

ARod Hypothesis Test

Dr. Ekstrom

Purpose: Did Alex Rodriguez shrink under the pressure of the post-season? Or is he the victim of a small sample size? In this activity, we explore ARod's post season struggles.

Sports Analytics: Test a claim regarding "clutch performance".

Statistics: Hypothesis testing using a simulation.

R: Simulating in *R*

To test the claim about ARod's post-season performance, we use a hypothesis test.

THE IDEA OF A HYPOTHESIS TEST

Imagine you are playing a dice game with the shady character Dr. X. You notice that when Dr. X rolls his die, the number six seems to appear too often. In fact, you record that the number six has appeared in 12 of his 20 rolls. You start to suspect that Dr. X isn't using a fair die. However, he won't let you inspect the die, so all you have to go on is your observations. Is the fact that you have observed the number six in 12 of his 20 rolls enough evidence to conclude that the die Dr. X is using is not fair?

Let us assume Dr. X is using a fair die. What is the probability of seeing an event this extreme (i.e. observing a six in at least 12 of 20 rolls)? If you know about the binomial distribution, you can compute the probability directly. We will compute this probability by simulating 200,000 experiments of 20 rolls.

We begin by simulating one of experiment of 20 rolls.

```
# vector representing the die
die=c(1,2,3,4,5,6)
# probability of each face occurring
prd=rep(1/6,6)
# sampling representing 20 rolls of the die - held in simd
simd=sample(die,20,prob=prd,replace=TRUE)
simd

## [1] 6 6 6 5 2 6 2 3 6 1 2 4 5 6 5 2 5 5 2 6

# counts the number of 6's that occur
sum(simd==6)

## [1] 7
```

Now we construct a `for` loop to simulate 200,000 experiments. We create the vector `simres` to store the results.

```
# ns is the number of simulation of 20 rolls
ns=200000
# simres is a place to hold the result of each simulation
simres=rep(-1,ns)
# main for loop begins
```

```

for(j in 1:ns){
  # sampling representing 20 rolls of the die - held in simd
  simd=sample(die,20,prob=prd,replace=TRUE)
  # stores the number of 6's from this simulation to simres
  simres[j]=sum(simd==6)
}
# main for loop ends
sum(simres>=12)/ns

## [1] 1.5e-05

```

We see that if Dr. X is using a fair die, such an observation is highly unlikely. Thus we conclude that Dr. X is not using a fair die. This reasoning is the heart of a hypothesis test. It should be noted that we cannot be certain that Dr. X is using an unfair die, he may be using a fair die and we observed something truly remarkable. (Consider how unlikely it is for anyone to win the lottery - yet people do win it!) However, it is so unlikely, we are willing to live with the chance that our conclusion is wrong.

SETTING UP OUR HYPOTHESIS TEST FOR AROD

The first step is to set up our hypotheses. In general, the null hypothesis is there is no difference in performance. In the example above, our null hypothesis was that Dr. X was using a fair (regular) die. The alternate is that there is some sort of difference (and we choose if it is less than, greater than, or not equal to).

Null Hypothesis: Alex Rodriguez's post-season performance is not significantly different from his regular season performance.

Alternate Hypothesis: Alex Rodriguez's post-season performance is significantly lower than his regular season performance.

The second step is to choose a significance level (or α level). I tend to use 5%, but 10% or 1% is also common. We will discuss the role of this significance level later, but it is important to make this choice now.

The data for this test is located in `Alex Rodriguez hypoth.csv`

```
z=read.csv("/Users/DoctorX/Desktop/Math 367 S19/Class Activities/07 ARod Hypoth/Alex Rodriguez
z

##      Season     G    PA    AB     R     H    X1B    X2B    X3B    HR    RBI    SB    CS    BB
## 1 Regular  2784 12207 10566 2021 3115 1840 548   31  696 2086 329  76 1338
## 2 Post     76    330   278   43    72   43   16    0   13   41    8   3   39
## 3 Total   2860 12537 10844 2064 3187 1883 564   31  709 2127 337  79 1377
##      SO HBP SH SF IBB
## 1 2287 176 16 111  97
## 2  77   9  1   3   4
## 3 2364 185 17 114 101
```

CHOOSING A METRIC

The metric we will use to evaluate Alex Rodriguez's performance is `OPS`. We have seen in a previous activity that `OPS` is a pretty effective measure for offensive production, and has been adopted as a standard metric in the baseball community.

Recall that $\text{OPS} = \text{OBP} + \text{SLG}$ and

$$\text{OBP} = \frac{\text{H} + \text{BB} + \text{HBP}}{\text{AB} + \text{BB} + \text{HBP} + \text{SF}}$$
$$\text{SLG} = \frac{\text{X1B} + 2 \cdot \text{X2B} + 3 \cdot \text{X3B} + 4 \cdot \text{HR}}{\text{AB}}$$

So Alex's career `OPS` (this includes both regular season and post-season) is:

```
attach(z)
OBP=(H+BB+HBP)/(AB+BB+HBP+SF)
SLG=(X1B+2*X2B+3*X3B+4*HR)/AB
OPS=OBP+SLG
OPS[3]
```

```
## [1] 0.9270815
```

We see that Alex's career `OPS` is 0.927 - including regular and post seasons. We will test whether Alex's post season `OPS` is significantly lower than his career `OPS`, so our hypotheses become:

$$H_0: \text{OPS}_{\text{Post}} = 0.927$$
$$H_a: \text{OPS}_{\text{Post}} < 0.927$$

An important question here: Why am I comparing his post-season `OPS` to his career `OPS`, rather than comparing it with just his regular season `OPS`? We are doing this because we are assuming the null hypothesis is true - that there is no difference between his performance in the regular season and the post-season - and that his post-season plate appearances should be treated as any other subset of plate appearances from ARod's career (such as plate appearances from July 2007).

We know Alex's post season `OPS`, it is:

```
OPS[2]
```

```
## [1] 0.8215762
```

Clearly it is less than his career `OPS`. However, Alex had very few at-bats in the post season (compared to the number of regular season at bats he had). `OPS` has natural fluctuations, and in a small sample, we will naturally see a wide variation of `OPS` values. Perhaps it was just unfortunate for Alex that this low fluctuation occurred in the post season. In the language of statistics, we ask *was this observed difference significant?*

So we are going to assume that Alex was the same player in the post-season as the regular season and that his lower `OPS` is simply due to chance.

In general, this is an important part of executing a hypothesis test - that you assume the null hypothesis is true - so the differences you see in the data are just due to natural variations one would expect to see. (Think of this as the presumption of innocence in our legal system.)

Then the question becomes - what is the likelihood seeing such a low `OPS` due to chance.

CREATING A SIMULATION

To find the answer to the question: what is the likelihood of seeing an OPS as low as 0.8215762 for ARod due to chance, we will simulate post-season experiences for Alex. To simplify our analysis, we will ignore sacrifice flies and sacrifice bunts since they would require a more advanced simulation than we are presenting here. However, since Alex had so few sacrifice flies and bunts in his career, we do not expect this to have much of an impact on our analysis.

First, I am going to create two new variables:

APA: (adjusted plate appearances) - which will be the sum of his at-bats, walks, and hit by pitches.

$$\text{APA} = \text{AB} + \text{BB} + \text{HBP}$$

Outs: Non-sacrifice outs.

$$\text{Outs} = \text{AB} - \text{X1B} - \text{X2B} - \text{X3B} - \text{HR}$$

APA=AB+BB+HBP

Outs=AB-X1B-X2B-X3B-HR

Next we will create a vector **poss** that contains the possible outcomes of an adjusted plate appearance, where 1 represents a single, 2 represents a double, 3 a triple, 4 a homerun, 5 represents either a walk or a hit by pitch, and 6 represents a non-sacrifice out. We also create the corresponding probability vector **pr**, using ARod's career statistics.

```
poss=c(1,2,3,4,5,6)
pr=c(X1B[3]/APA[3],X2B[3]/APA[3],X3B[3]/APA[3],HR[3]/APA[3],(BB[3]+HBP[3])/APA[3],Outs[3]/APA[3])
```

Note that ARod had 326 post-season adjusted plate appearances during his career. To simulate 326 adjusted plate appearances, we use

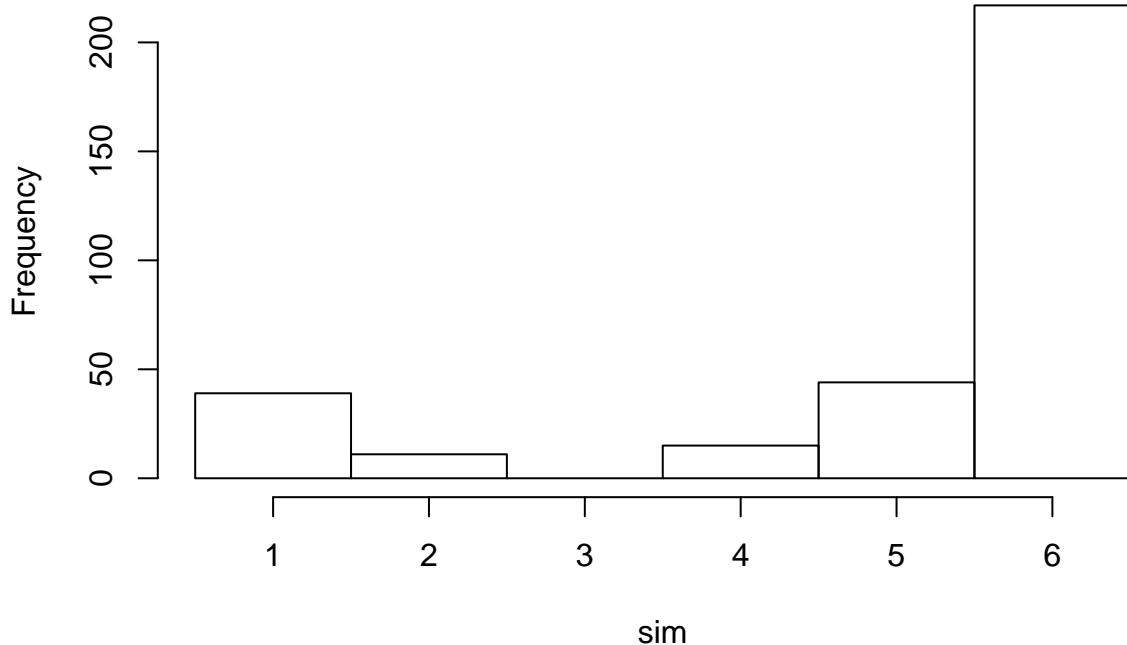
```
sample(poss,326,prob=pr,replace=TRUE)
```

```
## [1] 6 6 6 6 6 6 1 1 6 1 2 6 5 6 1 2 1 1 6 4 6 6 2 6 6 2 6 4 4 2 1 6 6
## [36] 6 1 5 5 6 1 5 6 2 6 6 6 4 5 6 1 6 6 5 6 6 1 5 6 5 6 6 1 6 6 6 6 6 4
## [71] 5 1 6 6 6 1 6 6 6 6 6 6 6 2 6 1 2 1 6 5 6 6 2 6 4 6 6 6 6 1 5 1 1 6
## [106] 6 6 1 6 5 6 6 6 1 1 5 6 5 6 6 1 5 6 6 1 6 6 6 6 1 4 6 5 2 5 4 6 6 1
## [141] 6 1 1 5 4 6 6 4 6 6 6 4 6 6 6 6 1 6 5 6 5 6 1 6 1 6 6 6 5 5 6 5
## [176] 6 1 5 6 6 6 2 6 6 6 1 6 6 6 5 6 6 1 6 1 6 6 6 6 6 5 1 6 6 6 6 5 6 6
## [211] 6 6 6 6 4 6 4 1 6 5 6 6 6 6 6 1 6 2 6 6 1 6 6 6 6 5 6 6 6 5 2 2 6
## [246] 6 1 6 2 6 6 4 6 5 6 6 6 6 5 4 6 1 6 6 1 1 2 6 6 6 6 6 5 5 6 6 6
## [281] 6 6 6 6 1 6 1 6 6 1 1 6 6 6 5 6 1 6 4 6 6 6 6 6 6 6 6 6 6 6 2 6
## [316] 6 6 1 6 5 1 6 5 6 6 1
```

Next, we want to know how many outs, singles, doubles, etc., happened during these simulated plate appearances. There are several ways to do this, I tend to use the histogram command, such as

```
sim=sample(poss,326,prob=pr,replace=TRUE)
q=hist(sim,breaks=c(0.5,1.5,2.5,3.5,4.5,5.5,6.5))
```

Histogram of sim



We then calculate the OPS that would result from these plate appearances (we are also calculating OBP and SLG along the way):

```

pobp=(q$counts[1]+q$counts[2]+q$counts[3]+q$counts[4]+q$counts[5])/(q$counts[1]+q$counts[2]+q$counts[3]+q$counts[4]+q$counts[5])
pobp

## [1] 0.3343558

pslg=(q$counts[1]+2*q$counts[2]+3*q$counts[3]+4*q$counts[4])/(q$counts[1]+q$counts[2]+q$counts[3]+q$counts[4])
pslg

## [1] 0.429078

pobp+pslg

## [1] 0.7634338

```

Of course, every time I ran this simulation, I could get completely different numbers. What is the likelihood I would get a number as low as 0.8215762? To answer this, we create a `for` loop so we can run this simulation n times. We create the vector `pops` to store the results.

```

# n is number of simulation
n=20000
# pops holds the ops for each simulated post-season
pops=rep(-1,n)
# begins main for loop
for(j in 1:n){
  # samples 326 plat appearances and stores them in sim
  sim=sample(poss,326,prob=pr,replace=TRUE)
}

```

```

# divides the results in sim into bins
q=hist(sim,breaks=c(0.5,1.5,2.5,3.5,4.5,5.5,6.5),plot=FALSE)
# computes the obp for this simulated post-season
pobp=(q$counts[1]+q$counts[2]+q$counts[3]+q$counts[4]+q$counts[5])/(q$counts[1]+q$counts[2]+q$counts[3])
# computes the slg for this simulated post-season
pslg=(q$counts[1]+2*q$counts[2]+3*q$counts[3]+4*q$counts[4])/(q$counts[1]+q$counts[2]+q$counts[3])
# computes the ops for this simulated post-season
pops[j]=pobp+pslg
}
# ends main for loop

```

Now that we have run this n times, we just need to count the number of times the simulated OPS was as low as 0.8215762, and divide by n

```
sum(pops<=.8215762)/n
```

```
## [1] 0.0939
```

When I ran this simulation with $n = 20000$, I got 0.0934.

INTERPRETING THE RESULTS

How do we interpret 0.0934 (or whatever number you got when you ran it)? We randomly created 20000 possible sets of 326 adjusted plate appearances based on ARod's career numbers. Of those, 9.34% had an OPS as low as 0.8215762. In other words, there is a 9.34% chance of seeing an OPS this low just by chance. Since I chose a significance level of 5%, this is not enough evidence to reject the null hypothesis in favor of the alternate. (Had I chosen 10%, it would have been borderline, and I would run the simulation a number of times before reaching any conclusion.) Or, in other words, based on this simulation, I do not have enough evidence to conclude that ARod's post-season performance was *significantly* worse than his regular season performance (or career performance). Or, the difference between ARod's post-season performance and regular season performance was not *statistically significant*.

FUTHER CONSIDERATIONS

Our simulation only considered ARod's offensive production, and had a few simplifying assumptions. It also did not take into consideration that ARod probably faced better pitching in the post-season. It would be interesting to run a more complete simulation and see if our results - and our conclusion - would be different.

GROUP ACTIVITY

Run a similar hypothesis test for a different athlete of your choosing. How does LeBron's performance in the post-season compare to his regular season performance? What about James Harden? Or Tom Brady? Or Derek Jeter?

- Pick an athlete.
- Find data for this athlete's performance. (Sites like www.baseball-reference.com have plenty of player data.)
- Determine a metric to measure their performance.
- Set-up your null and alternate hypothesis.

- Pick a level of significance.
- Create a simulation to determine the probability of such an extreme event as the one observed occurring.
- Reach an appropriate conclusion.

Use an *R* notebook for this assignment and turn in your knitted pdf document in the assignment dropbox in D2L.