

Biostatistics 140.622
Second Term, 2008-2009
Final Examination Answer Key

Instructions: You will have two hours for this examination. There are 20 problems. Please write on these pages. The formula pages are at the back for your use.

Questions 1 through 5 are general knowledge questions.

1. When using Stata to perform a one-way ANOVA with p groups, followed by pairwise comparisons of means, the p -value associated with each pairwise test from the Stata output is equal to: (*Circle only one response*).

- a) p -value divided by the number of possible pairwise comparisons
- b) p -value times the number of possible pairwise comparisons**

This question asks about the p -values listed in the Stata output after performing a one-way ANOVA. Stata adjusts the p -value by multiplying by the number of possible comparisons. Thus, the p -values provided by the Stata output are already adjusted and can always be compared to the overall significance level of interest (i.e., often 0.05).

When performing a Bonferroni correction by hand, one computes the p -value associated with a pairwise test and compares it to the adjusted significance level given by the desired overall significance level divided by the number of possible pairwise comparisons.

- c) p -value divided by the number of groups
- d) p -value times the number of groups
- e) the significance level

2. Suppose that Y is the dependent (response) variable of interest and X_1 and X_2 are independent variables (covariates). An ANCOVA (Analysis of Covariance) is performed to assess whether: (*Circle only one response*).

- a) The variable X_1 significantly confounds the relationship between Y and X_2 .
- b) The variable X_2 significantly confounds the relationship between Y and X_1 .
- c) The relationship between Y and X_1 differs significantly by the level of X_2 .**

In our course, we have presented ANCOVA models including the interaction term of $X_1 * X_2$. This allows one to assess whether the association between Y and one of the independent variables varies by the level of the second independent variable. A single ANCOVA could not be used to assess confounding: comparison with additional models would be required.

- d) The relationship between X_1 and X_2 differs significantly by the level of Y .
- e) The variable Y significantly confounds the relationship between X_1 and X_2

3. In a one-way ANOVA, Bartlett's test may be used to test the null hypothesis that: (*Circle only one response*).

- a) The population standard deviation equals zero.
- b) The population standard deviations for each group are the same.**

Bartlett's F-test is used to test the assumption of equal variance across all groups.

- c) The population correlation coefficient equals zero.
- d) None of the p covariates contribute to the explanation of variation in the outcome.
- e) At least one of the p covariates contributes to the explanation of variation in the outcome.

4. Suppose that the outcome of interest is disease incidence. A generalized linear model was used with family (binomial) and link (log) to assess the relationship between the outcome and age centered at the mean. The model is $\log(\text{rate}) = \beta_0 + \beta_1 X_1$. The interpretation of β_1 is: (*Circle only one response*).

- a) The expected Y when age is equal to zero.
- b) The change in expected Y between individuals who differ by one year of age.
- c) The difference in the log rate of disease between individuals who differ by one year of age.**

This interpretation is based on X_1 (age) being a continuous variable. As a proof:

Let $X_1 = X_1 + 1$. Then, $\log(\text{rate} | X_1 + 1) = \beta_0 + \beta_1 (X_1 + 1)$

Also, let $X_1 = X_1$ (one year younger). Then, $\log(\text{rate} | X_1) = \beta_0 + \beta_1 X_1$

Subtracting, we see that:

$\log(\text{rate} | X_1 + 1) - \log(\text{rate} | X_1) = \beta_0 + \beta_1 (X_1 + 1) - (\beta_0 + \beta_1 X_1) = \beta_1$

indicating that β_1 is the difference in log rate of disease associated with each one unit increase in age.

- d) The change in expected Y for individuals of average age.
- e) The difference in the log odds of disease between individuals who differ by one year of age.

5. Consider the following sentence: “A two-sample test of difference in disease incidence provides power=0.80 to detect a difference of 20 cases per 1000 person years.” Which of the following is the most correct interpretation of the sentence? (*Circle only one response*).
- a) If there is truly no difference, the test has an 8 in 10 chance of concluding that there is a difference.
 - b) If there is truly no difference, the test has a 2 in 10 chance of concluding that there is a difference.
 - c) **If there is truly a difference of 20 cases per 1000 person years, the test has an 8 in 10 chance of concluding that there is a difference.**

Power is the probability of detecting a difference when it truly exists. It is the probability of rejecting the null hypothesis of no difference in favor of a specified alternative hypothesis.

- d) If there is truly a difference of 20 cases per 1000 person years, the test has a 2 in 10 chance of concluding that there is a difference.
- e) The test has an 8 in 10 chance of concluding that there is a difference irrespective of the size of the true difference.

Questions 6 through 14 pertain to the results of a study assessing the continuity of health care for individuals who had been previously incarcerated (jail inmates). Inmates who were positive for HIV received discharge planning regarding health care whereas other inmates did not. (*AJPH* 2008;98(12):2182). There were **3 disease groups: 1= HIV Disease, 2 = Chronic Disease Other Than HIV, and 3= No Chronic Disease**

The **outcome of interest** was access to health care as measured by the self-reporting of a **regular source of care (Y=1)** or none (**Y=0**).

The data are given below in **Table 1**:

care	disease			Total
	1	2	3	
0	29 16.02	56 54.90	40 62.50	125 36.02
1	152 83.98	46 45.10	24 37.50	222 63.98
Total	181 100.00	102 100.00	64 100.00	347 100.00

Pearson chi2 (2) = 66.6597 Pr = 0.000

6. The chi-squared statistic suggests that: (*Circle only one response*).
- There is a statistically significant difference in the proportion of individuals reporting a regular source of care between all three disease status groups ($\chi^2 = 66.7$, $p=0.000$).
 - The three disease status groups are similar with respect to reporting a regular source of care ($\chi^2 = 66.7$, $p=1.00$).
 - There is a statistically significant difference in the proportion of individuals reporting a regular source of care in at least one of the disease status groups. ($\chi^2 = 66.7$, $p=0.000$).**
For this 2x3 table, the null hypothesis is that there is no association between regular source of care and disease group. In other words, the null hypothesis states that the proportion with a regular source of care is the same across disease groups. The large chi-squared statistic with associated small p-value leads us to reject this null hypothesis and conclude that at least one of the groups is different with respect to regular source of care.
 - There is no association between disease status and regular source of care. ($\chi^2 = 66.7$, $p=1.00$).
 - Regular source of care is independent of disease status. ($\chi^2 = 66.7$, $p=0.000$).

At the back of the exam, Table 1 is repeated; Models A through D also pertain to the data from this study.

7. Define the model $\Pr(Y=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ where $X_1=1$ for chronic disease other than HIV; 0 otherwise, and $X_2=1$ if no chronic disease; 0 otherwise. The sum of the coefficients, $\beta_0 + \beta_1$, can be interpreted as the: (*Circle only one response*).
- Proportion of individuals with a chronic disease other than HIV who have a regular source of care.**

This model corresponds to a generalized linear model in which the outcome is binary and the identity link is used. For this model,

$X_1=1$ and $X_2=0$ refers to individuals with chronic disease other than HIV

$X_1=0$ and $X_2=1$ refers to individuals with no chronic disease

$X_1=0$ and $X_2=0$ is the reference group of individuals with HIV

Thus, when $X_1=1$ and $X_2=0$, the model reduces to

$\Pr(Y=1 | X_1=1, X_2=0) = \beta_0 + \beta_1(1) + \beta_2(0) = \beta_0 + \beta_1$ which is the proportion with a regular source of care of individuals with chronic disease other than HIV.

- Difference in proportion with a regular source of care between individuals with HIV and individuals with no chronic disease.
- Difference in proportion with a regular source of care between individuals with chronic diseases other than HIV and individuals with no chronic disease.
- Proportion of individuals without chronic disease who have a regular source of care.
- Proportion of individuals with HIV who have a regular source of care.

8. Again consider the model defined in **question 7**. Based on the data in **Table 1**, the value of the estimated coefficient, b_0 , is approximately: (*Circle only one response*).

- a) 152/29
- b) **152/181**

When $X_1=0$ and $X_2=0$, the model reduces to

$\Pr(Y=1 | X_1=0, X_2=0) = \beta_0 + \beta_1(0) + \beta_2(0) = \beta_0$ which is the proportion with a regular source of care of individuals with HIV.

- c) $\frac{152/29}{24/40}$
- d) $\log_e \left(\frac{152}{29} \right)$
- e) $\log_e \left(\frac{46}{56} \right)$

9. Consider **Model C**. The exponentiated coefficient, e^{β_0} , can be interpreted as: (*Circle only one response*).

- a) The odds of having a regular source of care for female inmates with no chronic disease.
- b) The odds of having a regular source of care for male inmates with HIV disease.
- c) **The odds of having a regular source of care for female inmates with HIV disease.**

In Model C, $\log(\text{odds } Y=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$.

Thus, β_0 is the log odds of a regular source of care for female inmates with HIV disease when $X_1 = 0$ and $X_2 = 0$ (HIV disease) and $X_3 = 0$ (females).

- d) The log odds of having a regular source of care for female inmates with no chronic disease.
- e) The log odds ratio of having a regular source of care for female versus male inmates with no chronic disease.

10. In **Model D**, the difference in the coefficients, $\beta_2 - \beta_1$, can be interpreted as: (*Circle only one response*).

- a) **The difference in the log odds of regular source of care between female inmates with no chronic disease and female inmates with a chronic disease other than HIV.**

In Model D $\log(\text{odds } Y=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_2 X_3$

We know that:

When $X_3 = 0$ (females), the model reduces to

$\log(\text{odds } Y=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

So, among female inmates ,

$\log(\text{odds } Y=1 | X_1=0, X_2=1) = \beta_0 + \beta_2$ for those with no chronic disease

$\log(\text{odds } Y=1 | X_1=1, X_2=0) = \beta_0 + \beta_1$ for those with chronic disease other than HIV.

Subtracting, we see that $\beta_2 - \beta_1$ can be interpreted as the difference in log odds of a regular source of care between females inmates with no chronic disease and those with chronic disease other than HIV (the log OR).

- b) The difference in the log odds of regular source of care between males and females, controlling for disease status.
 c) The odds ratio of regular source of care for inmates with chronic disease versus without chronic disease, controlling for gender.
 d) The odds ratio of being male versus females, controlling for regular source of care.
 e) The odds of regular source of care among female inmates with no chronic disease.

11. A comparison of **Models A through D** suggests that: (*Circle only one response*).

- a) Gender confounds the relationship between disease status and regular source of care.
 b) Gender modifies the relationship between disease status and regular source of care.
 c) Regular source of care modifies the effect of gender on disease status.
 d) **Gender does not confound the relationship between disease status and regular source of care.**

We can see from Model C that the addition of gender to the model does not change the estimated coefficient for the disease status variables (based on a comparison to Model A). Model D indicates that gender does not modify the relationship between disease status and regular source of care. In this example, regular source of care is the dependent variable

- e) Regular source of care does not modify the relationship between disease status and gender.

12. Suppose that one is interested in constructing a 95% confidence interval for the percentage of inmates with HIV disease who have a regular source of care, in a future study. Using the percentage observed in Table 1, what approximate sample size would be needed in a future study in order to obtain a 95% confidence interval of 10 percentage points in width? (*Circle only one response*).

- a) 14 inmates
- b) 52 inmates
- c) 104 inmates
- d) **206 inmates**

We are given width= 0.10 such that d (half-width) =0.05.

Then, we set $0.05 = 1.96 \sqrt{(pq/n)}$.

Assuming $p=0.84$ and $q=0.16$, we solve for n:

$$n = \frac{1.96^2(0.84)(0.16)}{(0.05)^2} = 206$$

- e) 412 inmates

13. For a future study, one is interested in the sample size needed in order to detect a **20% difference** in self-report of regular source of care between the HIV disease group and the no chronic disease group. Assume the proportion with a regular source of care in the HIV group is that observed in Table 1. Also assume equal sample sizes, a 2-sided significance level of 0.05 and power of 90%. The **approximate** sample size needed in each group is: (*Circle only one response*).

- a) 20
- b) 50
- c) **100**

From Table 1, $p_1=0.86$ and $p_2=0.64$. Assuming equal sample sizes in each group,

$\bar{p} = (p_1 + p_2) / 2 = (0.84 + 0.64) / 2 = 0.74$. The, we can solve for $n_1=n_2=$

$$n = \frac{\left[z_{\alpha/2} \sqrt{2\bar{p}q} + z_{\beta} \sqrt{p_1q_1 + p_2q_2} \right]^2}{\Delta^2}$$

$$= \frac{\left[1.96 \sqrt{2(0.74)(0.26)} + 1.28 \sqrt{(0.84)(0.16) + (0.64)(0.36)} \right]^2}{(0.20)^2} = 100$$

- d) 200
- e) 400

In a new study, 50 pairs of inmates were sampled with one member of each pair having a regular source of care, and the other member not having a regular source of care. In each pair, inmates were matched to have the same duration of incarceration (being jailed) and observed to have chronic disease (HIV or otherwise) or to not have chronic disease. In 10 of the pairs, neither member had chronic disease; in 15 pairs, both members had chronic disease; in 20 pairs, the person with a regular source of care had chronic disease and the one without a regular source of care did not; and in 5 pairs, the person without a regular source of care had chronic disease and the one with a regular source of care did not.

14. Which of the following is an appropriate estimate of the odds ratio for having a regular source of care, comparing those with chronic disease to those without? (*Circle only one response*).
- a) 0.25
 - b) 0.67
 - c) 1.5
 - d) 1.75
 - e) 4.0

Since this is a matched case-control study in which a case had a regular source of care and the control did not, we see that:

	Controls with a chronic disease	Controls without a chronic disease	Total pairs
Cases with a chronic disease	15	20	35
Cases without a chronic disease	5	10	15
Total pairs	20	30	50

Then, the odds ratio of a regular source of care for those with versus without chronic disease = $b/c = 20/5 = 4$.

Similarly we can interpret this as the odds ratio of chronic disease for those with versus without a regular source of care.

Questions 15 through 20 concern the relationship between daily mortality data (Y =number of daily deaths), X_1 = average daily total suspended particle level measured as $\mu\text{g}/\text{m}^3$ (**tsp**) and X_2 = daily temperature (**temp**) in Fahrenheit degrees in Philadelphia over the years 1974 through 1988. Both X_1 and X_2 are centered at their medians. Models E through H display different linear models for these relationships.

15. Based on **Model G**, the difference in average daily mortality between two days that differ by $20 \mu\text{g}/\text{m}^3$, adjusting for temperature, is: (*Circle only one response*).

- a) 0.06 deaths, 95% CI (0.05 to 0.07 deaths)
- b) -0.17 deaths, 95% CI (-0.18 to -0.16 deaths)
- c) **1.19 deaths, 95% CI (1.03 to 1.35 deaths)**

After adjusting for temperature, the change in average daily mortality associated with a 20 unit increase in tsp is estimated as 20(0.059) deaths with associated 95% CI (20(0.0516), 20(0.0676)).

- d) -3.38 deaths (-3.62 to -3.14 deaths)
- e) 20.06 deaths, 95% CI (20.05 to 20.07 deaths)

16. If we fix the values of temperature and tsp, then we would estimate that approximately 95% of the population daily deaths would fall within what range of the true mean mortality on days of that temperature and tsp? Use the results of **Model G**. (*Circle only one response*).

- a) ± 0.004 deaths
- b) $\pm \frac{0.004}{\sqrt{5390}}$ deaths
- c) ± 7.8 deaths
- d) **± 15.7 deaths which equals $\pm 1.96 \cdot (\text{root MSE}) = \pm 1.96 \cdot (7.85)$**
- e) ± 61.6 deaths

17. In **Model H**, the **coefficient for the interaction term** can be interpreted as: (*Circle only one response*).

- a) The change in average mortality associated with each 1 degree increase in temperature.
- b) The change in average mortality associated with each 1 $\mu\text{g}/\text{m}^3$ increase in total suspended particulate level.
- c) **The difference, between days differing by one degree in temperature, in the change in average mortality associated with each 1 $\mu\text{g}/\text{m}^3$ increase in total suspended particulate level.**

In Model H, $E[\text{deaths}] = \beta_0 + \beta_1 \text{tspc} + \beta_2 \text{tempc} + \beta_3 \text{tspc} * \text{tempc}$

Alternatively, we could interpret the interaction term as the difference, between days differing by one unit in total suspended particulates, in the change in average mortality associated with each degree increase in temperature.

- d) The difference, between days differing by one death, in the change in average temperature associated with each 1 $\mu\text{g}/\text{m}^3$ increase in total suspended particulate level.
- e) The difference, between days differing by 1 $\mu\text{g}/\text{m}^3$ in total suspended particulate level, in the change in average temperature associated with each increase of one death.

18. In **Model H**, the **coefficient for the particulate variable, tspc**, can be interpreted as: (*Circle only one response*).

- a) The change in average mortality associated with each 1 $\mu\text{g}/\text{m}^3$ increase in total suspended particulate level, holding temperature constant.
- b) **The change in average mortality associated with each 1 $\mu\text{g}/\text{m}^3$ increase in total suspended particulate level, on days with temperature of 55.25 degrees.**

In Model H, $E[\text{deaths}] = \beta_0 + \beta_1 \text{tspc} + \beta_2 \text{tempc} + \beta_3 \text{tspc} * \text{tempc}$

The interpretation of β_1 can be easily obtained by setting tempc=0 (which means temp=55.25). Then, $E[\text{deaths}|\text{tempc}=0] = \beta_0 + \beta_1 \text{tspc}$.

Thus, β_1 is interpreted as the change in average mortality with each unit increase in tsp on days with temperature of 55.25 degrees.

- c) The average mortality on days with particulate of 0 $\mu\text{g}/\text{m}^3$ and temperature of 0 degrees.
- d) The average mortality on days with particulate of 63 $\mu\text{g}/\text{m}^3$ and temperature of 55.25 degrees.
- e) The difference, between days differing by one degree in temperature, in the change in average mortality associated with each 1 $\mu\text{g}/\text{m}^3$ increase in total suspended particulate level.

19. A colleague evaluating **Models E through H** states that the association between particulate concentration and mortality appears to be **confounded** by temperature. Which of the following is the most reasonable justification for the statement? (*Circle only one response*).

a) The tspc coefficient in Model G is 67% larger than in Model E.

There is evidence of confounding of the tsp-mortality relationship by temperature. This is based on the change in the tspc coefficient in the single variable model (Model E) versus the adjusted model (Model G). Since, temperature is associated with both tsp and mortality but not in the causal pathway between tsp and mortality, it appears to be a confounder.

- b) The tempc coefficient in Model G is 13% larger (in absolute value) than in Model F.
- c) In Model G, the tempc coefficient is statistically significantly different from 0.
- d) In Model G, the tspc coefficient is statistically significantly different from 0.
- e) In Model H, the interaction coefficient is statistically significantly different from 0.

20. An epidemiologist suggests that mortality on a given day strongly predicts mortality in the next few days independently of covariates. If this is true, which, if any, linear regression assumption would be violated as a result? (*Circle only one response*).

- a) Specification of the model for expected mortality (“L” in “LINE”)
- b) Statistical independence between outcomes (“I” in “LINE”)**

If mortality on one day can predict mortality on subsequent days, then these daily outcomes are correlated (associated or dependent) and not independent of each other.

- c) Normal distribution of residuals (“N” in “LINE”)
- d) Equal variances of residuals across all values of temperature and particulate (“E” in LINE)
- e) No assumption would be called into concern by the epidemiologist’s suggestion

Table 1

where **regular source of care (Y=1)** or none (**Y=0**):

disease 1= HIV disease; 2= chronic disease other than HIV; 3= no chronic disease

care	disease			Total
	1	2	3	
Y=0	29 16.02	56 54.90	40 62.50	125 36.02

Y=1	152	46	24	222
	83.98	45.10	37.50	63.98
Total	181	102	64	347
	100.00	100.00	100.00	100.00

Model A $\log(\text{odds } Y=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

where $X_1=1$ if disease status group is chronic disease other than HIV; 0 otherwise, and
 $X_2=1$ if disease status group is no chronic disease; 0 otherwise.

```
. xi: logit care i.disease
i.disease      _Idisease_1-3      (naturally coded; _Idisease_1 omitted)
```

```
Logistic regression      Number of obs =      347
LR chi2(2)              =      69.17
Prob > chi2             =      0.0000
Pseudo R2               =      0.1525
Log likelihood = -192.19662
```

care	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_Idisease_2	-1.853295	.2840002	-6.53	0.000	-2.409925 -1.296665
_Idisease_3	-2.16741	.3282172	-6.60	0.000	-2.810704 -1.524116
_cons	1.656585	.2026323	8.18	0.000	1.259433 2.053737

Model B $\log(\text{odds } Y=1) = \beta_0 + \beta_1 X_1$

where $X_1=1$ if male, 0 if female.

```
. xi: logit care gender
```

```
Logistic regression      Number of obs =      347
LR chi2(1)              =      0.06
Prob > chi2             =      0.8094
Pseudo R2               =      0.0001
Log likelihood = -226.753
```

care	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
gender	.0644189	.2666232	0.24	0.809	-.458153 .5869908
_cons	.5245245	.2342881	2.24	0.025	.0653282 .9837207

Model C $\log(\text{odds } Y=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

where $X_1=1$ if disease status group is chronic disease other than HIV; 0 otherwise, and
 $X_2=1$ if disease status group is no chronic disease; 0 otherwise.
 $X_3=1$ if male, 0 if female.

```
. xi: logit care i.disease gender
i.disease      _Idisease_1-3      (naturally coded; _Idisease_1 omitted)
```

```
Logistic regression      Number of obs =      347
LR chi2(3)              =      69.30
Prob > chi2             =      0.0000
Pseudo R2               =      0.1528
Log likelihood = -192.13208
```

care	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_Idisease_2	-1.852106	.2840587	-6.52	0.000	-2.40885 -1.295361

_Idisease_3	-2.172492	.3286992	-6.61	0.000	-2.816731	-1.528254
gender	.106925	.2972389	0.36	0.719	-.4756525	.6895025
_cons	1.574565	.3036286	5.19	0.000	.9794633	2.169666

Model D $\log(\text{odds } Y=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_2 X_3$

where $X_1=1$ if disease status group is chronic disease other than HIV; 0 otherwise, and
 $X_2=1$ if disease status group is no chronic disease; 0 otherwise.
 $X_3=1$ if male, 0 if female.

```
. xi:logit care i.disease*gender
i.disease      _Idisease_1-3      (naturally coded; _Idisease_1 omitted)
i.dise-e*gender  _IdisXgende_#    (coded as above)
```

Logistic regression Number of obs = 347
LR chi2(5) = 70.94
Prob > chi2 = 0.0000
Pseudo R2 = 0.1564
Log likelihood = -191.31011

care	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Idisease_2	-1.822531	.5674237	-3.21	0.001	-2.934661	-.710401
_Idisease_3	-1.417066	.6990253	-2.03	0.043	-2.78713	-.0470016
gender	.317535	.4597015	0.69	0.490	-.5834633	1.218533
_IdisXgend~2	-.0421231	.6558659	-0.06	0.949	-1.327597	1.243351
_IdisXgend~3	-.9535238	.7934898	-1.20	0.229	-2.508735	.6016877
_cons	1.417066	.3940851	3.60	0.000	.6446735	2.189459

Model E $E[\text{deaths}] = \beta_0 + \beta_1 \text{tspc}$

```
. gen tspc=tsp-63
. regress deaths tspc
```

Source	SS	df	MS	Number of obs = 5398		
Model	5002.21841	1	5002.21841	F(1, 5396)	= 71.33	
Residual	378410.612	5396	70.127986	Prob > F	= 0.0000	
				R-squared	= 0.0130	
				Adj R-squared	= 0.0129	
Total	383412.831	5397	71.0418437	Root MSE	= 8.3742	

deaths	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tspc	.0358235	.0042416	8.45	0.000	.0275082	.0441387
_cons	46.52931	.1154415	403.06	0.000	46.303	46.75562

Model F $E[\text{deaths}] = \beta_0 + \beta_1 \text{tempc}$

```
. gen tempc=temp-55.25
. regress deaths tempc
```

Source	SS	df	MS	Number of obs = 5471		
Model	38155.2896	1	38155.2896	F(1, 5469)	= 598.61	
Residual	348593.584	5469	63.739913	Prob > F	= 0.0000	
				R-squared	= 0.0987	
				Adj R-squared	= 0.0985	

Total | 386748.874 5470 70.7036332 Root MSE = 7.9837

deaths	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tempc	-.1487459	.0060796	-24.47	0.000	-.1606643	-.1368275
_cons	46.53834	.1080812	430.59	0.000	46.32646	46.75022

Model G $E[\text{deaths}] = \beta_0 + \beta_1 \text{tspc} + \beta_2 \text{tempc}$

. regress deaths tspc tempc

Source	SS	df	MS	Number of obs = 5390	
Model	51309.389	2	25654.6945	F(2, 5387) =	416.37
Residual	331920.953	5387	61.615176	Prob > F =	0.0000
Total	383230.342	5389	71.1134426	R-squared =	0.1339
				Adj R-squared =	0.1336
				Root MSE =	7.8495

deaths	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tspc	.0596032	.0040713	14.64	0.000	.0516218	.0675846
tempc	-.1688994	.0061597	-27.42	0.000	-.1809749	-.156824
_cons	46.26665	.1087268	425.53	0.000	46.0535	46.4798

Model H $E[\text{deaths}] = \beta_0 + \beta_1 \text{tspc} + \beta_2 \text{tempc} + \beta_3 \text{tspc} * \text{tempc}$

. gen x3=tspc*tempc
(89 missing values generated)

. regress deaths tspc tempc x3

Source	SS	df	MS	Number of obs = 5390	
Model	52175.66	3	17391.8867	F(3, 5386) =	282.95
Residual	331054.682	5386	61.4657783	Prob > F =	0.0000
Total	383230.342	5389	71.1134426	R-squared =	0.1361
				Adj R-squared =	0.1357
				Root MSE =	7.84

deaths	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
tspc	.0605831	.0040747	14.87	0.000	.052595	.0685712
tempc	-.1756506	.0064097	-27.40	0.000	-.1882161	-.1630851
x3	.00086	.0002291	3.75	0.000	.0004109	.001309
_cons	46.17185	.1114923	414.13	0.000	45.95328	46.39042