6.041/6.431 Lecture 26

December 10, 2008

Review Statistics

Stock Forecasting Problem

STATISTICS OVERVIEW

The following summary uses a simple uniform notation to make it easier to compare methods and results from different parts of statistics we have covered. The price of this simplicity is that in some places the notation is too narrow and in some places it differs from that used in the book.

This page gives the full notation and the notation used in the book, where the simple notation is inadequate. Nonetheless, you will find that the simple notation makes the summary on the next page much easier to follow and digest.

The notation in the summary assumes the data observation is always a continuous random variable Y. In fact, the data could equally well be a vector of random variables $Y_1, Y_2, \cdots, Y_n$, or a discrete random variable, or a vector of discrete random variables.

The notation in the summary always assumes a probability density for Y, e.g., $f_{Y|X}(y|x)$, $f_{Y;\theta}(y;\theta)$, $f_{Y|H}(y|H=H_1)$, or $f_{Y;H}(y;H=H_1)$. These could equally well be probability mass functions if Y is discrete, and joint densities or joint probability mass functions if Y is a vector of observations.

For Bayesian Hypothesis testing, the book allows for n hypotheses and uses the notation $\theta_1, \theta_2, \cdots, \theta_n$ for these hypotheses. The summary simplifies by assuming there are two hypotheses in the Bayesian case, and they are denoted H_0 and H_1 in the summary.

ESTIMATION

BAYESIAN ESTIMATION

GOAL: Given observation of random variable Y, create and evaluate estimator $\widehat{X}(Y)$ for random variable X.

GIVEN DATA: Prior distribution of unknown, $f_X(x)$, &
conditional density of observation, $f_{Y|X}(y|x)$.

ERROR METRIC: **Bias:** $E[(\widehat{X}(Y)-X)]$

Individual Trial: Squared Error = $e^2 \triangleq (\widehat{X}(y)-x)^2$

Average Behavior: Expected Squared Error $E[E^2]=E[(\widehat{X}(Y)-X)^2]$

METHODS: Conditional Mean, $\widehat{X}_{LMS}(y)\triangleq E[X\,|\,Y=y]$, is unbiased and minimizes the Expected Squared Error: $E[(\widehat{X}(Y)-X)^2]$.

Best Linear Estimator, $\widehat{X}_{LIN\ LMS}(y)=E[X]+\dfrac{\text{cov}(X,Y)}{\sigma_Y^2}(Y-E[Y])$,

has average error: $E[(\widehat{X}_{LIN\ LMS}-X)^2]=\sigma_X^2(1-\rho_{X,Y}^2)\ \geq\ E[E_{LMS}^2]$.

Maximum A Posteriori Estimator
MAP estimator, $\widehat{X}_{MAP}(y)=\arg\max_x \{f_{X|Y}(x|Y=y)$, has ***no*** special mean-squared error minimization properties.

HYPOTHESIS TESTING

BAYESIAN BINARY HYPOTHESIS TESTING

GOAL: Given observation of random variable Y, create and evaluate function $g(Y)=H_0$ or H_1 to judge whether hypothesis H_0 or H_1 is valid.

GIVEN DATA: Prior probabilities $p_0=P(H_0)$ and $p_1=P(H_1)$ &
conditional densities of observ. Y, $f_{Y|H}(y|H=H_0)$ and $f_{Y|H}(y|H=H_1)$.

ERROR METRIC:
Individual Trial: e =1 **(error!)** if $g(y)=H_0$, when H_1 is true or $g(y)=H_1$, when H_0 true. Otherwise **(no error!)** e=0.
Avg. Behavior: E[E] = P(error) = $P(g(Y)=H_1\,|\,H_0)p_0+P(g(Y)=H_0\,|\,H_1)p_1$

METHODS:
MAP Decision Rule minimizes the probability of error:

"If $P(H_1|Y=y)>P(H_0|Y=y)$, pick H_1.
If $P(H_1|Y=y)<P(H_0|Y=y)$, pick H_0."

Implementation:

If $f_{Y|H}(Y=y|H_1)p_1\ >\ f_{Y|H}(Y=y|H_0)p_0$, pick H_1,
If $f_{Y|H}(Y=y|H_1)p_1\ <\ f_{Y|H}(Y=y|H_0)p_0$, pick H_0.

CLASSICAL ESTIMATION

GOAL: Given observation of random variable Y, create and evaluate estimator $\hat{\theta}(Y)$ for deterministic but unknown constant θ.

GIVEN DATA: Distribution of observation Y for each value of θ, $f_{Y;\theta}(y;\theta)$.

ERROR METRIC: **Bias:** $E_\theta[\hat{\theta}(Y)-\theta]$ (may depend on θ)

(Not emphasized in text.)
Individual Trial: $e^2 \triangleq (\hat{\theta}(y)-\theta)^2$
Average Behavior: $E_\theta[(\hat{\theta}(y)-\theta)^2]$ (may depend on θ)

METHODS:

Maximum Likelihood: Given observation Y, the ML estimator is
$$\hat{\theta}_{ML}(y)\triangleq\arg\max_\theta f_{Y;\theta}(y;\theta).$$
Maximum likelihood may not be best. Other methods were considered.

Confidence Interval: Bounds $\hat{\theta}^-$ and $\hat{\theta}^+$ on probable error in estimator $\hat{\theta}(Y)$. For small α (that you choose):
$$P_\theta(\hat{\theta}^-(Y)\leq\theta\leq\hat{\theta}^-(Y))\ \geq 1-\alpha,\quad \text{for all } \theta.$$
t-distribution: Was used when $\text{var}_\theta(\hat{\theta}(Y))$ must be estimated.

Classical Least-Squares Regression (pp. 478-479, part (a)) was discussed only for the ***linear*** model with ***Gaussian*** additive noise using ***maximum likelihood*** estimation.

CLASSICAL BINARY HYPOTHESIS TESTING

GOAL: Given observation of random variable Y, create and evaluate function $g(Y)=H_0$ or H_1 to judge whether hypothesis H_0 or H_1 is valid.

GIVEN DATA: Densities of observation Y, $f_{Y;H}(y;H_0)$ when H_0 is true and $f_{Y;H}(y;H_1)$ when H_1 is true.

ERROR METRIC:
Individual Trial: A **false alarm** occurs if g(y) = H_1 and H_0 is true, & a **miss** occurs if g(y) = H_0 and H_1 is true.
Avg. Behavior: Probability P_{FA} of false alarm and probability P_M of a miss.

METHODS: Pick maximum value α for $P_{FA}:P_{FA}<\alpha$. Given this constraint, use **likelihood ratio** $L(y)=\dfrac{f_{Y;H}(y;H=H_1)}{f_{Y;H}(y;H=H_0)}$ to minimize P_M.

Likelihood Ratio Test: Find ξ such that $P(L(Y)>\xi;H_0)=\alpha$,

If L(y) > ξ, choose H_1, (i.e., g(y)=H_1),
if L(y) < ξ, choose H_0, (i.e., g(y)=H_0).

Significance testing: "Reject H_0 at significance level α" if $P_{FA}=\alpha$ and test rejects H_0.

The p-value of a rejection of H_0 for data Y=y: the lowest value of α for which $P_{FA}=\alpha$ and test rejects H_0

ESTIMATION

BAYESIAN ESTIMATION

GOAL: Given observation of random variable Y, create and evaluate estimator $\widehat{X}(Y)$ for random variable X.

GIVEN DATA: Prior distribution of unknown, $f_X(x)$, &
conditional density of observation, $f_{Y|X}(y|x)$.

ERROR METRIC: **Bias:** $E[(\widehat{X}(Y)-X)]$

Individual Trial: Squared Error = $e^2 \triangleq (\widehat{X}(y)-x)^2$

Average Behavior: Expected Squared Error $E[E^2]=E[(\widehat{X}(Y)-X)^2]$

METHODS: Conditional Mean, $\widehat{X}_{LMS}(y) \triangleq E[X \mid Y=y]$, is unbiased and minimizes the Expected Squared Error: $E[(\widehat{X}(Y)-X)^2]$.

Best Linear Estimator, $\widehat{X}_{LIN\ LMS}(y)=E[X]+\dfrac{\operatorname{cov}(X,Y)}{\sigma_Y^2}(Y-E[Y])$,

has average error: $E[(\widehat{X}_{LIN\ LMS}-X)^2]=\sigma_X^2(1-\rho_{X,Y}^2) \geq E[E_{LMS}^2]$.

Maximum A Posteriori Estimator

MAP estimator, $\widehat{X}_{MAP}(y)=\arg\max_x \{f_{X|Y}(x|Y=y)$, has ***no*** special mean-squared error minimization properties.

CLASSICAL ESTIMATION

GOAL: Given observation of random variable Y, create and evaluate estimator $\hat{\theta}(Y)$ for deterministic but unknown constant θ.

GIVEN DATA: Distribution of observation Y for each value of θ, $f_{Y;\theta}(y;\theta)$.

ERROR METRIC: **Bias:** $E_\theta[\hat{\theta}(Y)-\theta]$ (may depend on θ)

(Not emphasized in text.)
$$\begin{cases} \text{Individual Trial: } e^2 \triangleq (\hat{\theta}(y)-\theta)^2 \\ \text{Average Behavior: } E_\theta[(\hat{\theta}(y)-\theta)^2] \text{ (may depend on } \theta) \end{cases}$$

METHODS:

Maximum Likelihood: Given observation Y, the ML estimator is

$$\hat{\theta}_{ML}(y) \triangleq \arg\max_{\theta} f_{Y;\theta}(y;\theta).$$

Maximum likelihood may not be best. Other methods were considered.

Confidence Interval: Bounds $\hat{\theta}^-$ and $\hat{\theta}^+$ on probable error in estimator $\hat{\theta}(Y)$. For small α (that you choose):

$$P_\theta(\hat{\theta}^-(Y) \leq \theta \leq \hat{\theta}^-(Y)) \geq 1-\alpha, \quad \text{for all } \theta.$$

t-distribution: Was used when $\mathrm{var}_\theta(\hat{\theta}(Y))$ must be estimated.

Classical Least-Squares Regression (pp. 478-479, part (a)) was discussed only for the ***linear*** model with ***Gaussian*** additive noise using ***maximum likelihood*** estimation.

HYPOTHESIS TESTING

BAYESIAN BINARY HYPOTHESIS TESTING

GOAL: Given observation of random variable Y, create and evaluate function g(Y) = H_0 or H_1 to judge whether hypothesis H_0 or H_1 is valid.

GIVEN DATA: Prior probabilities $p_0 = P(H_0)$ and $p_1 = P(H_1)$ & conditional densities of observ. Y, $f_{Y|H}(y|H = H_0)$ and $f_{Y|H}(y|H = H_1)$.

ERROR METRIC:
Individual Trial: e =1 (**error!**) if $g(y) = H_0$, when H_1 is true or $g(y) = H_1$, when H_0 true. Otherwise (**no error!**) e=0.
Avg. Behavior: E[E] = P(error) = $P(g(Y) = H_1 | H_0) p_0 + P(g(Y) = H_0 | H_1) p_1$

METHODS:
MAP Decision Rule minimizes the probability of error:

"If $P(H_1|Y = y) > P(H_0|Y = y)$, pick H_1.
If $P(H_1|Y = y) < P(H_0|Y = y)$, pick H_0."

Implementation:

If $f_{Y|H}(Y = y|H_1)p_1 > f_{Y|H}(Y = y|H_0)p_0$, pick H_1,
If $f_{Y|H}(Y = y|H_1)p_1 < f_{Y|H}(Y = y|H_0)p_0$, pick H_0.

CLASSICAL BINARY HYPOTHESIS TESTING

GOAL: Given observation of random variable Y, create and evaluate function g(Y) $=H_0$ or H_1 to judge whether hypothesis H_0 or H_1 is valid.

GIVEN DATA: Densities of observation Y, $f_{Y;H}(y;H_0)$ when H_0 is true and $f_{Y;H}(y;H_1)$ when H_1 is true.

ERROR METRIC:

Individual Trial: A **false alarm** occurs if g(y) = H_1 and H_0 is true, & a **miss** occurs if g(y) = H_0 and H_1 is true.

Avg. Behavior: Probability P_{FA} of false alarm and probability P_M of a miss.

METHODS: Pick maximum value α for $P_{FA}: P_{FA} < \alpha$. Given this constraint, use **likelihood ratio** L(y) = $\dfrac{f_{Y;H}(y;H=H_1)}{f_{Y;H}(y;H=H_0)}$ to minimize P_M.

Likelihood Ratio Test: Find ξ such that $P(L(Y)>\xi;H_0)=\alpha$,

If L(y) > ξ, choose H_1, (i.e., g(y)=H_1),
if L(y) < ξ, choose H_0, (i.e., g(y)=H_0).

Significance testing: "Reject H_0 at significance level α" if $P_{FA}=\alpha$ and test rejects H_0.

The p-value of a rejection of H_0 for data Y=y: the lowest value of α for which $P_{FA}=\alpha$ and test rejects H_0

Stock Outcome and Market Trends

Stock	Sign of Change	August 29	November 30
IBM	- (-33.0%)	121.73	81.60
Microsoft	- (-25.9%)	27.29	20.22
Apple	- (-45.4%)	169.53	92.67
Intel	- (-44.6%)	23.97	13.29
Google	- (-36.8%)	463.29	292.69

Dow Jones Industrial	-23.5%	11,543.55	8829.04
Nasdaq	-35.1%	2,367.52	1535.57
S&P 500	-30.1%	1,282.83	896.24

Correlation coefficients between 3-month increments in price (since Google went public in August 2004)

	IBM	μ-SOFT	APPLE	INTEL	GOOGLE
IBM	1				
μ-SOFT	+0.72	1			
APPLE	+0.80	+0.86	1		
INTEL	+0.34	+0.45	+0.31	1	
GOOGLE	+0.61	+0.81	+0.94	+0.19	1

All correlation coefficients positive

To maximize one's chances of getting all five "direction of change" forecasts correct, a good approach would have been to bet they would all rise, or else all fall.

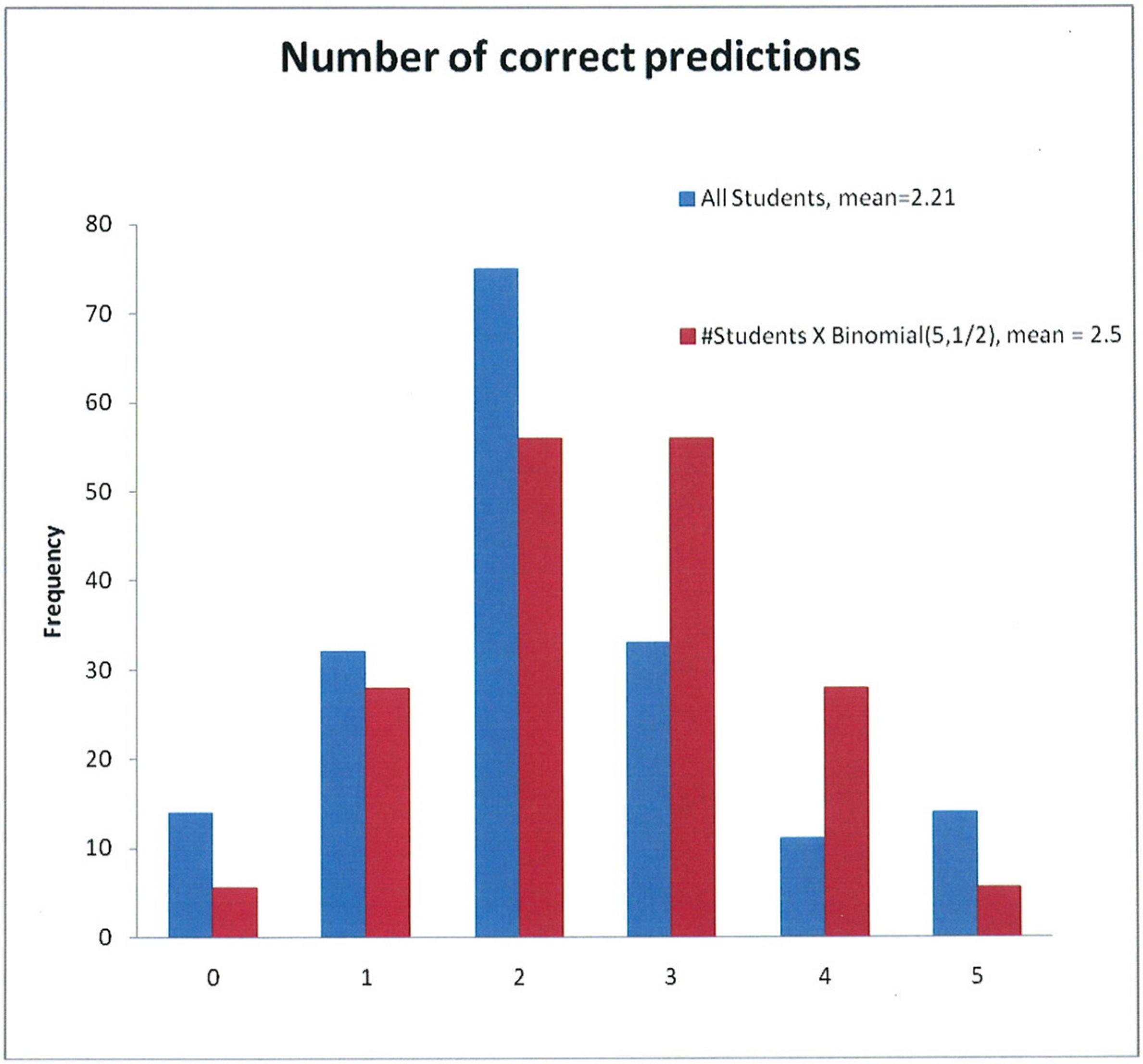
Stock Forecasting Problem

We have data from Problem Set #1 from

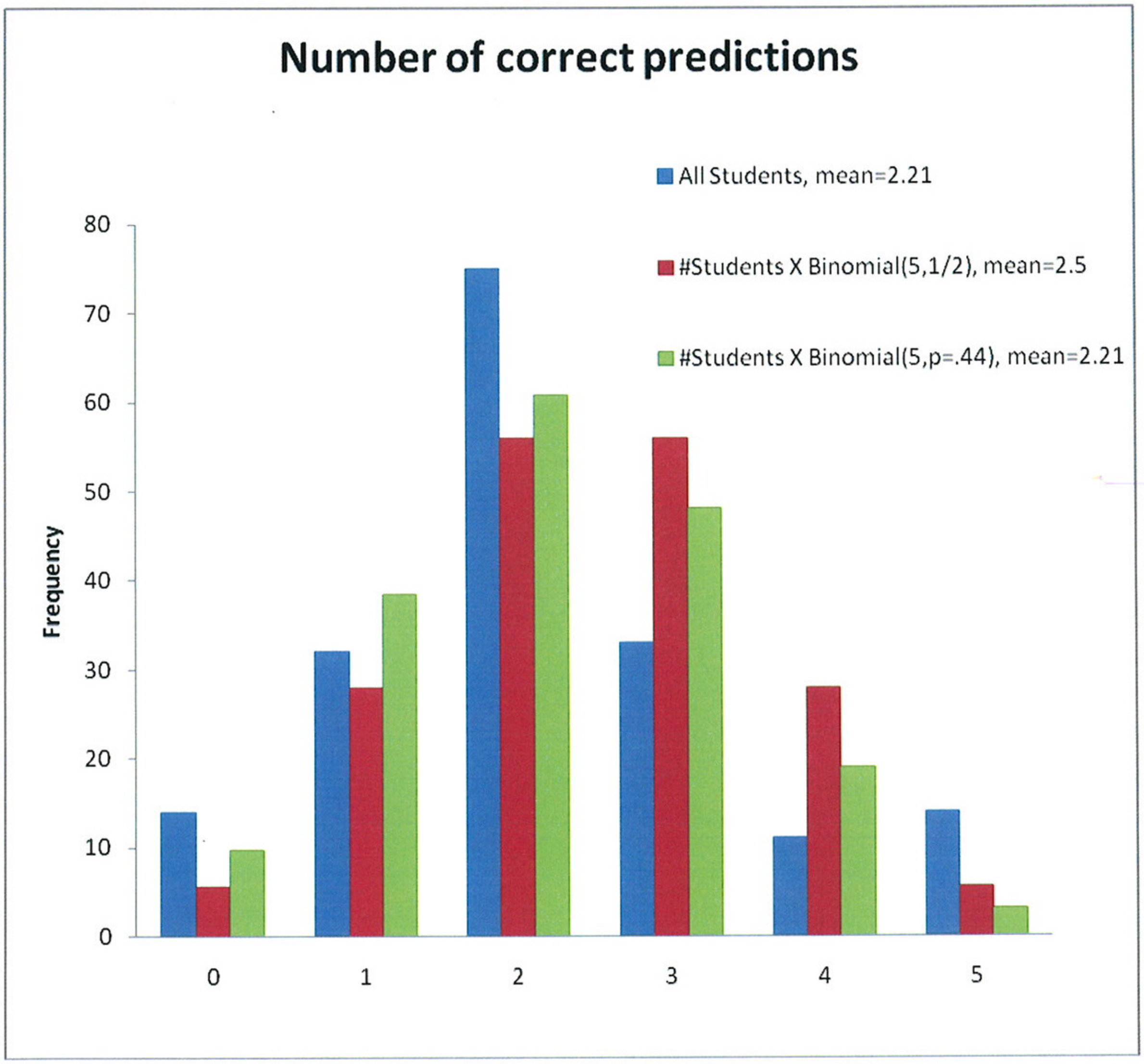
134 6.041 students, who correctly predicted the outcome, on average, for 43.1% of the stocks listed

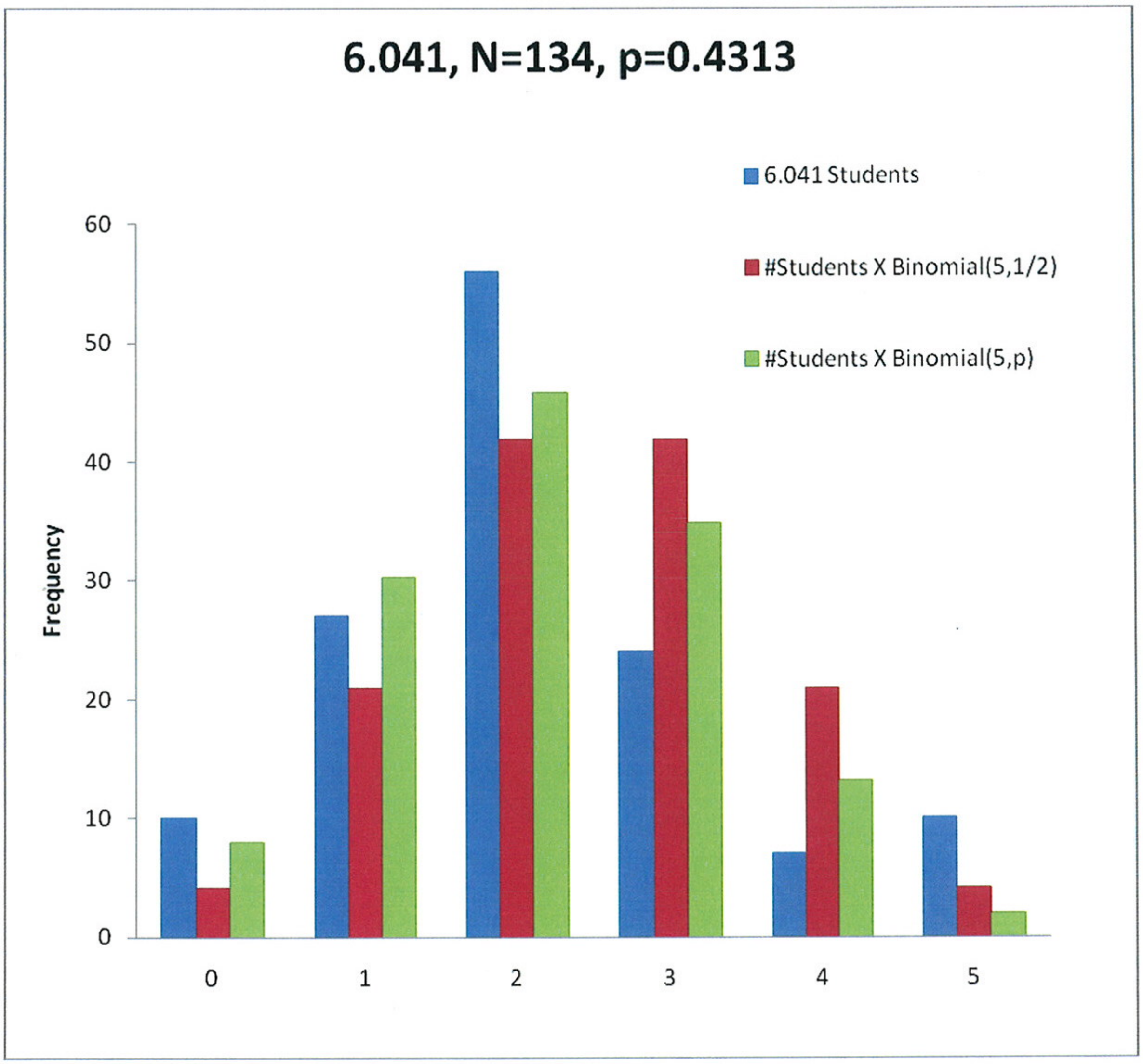
45 6.431 students, who correctly predicted the outcome, on average, for 47.1% of the stocks listed.

Model #1 All students get each stock's direction correctly with $p = 0.5$, independent of all other choices by that student or other students.

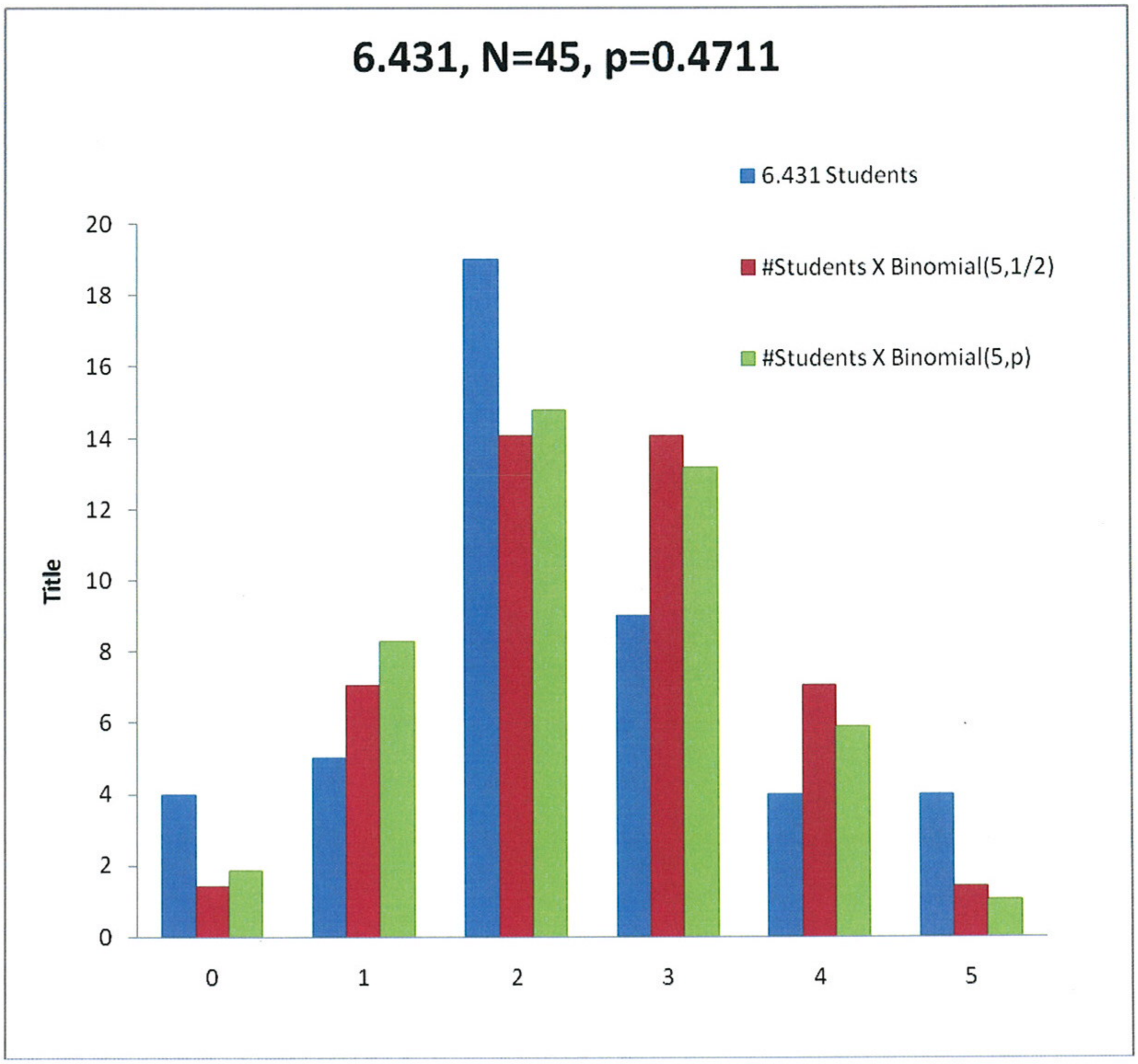


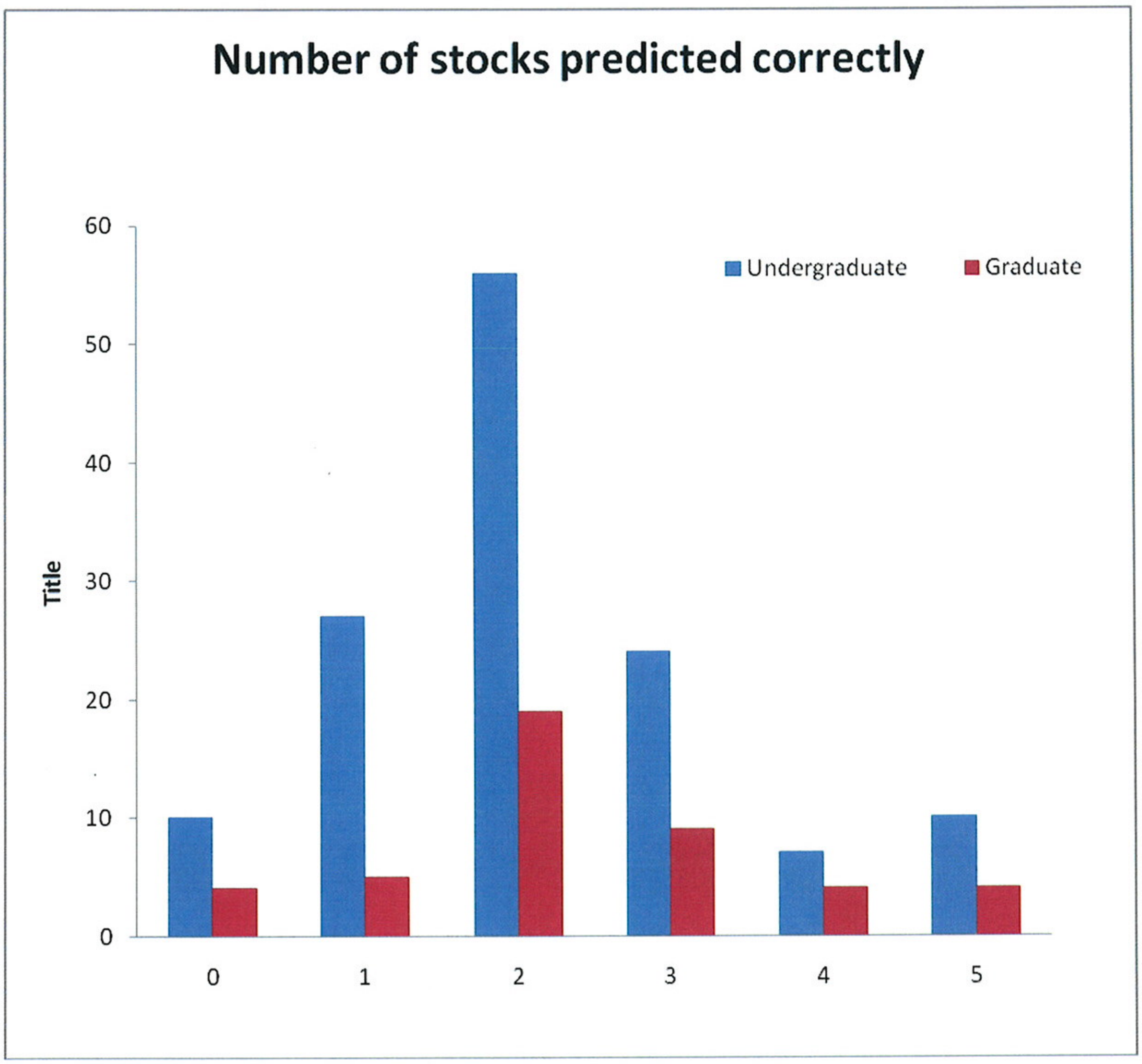
Number of correct predictions
All Students, mean=2.21
#Students X Binomial(5,1/2), mean = 2.5
Frequency
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5

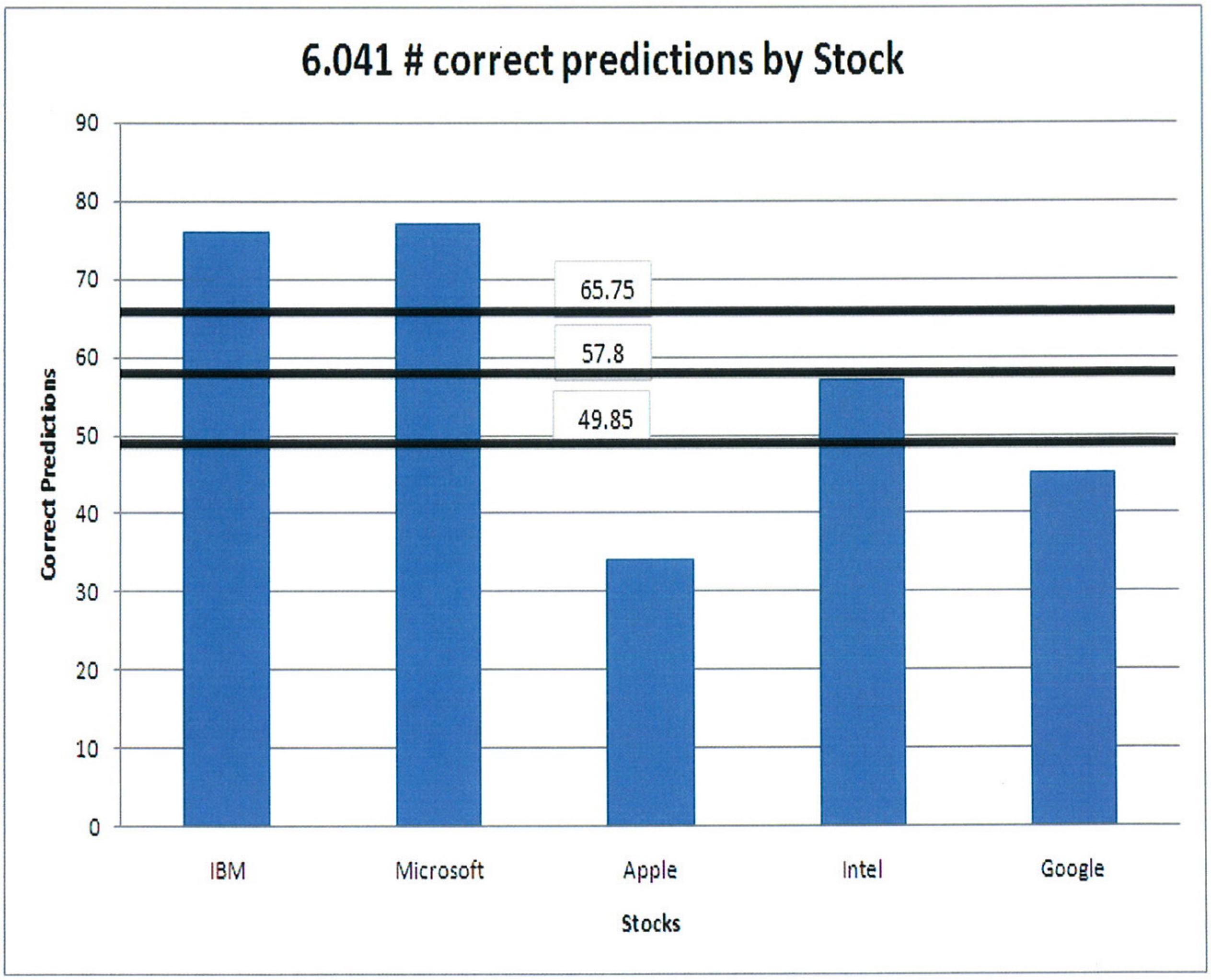

Number of correct predictions
All Students, mean=2.21
#Students X Binomial(5,1/2), mean=2.5
#Students X Binomial(5,p=.44), mean=2.21
Frequency
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5



6.041, N=134, p=0.4313
6.041 Students
#Students X Binomial(5,1/2)
#Students X Binomial(5,p)
Frequency
60
50
40
30
20
10
0
0
1
2
3
4
5


6.431, N=45, p=0.4711
6.431 Students
#Students X Binomial(5,1/2)
#Students X Binomial(5,p)
Title
0
2
4
6
8
10
12
14
16
18
20
0
1
2
3
4
5


Number of stocks predicted correctly
Undergraduate
Graduate
Title
60
50
40
30
20
10
0
0
1
2
3
4
5


6.041 # correct predictions by Stock
90
80
70
60
50
40
30
20
10
0
Correct Predictions
65.75
57.8
49.85
IBM
Microsoft
Apple
Intel
Google
Stocks

Stock Forecasting Winners

6.431

Albert Chang
HongSeok Cho
Hanbong Lee
Nidhi Santen

6.041

Hareem Ahmad
Douglas Albert
Eric Eisner
Eletha Flores
Kyle Fritz
Matthew Greyson
Benjamin Huan
Anjaney Kottapalli
Vrajesh Modi
Ryan Munoz