POS 5737
Fall 2001, Key for Homework #8

1. You want to determine if economic productivity is the same across regions in the world, or if there is some difference (such as a first world-third world disparity). You collect information on both variables using the 1995 World Survey. The first variable, Gross Domestic Product/per capita (GDPCAT), is an ordinal variable and takes on three possible values (1=low GDP/cap, 2=medium GDP/cap, 3=high GDP/cap). The second variable, region, is a nominal variable coded 1-6 for the following regions (OECD, East Europe, Pacific/Asia, Africa, Middle East, Latin America).

    a. State the null and alternative hypotheses.

    H_o: $\pi_i \pi_j = \pi_{ij}$ (GDP/capita and region are independent.)
    H_1: $\pi_i \pi_j \neq \pi_{ij}$ (GDP/capita and region are dependent, i.e. they are related.)

    b. Using the 1995 World Survey, test your hypothesis (hint: run a cross-tabulation and use chi-square for your test).

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	106.262[a]	10	.000
Likelihood Ratio	104.257	10	.000
Linear-by-Linear Association	50.906	1	.000
N of Valid Cases	109		

a. 9 cells (50.0%) have expected count less than 5. The minimum expected count is 2.18.

The critical χ^2 is 18.3 (df = 10), so we reject H_o because 106.262 > 18.3 (also because $p < \alpha$; $p < .0001$, which is smaller than .05). We conclude that there is a difference in economic productivity across various regions in the world.

2. One of your colleagues criticizes your approach, arguing that you should not have collapsed the GDP per capita variable into three categories. Instead, you should compare the mean GDP per capita levels across regions (using an interval measure of GDP/cap) to see if there is any difference. You re-analyze your data, using GDP_CAP to see if this makes any difference.

   a. State the null and alternative hypotheses.

H_o: $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5=\mu_6$ (The regions come from the same population)
H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5 \neq \mu_6$ (The regions come from different populations)

   b. Compare the means between regions. Do they seem different just by looking at the raw data?

Descriptives

Gross domestic product / capita

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
OECD	21	16610.86	3725.97	813.07	14914.82	18306.90	8060	23474
East Europe	14	5159.79	1708.70	456.67	4173.21	6146.36	2340	7400
Pacific/Asia	17	4263.00	6291.05	1525.80	1028.44	7497.56	202	19860
Africa	19	998.68	1178.26	270.31	430.78	1566.59	122	4283
Middle East	17	4957.41	4057.45	984.08	2871.26	7043.56	748	14193
Latn America	21	1997.67	1482.12	323.43	1323.01	2672.32	383	6950
Total	109	5859.98	6479.84	620.66	4629.73	7090.23	122	23474

Yes, the means are different when we examine the raw data. The mean GDP per capita for the richest region (OECD) is over $15,000 larger than the mean for the poorest region (Africa).

   c. Test your hypotheses for a difference in means across regions (hint: run a one-way ANOVA).

ANOVA

Gross domestic product / capita

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.3E+09	5	650711082.0	52.314	.000
Within Groups	1.3E+09	103	12438622.64		
Total	4.5E+09	108			

The calculated F (52.314) exceeds the critical F (between 2.29 and 2.37), so we reject H_o and conclude that there is a difference in economic productivity across the regions. This confirms what we found with the ordinal level measure in #1.

3. Your next task is to examine the life expectancy rates of men and women in these 109 countries. You want to determine if there is any difference between men and women in terms of their life expectancy (use the variables LIFEEXPF and LIFEEXPM).

   a. State the null and alternative hypotheses.

H_o: D = 0 where D = Female life expectancy - Male life expectancy
H_1: D ¹ 0

   b. Test the difference between male and female life expectancy (use a matched or paired samples t-test).

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Average female life expectancy	70.16	109	10.57	1.01
	Average male life expectancy	64.92	109	9.27	.89

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Average female life expectancy & Average male life expectancy	109	.982	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Average female life expectancy - Average male life expectancy	5.24	2.27	.22	4.81	5.67	24.109	108	.000

We can see that the calculated t value (24.109) exceeds the critical t value (between 1.98 and 2.00), so we reject H_0 and conclude that there is a significant difference in male and female life expectancy (also $p < \alpha$). Women live longer than men on average.

c. You decide to examine the correlation between these two variables as well. How do you interpret this number? How can you compare this to your answer in b?

The correlation between male and female life expectancy is .982, indicating a strong, positive relationship between these two variables. This means that even though women live longer than men (there is a significant difference), countries that have a high life expectancy for women will also have a high life expectancy for men (both will be above their means). Conversely, in countries where women have lower than average life expectancy, men will also have lower than average life expectancy (both will be below their means).

4. Finally, you want to develop a more complete model of life expectancy, taking into account factors that may account for longer or shorter life spans. Run a regression model using male life expectancy as your dependent variable (LIFEEXPM). Select at least 3 independent variables from the data set that you think will help to explain differences in life expectancy across these countries.

   a. For each independent variable selected, provide a brief theoretical rationale for including the variable. Present the hypothesis to be tested for each parameter (discuss what sign the parameter should have and why).

Model: Male Life Expectancy = $\beta_1 + \beta_2$*Calories + β_3*GDPCAP + β_4*Urban + ε

- **Daily calorie intake: People who have healthy diets will live longer on average, thus we expect that increased caloric intake will increase male life expectancy.**
  H_o: $b_2 = 0$
  H_1: $b_2 > 0$

- **Gross Domestic Product per capita: Nations with higher economic productivity will have better living conditions, improved health care, etc. Thus as GDP/capita increases, male life expectancy will increase.**
  H_o: $b_3 = 0$
  H_1: $b_3 > 0$

- **People living in cities: Urbanization is an indication of economic development, and people living in cities will have greater access to health care. Thus nations with a higher percentage of people living in cities will have higher male life expectancy.**
  H_o: $b_4 = 0$
  H_1: $b_4 > 0$

b. Present your regression results and evaluate your model in terms of 1) the overall fit or explanatory power of the model, 2) the individual parameters (are the variables statistically significant and in the predicted direction, and 3) the substantive significance of each variable (interpret the regression parameters in terms of their impact on life expectancy).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.830[a]	.690	.676	5.78

a. Predictors: (Constant), People living in cities (%), Gross domestic product / capita, Daily calorie intake

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5198.631	3	1732.877	51.867	.000[a]
	Residual	2338.721	70	33.410		
	Total	7537.351	73			

a. Predictors: (Constant), People living in cities (%), Gross domestic product / capita, Daily calorie intake

b. Dependent Variable: Average male life expectancy

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	33.367	4.553		7.329	.000
	Daily calorie intake	7.487E-03	.002	.414	3.632	.001
	Gross domestic product / capita	1.235E-04	.000	.087	.839	.404
	People living in cities (%)	.168	.038	.415	4.424	.000

a. Dependent Variable: Average male life expectancy

1) Overall fit of the model
The R^2 is .690, which indicates that daily calorie intake, GDP/capita, and the % of people living in cities can account for 69% of the variance in male life expectancy across the countries in our sample.

We can also conduct a F test for the model as a whole.
H_o: $\beta_2 = \beta_3 = \beta_4 = 0$
H_1: $\beta_2 \neq \beta_3 \neq \beta_4 \neq 0$

The calculated F (51.867) exceeds the critical F (between 2.68 and 2.76), so we reject H_o and conclude that the model overall is a good one (at least one of the parameters for the independent variables is not zero, i.e., has a significant relationship with the dependent variable).

2) Individual parameters
All of the parameters are in the predicted direction (positive), and daily calorie intake and people living in cities are statistically significant at the .05 level (because the reported p values are < .001). Thus as daily calorie intake and urbanization increase, male life expectancy increases. GDP/capita does not have a significant effect on male life expectancy.

3) Substantive significance of the parameters
We will focus only on the two parameters that are statistically significant (daily calorie intake and people living in cities).

- **Daily calorie intake: a one unit (calorie) increase in daily calorie intake leads to a .007487 increase in male life expectancy. It would take an increase of about 133 calories per day to increase male life expectancy by 1 year. This would be feasible given the large range of daily calorie intake in our sample (difference between largest and smallest is 2158 calories).**

- **People living in cities: a one percent increase in the amount of people living in cities increases male life expectancy by .168 years. This means that it would take close to a 6% increase in level of urbanization to increase male life expectancy by 1 year.**