STAT:2010/4200, Statistical Methods and Computing
Spring 2017, Instructor: Cowles
Midterm 3

`Name: _________________________________________ Course no. (30 or 105) _____`

1. An economist believes that women are more successful business people than men. He collects to data to assess whether the proportion of women-owned businesses that fail is smaller than the proportion of men-owned businesses that fail. He observes a random sample of 148 small businesses for a three-year period. During that time, 106 of the businesses headed by men and 7 of the 42 businesses headed by women fail. The attached SAS output can be used to answer some of the following questions. In the dataset used in SAS, the codings are:

```
gender:  M = men
         F = women
failed:  F = failed
         S = succeeded (did not fail)
```

   (a) Should the economist carry out a one-sided or two-sided hypothesis test? Briefly state why.

   (b) Write the null and alternative hypotheses that the economist wishes to test. Use conventional statistical symbols.

   (c) Calculate the expected count for the first cell of the contingency table (layout as in the SAS output). Numeric answer; show your work.

   (d) Should the economist reject the null hypothesis? Why or why not?

   (e) What does your answer mean in terms of business failures for women versus men?

(f) Why was the Chi-square test appropriate for this problem instead of the two-sample t test?

2. You are interested in estimating the proportion of failures in the population of women-owned businesses. You will use the economist's data on women-owned businesses from the previous problem (7 failures in 42 businesses).

   (a) Are the rules of thumb met for the use of the large-sample normal approximation? List the rules of thumb and assess each one for these data.

   (b) Use the plus-4 method to calculate a 95% confidence interval for the population proportion of failures in women-owned businesses. (Numeric answer; show your work.)

   (c) What was the margin of error in your confidence interval? (Numeric answer).

3. This problem is based on data on per capita gross domestic product and life expectancy from the Gap Minder project (gapminder.org). Data from 143 countries in 2007 are used.

   (a) I was interested in using per capita gross domestic product (`gdpPercap`) to predict life expectancyi (`lifeExp`). Refer to the SAS output to explain why I chose to use the log of per capita gross domestic product (`loggdppc`) instead of the untransformed version.

(b) I wish to test the null hypothesis that there is no linear relationship between `loggdppc` and `lifeExp` in the population of all countries. Write the null and alternative hypotheses using the statistical symbol we used in class.

(c) Cite two different parts of the SAS output to explain why I should or should not reject H_0 at significance level $\alpha = 0.05$.

(d) Country A has .2 log units higher `loggdppc` than Country B. What is the expected difference between `lifeExp` in Country A and CountryB? (Numeric answer)

(e) Give a 95% prediction interval for life expectancy in an individual new country with `loggdppc` equal to 10.45.

(f) Would the 95% confidence interval for the mean life expectancy in all countries with `loggdppc` equal to 10.45 be wider or narrower than the 95% prediction interval that you gave above? Briefly justify your answer.

SAS output for problem 1.

```
                   The FREQ Procedure

               Table of gender by failed

          gender     failed

          Frequency|
          Percent  |
          Row Pct  |
          Col Pct  |F       |S       |  Total
          ---------+--------+--------+
          F        |      7 |     35 |     42
                   |   4.73 |  23.65 |  28.38
                   |  16.67 |  83.33 |
                   |  31.82 |  27.78 |
          ---------+--------+--------+
          M        |     15 |     91 |    106
                   |  10.14 |  61.49 |  71.62
                   |  14.15 |  85.85 |
                   |  68.18 |  72.22 |
          ---------+--------+--------+
          Total          22      126      148
                      14.86    85.14   100.00


        Statistics for Table of gender by failed
```

Statistic	DF	Value	Prob
Chi-Square	1	0.1504	0.6981
Likelihood Ratio Chi-Square	1	0.1476	0.7008
Continuity Adj. Chi-Square	1	0.0173	0.8953
Mantel-Haenszel Chi-Square	1	0.1494	0.6991

SAS output for problem 3.

```
  Plot of lifeExp*gdpPercap.  Symbol used is '.'.


lifeExp |
     90 +
        |
        |
        |
        |                     .
        |                   .  . . .
     80 +                ........      ..
        |      .  .   . . . .   .  . .  .
        |    .. ..      .
        |  . ... . . .  .   .
        |  .. ....   . .
        |  .......
     70 +  .. .      .
        |
        | ..
        | ..
        | .
        | .
     60 +.. .
        |...
        |..       .
        | ..
        |.. .
        | .      .
     50 +..    . .
        |..
        |..
        |.
        |.. .
        | .
     40 +   .
        |
        -+------------+------------+------------+
         0          20000        40000        60000

                        gdpPercap
```

```
  Plot of lifeExp*loggdppc.  Symbol used is '.'.

lifeExp |
     90 +
        |
        |
        |
        |                              .
        |                             ...
     80 +                             ....
        |                        . ......
        |                      .. .  .
        |                  .    .......
        |                   . . ... ..
        |                   .......
     70 +                    .. .   .
        |
        |                .  .
        |              . . . .
        |              . ...
        |              ..
     60 +             ... .   .
        |            . .   .
        |            . . .         .
        |              .  .  .
        |            .  ..    .
        |              ..         .
     50 +          .     ..      ..
        |             .. .
        |        .   ..   .
        |          .  .
        |           . .. .    .
        |             . .
     40 +                     .
        |
        --+--------+--------+--------+--------+--
          4        6        8        10      12

                         loggdppc
```

The REG Procedure
Model: MODEL1
Dependent Variable: lifeExp

Number of Observations Read	142
Number of Observations Used	142

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	13450	13450	265.15	<.0001
Error	140	7101.71245	50.72652		
Corrected Total	141	20552			

Root MSE	7.12226	R-Square	0.6544
Dependent Mean	67.00742	Adj R-Sq	0.6520
Coeff Var	10.62905		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.94961	3.85768	1.28	0.2016
loggdppc	1	7.20280	0.44234	16.28	<.0001

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	-2.67724	12.57646
loggdppc	1	6.32827	8.07733

The REG Procedure
Model: MODEL1
Dependent Variable: lifeExp

Output Statistics

Obs	loggdppc	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Predict		Residual
1	6.88	43.8	54.5193	0.9723	40.3077	68.7310	-10.6913
2	8.69	76.4	67.5345	0.5986	53.4038	81.6652	8.8885
3	8.74	72.3	67.8738	0.6000	53.7428	82.0047	4.4272
4	8.48	42.7	65.9991	0.6009	51.8680	80.1302	-23.2681
5	9.46	75.3	73.0563	0.7037	58.9067	87.2060	2.2637
6	10.45	81.2	80.1961	1.0066	65.9751	94.4171	1.0389
7	10.49	79.8	80.5414	1.0237	66.3156	94.7672	-0.7124
8	10.30	75.6	79.1538	0.9559	64.9465	93.3611	-3.5188
9	7.24	64.1	57.0832	0.8536	42.9013	71.2651	6.9788
10	10.43	79.4	80.0391	0.9989	65.8202	94.2579	-0.5981
11	7.27	56.7	57.3377	0.8425	43.1584	71.5169	-0.6097
12	8.25	65.6	64.3624	0.6194	50.2282	78.4966	1.1916
13	8.92	74.9	69.1660	0.6122	55.0330	83.2990	5.6860
14	9.44	50.7	72.9373	0.6999	58.7884	87.0862	-22.2093
15	9.11	72.4	70.5834	0.6368	56.4462	84.7207	1.8066
16	9.28	73.0	71.7643	0.6653	57.6219	85.9066	1.2407
17	7.10	52.3	56.1195	0.8968	41.9273	70.3118	-3.8245
18	6.06	49.6	48.6270	1.2773	34.3213	62.9328	0.9530
19	7.45	59.7	58.5850	0.7904	44.4174	72.7525	1.1380
20	7.62	50.4	59.8474	0.7420	45.6901	74.0047	-9.4174
21	10.50	80.7	80.5798	1.0257	66.3534	94.8061	0.0732
22	6.56	44.7	52.1974	1.0883	37.9529	66.4419	-7.4564
23	7.44	50.7	58.5440	0.7921	44.3761	72.7119	-7.8930
24	9.49	78.6	73.2741	0.7109	59.1231	87.4251	5.2789
25	8.51	73.0	66.2381	0.5996	52.1072	80.3690	6.7229
26	8.85	72.9	68.7276	0.6070	54.5955	82.8597	4.1614
27	6.89	65.2	54.6043	0.9682	40.3937	68.8149	10.5477
28	5.63	46.5	45.4726	1.4513	31.1022	59.8431	0.9894
29	8.20	55.3	63.9960	0.6256	49.8607	78.1313	-8.6740
30	9.17	78.8	71.0296	0.6467	56.8906	85.1686	7.7524