NFL Point Spreads - Solutions

R. Lock

Solutions to 20 questions from the NFL Point Spreads and Game Scores module.

Get the dataset.

NFLPoints<-read.csv("NFLPoints.csv")

**1. What proportion of the time does the favored team actually win the game outright?**

Procedure: CI for a proportion (one sample)

Via formula

x=sum(NFLPoints$FavWin)  
n=length(NFLPoints$FavWin)  
phat=x/n  
SE=sqrt(phat\*(1-phat)/n)  
zstar=qnorm(0.975)  
ME=zstar\*SE  
lower=phat-ME  
upper=phat+ME

370/544=0.68 with SE=0.02

CI for p is 0.68 1.96 0.02 = (0.641,0.719)

We are 95% sure that the proportion of times the favored team wins the game is between 0.641 and 0.719.

Using an internal R function

prop.test(x,n,correct=FALSE)

1-sample proportions test without continuity correction  
  
data: x out of n, null probability 0.5  
X-squared = 70.618, df = 1, p-value < 2.2e-16  
alternative hypothesis: true p is not equal to 0.5  
95 percent confidence interval:  
 0.6398066 0.7179611  
sample estimates:  
 p   
0.6801471

Note that the method used for this CI is slightly different than the formula-based method above.

**2. Is the chance the favorite covers the spread discernibly different from 0.50?**

Procedure: Test for proportion (one sample)

vs.  where is the proportion of games the favored team covers the spread

Via formula

x=sum(NFLPoints$FavCover)  
n=length(NFLPoints$FavCover)  
phat=x/n  
SE=sqrt(phat\*(1-phat)/n)  
z=(phat-0.5)/SE  
pvalue=2\*pnorm(-abs(z))

293/544=0.539 with SE=0.0214

z=(0.539 - 0.5)/0.0214=1.81, p-value=0.071

At a 5% significance level we do not have convincing evidence that the proportion of times the favorite covers the spread is different from 0.50.

Using an internal R function

prop.test(x,n, correct=FALSE)

1-sample proportions test without continuity correction  
  
data: x out of n, null probability 0.5  
X-squared = 3.2426, df = 1, p-value = 0.07174  
alternative hypothesis: true p is not equal to 0.5  
95 percent confidence interval:  
 0.4965875 0.5800770  
sample estimates:  
 p   
0.5386029

Note that the R prop.test function uses a chi-square statistic, rather than a z-statistic, but the square root of that chi-square statistic should match the z-statistic up to the +/- sign.

**3. How different is the average point spread when the favored team is playing at home as compared to the favorite playing on the road?**

Procedure: CI for a difference in means (two samples)

(HomePts<-t.test(Pts~HomeDog, data=NFLPoints))

Welch Two Sample t-test  
  
data: Pts by HomeDog  
t = 3.7048, df = 503.38, p-value = 0.0002349  
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
95 percent confidence interval:  
 0.4430565 1.4435015  
sample estimates:  
mean in group 0 mean in group 1   
 5.119335 4.176056

The difference in means is and the 95% confidence interval for the difference in means is (0.44,1.44).

We are 95% sure that the average point spread when the home team is favored is between 0.44 and 1.44 more than when the road team is favored.

**4. Many football fans say that the home field advantage is about a field goal (three points). Is the average home margin (*HomeDiff*) discernibly different from three points?**

Procedure: Test for a mean (one sample using *HomeDiff*) OR Test for two means (paired data, using *HomeScore* and *AwayScore*)

vs.  where is the mean home field margin

(Home4<-t.test(NFLPoints$HomeDiff,mu=3))

One Sample t-test  
  
data: NFLPoints$HomeDiff  
t = -1.7374, df = 543, p-value = 0.08289  
alternative hypothesis: true mean is not equal to 3  
95 percent confidence interval:  
 0.681371 3.142158  
sample estimates:  
mean of x   
 1.911765

The mean home field advantage is 1.91.

The p-value for the test is 0.083 which is more than 5% so we don’t (quite) have enough evidence to tell that the mean home margin is different from 3 points. Note also that the 95% confidence interval for the mean difference includes 3.

**5. Is there convincing evidence that the average point spread assigned to the home team (HomePts) is different from three points?**

Procedure: Test for a mean (one sample)

vs.  where is the mean point spread

(Home5<-t.test(NFLPoints$HomePts,mu=3))

One Sample t-test  
  
data: NFLPoints$HomePts  
t = -6.5131, df = 543, p-value = 1.679e-10  
alternative hypothesis: true mean is not equal to 3  
95 percent confidence interval:  
 1.021286 1.938273  
sample estimates:  
mean of x   
 1.479779

The mean point spread for the home team is 1.48.

The p-value for the test is very close to zero so we have very strong evidence that the mean point spread assigned to the home team is less than 3 points.

*Interesting point:* The mean home point spread (1.48) is not a lot farther from 3 than the mean home margin (1.91), but the evidence of a difference is much stronger. One reason for this is that the standard deviation of the home point spreads (5.44) is much smaller than the standard deviation of the home margins (14.61).

**6. Estimate the mean difference between the point spread (*Pts*) and the actual margin for the favored team (*FavDiff*). Note that the margin will be negative if the favored team loses.**

Procedure: CI for difference in means (paired data)

(Q6<-t.test(NFLPoints$Pts,NFLPoints$FavDiff,paired=TRUE))

Paired t-test  
  
data: NFLPoints$Pts and NFLPoints$FavDiff  
t = -1.8678, df = 543, p-value = 0.06232  
alternative hypothesis: true mean difference is not equal to 0  
95 percent confidence interval:  
 -2.15726809 0.05432692  
sample estimates:  
mean difference   
 -1.051471

The mean difference between the point spread and the actual margin for the favored team is -1.05 and the 95% confidence interval for the difference mean difference is (-2.16,0.05).

We are 95% sure that the mean point spread is between 2.16 points less and 0.05 points more than the actual margin for the favored team.

**7. Is there convincing evidence that point spreads (*Pts*) tend, on average, to underestimate the margin for the favored team (*FavDiff*)?**

Procedure: Test for a difference in means (paired data)

vs.  where is the mean difference between the point spread and the actual margin for the favored team.

(Q7<-t.test(NFLPoints$Pts,NFLPoints$FavDiff,paired=TRUE,alternative="less"))

Paired t-test  
  
data: NFLPoints$Pts and NFLPoints$FavDiff  
t = -1.8678, df = 543, p-value = 0.03116  
alternative hypothesis: true mean difference is less than 0  
95 percent confidence interval:  
 -Inf -0.1239421  
sample estimates:  
mean difference   
 -1.051471

The mean difference between the point spread and the actual margin for the favored team is -1.05.

The t-statistic is -1.87 and p-value is 0.031 which is less than 5%.

At a 5% level we have enough evidence to conclude that the mean point spread is less than the mean actual margin for the favored team.

Note: The two-tailed CI for #6 barely includes zero (which would indicate not a discernible difference), but the one-tailed procedure in #7 has enough evidence to conclude the mean difference is negative.

**8. What is the average absolute value of the difference between the point spread (Pts) and the actual game margin (FavDiff)?**

Procedure: CI for a mean (one sample)

Via formula:

AbsDiff<-abs(NFLPoints$Pts-NFLPoints$FavDiff)  
xbar=mean(AbsDiff)  
s=sd(AbsDiff)  
n=length(AbsDiff)  
SE=s/sqrt(n)  
tstar=qt(0.975,n-1)  
ME=tstar\*SE  
lower=round(xbar-ME,2)  
upper=round(xbar+ME,2)

10.03 with SE=0.366

CI for the mean is 10.03 1.964 0.366 = (9.31,10.75)

We are 95% sure that the average distance from the point spread to the actual game margin is between 9.31 and 10.75 points.

Using an internal R function

t.test(AbsDiff)

One Sample t-test  
  
data: AbsDiff  
t = 27.419, df = 543, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 9.309169 10.745978  
sample estimates:  
mean of x   
 10.02757

**Do point spreads get more accurate as the season goes along? Address this in two ways:**

**9. Is the proportion of games where the favored team wins higher during the second half of the season (weeks 10-18) than the first half of the season (weeks 1-9)?**

Procedure: Test for a difference in proportions (two samples)

vs.  where and are the proportion of games the favored team wins in the first and second half of the season, respectively.

Via formula

Half<-ifelse(NFLPoints$Week<10,1,2)  
FirstCover<-NFLPoints$FavWin[Half==1]  
SecondCover<-NFLPoints$FavWin[Half==2]  
x1=sum(FirstCover)  
x2=sum(SecondCover)  
n1=length(FirstCover)  
n2=length(SecondCover)  
phat1=x1/n1  
phat2=x2/n2  
phat=(x1+x2)/(n1+n2)  
SE=sqrt(phat\*(1-phat)\*(1/n1+1/n2))  
z=(phat2-phat1)/SE  
pvalue=1-pnorm(z)

189/270 - 181/274 = 0.7 - 0.661 = 0.039 with SE=0.04

z=(0.7 0.661)/0.04 = 0.99, p-value=0.162

We do not have convincing evidence that the proportion of games predicted correctly by the point spread is higher in the second half of the season than in the first half.

Using an internal R function

prop.test(c(x2,x1),c(n2,n1),correct=FALSE,alternative="greater")

2-sample test for equality of proportions without continuity correction  
  
data: c(x2, x1) out of c(n2, n1)  
X-squared = 0.9712, df = 1, p-value = 0.1622  
alternative hypothesis: greater  
95 percent confidence interval:  
 -0.02629729 1.00000000  
sample estimates:  
 prop 1 prop 2   
0.7000000 0.6605839

**10. Refer to the absolute value of the difference between the point spread (*Pts*) and the actual game margin (*FavDiff*) from question #8. Is the average discrepancy smaller in the second half of the season than the first half?**

Procedure: Test for a difference in means (two samples)

vs. , where and are the mean absolute difference between the point spread and actual margin for the favored team in the first and second half of the season, respectively.

(Q10<-t.test(AbsDiff~Half,alternative="greater"))

Welch Two Sample t-test  
  
data: AbsDiff by Half  
t = 1.0309, df = 534.77, p-value = 0.1515  
alternative hypothesis: true difference in means between group 1 and group 2 is greater than 0  
95 percent confidence interval:  
 -0.4507001 Inf  
sample estimates:  
mean in group 1 mean in group 2   
 10.401460 9.648148

10.4 - 9.65 = 0.75

t = 1.03 and p-value = 0.152.

We do not have convincing evidence that the mean absolute difference between the point spreads and actual game margins is smaller in the second half of the season than in the first half.

**11. Some fans say they avoid choosing a favorite when the spread is double digits (more than 10 points). Is the proportion of favorites who cover discernibly less than 0.50 when the spread is more than ten points?**

Procedure: Test for a proportion (one sample)

vs.  where is the proportion of games teams favored by at least 10 points that cover the spread.

Via formula

BigSpread<-NFLPoints$FavCover[NFLPoints$Pts>10]  
x=sum(BigSpread)  
n=length(BigSpread)  
phat=x/n  
SE=sqrt(phat\*(1-phat)/n)  
z=(phat-0.5)/SE  
pvalue=pnorm(z)

18/34=0.529 with SE=0.0856

Note: The sample proportion of big favorites that cover the spread is not less than 0.50! Thus we know even before doing the details of the test that we will not be able to go with .

Checking the details for the test, z=(0.529 - 0.5)/0.0856=0.34, p-value=0.634

At a 5% significance level we do not have convincing evidence that the proportion of times the a 10+ point favorite covers the spread is less than 0.50.

Using an internal R function

prop.test(x,n, correct=FALSE,alternative="less")

1-sample proportions test without continuity correction  
  
data: x out of n, null probability 0.5  
X-squared = 0.11765, df = 1, p-value = 0.6342  
alternative hypothesis: true p is less than 0.5  
95 percent confidence interval:  
 0.0000000 0.6627735  
sample estimates:  
 p   
0.5294118

**12. How often is the away team favored to win the game?**

Procedure: CI for a proportion (one sample)

Via formula

x=sum(NFLPoints$HomeDog)  
n=length(NFLPoints$HomeDog)  
phat=x/n  
SE=sqrt(phat\*(1-phat)/n)  
zstar=qnorm(0.975)  
ME=zstar\*SE  
lower=phat-ME  
upper=phat+ME

213/544=0.392 with SE=0.0209

CI for p is 0.392 1.96 0.0209 = (0.351,0.433)

We are 95% sure that the proportion of times the away team is favored to win is between 0.351 and 0.433.

Using an internal R function

prop.test(x,n,correct=FALSE)

1-sample proportions test without continuity correction  
  
data: x out of n, null probability 0.5  
X-squared = 25.596, df = 1, p-value = 4.21e-07  
alternative hypothesis: true p is not equal to 0.5  
95 percent confidence interval:  
 0.3514256 0.4331837  
sample estimates:  
 p   
0.3915441

**13. Is the mean number of points scored by the favored team higher in 2024 than in 2023?**

Procedure: No statistical inference needed. We know the means in 2023 and 2024 exactly!

(Q13<-aggregate(FavScore~Year,FUN=mean,data=NFLPoints))

Year FavScore  
1 2023 23.88971  
2 2024 26.29044

23.89 and 26.29, so the mean points scored by the favorite is higher in 2024 than in 2023.

**14. What is the average number of points scored by the winning team in NFL games?**

Procedure: CI for a mean (one sample)

Via formula:

WinScore<-with(NFLPoints,ifelse(FavWin==1,FavScore,DogScore))  
xbar=mean(WinScore)  
s=sd(WinScore)  
n=length(WinScore)  
SE=s/sqrt(n)  
tstar=qt(0.975,n-1)  
ME=tstar\*SE  
lower=round(xbar-ME,2)  
upper=round(xbar+ME,2)

27.86 with SE=0.369

CI for the mean is 27.86 1.964 0.369 = (27.13, 28.58)

We are 95% sure that the mean points scored by the winning team in NFL games is between 27.13 and 28.58 points.

Using an internal R function

t.test(WinScore)

One Sample t-test  
  
data: WinScore  
t = 75.517, df = 543, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 27.13201 28.58123  
sample estimates:  
mean of x   
 27.85662

**15. The weather is often more of a factor later in the season. How much does the mean number of points scored (both teams combined) compare between the first half of the season (weeks 1-9) and the second half (weeks 10-18)?**

Procedure: CI for a difference in means (two samples)

(Q15<-t.test((FavScore+DogScore)~Half, data=NFLPoints))

Welch Two Sample t-test  
  
data: (FavScore + DogScore) by Half  
t = -1.0474, df = 526.5, p-value = 0.2954  
alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0  
95 percent confidence interval:  
 -3.462563 1.054236  
sample estimates:  
mean in group 1 mean in group 2   
 43.78102 44.98519

The difference in means is 43.78 - 44.99 = -1.2 and the 95% confidence interval for the difference in means is (-3.46, 1.05).

We are 95% sure that the mean combined points scored by both teams in NFL games in the first half of the season is between 3.46 less and 1.05 more than in the second half of the season.

**16. The most common scoring events in football are a field goal (3 points) and a touchdown with an extra point (7 points). What proportion of point spreads are within ½ point of either 3 or 7 (i.e., 2.5, 3.5, 6.5, or 7.5)?**

Procedure: CI for a proportion (one sample)

Via formula

Pts37<-ifelse(NFLPoints$Pts %in% c(2.5, 3.5, 6.5, 7.5),1,0 )  
x=sum(Pts37)  
n=length(Pts37)  
phat=x/n  
SE=sqrt(phat\*(1-phat)/n)  
zstar=qnorm(0.975)  
ME=zstar\*SE  
lower=phat-ME  
upper=phat+ME

308/544=0.566 with SE=0.0212

CI for p is 0.566 1.96 0.0212 = (0.525,0.608)

We are 95% sure that the proportion of times the point spread is within one-half point of 3 or 7 is between 0.525 and 0.608.

Using an internal R function

prop.test(x,n,correct=FALSE)

1-sample proportions test without continuity correction  
  
data: x out of n, null probability 0.5  
X-squared = 9.5294, df = 1, p-value = 0.002022  
alternative hypothesis: true p is not equal to 0.5  
95 percent confidence interval:  
 0.5242094 0.6072155  
sample estimates:  
 p   
0.5661765

**17. How much more (or less) often do favorites cover the spread when playing at home vs. favorites on the road?**

Procedure: CI for a difference in proportions (two samples)

Via formula

HomeCover<-with(NFLPoints,FavCover[HomeDog==0])  
AwayCover<-with(NFLPoints,FavCover[HomeDog==1])  
x1=sum(HomeCover)  
x2=sum(AwayCover)  
n1=length(HomeCover)  
n2=length(AwayCover)  
phat1=x1/n1  
phat2=x2/n2  
pdiff=phat1-phat2  
SE=sqrt(phat1\*(1-phat1)/n1+phat2\*(1-phat2)/n2)  
zstar=qnorm(0.975)  
ME=zstar\*SE  
lower=pdiff-ME  
upper=pdiff+ME

173/331 - 120/213 = 0.523 - 0.563 = -0.041 with SE=0.0437

CI for is -0.041 1.96 0.0437 = (-0.126,0.045)

We are 95% sure that the proportion of times that home favorites cover the spread is between 0.126 less and 0.045 more than the proportion of road favorites that cover.

Using an internal R function

prop.test(c(x1,x2),c(n1,n2),correct=FALSE)

2-sample test for equality of proportions without continuity correction  
  
data: c(x1, x2) out of c(n1, n2)  
X-squared = 0.8648, df = 1, p-value = 0.3524  
alternative hypothesis: two.sided  
95 percent confidence interval:  
 -0.12634732 0.04490398  
sample estimates:  
 prop 1 prop 2   
0.5226586 0.5633803

**18. How often does the favorite win the game, but fail to cover the point spread?**

Procedure: CI for a proportion (one sample)

Via formula

#Find the cases where favorite wins but doesn't cover  
WinNoCover<-NFLPoints$FavWin-NFLPoints$FavCover  
x=sum(WinNoCover)  
n=length(WinNoCover)  
phat=x/n  
SE=sqrt(phat\*(1-phat)/n)  
zstar=qnorm(0.975)  
ME=zstar\*SE  
lower=round(phat-ME,3)  
upper=round(phat+ME,3)

77/544=0.142 with SE=0.0149

CI for p is 0.142 1.96 0.0149 = (0.112, 0.171)

We are 95% sure that the proportion of times the favorite wins the game but fails to cover the spread is between 0.112 and 0.171.

Using an internal R function

prop.test(x,n,correct=FALSE)

1-sample proportions test without continuity correction  
  
data: x out of n, null probability 0.5  
X-squared = 279.6, df = 1, p-value < 2.2e-16  
alternative hypothesis: true p is not equal to 0.5  
95 percent confidence interval:  
 0.1147602 0.1733551  
sample estimates:  
 p   
0.1415441

**Is there a home field advantage? Address this in two ways:**

**19. Compare the mean points scored by home teams to the mean points scored by away teams.**

Procedure: Test for a difference in means (paired data using *HomeScore* and *AwayScore*) or  
Test for a single mean (using *HomeDiff*)

vs.  where and are the mean points scored when home or away, respectively.

Or

vs.  where is the mean difference of home minus away scores.

(Q19<-t.test(NFLPoints$HomeScore,NFLPoints$AwayScore,paired=TRUE,alternative="greater"))

Paired t-test  
  
data: NFLPoints$HomeScore and NFLPoints$AwayScore  
t = 3.0522, df = 543, p-value = 0.001192  
alternative hypothesis: true mean difference is greater than 0  
95 percent confidence interval:  
 0.8797266 Inf  
sample estimates:  
mean difference   
 1.911765

The mean difference between the home score and the away score is 1.91.

The t-statistic is 3.05 and p-value is 0.001 which is less than 5%.

We have very strong evidence to conclude that the mean points scored by home teams is more than the mean points scored by road teams.

**20. Use how often the home team wins the game outright.**

Procedure: Test for a proportion (one sample)

vs.  where is the proportion of games teams won by the home team.

Via formula

x=sum(NFLPoints$HomeWin)  
n=length(NFLPoints$HomeWin)  
phat=x/n  
SE=sqrt(phat\*(1-phat)/n)  
z=(phat-0.5)/SE  
pvalue=1-pnorm(z)

295/544=0.542 with SE=0.0214

Checking the details for the test, z=(0.542 - 0.5)/0.0214=1.98, p-value=0.024

At a 5% significance level we have convincing evidence that the proportion of times home team wins the game outright is more than 0.50, so there appears to be some home field advantage.

Using an internal R function

prop.test(x,n, correct=FALSE,alternative="greater")

1-sample proportions test without continuity correction  
  
data: x out of n, null probability 0.5  
X-squared = 3.8897, df = 1, p-value = 0.02429  
alternative hypothesis: true p is greater than 0.5  
95 percent confidence interval:  
 0.5070216 1.0000000  
sample estimates:  
 p   
0.5422794