Business Statistics 41000-81/82 – Spring 2011
University of Chicago Booth School of Business
FINAL EXAM - SOLUTION

QUESTION 1 (10 points):

vale: Vale S.A. daily returns in 2009 (223 obs.)
ibovespa: IBOVESPA daily returns in 2009 223 obs.)
Fitted regression: vale = a + b*ibovespa

Coefficient	Estimate	Standard Error	Test Statistic
a (intercept)	0.0000584	**[0.001123]**	0.052
b (slope)	**[1.698163]**	0.0555300	30.581

a)(4) Fill the blanks.
0.0000584/se=0.052 => se=0.001123077
estimate/0.0555300=30.581 => estimate=1.698163

b)(2) Obtain the approximate 95% confidence interval for the slope.
Lower bound = 1.698163-2*0.0555300 = 1.587103
Upper bound = 1.698163+2*0.0555300 = 1.809223

c)(2) Test the hypothesis that the intercept is equal to zero.
Since the Test Statistic for the intercept is 0.052 < 2, then we fail to reject the hypothesis at the 5% level.

d)(2) Test the hypothesis that the slope is equal to one.
The Test Statistic (1.698163-1)/0.0555300=12.57272 > 2, then we reject the hypothesis at the 5% level (or 1% level).

QUESTION 2 (10 points):

Two branches of a firm deliver daily advice regarding whether an asset is up or down. Their performances over the last several months are as follows: When the asset is up, branch A says it should be up 85% of the time, while branch B says it should be up 90% of the time. When the asset is down, branch A says it should be down 98% of the time, while branch B says it should be down 80% of the time. Suppose that the asset goes up 45% of the time and down 55% of the time. Suppose that branches A and B say that tomorrow the asset is going down. Which branch will produce more reliable forecast? Why?

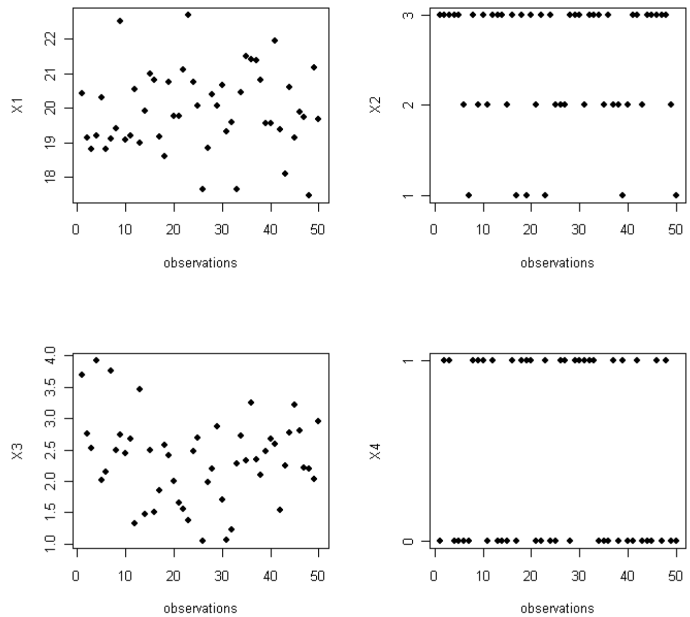
Let A=1 if branch A says the asset is up and A=0 if Branch A says the asset is down.
Let B=1 if branch B says the asset is up and B=0 if Branch B says the asset is down.
Let U=1 if asset goes up and U=0 if asset goes down.

Branch A	Branch B	Asset
Pr(A=1|U=1)=0.85	Pr(B=1|U=1)=0.90	Pr(U=1)=0.45
Pr(A=0|U=0)=0.98	Pr(B=0|U=0)=0.80	Pr(U=0)=0.55

Let us compute
Pr(U=0|A=0) = Pr(A=0|U=0)Pr(U=0)/Pr(A=0) = 0.98*0.55/Pr(A=0)
Pr(U=0|B=0) = Pr(B=0|U=0)Pr(U=0)/Pr(B=0) = 0.80*0.55/Pr(B=0)

We know that
Pr(A=0) = Pr(A=0|U=1)Pr(U=1)+Pr(A=0|U=0)Pr(U=0)=0.15*0.45+0.98*0.55=0.6065
Pr(B=0) = Pr(B=0|U=1)Pr(U=1)+Pr(B=0|U=0)Pr(U=0)=0.10*0.45+0.80*0.55=0.4850
Therefore,
Pr(U=0|A=0)= 0.98*0.55/0.6065 = 0.8887057
Pr(U=0|B=0)= 0.80*0.55/0.4850 = 0.9072165
Conclusion: There is a 11% chance branch A is wrong and 9% chance branch B is wrong. Therefore, we should be more confident on the forecast provided by branch B.

QUESTION 3 (10 points):
Below are time series plots of X1, X2, X3 and X4.



Do any of the above data sets look like sample from any of the following distributions?

a) (2) N(1.5,4) [--]

b) (2) N(20,1) [X1]

c) (2) Binomial(3,0.5) [--]

d) (2) Bernoulli(0.5) [X4]

e) (2) Binomial(3,0.8) [X2]

QUESTION 4 (10 points):
Suppose you run a regression based on a couple of thousand observations to explain earnings in dollars (Y) in a particular industry as a function of the number of years of experience (X1) and whether or not you have an MBA degree (X2). Here X2=1 means the employee has an MBA degree. The fitted regression is Y = a + b*X1 + c*X2. Explain the meanings of

a) (5) The intercept a.

a is the earnings in dollars of an employee with zero years of experience and without a MBA degree.

b) (5) The slope c.

c is the additional earnings in dollars an employee with zero years of experience but with a MBA degree.

QUESTION 5 (20 points):
The following table partially shows the regression of wages on education level, where wage is measured in dollars per hour and education is measured in school years.

SUMMARY OUTPUT

Regression Statistics	
R Square	
Standard Error	3.378389522
Observations	526

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1179.73205	1179.73205	103.3627	2.7826E-22
Residual	524		11.4135158		
Total	525	7160.41431			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.904851665	0.684967821		0.187073		
educ		0.053248037	10.1667459			

a) (18) Fill in the nine empty boxes.

R-square = SSRegression/SSTotal = 1179.73205/7160.41431 = 0.1647575

SSResidual = SSTotal-SSRegression = 7160.41431 - 1179.73205 = 5980.682

t-stat intercept = -0.904851665/0.684967821 = -1.321013

Slope = 0.053248037*10.1667459 = 0.5413593

P-value for slope is less than 0.01

95% C.I. for the intercept = (-2.274787;0.465084)

95% C.I. for the slope = (0.4348632;0.6478554)

b) (2) Provide a 95% predictive interval for the wage of an employee with 10-years of education.

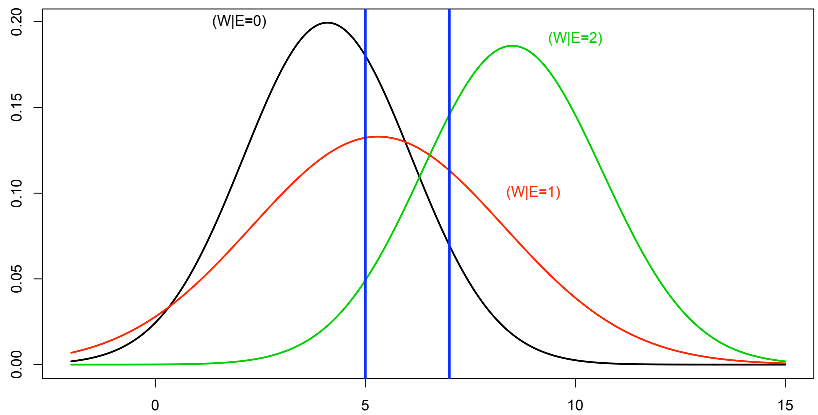
Point estimate: Wage = -0.904851665+0.5413593*10 = 4.508741

Lower bound: 4.508741 - 2*3.378389522 = -2.248038

Upper bound: 4.508741 + 2*3.378389522 = 11.26552

QUESTION 6: (10 points)

Let W be the wage (in dollars per hour) and E a categorical variable such that E=0 when years of education is between 0 and 9; E=1 when years of education is between 10 and 14; and E=2 when years of education is between 15 and 18. It is known that (W|E=0) ~ N(4.1,4.0), (W|E=1) ~ N(5.3,9.0) and (W|E=2) ~ N(8.5,4.6), where N(a,b) stands for the normal distribution with mean a and variance b. Therefore, the standard deviation is equal to sqrt(b). It is also known that 11% of the employees fall into category E=0, 66% fall into category E=1 and, consequently, the remainder 23% fall into category E=2. Question: Suppose the wage of a new employee is between 5 and 7 dollars per hour. In which category he/she is most likely to fall into?



We want to compare Pr(E=0|5<W<7), Pr(E=1|5<W<7) and Pr(E=2|5<W<7) and see which one bigger.

But first, we need to compute Pr(5<W<7|E=0), Pr(5<W<7|E=1) and Pr(5<W<7|E=2)

By using the normal table at the end of the exam, we can compute
Pr(5<W<7|E=0)=Pr(0.45<Z<1.45)=Pr(0<Z<1.45)-Pr(0<Z<0.45)=0.4265-0.1736=0.2529
Pr(5<W<7|E=1)=Pr(-0.1<Z<0.57)=Pr(0<Z<0.57)+Pr(0<Z<0.1)=0.2157+0.0398=0.2555
Pr(5<W<7|E=2)=Pr(-1.63<Z<-0.70)=Pr(0<Z<1.63)-Pr(0<Z<0.7)=0.4484-0.2580=0.1904

Therefore,
Pr(E=0|5<W<7)=Pr(5<W<7|E=0)*Pr(E=0)/Pr(5<W<7)
Pr(E=1|5<W<7)=Pr(5<W<7|E=1)*Pr(E=1)/Pr(5<W<7)
Pr(E=2|5<W<7)=Pr(5<W<7|E=2)*Pr(E=2)/Pr(5<W<7)

Since the denominators are the same, we just need to compute the numerators.
Pr(5<W<7|E=0)*Pr(E=0)=0.2529*0.11=0.027819
Pr(5<W<7|E=1)*Pr(E=1)=0.2555*0.66=0.168630
Pr(5<W<7|E=2)*Pr(E=2)=0.1904*0.23=0.043792

Therefore, he/she is most likely to fall into E=1.

In fact, the probability that he/she falls into E=1 given that W is in (5,7) is Pr(E=1|5<W<7)=0.168630/(0.027819+0.168630+0.043792)= 0.7019202 or 70.2%.

QUESTION 7: (10 points)

Suppose we are in the business of making "business cards", and that for a business card to be considered usable, say to fit your wallet or your business card holder, it must measure between 3.3 and 3.6 inches long. A total of 100 business cards were produced and lengths recorded: sample mean=3.495942 and sample variance=0.011786.

a)(4) Compute a 95% confidence interval for the true average length of a business card.

Lower bound = 3.495942 – 2*sqrt(0.011786/100) = 3.474229
Upper bound = 3.495942 + 2*sqrt(0.011786/100) = 3.517655

Let us now assume that X="length of business cards continuously produced" is normally distributed with mean 3.5 inches and standard deviation of 0.1 inches; that is X ~ N(3.5,0.01).

b)(4)Compute the probability that a business card is usable, i.e. compute p = Pr(3.3<X<3.6). *(Hint: use table 2)*

Let Z ~ N(0,1), i.e. a standard normal distribution with probabilities given by table 2 below. It is easy to see that Pr(3.3<X<3.6) = Pr(-2<Z<1). In words, 3.3 corresponds to 2 standard deviations to the left of the mean 3.5, while 3.6 corresponds to 1 standard deviation to the right of 3.5. Therefore, we just need to compute

Pr(-2<Z<1) = Pr(-2<Z<0) + Pr(0<Z<1)
= Pr(0<Z<2) + Pr(0<Z<1)
= 0.4772 + 0.3413
= 0.8186

Therefore, the probability that a business card is usable is 0.8186.

c) (2) Let D be the number of not usable business cards out of an i.i.d. sample of 1000 manufactured ones. What is the distribution of D?

Since the probability that a business card is not usable is 0.1814, it follows that D is binomially distributed with parameters 1000 and 0.1814, or D ~ BINOM(1000,0.1814).

QUESTION 8 (10 points):

Suppose we want to test H0: p=0.5 against the alternative Ha: H0 is false, where p is the proportion of people unhappy with Obama's health plan. We collected n=30 observations and observed 9 successes, i.e. 9 of the 30 persons were unhappy with Obama's health plan.

a)(4) Compute the P-value based on the normal approximation (table 2).

Under the null hypothesis, it follows that phat is approximately normal with mean 0.5, variance (0.5)(0.5)/30=0.008333333 and standard deviation 0.0912871. Since phat=0.3, it follows that the approximate P-value is equal to 2*Pr(Z<(0.3-0.5)/0.0912871)=2*Pr(Z<-2.19)=0.0286.

b)(4) Compute the exact P-value (table 1).

Let X=number of successes. Then, under the null hypothesis, it is known that X ~ Binomial(30,0.5). Since x=9 was observed, the P-value is equal to 2*Pr(X<=9) = 2*(0.000004+0.000553+0.001896+0.005451+0.013325) = 0.042458.

c)(2)Comment on the similarity/difference between a) and b).

The approximation is not so great mainly because n=30 is not a large sample size. Nonetheless, it is still a reasonable approximation.

QUESTION 9 (10 points):
Suppose you produce umbrellas and that in any given day you select 100 umbrellas from your production line to monitor the proportion of defective items. Your experience tells you that the defective rate is 2%. Nonetheless, you are not so sure since you have recently hired new employees and acquired new machines. Then, you decided to collect some data to learn about the process and update your experience. Below are sample percentages for each one of 20 consecutive days. The first 10 days are right before the new acquisitions, while the last 10 days are right after the new acquisitions:

First 10 days: 25 defective out of 1000 umbrellas
Last 10 days: 29 defective out of 1000 umbrellas

a)(4) Let p be the true population defective rate. Test the null hypothesis H0:p=0.02 based on the first 10 days.

Under the null hypothesis se=sqrt(0.02*0.98/1000)=0.004427189

Here phat = 25/1000=0.025.

The test statistic is (0.025-0.02)/0.004427189=1.29385.

Therefore, we fail to reject the null hypothesis at the 5% level.

b)(4) Repeat a) but now based on the last 10 days.

Under the null hypothesis se=sqrt(0.02*0.98/1000)=0.004427189

Here phat = 29/1000=0.029.

The test statistic is (0.025-0.02)/0.004427189=2.032893.

Therefore, we reject the null hypothesis at the 5% level.

c)(2) What happens to H0:p=0.02 when all 20 days are combined?

Under the null hypothesis se=sqrt(0.02*0.98/2000)=0.003130495

Here phat = 54/2000=0.027.

The test statistic is (0.027-0.02)/0.003130495=2.236068.

Therefore, we reject the null hypothesis at the 5% level.

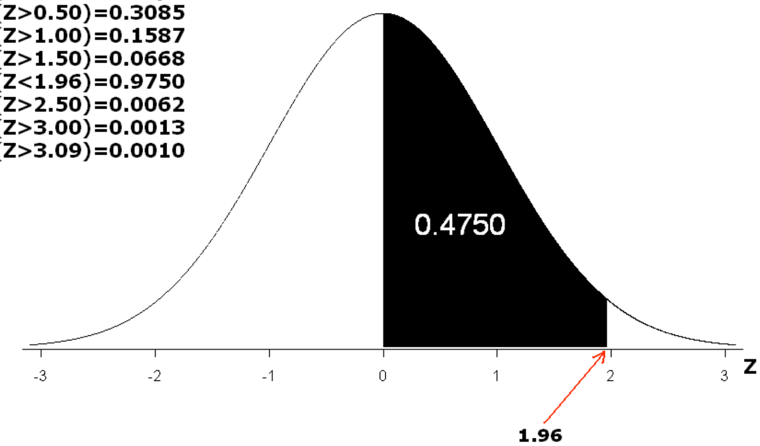
TABLE 1: BINOMIAL PROBABILITIES

X ~ Binomial(30,p)
Rows: number of successes
Columns: probability of success (p)
Table entry: Pr(x) for x=0,1,…,30.

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.042391	0.001238	0.000023	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.141304	0.009285	0.000290	0.000004	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.227656	0.033656	0.001801	0.000043	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.236088	0.078532	0.007203	0.000266	0.000004	0.000000	0.000000	0.000000	0.000000
4	0.177066	0.132522	0.020838	0.001197	0.000026	0.000000	0.000000	0.000000	0.000000
5	0.102305	0.172279	0.046440	0.004149	0.000133	0.000001	0.000000	0.000000	0.000000
6	0.047363	0.179457	0.082928	0.011524	0.000553	0.000008	0.000000	0.000000	0.000000
7	0.018043	0.153821	0.121854	0.026341	0.001896	0.000040	0.000000	0.000000	0.000000
8	0.005764	0.110559	0.150141	0.050487	0.005451	0.000173	0.000001	0.000000	0.000000
9	0.001565	0.067564	0.157291	0.082275	0.013325	0.000634	0.000006	0.000000	0.000000
10	0.000365	0.035471	0.141562	0.115185	0.027982	0.001997	0.000030	0.000000	0.000000
11	0.000074	0.016123	0.110308	0.139619	0.050876	0.005448	0.000126	0.000000	0.000000
12	0.000013	0.006382	0.074852	0.147375	0.080553	0.012938	0.000464	0.000002	0.000000
13	0.000002	0.002209	0.044418	0.136039	0.111535	0.026872	0.001498	0.000009	0.000000
14	0.000000	0.000671	0.023115	0.110127	0.135435	0.048945	0.004246	0.000042	0.000000
15	0.000000	0.000179	0.010567	0.078312	0.144464	0.078312	0.010567	0.000179	0.000000
16	0.000000	0.000042	0.004246	0.048945	0.135435	0.110127	0.023115	0.000671	0.000000
17	0.000000	0.000009	0.001498	0.026872	0.111535	0.136039	0.044418	0.002209	0.000002
18	0.000000	0.000002	0.000464	0.012938	0.080553	0.147375	0.074852	0.006382	0.000013
19	0.000000	0.000000	0.000126	0.005448	0.050876	0.139619	0.110308	0.016123	0.000074
20	0.000000	0.000000	0.000030	0.001997	0.027982	0.115185	0.141562	0.035471	0.000365
21	0.000000	0.000000	0.000006	0.000634	0.013325	0.082275	0.157291	0.067564	0.001565
22	0.000000	0.000000	0.000001	0.000173	0.005451	0.050487	0.150141	0.110559	0.005764
23	0.000000	0.000000	0.000000	0.000040	0.001896	0.026341	0.121854	0.153821	0.018043
24	0.000000	0.000000	0.000000	0.000008	0.000553	0.011524	0.082928	0.179457	0.047363
25	0.000000	0.000000	0.000000	0.000001	0.000133	0.004149	0.046440	0.172279	0.102305
26	0.000000	0.000000	0.000000	0.000000	0.000026	0.001197	0.020838	0.132522	0.177066
27	0.000000	0.000000	0.000000	0.000000	0.000004	0.000266	0.007203	0.078532	0.236088
28	0.000000	0.000000	0.000000	0.000000	0.000000	0.000043	0.001801	0.033656	0.227656
29	0.000000	0.000000	0.000000	0.000000	0.000000	0.000004	0.000290	0.009285	0.141304
30	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000023	0.001238	0.042391

TABLE 2: CUMULATIVE NORMAL PROBABILITIES

$P(\ 0.00<Z<1.96)=0.475$
$P(-1.96<Z<1.96)=0.950$
$P(Z>0.50)=0.3085$
$P(Z>1.00)=0.1587$
$P(Z>1.50)=0.0668$
$P(Z<1.96)=0.9750$
$P(Z>2.50)=0.0062$
$P(Z>3.00)=0.0013$
$P(Z>3.09)=0.0010$



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990