Public Affairs 818
Professor Wallace
December 15th, 2007

Final Examination

Instructions: You have 2 hours to complete the examination and there are 4 questions worth a total of 170 possible points. You may use a calculator, but you may not use pre-programmed formulas. The standard normal "Z" table from the second page of your book has been provided to you. Partial credit is possible on all questions so be sure to write clearly and show all of your work. If you have questions regarding the wording of a particular question feel free to ask for clarifications (we are not trying to trick you). Good luck!

Question 1. (40 Points) Go Pack Go.
Let's pretend that the Packers are in the Super Bowl. The game is tied and the Packers have the ball on their opponent's 10 yard line as time is running out. Brett Favre gets injured and his back-up, Aaron Rogers, is brought in for the last play of the game. Since Aaron Rogers has not gotten much playing time as a professional, the fans do not know whether he is a good quarter back or not. Let's say that the probability he is a good quarter back is 0.6. Let G denote the event that he is a good quarter back and let TD denote the event that he throws a game winning touchdown on the final play of the game. If he is a good quarter back, the probability that he throws a game winning touchdown is 0.9. If he is not a good quarter back, the probability that he throws a game winning touchdown is 0.3.

A) **(8 Points)** Say that Aaron Rogers throws a game winning touchdown on the final play of the game. What probability should we attach to him being a good quarter back? (Hint: Use Bayes Rule).

$$P(G|TD) = \frac{P(TD|G)P(G)}{P(TD|G)P(G) + P(TD|G^c)P(G^c)}$$

$$= \frac{0.9 \cdot 0.6}{0.9 \cdot 0.6 + 0.3 \cdot 0.4} \approx 0.82$$

For the rest of the question, we will pretend that we know certain statistics about the Green Bay Packers. The Packers are more likely to win when either of two related events occurs. The first is the event that Brett Favre completes more than 50% of his passes, denote this event by C. The second is the event that the Packers commit fewer turnovers than their opponents, denote this event by T. In any given football match, the probability that Brett Favre completes more than 50% of his passes is 0.5. If he completes more than 50% of his passes, then the probability that the Packers commit fewer turnovers than their opponents is 0.75. If he does not complete more than 50% of his passes, then the probability that the Packers commit fewer turnovers than their opponents is 0.6.

The Packers can either win or lose any given game. Let W denote the event that the Packers win. Whether they win or not depends on the two events described above. The probability of a win is:

i) 0.8 if that the Packers commit fewer turnovers than their opponents and Brett Favre completes more than 50% of his passes
ii) 0.4 if that the Packers commit fewer turnovers than their opponents and Brett Favre completes less than 50% of his passes
iii) 0.3 if that the Packers commit more turnovers than their opponents and Brett Favre completes more than 50% of his passes
iv) 0.1 if that the Packers commit more turnovers than their opponents and Brett Favre completes less than 50% of his passes.

B) **(7 Points)** What is the probability that the Packers win a particular game, Brett Farvre completes more than 50% of his passes and the Packers commit fewer turnovers than their opponents?

$$P(W \cap C \cap T) = P(W \cap (C \cap T)) = P(W \mid C \cap T)P(C \cap T)$$
$$= P(W \mid C \cap T)P(T \mid C)P(C) = 0.8 \cdot 0.75 \cdot 0.5 = 0.3$$

C) **(5 Points)** What is the probability that the Packers commit fewer turnovers than their opponents?

$$P(T) = P(T \mid C)P(C) + P(T \mid C^c)P(C^c) = 0.75 \cdot 0.5 + 0.6 \cdot 0.5 = 0.675$$

D) **(10 Points)** What is the probability that the Packers win?

$$P(W) = P(W \mid T \cap C)P(T \cap C) + P(W \mid T \cap C^c)P(T \cap C^c)$$
$$+ P(W \mid T^c \cap C)P(T^c \cap C) + P(W \mid T^c \cap C^c)P(T^c \cap C^c)$$
$$= P(W \mid T \cap C)P(T \mid C)P(C) + P(W \mid T \cap C^c)P(T \mid C^c)P(C)$$
$$+ P(W \mid T^c \cap C)P(T^c \mid C)P(C) + P(W \mid T^c \cap C^c)P(T^c \mid C^c)P(C^c)$$
$$= 0.8 \cdot 0.75 \cdot 0.5 + 0.4 \cdot 0.6 \cdot 0.5 + 0.3 \cdot 0.25 \cdot 0.5 + 0.1 \cdot 0.4 \cdot 0.5 = 0.4775$$

E) **(5 Points)** What is the probability that the Packers commit fewer turnovers than their opponents and they win?

$$\begin{aligned} P(W \cap T) &= P(W \cap T \cap C) + P(W \cap T \cap C^c) \\ &= 0.3 + P(W \mid T \cap C^c) P(T \cap C^c) \\ &= 0.3 + P(W \mid T \cap C^c) P(T \mid C^c) P(C^c) \\ &= 0.3 + 0.4 \cdot 0.6 \cdot 0.5 = 0.3 + 0.12 = 0.42 \end{aligned}$$

F) **(5 Points)** Are the events T and W independent?

No $- P(W \cap T) \neq P(W) P(T)$

Question 2. (40 Points) Fast Food and Adult Obesity.

In a paper titled "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System", Chou et al. (2004)[1] study the factors that may be responsible for the 50% increase in the number of obese adults in the US since the late 1970s. They employ the 1984–1999 Behavioral Risk Factor Surveillance System, augmented with state level measures pertaining to the per capita number of fast-food and full-service restaurants, the prices of a meal in each type of restaurant, food consumed at home, cigarettes, and alcohol, and clean indoor air laws. The variables used in their study are described below:

[1] "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System", Chou, Shin-Yi; Grossman, Michael; Saffer, Henry. *Journal of Health Economics*, vol. 23, no. 3, May 2004, pp. 565-87.

Table 2
Definitions, means, and standard deviations of variables[a]

Variable	Definition	Mean and standard deviation
Body mass index	Weight in kilograms divided by height in meters squared	26.015 (4.959)
Obese	Dichotomous variable that equals 1 if body mass index ≥ 30 kg/m^2	0.175 (0.380)
Black non-Hispanic	Dichotomous variable that equals 1 if respondent is Black but not Hispanic	0.092 (0.288)
Hispanic	Dichotomous variable that equals 1 if respondent is Hispanic	0.085 (0.279)
Other race	Dichotomous variable if respondent's race is other than White or Black	0.033 (0.179)
Male	Dichotomous variable that equals 1 if respondent is male	0.499 (0.500)
Some high school	Dichotomous variable that equals 1 if respondent completed at least 9 years but less than 12 years of formal schooling	0.092 (0.289)
High school graduate	Dichotomous variable that equals 1 if respondent completed exactly 12 years of formal schooling	0.330 (0.470)
Some college	Dichotomous variable that equals 1 if respondent completed at least 13 years but less than 16 years of formal schooling	0.262 (0.440)
College graduate	Dichotomous variable that equals 1 if respondent graduated from college	0.263 (0.440)
Married	Dichotomous variable that equals 1 if respondent is married	0.613 (0.487)
Divorced	Dichotomous variable that equals 1 if respondent is divorced or separated	0.089 (0.284)
Widowed	Dichotomous variable that equals 1 if respondent is widowed	0.066 (0.249)
Household income	Real household income in thousands of 1982–1984 dollars	29.460 (24.627)
Age	Age of respondent	43.381 (17.119)
Restaurants	Number of fast-food restaurants and full-service restaurants per 10,000 persons in respondent's state of residence[b]	13.252 (1.529)
Fast-food price	Real fast-food meal price in respondent's state of residence in 1982–1984 dollars[b]	2.903 (0.220)
Full-service restaurant price	Real full-service restaurant meal price in respondent's state of residence in 1982–1984 dollars[b]	5.971 (1.172)
Food at home price	Real food at home price in respondent's state of residence in 1982–84 dollars: weighted average of prices of 13 food items, weights are shares of each item in total food expenditures based on expenditure patterns of mid-management (middle-income) households[b]	1.258 (0.121)
Cigarette price	Real cigarette price in respondent's state of residence in 1982–1984 dollars[b]	1.287 (0.257)
Alcohol price	Real alcohol price in respondent's state of residence in 1982–1984 dollars: weighted average of prices of pure ounce of ethanol in beer, wine, and spirits; weights are shares of each item in total alcohol consumption[b]	1.065 (0.170)
Private	Dichotomous variable that equals 1 if smoking is prohibited in private workplaces in respondent's state of residence	0.343 (0.475)
Government	Dichotomous variable that equals 1 if smoking is prohibited in state and local government workplaces in respondent's state of residence	0.564 (0.496)
Restaurant	Dichotomous variable that equals 1 if smoking is prohibited in restaurants in respondent's state of residence	0.546 (0.498)
Other	Dichotomous variable that equals 1 if smoking is prohibited in other public places such as elevators, public transportation, and theaters in respondent's state of residence	0.688 (0.463)

[a] Standard deviations are in parentheses. Sample size is 1,111,074. BRFSS sample weights are used in calculating the mean and standard deviation.
[b] See text for more details.

They ran an OLS regression of body mass index (BMI) on the independent variables listed in the table above. The table below presents a selection of the coefficients with standard errors given in parentheses beside the coefficient estimates.

Table 3
Body mass index regression, persons 18 years of age and older

Independent variable	BMI	
Black	1.638*	(0.028)
Hispanic	0.737*	(0.028)
Other race	-0.406*	(0.057)
Male	0.890*	(0.016)
Some high school	-0.110*	(0.031)
High school graduate	-0.503*	(0.029)
Some college	-0.572*	(0.030)
College graduate	-1.150*	(0.032)
Married	0.187*	(0.016)
Divorced	-0.411*	(0.021)
Widowed	0.262*	(0.026)
Household income	-0.035*	(0.001)
Household income squared/100	0.020*	(0.001)
Age	0.346*	(0.002)
Age squared	-0.003*	(0.000)
Restaurants (full-service + fast-food)	0.631*	(0.067)
Restaurants squared	-0.011*	(0.002)
Fast-food restaurant price	-1.216	(0.728)
Fast-food restaurant price squared	0.135	(0.119)
Full-service restaurant price	-0.687*	(0.161)
Full-service restaurant price squared	0.050	(0.013)
Food at home price	-6.462*	(1.918)
Food at home price squared	2.244*	(0.719)
Cigarette price	0.486	(0.355)
Cigarette price squared	0.009	(0.113)
Alcohol price	1.140	(0.884)
Alcohol price squared	-0.734	(0.380)
Private	0.015	(0.039)
Government	0.115	(0.071)
Restaurant	-0.020	(0.056)
Other	0.054	(0.056)
R-squred	0.081	
F-statistic	1212.21	
Sample size	1,111,074	

Note: All regressions include state dummies. Standard errors are in parentheses. Intercept is not shown.

A) **(10 Points)** Provide an interpretation for the estimated coefficients associated with each of the 4 education variables. What can we conclude about the relationship between education and BMI from these coefficients?

Coefficient on:

- *Some high school – persons who have attend some high school, but do not have a diploma, are estimated to have a BMI that is 0.11 below otherwise observationally equivalent persons who have not attended high school*
- *High school graduate – persons who have graduated from high school, but not attended college, are estimated to have a BMI that is 0.50 below otherwise observationally equivalent persons who have not attended high school*
- *Some college – persons that have attended some college, but do not have a degree, are estimated to have a BMI that is 0.57 below otherwise observationally equivalent persons who have not attended high school*
- *College graduate – persons that have a college degree have are estimated to have a BMI that is 1.15 below otherwise observationally equivalent persons who have not attended high school*

From these coefficients we can conclude that persons who have attended high school have lower BMIs that persons who have not attended high school. Without conducting further hypothesis testing concerning differences between the coefficients on high school graduate and some college, and some college and college graduate, this is the extent of the conclusions we can draw.

B) **(10 Points)** Is the coefficient on Household Income statistically significant (i.e. statistically different from 0) at the 5% level? What about at the 1% level? Using the coefficients on Household Income and Household Income Squared together, what can we say about the relationship between Household Income and Obesity? Is the effect of an additional $1 in Household Income on BMI greater for rich or poor households?

The Z-statistic associated with the null hypothesis that the true value of the coefficient on family income is equal to zero is -33. on the basis of this Z-stat we would be able to reject this null the 5% level and the 1% level.

From the model the estimated effect of a 1000 change in family income on is equal to

$$\frac{\Delta BMI}{\Delta HHY} = -0.035 + \frac{2}{100} 0.020 \cdot \left(HHY\right) = -0.35 + 0.004 \cdot \left(HHY\right)$$

This implies a nonlinear relationship between HH income an BMI. It is estimated that for levels of HH income between zero and $8.75 (0.035/0.004) increases in HH income lead to reductions in BMI. For levels of income above $8.75 increases in HH income are estimated lead to increases in BMI.

C) **(5 Points)** How would your interpretation of the coefficients change if the dependent variable was measured in natural logs?

The coefficients would then be interpreted as the fractional change in BMI associated with a one unit increase in the associated independent variable, all else constant.

D) **(5 Points)** How would your interpretation of the coefficient on the food at home price change if both the dependent variable and the food at home price were measured in natural logs?

The coefficients on food at home price would then be interpreted as an elasticity so it would tell us how much of a percentage change in BMI to expect if food price at home increased by 1%, all else constant.

E) **(10 Points)** Use the coefficients on the variables Restaurants (full-service + fast-food), Restaurants squared, Fast-food restaurant price and Fast-food restaurant price squared to say something about how the availability and price of fast food affects adult obesity. In your answer, be sure to mention the significance of your results.

Persons that live in states with a lot of restaurants and/or where the price of restaurant food comparatively inexpensive are estimated to have higher BMIs than persons that live in states with fewer number of restaurants and/or where the price of restaurant food is comparatively more expensive. Thus, availability of low cost restaurant food appears to increase BMI. This said, there is a potential endogeniety problem in that low cost (fast food) restaurants may locate where there are the type of consumers that have a high demand for the type of food sell (i.e. high BMI folks).

F) **(10 Points)** Using the Standard Normal Tables provided, which independent variables do not have NOT statistically significant effects on BMI at the 1% level?

I've put asterisk on all next to the estimated coefficients associated with all of the independent variables that do have statistically significant effects on BMI at the 1% level.

Question 3. (40 Points) The Empirical Human Capital Model.

Consider the following specification for earnings:

$$\ln(wage) = \alpha_0 + \beta_{ed} ED + \beta_{ex} EX + \beta_{exsq} EXSQ + \alpha_{black} BLACK + \alpha_{hisp} HISP + \alpha_{87} Y87 + \varepsilon$$

where ED years of schooling, EX equals potential experience, EXSQ equals experience squared, BLACK equals 1 if black (0 otherwise), HISP equals 1 if HISP (0 otherwise), and Y87 equals 1 if year=1987 (0 if year=1979). The table below shows the results of estimating this specification on a sample of males from the 1979 and 1987 CPS Outgoing Rotation Group samples (the sample was drawn from the file cpsorg.dta that you were provided in conjunction with your homework assignment).

```
      Source |       SS       df       MS              Number of obs =    5924
-------------+------------------------------           F(  6,  5917) =  412.78
       Model |  392.798305     6  65.4663842           Prob > F      =  0.0000
    Residual |   938.42215  5917  .158597625           R-squared     =  0.2951
-------------+------------------------------           Adj R-squared =  0.2944
       Total |  1331.22046  5923  .224754424           Root MSE      =  .39824

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |    .069017   .0020337       ?       ?          ?           ?
          ex |    .037796   .0014979    25.23   0.000     .0348596    .0407323
        exsq |   -.000563   .0000329   -17.09   0.000    -.0006276   -.0004984
       black |  -.1787841   .0201763    -8.86   0.000     -.218337   -.1392311
        hisp |  -.0972316   .0232317    -4.19   0.000    -.1427743   -.0516889
         y87 |  -.0913855   .0104641    -8.73   0.000    -.1118989   -.0708721
       _cons |   1.561887   .0301004    51.89   0.000     1.502879    1.620894
------------------------------------------------------------------------------
```

A) **(10 Points)** Test the null hypothesis that one additional year of school increases wages by less than 6 percent. Use a significance level of 0.05.

$$H_0 : \beta_{ed} \le 0.06$$
$$H_a : \beta_{ed} > 0.06$$

$$\Rightarrow Z = \frac{0.069 - 0.06}{0.002} = \frac{0.009}{0.002} = 4.5$$

$$\Rightarrow p - value = P\left(\hat{\beta}_{ed} \ge 0.069 \mid \beta = 0.06\right) \approx 0 < 0.05$$

$$\Rightarrow \text{reject } H_0$$

B) **(10 Points)** There is good reason to believe that ability should be included in the above specification. Unfortunately most data do not contain information on ability. If ability were added to the above specification what might you reasonably expect to happen to the estimated coefficient on years of schooling? Why?

As it is easier for students with a high level of ability to attain high levels of schooling, ability and schooling are most likely positively correlated. The implication of this correlation is that the years of school variable partially reflecting differences in ability, and that the estimated coefficient reflects schooling and ability. If we included a measure of ability then the schooling variable would be measuring the effect of schooling in the natural log of wage, holding ability constant. This estimated effect would most likely be lower.

C) **(10 Points)** Suppose that you wanted the rate of return to a year of education to vary by level of education so that there was one rate of return on years of schooling for workers with less than a high school degree, another rate for the last year of high school, another rate for the first 3 years of college, and yet another rate for the last year of college and subsequent years of professionally or post-baccalaureate education. Briefly describe how you would specify a model that would allow for estimation of these different rates of return. (We are not looking for an exact specification, but just some idea of how you would solve this problem).

Such a model would require interactions between the years of schooling variable and education level dummy-indicator variables.

Next consider the following alternative specification where the 1987 year dummy variable is interacted with the schooling variable:

$$\ln(wage) = \alpha_0 + \beta_{ed} ED + \beta_{ed.y87} ED.Y87 + \beta_{ex} EX + \beta_{exsq} EXSQ$$
$$+ \alpha_{black} BLACK + \alpha_{hisp} HISP + \alpha_{87} Y87 + \varepsilon$$

The estimated coefficients associated with this model are shown in the table below (Note the 1987 year dummy – ed interaction term is ed_y87).

Source	SS	df	MS
Model	396.480974	7	56.6401392
Residual	934.739481	5916	.158001941
Total	1331.22046	5923	.224754424

Number of obs =	5924
F(7, 5916) =	358.48
Prob > F =	0.0000
R-squared =	?
Adj R-squared =	?
Root MSE =	.39749

lnwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ed	.0611403	.0026043	?	?	?	?
ed_y87	.017892	.003706	?	?	?	?
ex	.037695	.0014952	25.21	0.000	.0347639	.0406261
exsq	-.0005634	.0000329	-17.14	0.000	-.0006279	-.000499
black	-.1778687	.0201393	-8.83	0.000	-.2173491	-.1383884
hisp	-.0957991	.02319	-4.13	0.000	-.1412598	-.0503383
y87	-.3208827	.0486704	-6.59	0.000	-.4162945	-.225471
_cons	1.662049	.0365112	45.52	0.000	1.590474	1.733624

D) **(5 Points)** In which year (1979 or 1987) is the estimated rate of return to a year of education higher? Test the hypothesis that the rate of return to education is the same in both years. Use a significance level of 0.05.

The difference in the rate of return between 1987 and 1979 is given by $\beta_{ed.y87}$. .

$$H_0 : \beta_{ed.y87} = 0$$
$$H_a : \beta_{ed.y87} \neq 0$$

The Z-stat associated with this test is

$$Z = \frac{0.018 - 0}{0.004} \approx 4.5 \Rightarrow p - value = P\left(\hat{\beta}_{ed.y87} \geq 0.18 \mid \beta_{ed.y87} = 0\right) \approx 0 < 0.05$$
$$\Rightarrow \text{reject } H_0$$

E) **(5 Points)** What is the R-squared in the model that includes the interaction term?

$$R - squared = \frac{SSR}{SST} = \frac{396.48}{1331.22} \approx 0.30$$

Question 4. (30 Points) Texas SAT Scores.

Suppose you work for the government of the state of Texas in the Texas Education Agency. You have data on SAT scores at the school-district level from across the state and wish to analyze these scores. Suppose that historical data in the state has shown that the average SAT score by school district is 931. In the current year, you have a sample of 225 Texas school districts in which the average district-wide SAT score is 944. The standard deviation of the sample is 89.

A) **(3 points)** If you want to test the hypothesis that district-wide SAT scores have increased, what are the appropriate null and alternative hypotheses?

$$H_0 : \mu \leq 931$$
$$H_1 : \mu > 931$$

B) **(5 points)** Describe the distribution of the sample mean under the null hypothesis. A complete answer will specify the distributional form, mean and variance in terms of the population parameters under the null hypothesis.

$$\overline{X} \sim N\left(931, \frac{\sigma^2}{225}\right)$$

C) **(5 points)** Suppose you reject the null hypothesis, what is the probability that you make a type I error?

The probability of a type I error is equal to the significance level.

A member of the Texas Sate Board of Education believes that teacher experience is an important factor in the quality of education and therefore average SAT scores. This board member wishes to present information regarding her hypothesis at the next commission meeting. You also have data on average teacher experience in the school districts in your sample and gladly comply with her request to improve her presentation using your knowledge of statistics. In your sample, 169 school districts have an average teacher experience exceeding ten years. The average district-wide SAT score in these districts is 976 with a sample standard deviation of 70. The remaining 56 school districts have average teacher experience levels of 10 years or less. The district-wide average SAT score in these districts is 957 with a sample standard deviation of 71.

D) **(5 points)** What is the null hypothesis if you wish to provide conclusive statistical evidence that districts with more experienced teachers generate better SAT scores among their students?’

Let μ_h *equal the mean SAT scores from districts with an average teacher experience of more than 10 years and let* μ_l *equal the mean SAT score from* *districts with an average teacher attendance less than or equal to 10 years. The hypothesis we want to test is*

$$H_0 : \mu_h - \mu_l \leq 0$$
$$H_a : \mu_h - \mu_l > 0$$

E) **(7 points)** Test this hypothesis at the 5% level of significance.

The Z-stat for this test is

$$Z = \frac{(976 - 957) - 0}{\sqrt{\frac{70^2}{169} + \frac{71^2}{56}}} \approx 1.74 \Rightarrow$$
$$p - value = P\left(\Delta\overline{X} \geq 19 \mid \Delta\mu = 0\right) = P\left(Z \geq 1.75\right) = 0.409 < 0.05 = \alpha$$
$$\Rightarrow \text{reject } H_0$$

F) **(5 points)** Discuss your findings from part E), particularly how the level of significance may have affected your conclusion.

We are going to be able to reject the null at any significance level greater than 0.05. At standard significance levels below 0.05 we will not be able to reject the null.