

Group Names: _____

**Georgia-Florida Football Rivalry Problem Set
P-Hacking and Pitfalls of P-Values**

The data below show the result of each game in the Georgia-Florida football rivalry from 1904 to 2015. Use this information to proceed through this problem set.

Go
Dawgs!



Year	Winner	Year	Winner	Year	Winner
1904	Georgia	1954	Georgia	1986	Florida
1915	Georgia	1955	Florida	1987	Georgia
1916	Georgia	1956	Florida	1988	Georgia
1919	Georgia	1957	Florida	1989	Georgia
1920	Georgia	1958	Florida	1990	Florida
1926	Georgia	1959	Georgia	1991	Florida
1927	Georgia	1960	Florida	1992	Florida
1928	Florida	1961	Florida	1993	Florida
1929	Florida	1962	Florida	1994	Florida
1930	TIE	1963	Florida	1995	Florida
1931	Georgia	1964	Georgia	1996	Florida
1932	Georgia	1965	Florida	1997	Georgia
1933	Georgia	1966	Georgia	1998	Florida
1934	Georgia	1967	Florida	1999	Florida
1935	Georgia	1968	Georgia	2000	Florida
1936	Georgia	1969	TIE	2001	Florida
1937	Florida	1970	Florida	2002	Florida
1938	Georgia	1971	Georgia	2003	Florida
1939	Georgia	1972	Georgia	2004	Georgia
1940	Florida	1973	Florida	2005	Florida
1941	Georgia	1974	Georgia	2006	Florida
1942	Georgia	1975	Georgia	2007	Georgia
1944	Georgia	1976	Georgia	2008	Florida
1945	Georgia	1977	Florida	2009	Florida
1946	Georgia	1978	Georgia	2010	Florida
1947	Georgia	1979	Georgia	2011	Georgia
1948	Georgia	1980	Georgia	2012	Georgia
1949	Florida	1981	Georgia	2013	Georgia
1950	Georgia	1982	Georgia	2014	Florida
1951	Georgia	1983	Georgia	2015	Florida
1952	Florida	1984	Florida		
1953	Florida	1985	Georgia		



It's Great
to Be a
Florida
Gator!

Group Names: _____

1. Start by examining the results of the first 30 games of the rivalry (from 1904 to 1951). Which team appears to be superior?

Superior team: _____

2. Suppose we want to examine if there is statistically significant evidence showing one team is superior to the other. Give the appropriate null and alternative hypotheses. (Hint: For the Null Hypothesis, what would the proportion of wins be for Georgia if two teams were evenly matched. Use a two-tailed one-proportion z-test.)

Let p = The proportion of games won by UGA

H_0 : _____

H_a : _____

3. Calculate the proportion of games won by Georgia in this 30-year period. This will be the sample proportion.

\hat{p} : _____

4. Assume the conditions for a 1-proportion Z-test are met. Take a standard deck of playing cards and shuffle several times. Let the red cards represent Georgia wins, and let the black cards represent Georgia loses. Deal out 30 cards and count the number of Georgia wins (red cards). Calculate the proportion of Georgia wins. Repeat until you have a total of 10 proportions. Record those below.

5. Place your 10 proportions along with your classmates' on the dotplot on the board in the classroom. What proportion of dots on the dotplot were at the sample proportion (\hat{p}) or higher? Because this is a two-tailed test, double this proportion and this will be our simulated p-value.

6. Interpret the simulated p-value from above.

Group Names: _____

7. Use your p-value to make a conclusion about the Georgia-Florida rivalry based off of this sample of the first 30 games.

8. One or more of the conditions for a 1-Proportion Z test were not met. Explain.

9. Do you feel that the conclusion from number 7 is valid? Why or why not?

10. Is it possible to get a random sample like this one where Georgia wins 24 of 30 games?

11. Now examine the results of the last 30 games of the rivalry (from 1986 to 2015). Which team appears to be superior?

Superior team: _____

12. Again, we want to see if there is statistically significant evidence showing one team is superior to the other. Give the appropriate null and alternative hypotheses. (Hint: For the Null Hypothesis, what would the proportion of wins be for Georgia if two teams were evenly matched. Use a two-tailed one-proportion z-test.)

Let p = The proportion of games won by Georgia

H_0 : _____

H_a : _____

Group Names: _____

13. Calculate the proportion of games won by Georgia in this 30-year period. This will be the sample proportion.

\hat{p} : _____

14. Assume the conditions for a 1-proportion Z-test are met. Look at the previously constructed dotplot on the board in the classroom. What proportion of dots on the dotplot were at the sample proportion (\hat{p}) or higher? Because this is a two-tailed test, double this proportion and this will be our simulated p-value.

15. Interpret the simulated p-value from above.

16. Use your p-value to make a conclusion about the Georgia-Florida rivalry based off of this sample of the last 30 games.

17. Do you feel that the conclusion from number 16 is valid? Why or why not?

18. Is it possible to get a random sample like this one where Florida wins 21 of 30 games?

19. Compare and contrast the results of this sample and the first sample.

Group Names: _____

20. Now use technology to generate a simple random sample of 30 games from the rivalry. The games haven been numbered in the table below.

Year	Winner	Year	Winner	Year	Winner
1904 (1)	Georgia	1954 (33)	Georgia	1986 (65)	Florida
1915 (2)	Georgia	1955 (34)	Florida	1987 (66)	Georgia
1916 (3)	Georgia	1956 (35)	Florida	1988 (67)	Georgia
1919 (4)	Georgia	1957 (36)	Florida	1989 (68)	Georgia
1920 (5)	Georgia	1958 (37)	Florida	1990 (69)	Florida
1926 (6)	Georgia	1959 (38)	Georgia	1991 (70)	Florida
1927 (7)	Georgia	1960 (39)	Florida	1992 (71)	Florida
1928 (8)	Florida	1961 (40)	Florida	1993 (72)	Florida
1929 (9)	Florida	1962 (41)	Florida	1994 (73)	Florida
1930 (10)	TIE	1963 (42)	Florida	1995 (74)	Florida
1931 (11)	Georgia	1964 (43)	Georgia	1996 (75)	Florida
1932 (12)	Georgia	1965 (44)	Florida	1997 (76)	Georgia
1933 (13)	Georgia	1966 (45)	Georgia	1998 (77)	Florida
1934 (14)	Georgia	1967 (46)	Florida	1999 (78)	Florida
1935 (15)	Georgia	1968 (47)	Georgia	2000 (79)	Florida
1936 (16)	Georgia	1969 (48)	TIE	2001 (80)	Florida
1937 (17)	Florida	1970 (49)	Florida	2002 (81)	Florida
1938 (18)	Georgia	1971 (50)	Georgia	2003 (82)	Florida
1939 (19)	Georgia	1972 (51)	Georgia	2004 (83)	Georgia
1940 (20)	Florida	1973 (52)	Florida	2005 (84)	Florida
1941 (21)	Georgia	1974 (53)	Georgia	2006 (85)	Florida
1942 (22)	Georgia	1975 (54)	Georgia	2007 (86)	Georgia
1944 (23)	Georgia	1976 (55)	Georgia	2008 (87)	Florida
1945 (24)	Georgia	1977 (56)	Florida	2009 (88)	Florida
1946 (25)	Georgia	1978 (57)	Georgia	2010 (89)	Florida
1947 (26)	Georgia	1979 (58)	Georgia	2011 (90)	Georgia
1948 (27)	Georgia	1980 (59)	Georgia	2012 (91)	Georgia
1949 (28)	Florida	1981 (60)	Georgia	2013 (92)	Georgia
1950 (29)	Georgia	1982 (61)	Georgia	2014 (93)	Florida
1951 (30)	Georgia	1983 (62)	Georgia	2015 (94)	Florida
1952 (31)	Florida	1984 (63)	Florida		
1953 (32)	Florida	1985 (64)	Georgia		

Group Names: _____

21. Suppose we want to see if there is statistically significant evidence showing one team is superior to the other. Give the appropriate null and alternative hypotheses. (Hint: For the Null Hypothesis, what would the proportion of wins be for Georgia if two teams were evenly matched. Use a two-tailed one-proportion z-test.)

Let p = The proportion of games won by Georgia

H_0 : _____

H_a : _____

22. Calculate the proportion of games won by Georgia in the simple random sample. Once again, this will be our sample proportion.

\hat{p} : _____

23. Assume the conditions for a 1-proportion Z-test are met. Look at the previously constructed dotplot on the board in the classroom. What proportion of dots on the dotplot were at the sample proportion (\hat{p}) or higher? Because this is a two-tailed test, double this proportion and this will be our simulated p-value.

24. Interpret the simulated p-value from above.

25. Are all of the conditions for a 1-Proportion Z-test met? Explain.

Group Names: _____

26. Use your p-value to make a conclusion about the Georgia-Florida rivalry based off of this random sample of 30 games.

27. Does this conclusion differ from your previous two conclusions? If so, how does it differ?

28. So far you have looked at three different samples of 30 games from the rivalry. Now examine the results of all 94 games. Calculate the proportion of games won by Georgia in this 94-year period. This will be the sample proportion.

\hat{p} : _____

29. Suppose we want to examine if there is statistically significant evidence showing one team is superior to the other. Give the appropriate null and alternative hypotheses. (Hint: For the Null Hypothesis, what would the proportion of wins be for Georgia if two teams were evenly matched. Use a two-tailed one-proportion z-test.)

Let p = The proportion of games won by Georgia

H_0 : _____

H_a : _____

30. Assume the conditions for a 1-proportion Z-test are met. Calculate the z-score and p-value for these 94 games.

z-score: _____

p-value: _____

Group Names: _____

31. Use your p-value to make a conclusion to determine if one team is superior in the Georgia-Florida rivalry based off of this sample of 94 games.

32. Type I Error is defined to be incorrectly rejecting the null hypothesis. It is possible to collect a sample that leads to statistically significant results in error. Which of the first 3 samples in this activity are examples of Type I Error? Also, explain Type I Error in the context of this problem.

33. Reflect on the different results the came from these four data sets and their resulting p-values. What does this reveal about statistical significance, p-values, and errors in conclusions based off of sampling?