STATISTICS REPORT ON THE GAMES PLAYED

I. INTRODUCTION

A college basketball player is drafted during NBA 2022 drafts. He believes if he is given the chance, he can contribute a great deal to the team's success yet since it is his first year, he is afraid he will not get enough time to demonstrate his performance. He believes that as long as he shows his talent during the practice, he will play up to 45 games, and he believes he has at most 40% chance of playing for more than 45 games. To check his belief of playing up to 45 games, a dataset (Cirtautas, 2022) consisting of the names of the players that are drafted in 2020 and 2021, and how many games they played in their first season will be used. The dataset contains two variables, the first one is the names of the players – which cannot be used as a parameter – and the second one is the number of games played by the player during his first season which is a discrete numerical variable and is a ratio data. The sample has 110 players in total. In this report, we will test the hypothesis that this player will play for less than or equal to 45 games as he guesses and using the knowledge that he has 40% chance of playing for more than 45 games, we will test whether he will play for more or not.

II. HYPOTHESIS TEST FOR THE MEAN

In this part, we will test the hypothesis on the number of games the player may take part in. For convenience we use SPSS analysis to get the sample mean and sample standard deviation as follows (check table-1 below),

$n = 110$

$\bar{x} = 42.6$

$s = 21.1$

which we can use for both one-sided and two-sided approaches. We will test the dataset for the null hypothesis that the player will get less than or equal to 45 games using a one-sided approach, and the hypothesis that he will play for 45 games using a two-sided approach respectively, and we will use t-score as we do not know the population variance. Then, we will check the p-value and confidence interval for two-sided approach. Lastly, we will interpret the SPSS output.

A. One-Sided Approach

We want to test the player's belief that he will play up to 45 games. Hence, our null hypothesis is

H_0: μ ≤ 45

against the alternative

H_1: μ > 45

Hence, this is an upper-tail test, and we need to take the positive value of the t-score.

We have

$d.f. = 110 - 1 = 109$

and we will take 5% significance level

$\alpha = 0.05$

Using the equation for hypothesis test for the mean with unknown population standard deviation, we get the equation

$t = \frac{42.6 - 45}{21.1/\sqrt{110}} = -1.19$

$t_{109,0.05} = 1.659$

(However, since the sample size is larger than 100, we can also use Z-table which would give $z_\alpha = 1.645$ which is close to the t-value.)

The decision rule is that if $\bar{x} > \bar{x}_{crit}$, reject the null hypothesis. Hence, we will reject the null hypothesis, if our value is above t-score.

$$\bar{x}_{crit} = 1.659 * \frac{21.1}{\sqrt{110}} + 45 = 48.3$$

Since $t = -1.19 < t_{109,0.05}$, and $\bar{x} = 42.6 < \bar{x}_{crit} = 48.3$, we fail to reject H_0. We do not have sufficient evidence to claim that the player will play for more than 45 games. The player will play for less than or equal to 45 games.

B. Two-Sided Approach

For two-sided approach we use the same μ with different hypothesis and different significance level. We want to see if the player will play for 45 games or not. Below are the null and alternative hypothesis with the significance level we choose.

H_0: $\mu = 45$

H_1: $\mu \neq 45$

$d.f. = 110 - 1 = 109$

$\alpha/2 = 0.05$

Following we get,

$$t = \frac{42.6 - 45}{21.1/\sqrt{110}} = -1.19$$

Since this is a two-sided test, we take need positive and negative values of t-score.

$t_{109,0.1} = \mp 1.659$

The decision rule is that if $-\bar{x}_U > \bar{x} > \bar{x}_L$, reject the null hypothesis. Hence, we will reject the null hypothesis, if our value is above the positive t-score or below the negative score.

Since $-1.659 < t = -1.19 < 1.659$, we fail to reject H_0. The player may play equal to 45 games.

C. P-Value Test

For one sided approach

$p - value = P(Z < -1.19) = 1 - P(Z < 1.19) = 1 - 0.88298 = 0.11072 > \alpha = 0.05,$

hence, we fail to reject the null hypothesis.

For two-sided approach

$p - value = 2P(Z < -1.19) = 0.23404 > \alpha = 0.1$

We fail to reject the null hypothesis.

For alpha values of 1%, 5%, and 10%, we fail to reject the null hypothesis for both approaches as the p-values are higher in each case and we do not have sufficient evidence to reject. Yet as the α increases, the strength of evidence against the null hypothesis increases, and we get closer to rejecting it.

D. Confidence Interval Estimation

For α = 1%

$42.6 \mp \frac{21.1}{\sqrt{110}} * 2.622 = 42.6\ \mp 3.34$ = (37.33, 47.87)

For α = 5%

$42.6 \mp \frac{21.1}{\sqrt{110}} * 1.982 = 42.6\ \mp 3.34$ = (38.7, 46.59)

For α = 10%

$42.6 \mp \frac{21.1}{\sqrt{110}} * 1.659 = 42.6\ \mp 3.34$ = (39.26, 45.94)

The intervals show us that the larger the α, the smaller the interval. This is because by increasing the α, we narrow down the accepted region.

E. SPSS Demonstration

We will test our dataset using SPSS. Since SPSS only allows two-sided approach, we will use α = 0.10 as we used in the two-sided approach above.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Games played	110	42.5545	21.08810	2.01067

Table-1. One sample statistic.

Table-1 gives us the necessary information to do our manual calculations.

One-Sample Test

	Test Value = 45					
					90% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Games played	-1.216	109	.227	-2.44545	-5.7811	.8902

Table-2: One-sample test.

Table-2 tells us about the t-value, p-value, and the confidence interval. We had

$t = -1.19$,

$p - value = 0.234$,

Confidence interval = (39.26, 45.94), and SPSS gives,

$t = -1.216$,

$p - value = 0.227$,

Confidence interval = $(45 - 5.7811, 45 + 0.8902) = (39.2189, 45.8902)$.

As seen, the difference of values in manual calculations and SPSS calculation are negligible, and the conclusion that we do not have sufficient evidence to reject the null hypothesis stands in SPSS calculations as well.

III. HYPOTHESIS TEST FOR THE POPULATION PROPORTION

In this part, we will test the player's belief on the probability of him playing for more than 45 games. We will test this claim as a one-sided claim with the null hypothesis that the player has less than or equal to 40% chance, and as a two-sided claim with the null hypothesis that his chance is equal to 40%. We need the proportion of players that played more than 45 games against the number of players in the sample:

$$\hat{p} = \frac{53}{110} = 0.482$$

Following, we get the z-value,

$$z = \frac{0.482 - 0.4}{\surd(0.4 * 0.6/110)} = 1.76$$

A. One-Sided Approach

We will test whether he has more than or less than 40% chance of playing for more than 45 games:

$H_0: p \leq 0.4$

$H_1: p > 0.4$

$\alpha = 0.05$

$Z_\alpha = 1.645$

The decision rule for one-sided test is to reject the null hypothesis if $z > z_\alpha$. Since it is an upper-tail test, we need the positive value of Z_α. As, 1.76 > 1.645=Z_α, we have sufficient evidence to reject the null hypothesis. The player has higher chance of playing for more than 45 games.

B. Two-Sided Approach

We will test if player's chance of playing for more than 45 games is equal to 40% or not. Hence,

H_0: $\mu = 0.4$

H_1: $\mu \neq 0.4$

$d.f. = 110 - 1 = 109$

$\alpha/2 \ = \ 0.05$

$z_{\alpha/2} = \mp 1.645$

The decision rule for two-sided test is to reject the null hypothesis if $-z_{\alpha/2} > z > z_{\alpha/2}$. Since $-1.645 < 1.645 < 1.76$, we reject the null hypothesis. The player does not have equal to 0.4 chance of playing for more than 45 games but has higher chance.

C. P-Value Test

We need to first check whether normal distribution is applicable:

$110 * 0.4 * 06 \ = 26.4 > 5$

Hence, normal distribution is applicable.

$$p - value = P(Z > 1.76) = 1 - Fz(1.76) = 1 - 0.99608 = 0.0392 < 0.05$$

We reject the null hypothesis since p-value is less than α.

$$p - value = 2P(Z > 1.76) = 2 - 2Fz(1.76) = 0.0784 < 0.1$$

We reject the null hypothesis since p-value is less than α.

For α=1%, we fail to reject the null hypothesis for both approaches. For α=5%, we reject for the one-sided approach, yet fail to reject for two-sided. As for α = 10%, we reject the null hypothesis for both approaches as the p-values are less than α in each case and we have sufficient evidence to reject. As seen, the larger the α, the more inclined we are to find enough evidence to reject the null hypothesis.

D. Confidence Interval Estimation

We will check the confidence interval for two-sided approach at different alpha levels.

For $\alpha = \ 1\%$

$$0.482 \ \mp \ 2.575 * \sqrt{\frac{0.482 * 0.518}{110}} = (0.360, 0.605)$$

For $\alpha = 5\%$

$$0.482 \mp 1.96 * \sqrt{\frac{0.482 * 0.518}{110}} = (0.389, 0.576)$$

For $\alpha = 10\%$

$$0.482 \mp 1.645 * \sqrt{\frac{0.482 * 0.518}{110}} = (0.404, 0.560)$$

As in the hypothesis test for the mean, the larger the α, the narrower the confidence interval. This is because α value is a big factor in deciding the critical values and hence, it narrows down the interval.

IV. CONCLUSION

In this report, we aimed to test the hypothesis on a new NBA draft's possibility of taking part in games. We used a sample of previous drafts' information and evaluated it for the new draft. The results show that within the chosen confidence levels the player is right about playing up to 45 games, but he is wrong about his chance of playing for more than 45 games. There is a slightly larger probability for him to play for more than 45 games. As this dataset contains only the last two year's draft data, it is not a completely conclusive report, yet it is suggestive. If we had a larger dataset with other statistical data that contains more information that we could connect to our player, we could have a more conclusive report. Moreover, with the right performance the player may have a higher chance of refuting the conclusion of this report. In other words, this report it to describe the past events, not to dictate the future. We wish the best to this player who has hired us for free and are looking forward to working with him again but with a payment plan.

BIBLIOGRAPHY

- Cirtautas, J. (2022). *NBA Players*. Retrieved 12 28, 2022, from Kaggle: https://www.kaggle.com/datasets/justinas/nba-players-data?select=all_seasons.csv