Lecture : Hypothesis Testing and Midterm Review (POLS 3316 - Statistics for Political Scientists)

Tom Hanna

##Today

+Part 1: What is hypothesis testing? (No formulas/No R) +Part 2: Midterm Review and Practice

Part 1: Hypothesis Testing

- Friday I told you we needed these things to get to hypothesis testing:

```
+ The *68-95-99.7 Rule as a normal distribution probability shorthand
+ To understand samples and populations
+ The Central Limit Theorem to tie things to a normal distribution
+ The Law of Large Numbers to tie the sample to the population
+ Two new, related statistics: standard error and Z-score
+ The goal was **Hypothesis Testing**
```

So what is hypothesis testing?

```
+ We have a theory and we want to see if it's valid
+ We formulate hypotheses (pl.) that would be true if the theory was valid
+ We try to disprove or *falsify* them
+ How? By statistically testing
```

##For each hypothesis we test, there are actually two: The null hypothesis and the alternative hypothesis

```
+ The Alternative Hypothesis: This is actually the hypothesis that agrees with our theory. W
+ The default assumption is that our theory is that there is no effect: The Null Hypothesis
+ We can’t prove our hypothesis, but we can get sufficient evidence to **reject the null hyp
+ How? Data and Statistics
```

How do we reject or confirm the null hypothesis?

```
+ We conduct experiments or make observations to gather data
```

https://www.youtube.com/watch?v=3P8shnNEXb4&t=18s

```
+ We compare the results to what we would expect if they were due to random chance
+ If they match what we would expect from random chance: The null hypothesis is retained, th
+ If the statistics show they aren't due to random chance: The null hypothesis is rejected,
+ How do we do that? Our knowledge of probability and statistics
```

We use test statistics: Z-Scores, Chi-square, ANOVA, t-tests, and many others to determine the likelihood that the effect we are seeing is due to random chance.

I told you Z-scores serve dual purposes. One of them is to tie serve as a hypothesis test for large, normally distributed results. To give us probability that place us on this curve:

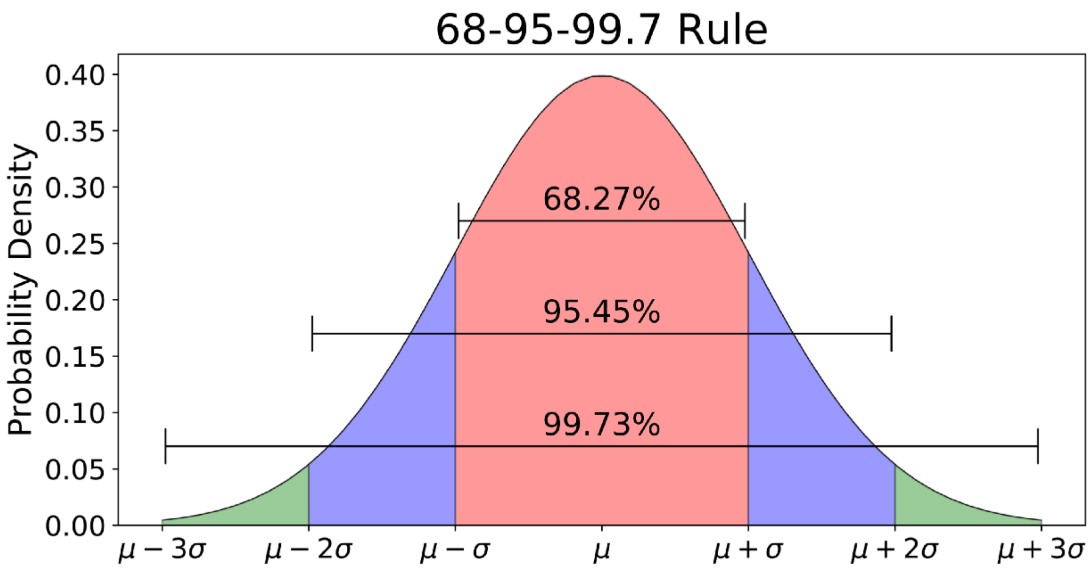

Figure 1: 68-95-99.7 rule

Now that's actually the easy one. The other tests we use apply where the normal distribution won't help us, at least without help. And that's where we will pick up after we finish with Z-scores on Wednesday.

Part 2: Midterm Review

```
\section*{Worksheet for Section 1}



\begin{table}[H]

\resizebox{\textwidth}{!}{
\begin{tabular}{|c|c|c|c|}
\hline
Team & Wins percentage &  Deviations from the Mean  & Squared Deviations from the Mean \\
(Observation)    & (Value)  &  &  \\
\hline
 & $x_i$  & $(x_i - \bar{x})$  & $(x_i - \bar{x})^2$ \\
\hline
&  &  &  \\
\hline
&  &  &  \\
\hline
&  &  &  \\
\hline
```

```
& & & \\
\hline
& & & \\
\hline
& & & \\
\hline
& & & \\
\hline
& & & \\
\hline
& & & \\
\hline
& & & \\
\hline
\hline
Item & Symbol & \multicolumn{2}{c|}{\textit{Value:}}  \\
\hline
Sum & $\Sigma x_i$ & \multicolumn{2}{c|}{}  \\
    \hline
Observations & \textit{n} & \multicolumn{2}{c|}{}  \\
        \hline
Observations - 1 & \textit{n - 1} & \multicolumn{2}{c|}{}  \\
        \hline
Mean & $\bar{x} = \frac{\Sigma x_i}{n}$ & \multicolumn{2}{c|}{}  \\
        \hline
Sum Squared Deviations  & $\Sigma (x_i - \bar{x})^2$ & \multicolumn{2}{c|}{}  \\
        \hline
Sample Variance & $\frac{\Sigma (x_i - \bar{x})^2}{n-1}$ & \multicolumn{2}{c|}{}  \\
        \hline
Sample Standard Deviation & $\sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$ & \multicolumn{2}{c|}{}  \\
            \hline
Population Variance & $\frac{\Sigma (x_i - \bar{x})^2}{n}$ & \multicolumn{2}{c|}{}  \\
        \hline
Population Standard Deviation &   $\sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$ & \multicolumn{2}{c|}{}  \
            \hline
Median & Middle or $\frac{lower\:middle + higher\:middle}{2}$ & \multicolumn{2}{c|}{}  \\
                \hline
Mode &   Just count & \multicolumn{2}{c|}{}  \\
                    \hline
\end{tabular}
}


\end{table}
```

```
\section*{Formulas}

\begin{table}[H]

\resizebox{\textwidth}{!}{
\begin{tabular}{|c|c|c|c|}
\hline


Mean & $\bar{x} = \frac{\Sigma x_i}{n}$ & \multicolumn{2}{c|}{}  \\
        \hline
Sum Squared Deviations  & $\Sigma (x_i - \bar{x})^2$ & \multicolumn{2}{c|}{}  \\
        \hline
Sample Variance & $\frac{\Sigma (x_i - \bar{x})^2}{n-1}$ & \multicolumn{2}{c|}{}  \\
        \hline
Sample Standard Deviation & $\sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$ & \multicolumn{2}{c|}{}  \\
            \hline
Population Variance & $\frac{\Sigma (x_i - \bar{x})^2}{n}$ & \multicolumn{2}{c|}{}  \\
        \hline
Population Standard Deviation &   $\sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}$ & \multicolumn{2}{c|}{}  \
            \hline
Median & Middle or $\frac{lower\:middle + higher\:middle}{2}$ & \multicolumn{2}{c|}{}  \\
                \hline
Mode &   Just count & \multicolumn{2}{c|}{}  \\
                    \hline
      Probability Formulas:  &    \multicolumn{3}{c|}{}  \\
                            \hline
      P(A)  & P(A) = $\frac{favorable outcomes}{possible outcomes}$    & \multicolumn{2}{c|}{}  \\
                            \hline
  Mutually exclusive  &    \multicolumn{3}{c|}{}  \\
                        \hline
      P(A $\cap$ B)  &  0  & \multicolumn{2}{c|}{}  \\
                            \hline
  P(A $\cup$ B)  & P(A $\cup$ B) = P(A) + P(B)   & \multicolumn{2}{c|}{}  \\
                        \hline
  Non-Mutually exclusive  &    \multicolumn{3}{c|}{}  \\
                        \hline
  P(A $\cup$ B)  &  P(A $\cup$ B) = P(A) + P(B) - P(A $\cap$ B)  & \multicolumn{2}{c|}{}  \\
                            \hline

\end{tabular}
}


\end{table}
```