Homework 8, due Thursday November 3 in lab ---- **Solutions**

(assignment for these solutions can be found on the last page)

1. *Compare the observed proportions across months qualitatively and through a formal goodness-of-fit test.*

Qualitatively: From the table below with observed and expected values, and the first chart below, it appears that some months' observed values are not very close to the expected values. In particular, February and May stand out as different.

Quantitative goodness-of-fit test:
H0: the proportion of suicides is the same each month
HA: the proportion of suicides is different for at least one month

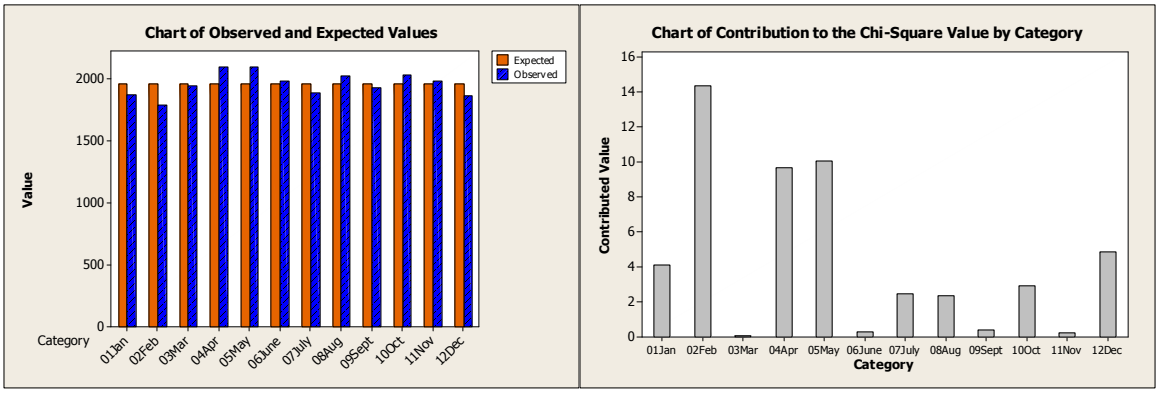
The *p*-value for the goodness-of-fit test below is 0.000, so we reject H0 in favor of HA, concluding that the proportion of suicides is dfferent for at least one month. In the chart of contribution to the chi-square value below we see that February is the most different, and because the observed value is less than the expected value, it is underrepresented (what an awful thing to say). This may be due to Feb actually having the fewest number of days. We also see that both April and May are overrepresented. Decemeber and January are also slightly underrepresented.

Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: Suicides

Using category names in Month

Category	Observed	Test Proportion	Expected	Contribution to Chi-Sq
01Jan	1867	0.0833333	1956.67	4.1091
02Feb	1789	0.0833333	1956.67	14.3673
03Mar	1944	0.0833333	1956.67	0.0820
04Apr	2094	0.0833333	1956.67	9.6391
05May	2097	0.0833333	1956.67	10.0648
06June	1981	0.0833333	1956.67	0.3026
07July	1887	0.0833333	1956.67	2.4805
08Aug	2024	0.0833333	1956.67	2.3171
09Sept	1928	0.0833333	1956.67	0.4200
10Oct	2032	0.0833333	1956.67	2.9004
11Nov	1978	0.0833333	1956.67	0.2326
12Dec	1859	0.0833333	1956.67	4.8750

N	DF	Chi-Sq	P-Value
23480	11	51.7905	**0.000**



2. *Is there a difference in the death rate for infants whose mothers had less than one month of pre-natal care as compared with infants whose mothers had at least one month of pre-natal care?*
    a. *Conduct an hypothesis test for differences between two binomial proportions. Clearly state populations, parameters, hypotheses and conclusion (and clearly show what you used to get to the conclusion).*

Population 1 are all mothers having less than one month pre-natal care, while population 2 are all mothers having at least one month pre-natal care.

Parameters are p1, for the population of infants who experience death within one month of birth for population 1, and similarly p2 for population 2.

H0: p1=p2
HA: p1≠p2
Is the population proportion of deaths different between mothers having less than one month compared to those having at least one month pre-natal care?

By the hypothesis test p-value=0.016<α=0.05 in the table below, we reject H0 in favor of HA, concluding that there is a difference in the death rate for infants from our two populations. The same conclusion is drawn using the confidence interval not containing 0.

    b. *Construct a 95% confidence interval for the difference of two binomial proportions. Clearly interpret this interval.*

We are 95% confident that the true population difference in proportions is in the interval (0.00598384, 0.0585302).

Test and CI for Two Proportions

```
Sample   X    N  Sample p
1       20  393  0.050891
2        6  322  0.018634

Difference = p (1) - p (2)
Estimate for difference:  0.0322570
95% CI for difference:  (0.00598384, 0.0585302)
Test for difference = 0 (vs not = 0):  Z = 2.41  P-Value = 0.016
```

3. Blood type vs. Ethnicity data
   a. *Summarize these data, focusing on descriptive and graphical of the 4 ethnic groups on the proportions in the 4 blood groups. You should make a bar graph showing the distribution of blood type for each of the ethnic groups (make sure you use proportions, not counts).*

Below is a frequency table that gives the original cell counts and the column percentages. We can compare these percentages of blood type across the ethnicities. We observe the percentages for A is high for Hawaiian, AB is low for Hawaiian and Hawaiian-White, B is low for Hawaiian, and O is high for Hawaiian-White and White. That is one way to summarize the data below.

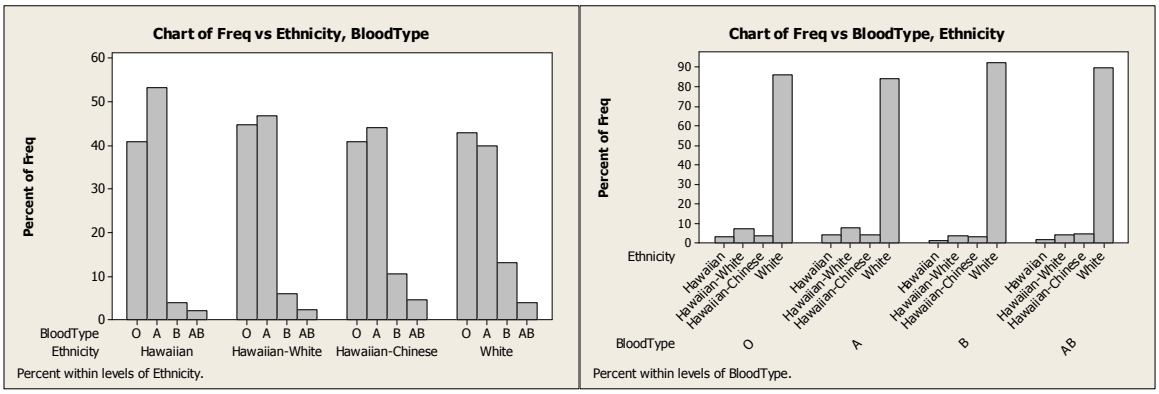
Looking at the graphs, B and AB appear lower in Hawaiian than in White, and A is higher in Hawaiian than in others.

Tabulated statistics: BloodType, Ethnicity

```
Using frequencies in Freq
Rows: BloodType   Columns: Ethnicity
```

	Hawaiian	Hawaiian-Chinese	Hawaiian-White	White	All
A	2490	2368	4671	50008	59537
	53.32	43.97	46.79	40.00	41.04
AB	99	243	236	5001	5579
	2.12	4.51	2.36	4.00	3.85
B	178	568	606	16252	17604
	3.81	10.55	6.07	13.00	12.14
O	1903	2206	4469	53759	62337
	40.75	40.97	44.77	43.00	42.97
All	4670	5385	9982	125020	145057
	100.00	100.00	100.00	100.00	100.00

```
Cell Contents:      Count
                    % of Column
```



b. Is there evidence that blood type and ethnicity are associated in Hawaii? Explain.

H0: proportions of bloodtype are the same for each ethnicity (no association).
HA: proportions of bloodtype are different for at least one ethnicity (an association).

The table below gives for each cell the counts, expected counts, residuals, and contribution to Chi-square. The Pearson chi-square p-value=0.000, so we reject H0 in favor of HA, concluding that the proportions of bloodtype are different for at least one ethnicity.

Looking at the contribution to Chi-square, cells that are particularly large include Hawaiian-White-B, Hawaiian-B, and Hawaiian-A. Several others are large, as well.

Tabulated statistics: BloodType, Ethnicity

Using frequencies in Freq
Rows: BloodType Columns: Ethnicity

	Hawaiian	Hawaiian-Chinese	Hawaiian-White	White	All
A	2490	2368	4671	50008	59537
	1917	2210	4097	51313	59537
	573.3	157.8	574.0	-1305.0	*
	171.45	11.26	80.42	33.19	*
AB	99	243	236	5001	5579
	180	207	384	4808	5579
	-80.6	35.9	-147.9	192.6	*
	36.18	6.22	56.99	7.72	*
B	178	568	606	16252	17604
	567	654	1211	15172	17604
	-388.7	-85.5	-605.4	1079.7	*
	266.65	11.19	**302.56**	76.83	*
O	1903	2206	4469	53759	62337
	2007	2314	4290	53726	62337
	-103.9	-108.2	179.3	32.7	*
	5.38	5.05	7.50	0.02	*
All	4670	5385	9982	125020	145057
	4670	5385	9982	125020	145057
	*	*	*	*	*
	*	*	*	*	*

Cell Contents: Count
Expected count
Residual
Contribution to Chi-square

Pearson Chi-Square = 1078.604, DF = 9, P-Value = **0.000**
Likelihood Ratio Chi-Square = 1242.179, DF = 9, P-Value = 0.000

c. Carry out any additional analyses that you deem relevant, and summarize your findings. This is open ended. For example, you might consider comparing all pairs of ethnicities on blood types, see which differ.

If we reduce our table to 6 tables, each containing a unique combination of two of the ethnicities, we can perform chi-square tests on each table to test if the blood-type proportions are the same. Below are the pairwise comparisons. Note that the p-value for each of the six tests is 0.000, so we conclude that the blood type proportions for each population is unique.

Tabulated statistics: BloodType_1, Ethnicity_1

```
Using frequencies in Freq_1
       Hawaiian  Hawaiian-Chinese    All
Pearson Chi-Square = 240.298, DF = 3, P-Value = 0.000
Likelihood Ratio Chi-Square = 251.476, DF = 3, P-Value = 0.000
```

Tabulated statistics: BloodType_2, Ethnicity_2

```
Using frequencies in Freq_2
       Hawaiian  Hawaiian-White    All
Pearson Chi-Square = 70.726, DF = 3, P-Value = 0.000
Likelihood Ratio Chi-Square = 72.378, DF = 3, P-Value = 0.000
```

Tabulated statistics: BloodType_3, Ethnicity_3

```
Using frequencies in Freq_3
       Hawaiian   White     All
Pearson Chi-Square = 543.057, DF = 3, P-Value = 0.000
Likelihood Ratio Chi-Square = 644.071, DF = 3, P-Value = 0.000
```

Tabulated statistics: BloodType_4, Ethnicity_4

```
Using frequencies in Freq_4
       Hawaiian-Chinese  Hawaiian-White    All
Pearson Chi-Square = 161.290, DF = 3, P-Value = 0.000
Likelihood Ratio Chi-Square = 155.046, DF = 3, P-Value = 0.000
```

Tabulated statistics: BloodType_5, Ethnicity_5

```
Using frequencies in Freq_5
       Hawaiian-Chinese   White     All
Pearson Chi-Square = 531.214, DF = 3, P-Value = 0.000
Likelihood Ratio Chi-Square = 613.512, DF = 3, P-Value = 0.000
```

Tabulated statistics: BloodType_6, Ethnicity_6

```
Using frequencies in Freq_6
       Hawaiian-White   White     All
Pearson Chi-Square = 531.214, DF = 3, P-Value = 0.000
Likelihood Ratio Chi-Square = 613.512, DF = 3, P-Value = 0.000
```

Biostatistics 1 Fall 2005 Homework 8 due Nov 3 in lab

Problem 1

The National Center for Health Statistics gave the following data on the distribution of suicides in the U.S. by month in 1990. Is there any evidence that the suicide rate varies monthly, or are the data consistent with the hypothesis that the rate is constant?

To simplify your analysis, assume the months have the same numbers of days. Compare the observed proportions across months qualitatively and through a formal goodness-of-fit test. Write a short and coherent summary to this problem.

Month	Number of Suicides
Jan	1867
Feb	1789
Mar	1944
Apr	2094
May	2097
June	1981
July	1887
Aug	2024
Sept	1928
Oct	2032
Nov	1978
Dec	1859

Problem 2:

A study published in 1969 was conducted to investigate the effect of the duration of pre-natal care and the place where that care is received on the survival of infants. The duration of care is classified as less than one month or at least one month. In our analysis we will look at the overall results, ignoring clinics. The results of the study are summarized in the table below:

	Duration of pre-natal care	
Status of infant one month after birth	< 1 Month	at least 1 month
DEAD	20	6
ALIVE	373	316

Is there a difference in the death rate for infants whose mothers had less than one month of pre-natal care as compared with infants whose mothers had at least one month of pre-natal care?

To answer the question, do the following:

a. Conduct an hypothesis test for differences between two binomial proportions. Clearly state populations, parameters, hypotheses and conclusion (and clearly show what you used to get to the conclusion).

b. Construct a 95% confidence interval for the difference of two binomial proportions. Clearly interpret this interval.

Problem 3:

There are four major blood groups in humans: O, A, B, and AB. A sample of individuals with records at the Blood Bank of Hawaii was selected. Each individual was classified according to blood type and ethnicity.

The following two-way table of counts was obtained.

Blood Type	Ethnicity			
	Hawaiian	Hawaiian-White	Hawaiian-Chinese	White
O	1903	4469	2206	53759
A	2490	4671	2368	50008
B	178	606	568	16252
AB	99	236	243	5001

a. Summarize these data, focusing on descriptive and graphical comparisons of the 4 ethnic groups on the proportions in the 4 blood groups. You should make a bar graph showing the distribution of blood type for each of the ethnic groups (make sure you use proportions,not counts). Graph > Bar Chart > Values from a table > Cluster does a nice job of this.

b. Is there evidence that blood type and ethnicity are associated in Hawaii? Explain.

c. Carry out any additional analyses that you deem relevant, and summarize your findings. This is open ended. For example, you might consider comparing all pairs of ethnicities on blood types, to see which differ.

Data:

```
Freq    Ethnicity       BloodType
1903    Hawaiian        O
2490    Hawaiian        A
178     Hawaiian        B
99      Hawaiian        AB
4469    Hawaiian-White  O
4671    Hawaiian-White  A
606     Hawaiian-White  B
236     Hawaiian-White  AB
2206    Hawaiian-Chinese        O
2368    Hawaiian-Chinese        A
568     Hawaiian-Chinese        B
243     Hawaiian-Chinese        AB
53759   White   O
50008   White   A
16252   White   B
5001    White   AB
```