By Mark Hyman

Baseball: Money Can't Buy Me Wins

As Major League's post-season begins, it's clear that the biggest spenders on salaries aren't always the
contenders for the crown . . .

... With the regular season now a memory, MLB owners once again have proved that the most
successful teams on the field aren't necessarily the ones with the biggest bank vaults. Sure, franchises with
money to burn have an edge over those forced to live on a tight budget. The New York Yankees and
Boston Red Sox, ranked No. 1 and 2, respectively, in the payroll sweepstakes this year, could underscore
that point if either makes it to the World Series.

However, among the other teams in the playoffs, the San Diego Padres ranked 16th in salaries, the
Chicago White Sox 13th, the Houston Astros were 12th, and the Los Angeles Angels were 4th.

UP FROM DOWNTRODDEN. But don't count on money making the difference. To be a winner,
franchises also need savvy management, discerning talent evaluators, and when all else fails, a large dose
of luck. Examples are easy to find this season, but no easier than in recent seasons, when the last five
World Series have been won by four different teams.

In 2005, baseball had the usual rags-to-respectability stories. The formerly downtrodden Milwaukee
Brewers finished 81-81 with a $40 million payroll, smaller than all but three teams. The Oakland A's, with
a $55 million payroll were pennant contenders until the season's final week despite a payroll that ranked
22nd out of MLB's 30 franchises.

The most stunning example of a cheapo franchise playing like a champ was this year's Cleveland
Indians, the Cinderella story of 2005 until a last-minute swoon bumped the surprising Tribe from what
appeared to be a safe berth in the playoffs. . . .

Let's use a simple linear regression model $y = b_0 + b_1 x$ to examine how much
information the team payroll variable tells us about the team wins variable. Shown
below are the 2005 team payrolls (the $x$ variable, in $\text{millions}$) and the number of wins
during the 2005 season (the $y$ variable) for the 30 major league baseball teams. The
“Win Rank” in column 1 is from best to worst; the “Cost Per Win Rank” in column 4 is
from highest to lowest.

<table>
<thead>
<tr>
<th>Team</th>
<th>2005 Payroll (Win rank)</th>
<th>Wins (y)</th>
<th>Cost per win (Win rank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>StL (1)</td>
<td>100</td>
<td>92.107</td>
<td>0.921 (13)</td>
</tr>
<tr>
<td>Chi (A) (2)</td>
<td>99</td>
<td>75.178</td>
<td>0.759 (18)</td>
</tr>
<tr>
<td>NYY (3)</td>
<td>95</td>
<td>208.307</td>
<td>2.193 (1)</td>
</tr>
<tr>
<td>Bos (3)</td>
<td>95</td>
<td>123.505</td>
<td>1.300 (2)</td>
</tr>
<tr>
<td>LAA (3)</td>
<td>95</td>
<td>97.725</td>
<td>1.029 (9)</td>
</tr>
<tr>
<td>Clev (6)</td>
<td>93</td>
<td>41.503</td>
<td>0.446 (29)</td>
</tr>
<tr>
<td>Atl (7)</td>
<td>90</td>
<td>86.457</td>
<td>0.961 (12)</td>
</tr>
<tr>
<td>Hous (8)</td>
<td>89</td>
<td>76.779</td>
<td>0.863 (14)</td>
</tr>
<tr>
<td>Phil (9)</td>
<td>88</td>
<td>95.522</td>
<td>1.085 (8)</td>
</tr>
<tr>
<td>Oak (9)</td>
<td>88</td>
<td>55.426</td>
<td>0.630 (24)</td>
</tr>
<tr>
<td>NYM (11)</td>
<td>83</td>
<td>101.306</td>
<td>1.220 (4)</td>
</tr>
<tr>
<td>Minn (11)</td>
<td>83</td>
<td>56.186</td>
<td>0.677 (22)</td>
</tr>
<tr>
<td>FLA (11)</td>
<td>83</td>
<td>60.409</td>
<td>0.728 (19)</td>
</tr>
<tr>
<td>SD (14)</td>
<td>82</td>
<td>63.291</td>
<td>0.772 (17)</td>
</tr>
<tr>
<td>Mil (15)</td>
<td>81</td>
<td>39.935</td>
<td>0.493 (28)</td>
</tr>
</tbody>
</table>

We would like to fit a least squares line $y = b_0 + b_1 x$ to the above $(x, y)$ data.

1. In the $EXCEL$ output on the next page find the least squares estimate $b_1$ of the slope and
the least squares estimate $b_0$ of the intercept and write their values below. Write the
equation for the least squares prediction line

$$b_1 = 0.1562 \quad b_0 = 69.5896 \quad \hat{y} = 69.5896 + 0.1562x$$
2. What are the predicted number of wins and the corresponding residual for Baltimore? (no calculations needed; use output)

\[ \hat{y}_{\text{Baltimore}} = 81.13 \text{ wins} \]

\[ \text{residual}_{\text{Baltimore}} = -7.13 \]

3. What proportion of the variation in wins is explained by the linear relationship between wins and team payroll? From the Excel output below, \( r^2 = R^2 \text{ Square} = 0.244 \)

4. Which team overachieved the most? Chi (A)

Which team underachieved the most? KC

Residual = observed wins - predicted wins; biggest overachiever is team with biggest positive residual; biggest underachiever is team with most negative residual.

5. How many more wins could a team expect if the team payroll is increased by $10 million? 1.562 more wins. Slope value 0.1562 means that for each additional $1 million in payroll, a team will win 0.1562 more games. So an additional $10 million in payroll results in 10*0.1562 = 1.562 more wins.

6. If team payroll is 2 standard deviations above the mean payroll, on average the number of wins will be how many standard deviations above the mean number of wins? 0.988 standard deviations above the mean number of wins.

\[ b_1 = r \cdot \frac{s_y}{s_x} \]

means that when the \( x \)-variable is 2 \( x \)-standard deviations above the \( x \) mean, the \( y \)-variable is 2\( r = 2 \cdot 0.4940 = 0.988 \) \( y \)-standard deviations above the \( y \) mean. Note that

\[ r = \sqrt{r^2} = \sqrt{0.244} = 0.4940 \]

7. Fans of the Philadelphia Phillies are not pleased since their team did not make the playoffs. The mean number of wins (to the nearest integer) by the eight 2005 playoff teams is 93. If Philadelphia wants to win 93 games next season, by how much should they increase their team payroll? Philadelphia won 88 games, so they need to win 5 more games. Each additional $1 million in payroll results in 0.1562 more wins, so they need to spend an additional \( \frac{5}{0.1562} = 32.01 \) million dollars.

**EXCEL output**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Payroll</th>
<th>Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>69.5896</td>
<td>73.0524</td>
</tr>
<tr>
<td>Slope</td>
<td>0.1562</td>
<td>34.2477</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>829.8317</td>
<td>829.8317</td>
</tr>
<tr>
<td>Residual</td>
<td>28</td>
<td>2574.168</td>
<td>91.93458</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>3404</td>
<td></td>
</tr>
</tbody>
</table>

**Regression Statistics**

- Multiple R: 0.494
- R Square: 0.244
- Adjusted R Square: 0.217
- Standard Error: 9.588
- Observations: 30

**Payroll Residual Plot**

**Residual OUTPUT**

<table>
<thead>
<tr>
<th>Team</th>
<th>Observed Wins</th>
<th>Predicted Wins</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIL</td>
<td>100</td>
<td>83.98</td>
<td>16.02</td>
</tr>
<tr>
<td>Chi (A)</td>
<td>99</td>
<td>81.33</td>
<td>17.67</td>
</tr>
<tr>
<td>NY Y</td>
<td>95</td>
<td>102.13</td>
<td>-7.13</td>
</tr>
<tr>
<td>Bos</td>
<td>95</td>
<td>88.88</td>
<td>6.12</td>
</tr>
<tr>
<td>LAA</td>
<td>95</td>
<td>84.85</td>
<td>10.15</td>
</tr>
<tr>
<td>Clev</td>
<td>93</td>
<td>76.07</td>
<td>16.93</td>
</tr>
<tr>
<td>Atl</td>
<td>90</td>
<td>83.09</td>
<td>6.91</td>
</tr>
<tr>
<td>Hous</td>
<td>89</td>
<td>81.58</td>
<td>7.42</td>
</tr>
<tr>
<td>Phil</td>
<td>88</td>
<td>84.51</td>
<td>3.49</td>
</tr>
<tr>
<td>Oak</td>
<td>88</td>
<td>78.25</td>
<td>9.75</td>
</tr>
<tr>
<td>NY M</td>
<td>83</td>
<td>85.41</td>
<td>-2.41</td>
</tr>
<tr>
<td>Minn</td>
<td>83</td>
<td>78.37</td>
<td>4.63</td>
</tr>
<tr>
<td>FLA</td>
<td>83</td>
<td>79.03</td>
<td>3.97</td>
</tr>
<tr>
<td>SD</td>
<td>82</td>
<td>79.48</td>
<td>2.52</td>
</tr>
<tr>
<td>Mil</td>
<td>81</td>
<td>75.83</td>
<td>5.17</td>
</tr>
<tr>
<td>Wash</td>
<td>81</td>
<td>77.18</td>
<td>3.82</td>
</tr>
<tr>
<td>Tor</td>
<td>80</td>
<td>76.73</td>
<td>3.27</td>
</tr>
<tr>
<td>Tex</td>
<td>79</td>
<td>78.31</td>
<td>0.69</td>
</tr>
<tr>
<td>Chi (N)</td>
<td>79</td>
<td>83.18</td>
<td>-4.18</td>
</tr>
<tr>
<td>Ariz</td>
<td>77</td>
<td>79.33</td>
<td>-2.33</td>
</tr>
<tr>
<td>SF</td>
<td>75</td>
<td>83.68</td>
<td>-8.68</td>
</tr>
<tr>
<td>Balt</td>
<td>74</td>
<td>81.13</td>
<td>-7.13</td>
</tr>
<tr>
<td>Cin</td>
<td>73</td>
<td>79.26</td>
<td>-6.26</td>
</tr>
<tr>
<td>Det</td>
<td>71</td>
<td>80.38</td>
<td>-9.38</td>
</tr>
<tr>
<td>LAD</td>
<td>71</td>
<td>82.56</td>
<td>-11.56</td>
</tr>
<tr>
<td>St le</td>
<td>69</td>
<td>83.30</td>
<td>-14.30</td>
</tr>
<tr>
<td>TB</td>
<td>67</td>
<td>74.18</td>
<td>-7.18</td>
</tr>
<tr>
<td>Col</td>
<td>67</td>
<td>77.11</td>
<td>-10.11</td>
</tr>
<tr>
<td>Pitt</td>
<td>67</td>
<td>75.55</td>
<td>-8.55</td>
</tr>
<tr>
<td>KC</td>
<td>56</td>
<td>75.35</td>
<td>-19.35</td>
</tr>
</tbody>
</table>

Note from above scatterplot and least squares line that teams who fired their managers immediately after the season (the squares) are all below the least squares line.