz 3.

According to data from the Census Bureau, the population of Oregon, by regional percentage, is given by the bar chart below left. The regional number of H1N1 cases this fall is shown below right (from flu.oregon.gov). Are the numbers of regional flu cases inconsistent with the population percentages?





Quiz 4.

For years, I played a game in my intro probability class that involved pulling blue and green blocks out of a bag (you might remember playing it). In this game, you had to pull four blocks out of the bag, and all possible color combinations, along with the number of observations of each color combination I noted over the years, were as follows:

Outcome	all Blue	3 Blue, 1 Green	2 Blue, 2 Green	1 Blue, 3 Green
Observations	96	156	88	30

Theoretically, the probability of each color outcome was as follows:

Outcome	all Blue	3 Blue, 1 Green	2 Blue, 2 Green	1 Blue, 3 Green
Probability	1/6	1/2	3/10	1/30

Were the observations I noticed in class were consistent with the expected probabilities?

(**extra 5 points**) (**w**) How many blue blocks were in the bag? This is most definitely a MTH 243 question, and a ***good*** one; hence the extra points. Good luck!