NAME: ___________________________

AP STATISTICS ASSIGNMENT:

SECTION 1: MONEYBALL

1. In 1991, what was the total payroll for:
   New York Yankees? Oakland Athletics?

2. The three players that the Oakland Athletics lost to free agency are:

3. What was Billy Beane (General Manager) told to do by the team owner of the Athletics?

4. What is the real problem that Billy and his team of scouts have to solve?

5. As a High School baseball star, why was Billy projected to be picked in the first round of the player draft?

6. According to Pete Brandt, "... the goal of people who run ball clubs is not to buy players, it is to ____________. In order to ____________, they need to ____________."

7. The equation on the whiteboard: $\frac{(Runs\ scored)^2}{(Runs\ scored)^2 + (Runs\ allowed)^2}$, is known as the 'Pythagorean Expectation Formula'. What does it allow Pete to project?

8. How many runs are needed, and how many runs can be allowed, in order for the team to reach this goal?

9. Pete explains to Billy that the code that is shown on his computer is an algorithm that analyzes players based on a number of statistics. It is all about getting everything down to _____________.
   What is it used for?

10. Pete knows that there are 25 potential players that must be out there and available for Oakland because ____________________________________.

11. Why is Chad Bradford (a pitcher), undervalued by other teams?

12. How is Billy going to replace players like Jason Giambi, Johnny Damon and Olmedo Saenz?

13. The three replacement players that Billy is interested in all have 'problems' but are cheap. Name the players:

    _________________________ _________________________ _________________________

    What do they all have in common?

14. Why did Billy not succeed at the major-league level as a player, despite having 'all the tools'?

15. The team that Billy has assembled is based on a statistical theory known colloquially as 'Moneyball'. Who invented this idea, and what was his real job at the time?

16. How many games back from the AL West lead are the Athletics as of May 23, 2002?

    How many games had they lost of their last 17 games?

17. Why does Billy prefer Scott Hatteburg at 1st base instead of Carlos Pena (despite Hatteburg's lack of experience and fielding skills in the position)?

18. How does Billy solve the problem of his manager (Art Howe) preferring to play Pena at 1st instead of Hatteburg?

19. Even though Jeremy Giambi was a player acquired using the 'Moneyball' formula, why do you think Billy felt the need to trade him?

20. What is 'The Streak"?

21. Billy states that if the A's win the last (championship) game of the season, they would have re-invented the game. What does he mean by this?

22. At the end of the 2002 season, the Athletics had exactly the same number of wins as the Yankees. How much did the Yankees pay on average for each win?

    How much did the A's pay on average for each win?

SECTION 2: PYTHAGOREAN EXPECTATION

Background:

Bill James, a pioneer in 'sabermetrics', developed an empirical formula to predict a baseball team's expected number of wins and losses called the *Pythagorean Expectation*:

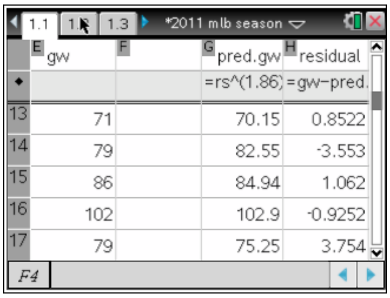
$$win\% = \frac{(Runs\ scored)^2}{(Runs\ scored)^2 + (Runs\ allowed)^2}$$

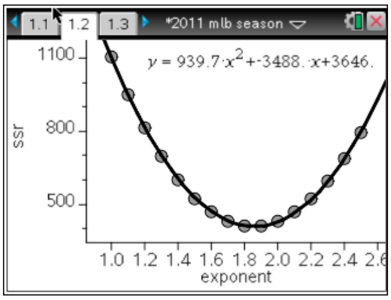
(It was named because the denominator bore similarities to the Pythagorean Theorem.)

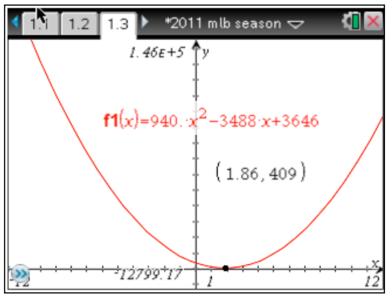
Thus the predicted number of games that a team will win in a season is:

#wins = $win\% \times number\ of\ games$ (typically 162 games in a season)

The exponent of '2' was later refined by Bill James to 1.83, which seemed to give a better prediction when compared with actual wins. A more recent analysis of the runs scored, runs allowed and actual number of wins for all thirty major league teams for the 2011 season seems to indicate that an exponent of 1.86 fits that particular season's data even better.







If we consider the regular season data for the St Louis Cardinals, the 2011 Champions:

runs scored = 762 runs allowed = 692 games played = 162 number of games won = 90

the Pythagorean expectation formula above (using an exponent of 1.86) predicts that the Cardinals should have won 88.24 games. Some Sabermatricians attribute the positive residual (actual - predicted games won) of 1.76 to *random chance*. Others argue that this reflects the ability of the manager in managing his players to win more games than expected.

Note: The Pythagorean Expectation has been successfully applied to other sports to predict win percentage:

Sport:	**Exponent** in 'Pythagorean Expectation Formula':
NBA Basketball	14
WNBA Basketball	12
NCAA Basketball	10
NFL (American Football)	2.37
NHL (Ice Hockey)	2

Try to calculate the exponent that works best at predicting *win*% in other sports:
Football (soccer), Rugby, Cricket, etc.

1. Choose a major league team (preferably one that is not winning it's league for now) and find the following statistics for their most recently completed season:
(try mlb.com or espn.com/mlb)

- Name of Team:
- Runs scored (RS)
- Runs allowed (RA)
- Games played (G)

$$win\% = \frac{(RS)^{1.86}}{(RS)^{1.86} + (RA)^{1.86}}$$

$$= \frac{(\_\_)^{1.86}}{(\_\_)^{1.86} + (\_\_)^{1.86}} \qquad =$$

Predicted number of wins for 2011 = (   ) x (   )

=

- Actual games won in 2011 was:

2. For the same team you chose above, find the most up-to-date data for this year's season.

- Name of Team:
- Runs scored to date (RS):
- Runs allowed to date (RA):
- Games played to date(G):

$$win\% = \frac{(RS)^{1.86}}{(RS)^{1.86} + (RA)^{1.86}}$$

$$= \frac{(\_\_)^{1.86}}{(\_\_)^{1.86} + (\_\_)^{1.86}}$$

=

Predicted number of wins for season to date = (   ) x (   )

=

- Actual number of games won to date this season:

How do the predicted number of wins compare with the actual number of wins to date?

Is the team performing above or below expectations?

3. Use the *win%* calculated above to predict the expected number of games that the team will win this season.

Predicted number of wins for this season = (   ) x (162)   =

Is this number realistic?

(Compare this to how the team performed last season and player personnel changes.)

SECTION 3: PLAYER TRADES

Background:

Extrapolated runs, abbreviated ***XR***, is a sabermetric that measures the number of runs a player contributes to his team. Developed by Jim Furtado, the beauty of this formula is that it is linear. If you add the *XR* values of each player on a team, you arrive at the total runs estimator for the team.

Although the extrapolated runs formula is linear, it is still somewhat complex because it includes over a dozen variables:

- *1B* - singles (can be calculated by subtracting *2B*, *3B* and *HR* from *H* (number of hits))
- *2B* - doubles
- *3B* - triples
- *HR* - home runs
- *HBP* - hit by a pitch
- *BB* - bases on balls (walks)
- *IBB* - intentional walks
- *SB* - stolen bases
- *CS* - caught stealing
- *AB* - at bats
- *SO* - Strike outs
- *GIDP* - ground into double play
- *SF* - sacrifice flyballs
- *SH* - sacrifice bunts

$$XR = 0.59(1B) + 0.81(2B) + 1.13(3B) + 1.53(HR) + 0.34(HBP) + 0.34(BB) - 0.09(IBB)$$
$$+ 0.18(SB) - 0.32(CS) - 0.09(AB) - 0.008(SO) - 0.37(GIDP) + 0.37(SF) + 0.04(SH)$$

Example: In 2011, Jose Bautista had the following individual statistics:

1B = 86	2B = 24	3B = 2	HR = 43	AB = 513	BB = 132	IBB = 24
HBP = 6	SB = 9	CS = 5	SH = 0	SF = 4	GIDP = 8	SO = 111

His extrapolated runs for that year was then *XR* = 134 (rounded to the nearest integer).
Jose Bautista actually scored 105 runs in the 2011 season for the Toronto Blue Jays.

1. Using the same team chosen for the previous exercise and statistics from the most recently completed season, add TWO positional players who you think would improve the total number of runs scored by your team.
   Calculate the *XR* for both players and add this to the team's total number of runs

   'Trade away' two other positional players on your team who you consider to have underperformed. Subtract the runs scored by these two players from your team's total.

   Use this 'extrapolated runs scored' for your team to predict the team's win% and number of games won for last season. (Assume that the runs allowed have stayed the same).

   Have your 'trades' made much of a difference?

Name of Team: Team Runs Scored: Team Runs Allowed:

Two Positional Players Added:

Name:		
1B		
2B		
3B		
HR		
HBP		
BB		
IBB		
SB		
CS		
AB		
SO		
GIDP		
SF		
SH		
Calculated RX:		

Names of two positional players to be traded away and why they were chosen to be traded away:

Combined runs scored by these two players:

New team runs scored: (Original RS plus total calculated RX minus RS of players traded away)

New teams win%: $= \frac{(\_\_\_)^{1.86}}{(\_\_\_)^{1.86} + (\_\_\_)^{1.86}} =$

Predicted number of wins for new team for the season = (   ) x (162) =

Have your 'trades' made much of a difference? (Why / Why not?)

Feedback to Student

Name:

Title of assignment: **MONEYBALL (Post-AP Exam Assignment)**

Date Due:

Section 1: Moneyball: **/4**

- All 22 questions answered accurately and comprehensively (4 mks)
- Most of the 22 questions answered accurately and comprehensively (3 mks)
- All or most of the questions answered, but details missing in many answers (2 mks)
- No detail to answers provided (1 mks)
- Incomplete (0 mks)

Section 2: Pythagorean Expectation: **/4**

- All questions answered, with clear details and calculations accurate (4 mks)
- All questions answered but details/calculations missing or inaccurate (2 mks)
- Incomplete (0 mks)

Section 3: Player Trades: **/4**

- All questions answered, with clear details and calculations accurate (4 mks)
- All questions answered but details/calculations missing or inaccurate (2 mks)
- Incomplete (0 mks)

Other Marks: (Including Extra Credit) **/2**

- Academic Honesty statement (giving credit to those who assisted you) (1 mk)
- URL or Screen shots of webpages used (1 mk)
- Evidence of sabermetric calculations beyond just those asked for (+2 mks extra)
- Application of sabermetrics to another sport / discipline (+3 mks extra)

Comment: **Total: 14 marks**