War, Enmity, and Statistical Tables
When Is the Lead Safe in a College Basketball Game?

Brian Schmotzer
Take the number of points one team is ahead.
Subtract three.
Add a half-point if the team that is ahead has the ball, and subtract a half-point if the other team has the ball (numbers less than zero become zero).
Square that.
If the result is greater than the number of seconds left in the game, the lead is safe.

Bill James’s rule
Figure 1. Bill James suggested boundaries between safe and not-safe leads based on margin of lead and time remaining. The end of the game is on the left with no time remaining. The lower line represents the boundary when the currently leading team has possession of the ball, while the upper line represents the boundary when the currently trailing team has the ball.
Statistical analysis:

Collected data from 2007-2008 Division 1 college games (1357 games out of 5540 total played)

For each combination of margin and time remaining, compute “proportion of safe leads” (Fig 3)

Estimate a 95% safety contour and a “LOWESS smooth” of that contour (Fig 6)

Repeat for various other margins (100%, 90%, 80%, 70%, 60% — Fig 7)

Finally represent a simplified version of the rule to compare with James’s rule (Fig 8)
Figure 3. Proportion of safe leads at each point of margin versus time remaining for data from 1,355 games. Lighter colors represent safer leads.
Figure 6. Duplicate of Figure 3, but with a 95% safety contour overlaid (solid line) and a LOWESS-smoothed version of the contour (dashed line).
Figure 7. Several LOWESS-smoothed safety contours. From the top to the bottom, the safety levels are 100%, 95%, 90%, 80%, 70%, and 60%.
Figure 8. Plot of safety boundaries comparing James’ suggestions (solid), LOWESS-smoothed contours of 100% and 95% safety (dashed), and simplified algorithmic boundaries of 100% and 95% safety (heavy solid)
Review of Last Class: Testing a Proportion

Suppose the data are a sample proportion \( \hat{p} \) from the sample of size \( n \) where the true population proportion is an unknown quantity \( p \). The null hypothesis is

\[
H_0 : p = p_0
\]

where \( p_0 \) is some given proportion.

The alternative hypothesis is almost always one of

\[
H_a : p > p_0, \quad \text{or} \quad (1)
\]
\[
H_a : p < p_0, \quad \text{or} \quad (2)
\]
\[
H_a : p \neq p_0. \quad \text{or} \quad (3)
\]

where the choice among (1)–(3) depends on the context of the problem.
Test statistic

\[ z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \]  \hspace{1cm} (4)

If \( H_0 \) is true, then \( z \) has a standard normal distribution with mean 0 and standard deviation 1.
Side comment:

The formula for the standard error is

\[ \sqrt{\frac{p_0(1 - p_0)}{n}}. \]

This is different from the confidence interval calculation where it’s

\[ \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}. \]

The difference is that in a hypothesis testing problem, we already have a specific value \( p_0 \) that we’re testing against. Therefore, it makes sense to use that as the basis for calculating a standard error.
Computing the P-value

The calculation of the P-value depends on the alternative hypothesis.

For (1), compute \( \Pr\{Z > z\} \) where \( Z \) is a standard normal random variable and \( z \) is the number computed in (4).

For (2), compute \( \Pr\{Z < z\} \).

For (3), compute \( \Pr\{Z < -|z|\} + \Pr\{Z > |z|\} \) where \( |z| \) means the magnitude of \( z \) (ignoring the sign). In practice, this almost always results in twice the P-value computed for (1) or (2).
Another side comment:

In the example with $n = 31, p_0 = 0.067, X = 6$, the usual condition $np_0 > 15, n(1 - p_0) > 15$ is not satisfied (because $np_0 = 2.077$). We could do an exact calculation, as follows.

The problem is to calculate $\Pr\{X \geq 6\}$ when $X$ has a binomial distribution with $n = 31, p = .067$. This is $1 - \Pr\{X \leq 5\}$.

1. In Excel, go to Formulas $\rightarrow$ More Functions $\rightarrow$ Statistical $\rightarrow$ BINOMDIST

2. Enter “Number” =5, “Trials” =31, “Probability” =.067, “Cumulative” =TRUE.

3. The answer comes up as 0.9844. Therefore, $1 - .9844 = .0156$ is the required (one-sided) P-value.
Interpreting the P-value

1. It's usual to interpret a P-value less than .05 as “significant”. However if it’s very important to make sure we don’t get a spurious result, we may adopt a more stringent criterion, e.g. $P < .01$.

2. In presenting research results, a common practice is just to ignore results for which $P > .05$, but when $P < .05$, state the exact P-value. That way, the reader can judge for herself just how strong the result is. This is what they did in the paper about skin cancer in marathon runners.
Testing the mean of a quantitative variable

An example: The mean height of male students at the University of Georgia (page 63) was 71 inches.

In a sample of 10 male students in this class, the mean $\bar{x}$ was 71.42 inches and the standard deviation $s$ was 2.93 inches.

*Is this a statistically significant difference?*

Let’s set this up as a formal hypothesis test.
Assume that the students in this class are a random sample of the population of all male students at UNC, and let $\mu$ be the mean height of that population. Also write $\mu_0 = 71$. The null hypothesis is

$$H_0 : \mu = \mu_0.$$  

For the alternative hypothesis, we again have three possibilities analogous to (1)–(3), namely

$$H_a : \mu > \mu_0,$$  

$$H_a : \mu < \mu_0,$$  

$$H_a : \mu \neq \mu_0.$$ 

In this case, there is no a priori reason to think that the students at UNC are either taller or shorter than the students at the University of Georgia, so it is most logical to choose (7) as our alternative hypothesis.
We are given $\bar{x} = 71.44$, and its standard error is $\frac{s}{\sqrt{n}} = \frac{2.93}{\sqrt{10}} = 0.9265$. The $t$ statistic is 

$$t = \frac{\bar{x} - \mu}{S.E.} = \frac{71.44 - 71}{0.9265} = 0.475.$$ 

If $H_0$ is true, then $t$ has a $t$ distribution with degrees of freedom $df = n - 1 = 9$, by the same theory as used for confidence intervals in Chapter 6. Therefore the P-value is the probability that a random variable with the $t_9$ distribution is greater than 0.475, multiplied by 2 (because it is a two-tailed test — therefore $t < -0.475$ has the same meaning as $t > 0.475$).

At this point, I'm going to differ a little from what the book tells you to do (on page 391). They recommend the use of the Minitab software package to find the exact P-value. I'm going to show you how to get the approximate value from the table on page A3.
<table>
<thead>
<tr>
<th>df</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
<th>99.8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td>63.657</td>
<td>318.309</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.920</td>
<td>4.303</td>
<td>6.965</td>
<td>9.925</td>
<td>22.327</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td>5.841</td>
<td>10.215</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>7.173</td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>5.893</td>
</tr>
<tr>
<td>6</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>3.143</td>
<td>3.707</td>
<td>5.208</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.785</td>
</tr>
<tr>
<td>8</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>4.501</td>
</tr>
<tr>
<td>9</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
<td>4.297</td>
</tr>
<tr>
<td>10</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.169</td>
<td>4.144</td>
</tr>
<tr>
<td>11</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.718</td>
<td>3.106</td>
<td>4.025</td>
</tr>
<tr>
<td>12</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.681</td>
<td>3.055</td>
<td>3.930</td>
</tr>
<tr>
<td>13</td>
<td>1.350</td>
<td>1.771</td>
<td>2.160</td>
<td>2.650</td>
<td>3.012</td>
<td>3.852</td>
</tr>
<tr>
<td>14</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.624</td>
<td>2.977</td>
<td>3.787</td>
</tr>
<tr>
<td>15</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.602</td>
<td>2.947</td>
<td>3.733</td>
</tr>
<tr>
<td>16</td>
<td>1.337</td>
<td>1.746</td>
<td>2.120</td>
<td>2.583</td>
<td>2.921</td>
<td>3.686</td>
</tr>
<tr>
<td>17</td>
<td>1.333</td>
<td>1.740</td>
<td>2.110</td>
<td>2.567</td>
<td>2.898</td>
<td>3.646</td>
</tr>
<tr>
<td>18</td>
<td>1.330</td>
<td>1.734</td>
<td>2.101</td>
<td>2.552</td>
<td>2.878</td>
<td>3.610</td>
</tr>
<tr>
<td>19</td>
<td>1.328</td>
<td>1.729</td>
<td>2.093</td>
<td>2.539</td>
<td>2.861</td>
<td>3.579</td>
</tr>
<tr>
<td>20</td>
<td>1.325</td>
<td>1.725</td>
<td>2.086</td>
<td>2.528</td>
<td>2.845</td>
<td>3.552</td>
</tr>
<tr>
<td>21</td>
<td>1.323</td>
<td>1.721</td>
<td>2.080</td>
<td>2.518</td>
<td>2.831</td>
<td>3.527</td>
</tr>
<tr>
<td>22</td>
<td>1.321</td>
<td>1.717</td>
<td>2.074</td>
<td>2.508</td>
<td>2.819</td>
<td>3.505</td>
</tr>
<tr>
<td>23</td>
<td>1.319</td>
<td>1.714</td>
<td>2.069</td>
<td>2.500</td>
<td>2.807</td>
<td>3.485</td>
</tr>
<tr>
<td>24</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.492</td>
<td>2.797</td>
<td>3.467</td>
</tr>
<tr>
<td>25</td>
<td>1.316</td>
<td>1.708</td>
<td>2.060</td>
<td>2.485</td>
<td>2.787</td>
<td>3.450</td>
</tr>
<tr>
<td>26</td>
<td>1.315</td>
<td>1.706</td>
<td>2.056</td>
<td>2.479</td>
<td>2.779</td>
<td>3.435</td>
</tr>
<tr>
<td>27</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.473</td>
<td>2.771</td>
<td>3.421</td>
</tr>
<tr>
<td>28</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.467</td>
<td>2.763</td>
<td>3.408</td>
</tr>
<tr>
<td>29</td>
<td>1.311</td>
<td>1.699</td>
<td>2.045</td>
<td>2.462</td>
<td>2.756</td>
<td>3.396</td>
</tr>
<tr>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.457</td>
<td>2.750</td>
<td>3.385</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.423</td>
<td>2.704</td>
<td>3.307</td>
</tr>
<tr>
<td>50</td>
<td>1.299</td>
<td>1.676</td>
<td>2.009</td>
<td>2.403</td>
<td>2.678</td>
<td>3.261</td>
</tr>
<tr>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2.000</td>
<td>2.390</td>
<td>2.660</td>
<td>3.232</td>
</tr>
<tr>
<td>80</td>
<td>1.292</td>
<td>1.664</td>
<td>1.990</td>
<td>2.374</td>
<td>2.639</td>
<td>3.195</td>
</tr>
<tr>
<td>100</td>
<td>1.290</td>
<td>1.660</td>
<td>1.984</td>
<td>2.364</td>
<td>2.626</td>
<td>3.174</td>
</tr>
<tr>
<td>Inf</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.327</td>
<td>2.576</td>
<td>3.091</td>
</tr>
</tbody>
</table>
In this table, corresponding to $df = 9$, we find $t$ values of 1.383, 1.833, 2.262, 2.821, 3.250, 4.297 corresponding to confidence levels of 80%, 90%, 95%, 98%, 99%, 99.8%.

The two-sided P-value is one minus the confidence level, expressed as a decimal fraction.

Thus, only a $t$ value bigger than 2.262 would be considered “significant” with a P-value of .05 or less.

Since in this case $t = 0.475$, this is not significant. We accept the null hypothesis $\mu = 71$. The students at the University of North Carolina are not significantly different from those at the University of Georgia (as we would have expected).
An alternative calculation

Suppose, however, the mean of the UNC students was 73.5 inches. In this case we would calculate $t = \frac{73.5 - 71}{0.9265} = 2.698$. In the table, this comes out between confidence levels 95% and 98% — in other words, the 2-sided P-value is between $1 - 0.98 = 0.02$ and $1 - 0.95 = 0.05$.

If we had set it up as a 1-sided test in the first place, the P-value would be between 0.01 and 0.025 (half the 2-sided P-value).

Either way, the conclusion would be that there is statistically significant evidence that the UNC students are taller than the Georgia students, though even then, the evidence is not overwhelmingly strong. We would probably not attach too much importance to the conclusion unless it were replicated by further studies.
Suppose, however, the study were not of all male students at UNC and UGa, but specifically of male basketball players. Then, it would not be at all surprising that the UNC players were taller (because the best players want to come to UNC...)

This illustrates the point that the context of a study often plays a role in interpreting its outcome — it is more likely to be accepted if it accords with prior expectation. Even then, however, the P-value calculation provides a “reality check” in helping us to identify spurious results.
Footnote to this section. Even though we don’t have Minitab at our disposal, it is still possible to mimic the book’s method of calculation using Excel.

1. In Excel, go to Formulas → More Functions → Statistical → TDIST

2. Enter “X” =2.698, “Deg-freedom” =9, “Tails” — either 1 or 2 for a 1-sided or 2-sided test.

3. The 1-sided P-value is 0.0122 and the 2-sided P-value is 0.0245.