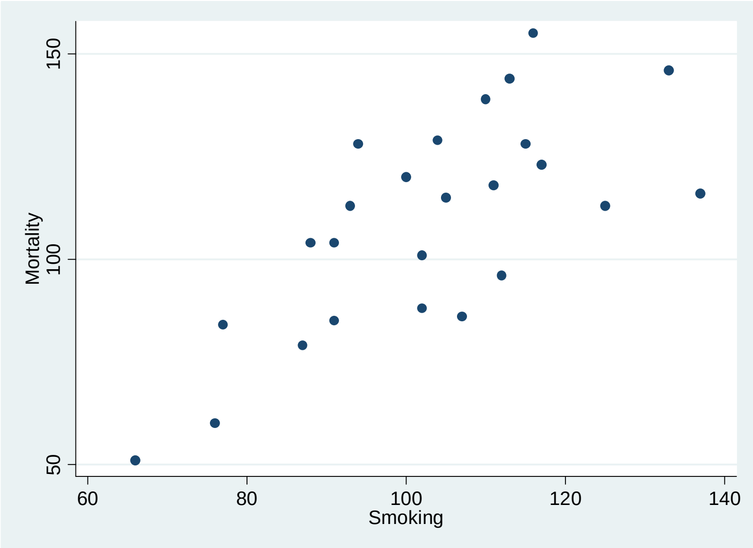
EE 325 ☺☺ **STATA Session I ANSWER**

1. Table 5.11 provides data on the lung cancer mortality index (100 = average) and the smoking index (100 = average) for 25 occupational groups.

   a. Plot the cancer mortality index against the smoking index. What general pattern do you observe?



   b. Letting Y= cancer mortality index and X = smoking index, estimate a linear regression model

Source	SS	df	MS
Model	8395.74904	1	8395.74904
Residual	7970.25096	23	346.53265
Total	16366	24	681.916667

Number of obs =	25
F(1, 23) =	24.23
Prob > F =	0.0001
R-squared =	0.5130
Adj R-squared =	0.4918
Root MSE =	18.615

mortality	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
smoking	1.087532	.2209452	4.92	0.000	.6304724	1.544592
_cons	-2.885319	23.03372	-0.13	0.901	-50.5342	44.76356

$$Y_i = -2.8853 + 1.0875X_i$$

   c. Test the hypothesis that smoking has no influence on lung cancer at $\alpha = 5\%$

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

$$t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)} = 4.92$$

$t > critical\ t \quad reject\ H_0$

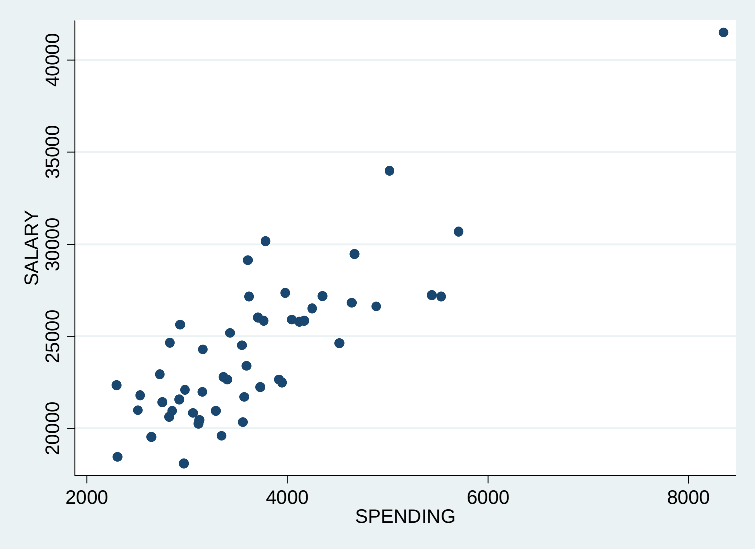
There is enough evidence to say that $\beta_2 \neq 0$

2. Table 5.5 gives data on average public teacher pay (annual salary in dollars) and spending on public school per pupil (dollars) in 1985 for 50 states and the District of Columbia
To find out if there is any relationship between teacher's pay and per pupil expenditure in public schools, the following model was suggested:

$$Pay_i = \beta_1 + \beta_2 Spend_i + u_i$$

, where Pay stands for teacher's salary and Spend stands for per pupil expenditure.

a. Plot the data and eyeball a regression line.



b. Suppose on the basis of (a) you decide to estimate the above regression model. Obtain the estimates of the parameters, their standard errors, r^2, RSS and ESS.

Source	SS	df	MS
Model	608555015	1	608555015
Residual	264825250	49	5404596.94
Total	873380265	50	17467605.3

Number of obs =	51
F(1, 49) =	112.60
Prob > F =	0.0000
R-squared =	0.6968
Adj R-squared =	0.6906
Root MSE =	2324.8

salary	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
spending	3.307585	.3117043	10.61	0.000	2.681192	3.933978
_cons	12129.37	1197.351	10.13	0.000	9723.204	14535.54

c. Interpret the regression. Does it make economic sense?
If the spending per pupil increases by a dollar, the average pay increases by about $3.31. The intercept term has no viable economic meaning.

d. Establish a 95 percent confidence interval for β_2. Would you reject the hypothesis that the true slope coefficient is 3.0?
Based on this CI you will not reject the null hypothesis that the true slope coefficient is 3.

3. **Construct regression model and hypothesis testing (p-value method)**
Table 3.3 gives data on the number of cell phone subscribers and the number of personal computers (PCs), both per 100 persons, and the purchasing-power adjusted per capita income in dollars for a sample of 34 countries.

3.1 To see if per capita income is a factor in the use of cell phones, we regressed each of these means of communication on per capita income using the sample of 34 countries. Construct a regression line and interpret the meaning. Is the estimated intercept coefficient different from zero at the 5 percent significance level? Is the estimated slope coefficient different from zero at the 5 percent significance level?

Source	SS	df	MS
Model	18494.4615	1	18494.4615
Residual	12702.6961	31	409.764392
Total	31197.1577	32	974.911177

Number of obs	33
F(1, 31)	45.13
Prob > F	0.0000
R-squared	0.5928
Adj R-squared	0.5797
Root MSE	20.243

cellphone	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pcapincome	.0021411	.0003187	6.72	0.000	.0014911	.0027911
_cons	14.8348	6.06433	2.45	0.020	2.466517	27.20308

3.2 To see if per capita income is a factor in the use of PCs, we regressed each of these means of communication on per capita income using the sample of 34 countries. Construct a regression line and interpret the meaning. Is the estimated intercept coefficient different from zero at the 5 percent significance level? Is the estimated slope coefficient different from zero at the 5 percent significance level?

Source	SS	df	MS
Model	13376.0083	1	13376.0083
Residual	2350.51746	31	75.8231438
Total	15726.5257	32	491.453928

Number of obs	33
F(1, 31)	176.41
Prob > F	0.0000
R-squared	0.8505
Adj R-squared	0.8457
Root MSE	8.7076

pcs	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pcapincome	.0018209	.0001371	13.28	0.000	.0015413	.0021005
_cons	-6.8197	2.608655	-2.61	0.014	-12.14009	-1.499314

4. **Construct regression model (The Log-Linear Model)**
Expenditure on Durable Goods in relation to total personal consumption expenditure
Table 6.3 presents data on total personal consumption expenditure (PCEXP), expenditure on durable goods (EXPDUR), expenditure on nondurable goods (EXPNONDUR), and expenditure on services (EXPSERVICES), all measured in 2000 billions of dollars.

Suppose we wish to find **the elasticity of expenditure on durable goods** with respect **to total personal consumption expenditure**. . Construct a regression line and interpret the meaning

Source	SS	df	MS
Model	.056007624	1	.056007624
Residual	.001764196	13	.000135707
Total	.057771819	14	.004126559

Number of obs	=	15
F(1, 13)	=	412.71
Prob > F	=	0.0000
R-squared	=	0.9695
Adj R-squared	=	0.9671
Root MSE	=	.01165

lnexpdur	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnpcexp	1.626604	.0800682	20.32	0.000	1.453627	1.79958
_cons	-7.541638	.7161477	-10.53	0.000	-9.088781	-5.994495

5. The demand for roses. Table 7.6 gives quarterly data on these variables:

$Y = quantity\ of\ roses\ sold,\ dozens$

$X_2 = average\ wholesale\ price\ of\ roses, \$/dozen$

$X_3 = average\ wholesale\ price\ of\ carnations, \$/dozen$

$X_4 = average\ weekly\ family\ disposable\ income, \$/week$

$X_5 = the\ trend\ Variable\ taking\ values\ of\ 1,\ 2, and\ so\ on,, for\ the\ period$
$1971 - III\ to\ 1975 - II$ in $the\ Detroit\ Metropolitian\ area$

You are asked to consider the following demand functions:

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + \alpha_5 X_{5t} + u_t$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 X_{5t} + u_t$$

a. Estimate the parameters of the linear model and interpret the results.

Source	SS	df	MS
Model	52249133.2	4	13062283.3
Residual	10347222.8	11	940656.615
Total	62596356	15	4173090.4

Number of obs	=	16
F(4, 11)	=	13.89
Prob > F	=	0.0003
R-squared	=	0.8347
Adj R-squared	=	0.7746
Root MSE	=	969.87

y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x2	-2227.704	920.4659	-2.42	0.034	-4253.636	-201.772
x3	1251.141	1157.021	1.08	0.303	-1295.445	3797.726
x4	6.283002	30.62166	0.21	0.841	-61.11482	73.68083
x5	-197.4	101.5612	-1.94	0.078	-420.9348	26.13479
_cons	10816.04	5988.35	1.81	0.098	-2364.229	23996.31

$$\hat{Y}_t = 10816.04 - 2227.704 X_{2i} + 1251.141 X_{3i} + 6.283 X_{4i} - 197.399 X_{5i}$$

$se \quad (5988.348)(\ 920.538) \quad (1157021) \quad (29.919) \quad (101.156)$

$$R^2 = 0.835$$

In this model the slope coefficients measure the rate of change of Y with respect to the relevant variable.

b. Estimate the parameters of the log-linear model and interpret the results.

Source	SS	df	MS
Model	1.12835383	4	.282088457
Residual	.284245018	11	.025840456
Total	1.41259884	15	.094173256

Number of obs =	16
F(4, 11) =	10.92
Prob > F =	0.0008
R-squared =	0.7988
Adj R-squared =	0.7256
Root MSE =	.16075

lny	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnx2	-1.170727	.4883241	-2.40	0.035	-2.245521	-.0959329
lnx3	.7379375	.6528623	1.13	0.282	-.6990027	2.174878
lnx4	1.153217	.9019901	1.28	0.227	-.8320496	3.138484
x5	-.0301109	.0164188	-1.83	0.094	-.0662484	.0060266
_cons	3.572134	4.695165	0.76	0.463	-6.761856	13.90612

$$\ln \hat{Y}_t = 0.627 - 1.274 \ln X_{2i} + 0.937 \ln X_{3i} + 1.713 \ln X_{4i} - 0.182 \ln X_{5i}$$

se (6.148) (0.527) (0.659) (1.201) (0.128)

$R^2 = 0.778$

In this model all the partial slope coefficients are partial elasticities of Y with respect to the relevant variable.

c. $\beta_2, \beta_3, and\ \beta_4$ give, respectively, the own-price, cross-price, and income elasticities of demand. What are their a priori signs? Do the results concur with the a priori expectations?

The own-price elasticity is expected to be negative, the cross price elasticity is expected to be positive for substitute goods and negative for complimentary goods, and the income elasticity is expected to be positive, since roses are a normal good.

6. Table 7.12 gives data for real consumption expenditure, real income, real wealth, and real interest rates for the U.S. for the years 1947-2000.

a. Given the data in the table, estimate the linear consumption function using income, wealth, and interest rate. What is the fitted equation?

Source	SS	df	MS
Model	119322125	3	39774041.8
Residual	71437.3791	50	1428.74758
Total	119393563	53	2252708.73

Number of obs =	54
F(3, 50) =	27838.40
Prob > F =	0.0000
R-squared =	0.9994
Adj R-squared =	0.9994
Root MSE =	37.799

c	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
yd	.7340283	.0137519	53.38	0.000	.7064068	.7616497
wealth	.0359756	.0024831	14.49	0.000	.0309881	.040963
interest	-5.521116	2.306643	-2.39	0.020	-10.15415	-.8880871
_cons	-20.63328	12.82697	-1.61	0.114	-46.397	5.130442

$$C_t = -20.6327 + 0.7340 Y_t + 0.0360 Wealth_t - 5.5212 Interest_t$$

b. What do you estimated coefficients indicate about the variables' relationships to consumption expenditure?

The three independent variables are statistically significant at the 5% level. It seems that increases in income and wealth are related to increases in consumption, whereas an increase in the interest rate corresponds to a decrease in consumption level.

7. Table 7.11 gives data for the manufacturing sector of the Greek economy for the period 1961-1987.

a. See if the Cobb-Douglas production function fits the data given in the table and interpret the results. What general conclusion do you draw?

Source	SS	df	MS
Model	5.37753949	2	2.68876975
Residual	.158356562	24	.00659819
Total	5.53589605	26	.212919079

Number of obs	27
F(2, 24)	407.50
Prob > F	0.0000
R-squared	0.9714
Adj R-squared	0.9690
Root MSE	.08123

lnoutput	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lncapital	.1398108	.1653906	0.85	0.406	-.2015386	.4811603
lnlabor	2.328398	.5994894	3.88	0.001	1.091112	3.565683
_cons	-11.93657	3.211061	-3.72	0.001	-18.56388	-5.30927

b. Now consider the following model:

$$Output / labor = A(K / L)^{\beta} e^{u}$$

where the regressand represents labor productivity and the regressor represents the capital labor ratio. What is the economic significance of such a relationship, if any? Estimate the parameters of this model and interpret your results.

Source	SS	df	MS
Model	2.17537967	1	2.17537967
Residual	.232757738	25	.00931031
Total	2.40813741	26	.09262067

Number of obs	27
F(1, 25)	233.65
Prob > F	0.0000
R-squared	0.9033
Adj R-squared	0.8995
Root MSE	.09649

lnproducti~y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnclratio	.6807562	.0445355	15.29	0.000	.5890337	.7724788
_cons	-1.155956	.0742171	-15.58	0.000	-1.308809	-1.003103

The elasticity of output/ labor ratio (labor productivity) with respect to capital/labor ratio is about 0.68, meaning that if the latter increases by 1% labor productivity, on average, goes up by about 0.68%. A key characteristic of developed economies is a relatively high capital/labor ratio.

8. Table 8.9 Savings and Personal Disposable income (billions of dollars), United States, 1970-1995.

a. Given the data in the table, estimate the following linear savings function using personal disposable income

Time period 1970-1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12$

Time period 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad n_2 = 14$

Time period 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_{3t} \quad n_1 + n_2 = 26$

Time period 1970-1981: $Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \quad n_1 = 12$

Source	SS	df	MS
Model	16456.2587	1	16456.2587
Residual	1785.03254	10	178.503254
Total	18241.2912	11	1658.2992

Number of obs	12
F(1, 10)	92.19
Prob > F	0.0000
R-squared	0.9021
Adj R-squared	0.8924
Root MSE	13.361

savings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
income	.0803319	.0083665	9.60	0.000	.0616901	.0989737
_cons	1.016115	11.63771	0.09	0.932	-24.91432	26.94655

Time period 1982-1995: $Y_t = \gamma_1 + \gamma_2 X_t + u_{2t} \quad n_2 = 14$

Source	SS	df	MS
Model	2614.39647	1	2614.39647
Residual	10005.2214	12	833.768451
Total	12619.6179	13	970.739837

Number of obs	14
F(1, 12)	3.14
Prob > F	0.1020
R-squared	0.2072
Adj R-squared	0.1411
Root MSE	28.875

savings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
income	.0148624	.0083932	1.77	0.102	-.0034248	.0331496
_cons	153.4947	32.71227	4.69	0.001	82.22075	224.7686

Time period 1970-1995: $Y_t = \alpha_1 + \alpha_2 X_t + u_{3t} \quad n_1 + n_2 = 26$

Source	SS	df	MS
Model	76621.7867	1	76621.7867
Residual	23248.3	24	968.679166
Total	99870.0867	25	3994.80347

Number of obs	26
F(1, 24)	79.10
Prob > F	0.0000
R-squared	0.7672
Adj R-squared	0.7575
Root MSE	31.124

savings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
income	.0376791	.0042366	8.89	0.000	.0289353	.046423
_cons	62.42267	12.76075	4.89	0.000	36.08578	88.75957

b. On the basis of the Chow test that there was a difference in the regression of savings on income between the two periods. Consider and estimate the following model with the dummy variable:

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t$$

Where Y = Savings

X= Personal disposable income

t = time

D = 1 for observations in 1982-1995

= 0 otherwise

(i) Test the coefficients individually statistically significant at the 5 percent level? From this test, how would you describe the difference in the two regressions (coincident regression, parallel regression, concurrent regression, dissimilar regression)?

Source	SS	df	MS
Model	88079.8327	3	29359.9442
Residual	11790.2539	22	535.920634
Total	99870.0867	25	3994.80347

Number of obs =	26
F(3, 22) =	54.78
Prob > F =	0.0000
R-squared =	0.8819
Adj R-squared =	0.8658
Root MSE =	23.15

savings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
d	152.4786	33.08237	4.61	0.000	83.86992	221.0872
income	.0803319	.0144968	5.54	0.000	.0502673	.1103964
dx	-.0654694	.0159824	-4.10	0.000	-.098615	-.0323239
_cons	1.016115	20.16483	0.05	0.960	-40.80319	42.83542

$H_0 : \alpha_2 = 0$
$H_1 : \alpha_2 \neq 0$
$P - value = 0.000$
$reject\ H_0$

$H_0 : \beta_2 = 0$
$H_1 : \beta_2 \neq 0$
$P - value = 0.000$
$reject\ H_0$

There is enough evidence suggest that the coefficients individually statistically significant at the 5 percent level.

The two regressions are dissimilar regressions.

(ii) Write down the mean personal savings function for 1970-1981 and the mean personal savings function for 1982-2005.

Mean savings function for 1970-1981

$$E(Y_t \mid D_t = 0, X_t) = 1.1061 + 0.0803 X_t$$

Mean savings function for 1982-1995

$$E(Y_t \mid D_t = 1, X_t) = (1.0161 + 152.4786) + (0.0803 - 0.0655) X_t = 153.4947 + 0.0148 X_t$$