Name:_____ Section_____ Date:_____

Use the following prompts to guide you through working with the cemetery data today.

1. *Research Hypothesis:*

2. *Two groups being compared:*

3. *Statistical Test Chosen:*

4. *Null Hypothesis of the Test:*

5. *Alternate Hypothesis of the Test:*

Please fill in the following tables based on your hypotheses and randomly-generated subset of data. Use the example provided to calculate your numbers. You may also choose to code these in R.

Variables of Note:

x: This is the time-specific interval. We refer to each interval by the first value listed, e.g. "0-5" is the x=0 time interval.

Obs #: This is the raw data for d_x, the number of deaths that occur in time x. Count these in your data.

n_x: This is the number of individuals alive at the beginning of each time x. We derive this by summing all individuals observed in our dataset (we make an assumption here that the frequency of individuals in each time interval approximates the lifespan).

l_x: This is the cumulative survival probability, n_x/n_0, after scaling to a per 1000 basis (standard for life table analysis to facilitate comparisons between datasets)

d_x: This is the number of deaths which occur in time x, after scaling to a per 1000 basis.

S_x: This is the proportion of individuals who survive to the next time interval, n_{x+1}/n_x

q_x: The is the probability of death in age class x, d_x/n_x

log (l_x): This is the $\log_{10}$ of the l_x values.

Ex. Human Life Table						
Time (yr)	**Obs # deaths in the time interval**	**Obs # Alive**	**Survivorship**	**# Surviving**	**Surv. Rate**	**For Graph**
x	**(counts)**	**n_x**	**l_x**	**L_x**	**S_x**	**log(L_x)**
0-9	1	115	1.00	1000	0.99	3.00
10-19	0	114	0.99	991	1.00	3.00
20-29	10	114	0.99	991	0.91	3.00
30-39	5	104	0.90	904	0.95	2.96
40-49	2	99	0.86	860	0.98	2.93
50-59	14	97	0.84	843	0.86	2.93
60-69	35	83	0.72	721	0.58	2.86
70-79	25	48	0.42	417	0.48	2.62
80-89	20	23	0.20	200	0.13	2.30
90-99	2	3	0.03	26	0.33	1.41
100-109	1	1	0.01	8	0.00	0.90

Group 1: _____

Human Life Table 1						
Time (yr)	**Obs # deaths in the time interval**	**Obs # Alive**	**Survivorship**	**# Surviving**	**Surv. Rate**	**For Graph**
x	**(counts)**	**n_x**	**l_x**	**L_x**	**S_x**	**log(L_x)**
0-9						
10-19						
20-29						
30-39						
40-49						
50-59						
60-69						
70-79						
80-89						
90-99						
100-109						

Group 2: _____

Human Life Table 2						
Time (yr)	**Obs # deaths in the time interval**	**Obs # Alive**	**Survivorship**	**# Surviving**	**Surv. Rate**	**For Graph**
x	**(counts)**	**n_x**	**l_x**	**L_x**	**S_x**	**log(L_x)**
0-9						
10-19						
20-29						
30-39						
40-49						
50-59						
60-69						
70-79						
80-89						
90-99						
100-109						

Results from the statistical test you ran:

Chi-squared works best on larger count datasets, hence why the example code gave an error message when performing calculations with our dataset. A very similar test, called a Log likelihood ratio (G-test) test of independence, can sometimes be more appropriate, particularly if you plan to investigate more complex comparisons. Run both tests, and comment on the results. Are both significant (or not)? Without the R-generated error alerting you to the possibility of a better test choice, what errors in your experimental conclusions might you likely make in using Excel?