Professional League of Legends Player Champion Diversity

Dylan DeJean, Justin Huang, Jong Hee Lee

2022-12-14

Executive Summary

Using data from the official League of Legends Worlds 2022 tournament group stages and through post hoc analysis, we found that Mid lane players use less unique champions over their games compared to Top lane players and Support players. We also observed from an omnibus test that there was no significant effect on unique champions played from region or the interaction of role and region. We discuss limitations of this observational study as well.

Introduction and Background

League of Legends is an online video game with an active professional scene in multiple regions around the world. A match consists of two teams of five players, each with their own character in-game (called champions) and assigned role. There are currently 162 unique champions available to pick in the game and 5 roles: Top, Jungle, Mid, Bottom, and Support. Champion pool is a statistic used to determine the health of the current game meta. A large champion pool indicates that a large number of different strategies can be viable in professional play, while a small champion pool indicates that players are being forced to stick to certain strategies to see success. We will be using a rate of unique champions picked per game as a measurement of champion pool. Professional play is divided into many regions, but the main four are North America, European Union, China, and Korea.

Research Questions

For this study we have the following research questions:

1) Does a player's geographical region impact their number of unique champions played?
    - $H_{1,0}$: There is no statistically significant impact on a player's number of unique champions played due to geographical region.
    - $H_{1,A}$: There is a statistically significant impact on a player's number of unique champions played due to geographical region.
2) Does a player's in-game role impact their number of unique champions played?
    - $H_{2,0}$: There is no statistically significant impact on a player's number of unique champions played due to in-game role.
    - $H_{2,A}$: There is a statistically significant impact on a player's number of unique champions played due to in-game role.
3) Does the interaction between a player's in-game role and geographical region impact their number of unique champions played?

- $H_{3,0}$: There is no statistically significant impact on a player's number of unique champions played due the interaction between geographical region and in-game role.
- $H_{3,A}$: There is a statistically significant impact on a player's number of unique champions played due the interaction between geographical region and in-game role.

If we find significant impacts, we will inquire further with the following:

- Which region has the smallest number of unique champions per game?
- Is there a significant difference in unique champions per game between geographical regions in the East vs. in the West?

Exploratory Data Analysis

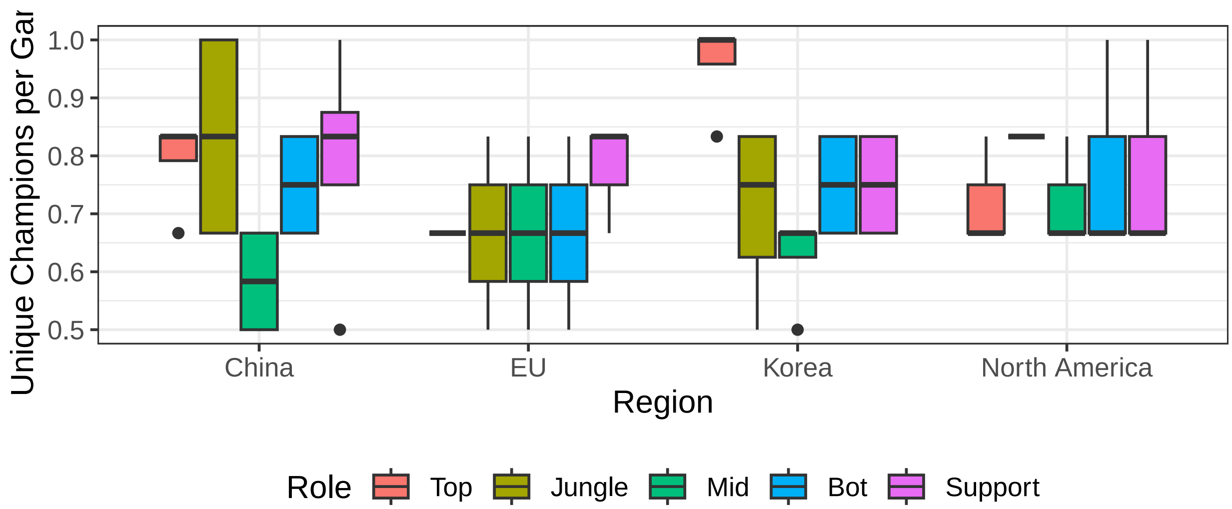

Figure 1: Box Plots of Unique Champions Played

Figure 1 shows the box plots of the twenty observed groups by region and role. We see four outliers (the black dots) in sample minimum from Chinese Top players, Chinese Support players, Korean Top players, and Korean Mid players, by the 1.5-IQR rule. We also see that European players have a very similar sample distribution across roles, with the exception being European supports. On the other hand, Chinese players have multiple differences in observations by role, with Chinese Mid players having the lowest observed unique champions picked. By contrast, North American players seem to outperform the most compared to other regions, as evidenced by their box plots. Although it's difficult to visually decide which role outperforms the most, it seems that Mid players are more conservative in their champion pool.

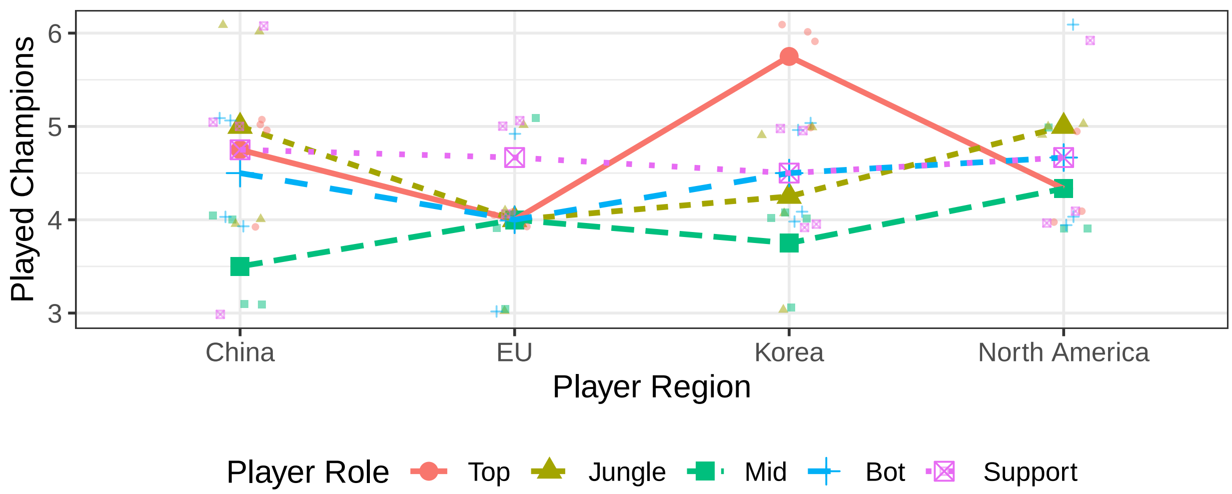

Figure 2: Interaction Plot between Region and Role for Unique Champions Played

Figure 2 displays a scatter plot of the 70 players' observed champions per game, broken down by region and role. The overlaid lines connect the Sample Arithmetic Means across roles for each region. In the scatter plot, we observe a wild overlap in data positions based on region for each role. This is confirmed by the trend lines, which show serious overlap. Based on the non-parallel lines, the data suggests an observed interaction between role and region.

On the other hand, Figure 3 displays an interaction plot between player role and player team. Again, the trend lines show serious overlap which indicates the possibility of an interaction, but this is harder to interpret since only one observation exists per group (only one player per team per role). This analysis will be used for our model decision.

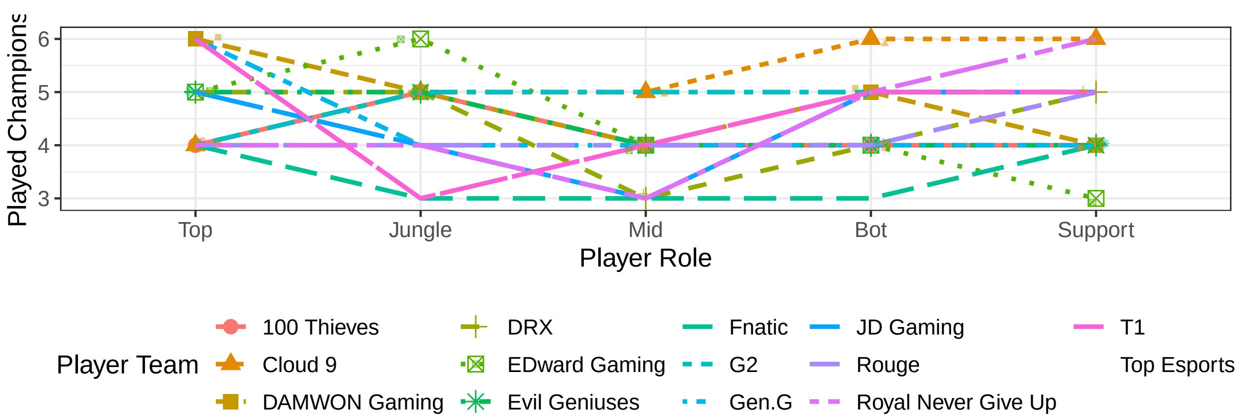

Figure 3: Interaction Plot between Role and Team for Unique Champions Played

Methods

To answer these questions, we pulled data from the group stages of the most recent major tournament, Worlds 2022. Our measurement units are players in the tournament and we are only observing the 6 group stage games of the tournament. Our response is unique champions played throughout these 6 matches, meaning a response of 6 would indicate every game a new champion was played. Our focus is on the four main regions in professional League of Legends: North America, EU, Korea, and China. The data we will be observing are of 20 players from Korea and China and 15 players from North America and EU that were in the tournament. Each player has one of the five roles and stay in that role for the entire tournament.

Analytical Methods

To analyze our data and answer our research questions we will use R (version 4.2.2) and make use of a two-way factorial ANOVA model.

Appropriateness of ANOVA

Our response of interest is quantitative and our two factors (role and region) are categorical.

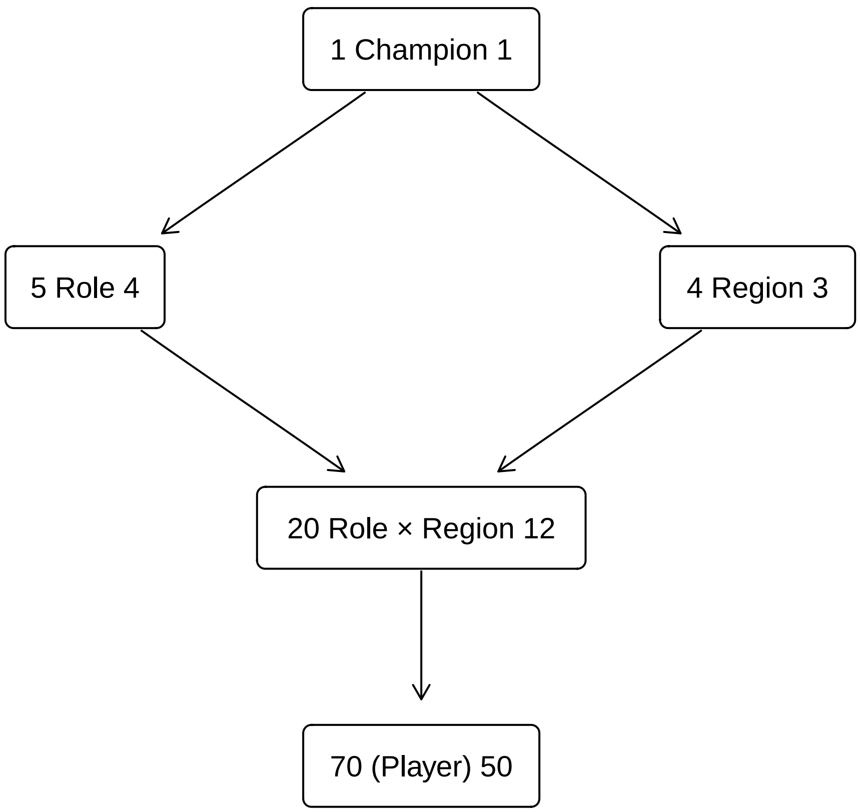

Figure 4: Hasse Diagram

Figure 4 shows the Hasse diagram for our study. We can see our two factors (fixed effects) as well as their interaction. Further, given our sample size of 70 players, we have plenty of degrees of freedom to estimate all main effects, interactions, and error terms. These elements point towards the appropriateness of ANOVA methods in answering our research questions. We will proceed with a two-way ANOVA design since we want to test the effects of both of these factors.

Addressing the Multiple Comparisons Problem

First, we will have an omnibus test on the factor effects of region on the response to answer our main research questions. The following research questions will be answered through post-hoc analysis. Our post-hoc analysis will be comprised of pairwise comparisons as well as contrasts. We will control our False Discovery Rate at a 10% level using the Benjamini-Hochberg method. Each hypothesis test will be evaluated at a 0.1 Unusualness Threshold.

Results

Assumptions

We will be checking three assumptions that, if met, would allow us to use a parametric shortcut for analysis. These assumptions are Gaussian residuals, homoscedasticity, independence of observations.

Based on plots that can be found in the appendix section of the report, we have decided that the assumptions of Gaussian residuals and homoscedaticity have been met by the model. However, we cannot fully guarantee independence of observations. This will be discussed in the limitations section, yet we will proceed with caution.

Omnibus

Table 1 shows the ANOVA table for our two-way ANOVA model, using the parametric shortcut. For our three omnibus research questions, we will focus on the first three rows. From the table we can see that region and the interaction term do not have unusual F ratios. In this case, we fail to reject the null hypothesis for these terms and can move on assuming that they are not statistically significant. The role term appears to have statistical significance (p-value less than our Unusualness Threshold of 0.1), and also appears to have a large effect size in explaining around 15% of the variation in the number of unique champs played.

Table 1: ANOVA Table for Unique Champions Played Study

	Source	SS	df	MS	F	p-value	Partial Omega Sq.	Partial Eta Sq.	Partial Epsilon Sq.
1	Region	2.0881	3	0.6960	1.1196	0.3500	0.0629	0.0051	0.0067
2	Role	5.8405	4	1.4601	2.3487	0.0669	0.1582	0.0716	0.0908
3	Region:Role	8.9714	12	0.7476	1.2026	0.3076	0.2240	0.0336	0.0377
5	Residuals	31.0833	50	0.6217					

Post Hoc Analysis

Given that we have significant results for role as a factor only, we will proceed with the following research questions:

- Which role has the least number of unique champions picked?
- Is there a significant difference in number of unique champions picked between solo laners and bottom lane?
- Is there a significant difference in number of unique champions picked between Jungle and all other roles?

We will break these into two testing families, still controlling FDR using Benjamini-Hochberg at 0.1. Both of these tests will be pairwise comparisons between the levels within the role variable. The first will focus on the pairwise comparisons of role. The second will focus on comparing the solo laners and bottom lane (first grouping) to Jungle and all other roles (second grouping).

Pairwise Comparisons

Table 2: Pairwise Comparisons of Role - Benjamini and Hochberg 90% Adjustment

Comparison	Estimate	SE	DF	t Statistic	p-value
Top - Jungle	0.1458	0.3011	50	0.4843	0.7878
Top - Mid	0.8125	0.3011	50	2.6985	0.0805
Top - Bot	0.2917	0.3011	50	0.9687	0.6747
Top - Support	0.0625	0.3011	50	0.2076	0.8364
Jungle - Mid	0.6667	0.3011	50	2.2141	0.1047
Jungle - Bot	0.1458	0.3011	50	0.4843	0.7878
Jungle - Support	-0.0833	0.3011	50	-0.2768	0.8364
Mid - Bot	-0.5208	0.3011	50	-1.7298	0.2246
Mid - Support	-0.7500	0.3011	50	-2.4909	0.0805
Bot - Support	-0.2292	0.3011	50	-0.7611	0.7503

Table 2 shows the post hoc pairwise comparisons of role using the Benjamini-Hochberg method to control our False Discovery Rate. With our Unusualness Threshold of 0.1, we find that the Mid lane appears to use less unique champions compared to both the Support role and Top lane. Table 3 shows that when we have a Mid lane player and a Top lane player, the Top lane player will pick a higher number of unique champions 76.7% of the time. It also shows that for a Mid lane player and a Support player, the Mid lane player will pick a higher number of unique champions only 25.1% of the time. This shows that there is a large effect of being a Mid lane player compared to Top lane and Support.

Table 3: Effect Sizes for Player Role Comparisons

Comparison	Cohen's d	Probability of Superiority
(Top - Jungle)	0.185	0.552
(Top - Mid)	1.030	0.767
(Top - Bot)	0.370	0.603
(Top - Support)	0.079	0.522
(Jungle - Mid)	0.846	0.725
(Jungle - Bot)	0.185	0.552
(Jungle - Support)	-0.106	0.470
(Mid - Bot)	-0.661	0.320
(Mid - Support)	-0.951	0.251
(Bot - Support)	-0.291	0.419

Contrasts

In League of Legends, each role plays differently in-game. For example, Top and Mid players play by themselves in "lane" (regions of the map that players start in early) while Bot and Support players play together in lane during the early stages of each game. Top and Mid are called "solo laners" while Bot and Support are grouped as "bot lane." Additionally, Jungle players specifically play in "jungle" regions of the map (everywhere outside of lanes, essentially) while the other four roles restrict themselves to lanes in the beginning. We can separate their roles as "jungler" and "laners." These two behaviors motivate our contrasts.

Table 4: ANOVA Table for Unique Champions Played with Role Contrasts

	SS	DF	F	Raw p-value	Adj. p-value
Region	2.0881	3	0.9388	0.4278	0.4278
T and M vs. B and S	0.7202	1	0.9715	0.3284	1.0000
J vs. T, M, B, and S	0.2333	1	0.3147	0.5770	1.0000
Region * Role: T and M vs. B and S	1.2738	3	0.5727	0.6352	1.0000
Region * Role: J vs. T, M, B, and S	2.1381	3	0.9613	0.4172	1.0000
Residuals	43.0000	58			

Table 4 displays the results of our post-hoc analysis with contrasts using Benjamini & Hochberg. T, J, M, B, S stand for Top, Jungle, Mid, Bot, and Support roles, respectively. In our contrasts, every single F-ratio is below one, which means that no comparison contributed more variation to the number of unique champions played than the residuals. In context, this means that all other factors that are left unexplained contribute just as much or more to the variation in each comparison, so the effect of role is barely seen. After adjusting our p-values, they actually come out to be 1 each (or a value very close to 1). Thus, we decide to fail to reject both null hypotheses. As a result, we observe no difference in number of unique champions picked between solo laners and bot lane, and no difference between junglers and laners.

Discussion

In our omnibus test we found role to be a significant factor in our model but we did not make that same discovery for region or the interaction between region and role. In our post hoc analysis we found that Mid lane players had a less diverse champion pool than both Top lane players and Support players and that these differences had large effect sizes. However, we did not find that Mid lane players had a less diverse champion pool than every other role. We did not find a significant difference between solo lane players and bottom lane players or between Jungle players and lane players.

Limitations

Regarding the assumption of independence of observations, the assessment becomes difficult because of the complexity of the game, and we have no provided measurement order to assess graphically. There are a few problems, two of which are:

1. When a player picks a champion in a game, they are picked without replacement, i.e. no two players can play the same champion.
2. Usually in a tournament, different teams play a different amount of games based on the elimination bracket.

We can address these two concerns. For the first problem, there exists a "meta" pool of champions per role; Top players play champions that a Bot player would not play, and the same is true for any two roles. By using role as a factor, we indirectly account for this problem. For the second problem, we restrict our study to the group stage of the tournament, where each team plays six games. This introduces a natural, uniform scope of analysis to our study. There were other model considerations to get around this issue.

First, if we blocked by player team, it could help account for independence between players, since it helps deal with nuisance factors like the number of games played and champion picks without replacement. However, this option isn't ideal because of our other factors: player region and player role. If we consider all three, we only have one player per region per team per role, which isn't enough data for analysis. If we consider a model with region and team, we do have more observations to work with, but teams are unique to their region. For example, one team in Korea would not have a sister team in North America. Although they could be considered as nested factors, we believe that this model is awkward to interpret and not worth analysis, since role is an important factor in deciding which champions a player chooses. If we consider role and team, we now have one player per team per role, which again isn't enough data per group to warrant analysis. Team can't be treated as a block since it has interactions with role. Thus, we decided to proceed with a factorial model using region and role.

References

Data from: https://liquipedia.net/leagueoflegends/World_Championship/2022/Group_Stage

Appendix

Additional EDA

Table 5: Summary Statistics for Unique Champions Played by Region

Region	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
China	20	3	4	4.5	5	6	0.74	4.50	0.95	0.00	-1.05
EU	15	3	4	4.0	5	5	1.48	4.13	0.74	-0.18	-1.30
Korea	20	3	4	4.5	5	6	0.74	4.55	0.89	0.07	-0.91
North America	15	4	4	4.0	5	6	0.00	4.60	0.74	0.68	-1.00

Table 5 shows the values for descriptive statistics of played champions per game broken down by the major regions: China, the European Union, South Korea, and North America. This describes the region that the player's team competes in during regional splits. Notably, the number of players per region is not the same across the regions, so we have to be careful about our assumptions and the type of sum of squares we use.

There are a few things to note from Table 5. At first glance, the quantiles seem similar between regions, as are Sample Arithmetic Mean (SAM), Sample Arithmetic Standard Deviation (SASD), and Sample Kurtosis. However, the North American players are much more skewed right in terms of absolute champion pool than the others, indicating that they have more observed unique champions picked. Additionally, the Median Absolute Deviation for EU players is higher than the other regions, which indicates that the distance of half the observations from the median is higher for EU players compared to other regions. The reason behind this will be addressed when looking at our box plot.

Table 6: Summary Statistics for Unique Champions Played by Role

Role	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
Top	14	4	4.00	5	5	6	1.48	4.79	0.80	0.35	-1.48
Jungle	14	3	4.00	5	5	6	1.48	4.57	0.94	-0.19	-1.04
Mid	14	3	3.25	4	4	5	0.00	3.86	0.66	0.12	-0.94
Bot	14	3	4.00	4	5	6	0.74	4.43	0.76	0.22	-0.59
Support	14	3	4.00	5	5	6	1.48	4.64	0.84	-0.06	-0.86

Table 6 breaks down the descriptive statistics for unique observed champion picks by role. There are always five players per team, each with a unique role (Top, Jungle, Mid, Bot, Support), which is why the role factor is balanced (confirmed by n).

Similar to Table 5, we have similar statistics across all roles, save two. The MAD of Mid players is 0 unique champions, which indicates that the difference in unique champions picked for Mid players is extremely low (i.e. similar values observed). Morever, Top, Mid, and Bot champion picks are more skewed right while Jungle and Support champion picks are slightly skewed left. Although the skewness is not large, Top in particular stands out at 0.35 compared to Jungle's -0.19 sample skewness, which may indicate that Top players tend to pick more unique champions.

Table 7: Summary Statistics for Champions per Game by Region and Role

Region	Role	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
China	Top	4	4	4.75	5.0	5.00	5	0.00	4.75	0.50	-0.75	-1.69
EU	Top	3	4	4.00	4.0	4.00	4	0.00	4.00	0.00		
Korea	Top	4	5	5.75	6.0	6.00	6	0.00	5.75	0.50	-0.75	-1.69
North America	Top	3	4	4.00	4.0	4.50	5	0.00	4.33	0.58	0.38	-2.33
China	Jungle	4	4	4.00	5.0	6.00	6	1.48	5.00	1.15	0.00	-2.44
EU	Jungle	3	3	3.50	4.0	4.50	5	1.48	4.00	1.00	0.00	-2.33
Korea	Jungle	4	3	3.75	4.5	5.00	5	0.74	4.25	0.96	-0.32	-2.08
North America	Jungle	3	5	5.00	5.0	5.00	5	0.00	5.00	0.00		
China	Mid	4	3	3.00	3.5	4.00	4	0.74	3.50	0.58	0.00	-2.44
EU	Mid	3	3	3.50	4.0	4.50	5	1.48	4.00	1.00	0.00	-2.33
Korea	Mid	4	3	3.75	4.0	4.00	4	0.00	3.75	0.50	-0.75	-1.69
North America	Mid	3	4	4.00	4.0	4.50	5	0.00	4.33	0.58	0.38	-2.33
China	Bot	4	4	4.00	4.5	5.00	5	0.74	4.50	0.58	0.00	-2.44
EU	Bot	3	3	3.50	4.0	4.50	5	1.48	4.00	1.00	0.00	-2.33
Korea	Bot	4	4	4.00	4.5	5.00	5	0.74	4.50	0.58	0.00	-2.44
North America	Bot	3	4	4.00	4.0	5.00	6	0.00	4.67	1.15	0.38	-2.33
China	Support	4	3	4.50	5.0	5.25	6	0.74	4.75	1.26	-0.42	-1.82
EU	Support	3	4	4.50	5.0	5.00	5	0.00	4.67	0.58	-0.38	-2.33
Korea	Support	4	4	4.00	4.5	5.00	5	0.74	4.50	0.58	0.00	-2.44
North America	Support	3	4	4.00	4.0	5.00	6	0.00	4.67	1.15	0.38	-2.33

Table 7 shows the values for various statistics broken out by the player's role and region. Again, we currently do not have a balanced design because of an unequal number of teams from each reigon, which makes a choice of Sum of Squares and assumptions slightly more challenging.

Looking at the descriptive statistics, we see multiple combinations of Region and Role have 0 MAD, which indicates, based on the quantiles, that there are not many unique observations in them (e.g. Chinese Top players, European Support players, etc). Most notably, European Top players and North American Jungle players have a SASD and MAD of 0 champions, and their quantiles are the same, which indicates that they all picked the same amount of unique champions. This may pose a challenge during analysis. Excluding those two combinations, there are many varying values of skewness and SASD, which may indicate that the number of unique champions played per player is dependent on their role and team region.

Assumption Plots

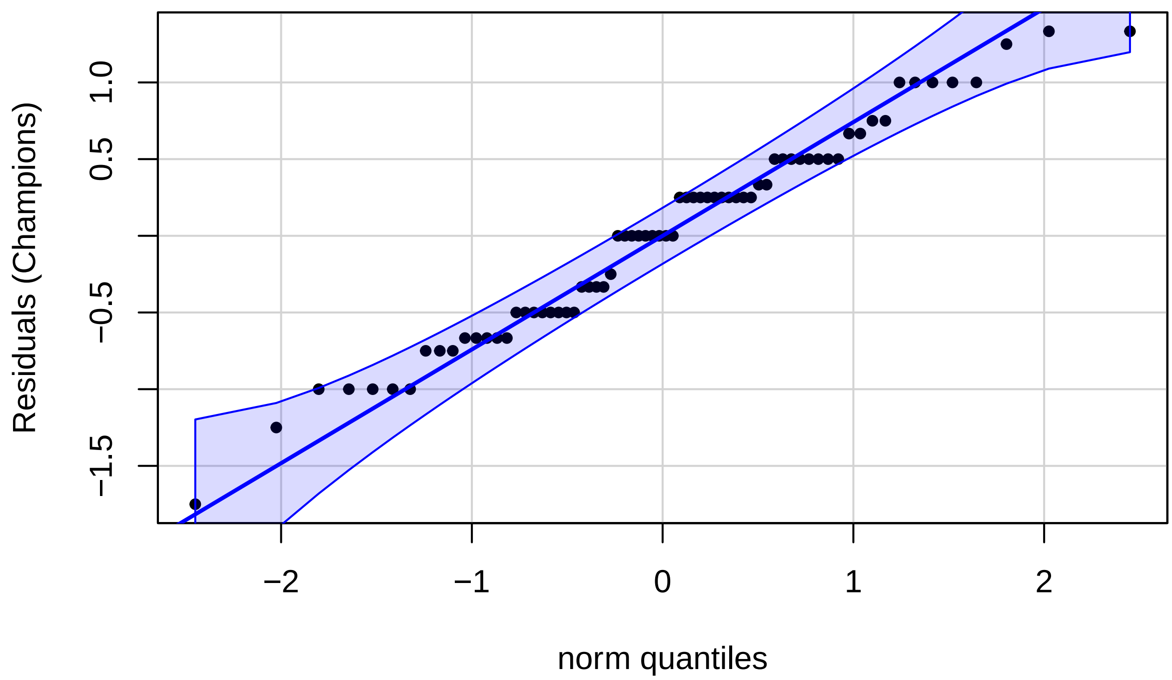

Figure 5: Normal QQ Plot for Unique Champions Played

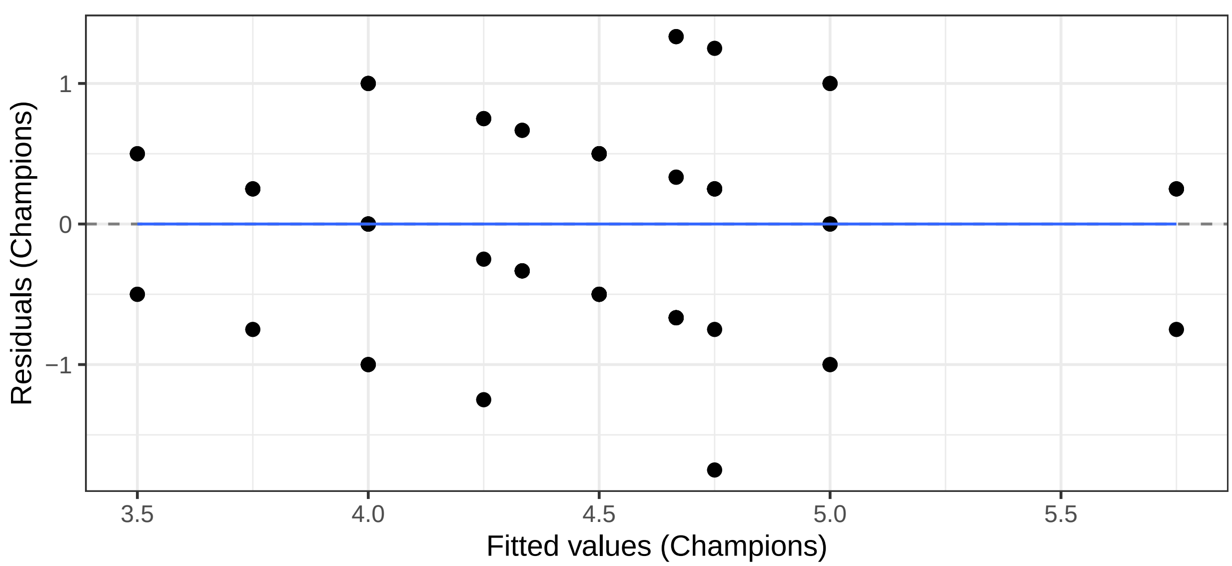

Figure 6: Tukey-Anscombe Plot for Unique Champions Played

Additional Model Information

$$Y_{ijk} = \mu_{\dots} + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk} \tag{1}$$

Equation 1 and Figure 4 both represent the fixed-effects, two-way ANOVA model we are interested in. In particular, the $\mu_{\dots}$ represents the baseline effect of playing professional League of Legends games, α_i represents the main effect of role, β_j represents the main effect of region, and $\alpha\beta_{ij}$ the interaction of role and region. Our residual/error term (ϵ_{ijk}) represents the variability present within the indiviudal professional players and any other sources of variability that we do not account for in our model.

Additional Contrasts Information

Let $\mu_T, \mu_J, \mu_M, \mu_B, \mu_S$ describe the true population sample arithmetic mean of Top, Jungle, Mid, Bot, and Support players, respectively. Note that our contrasts are limited to the Role factor since the observations are balanced and help with orthogonality.

Our first contrast examines the difference in the number of unique champions picked between Top-Mid players and Bot-Support players, group by their early-game playstyle. Thus, the contrast is set up as $(-\frac{1}{2}, 0, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Our hypotheses are:

$$H_0^{TM-BS} : \frac{\mu_T + \mu_M}{2} = \frac{\mu B + \mu_S}{2}$$
$$H_1^{TM-BS} : \frac{\mu_T + \mu_M}{2} \neq \frac{\mu B + \mu_S}{2}$$

Our second contrast examines the difference in the number of unique champions picked between Jungle players and all other roles, again grouped by their early-game playstyle. Thus, the contrast is set up as $(\frac{1}{4}, -1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$. Our hypotheses are:

$$H_0^J : \frac{\mu_T + \mu_M + \mu_B + \mu_S}{4} = \mu_J$$
$$H_1^J : \frac{\mu_T + \mu_M + \mu_B + \mu_S}{4} \neq \mu_J$$

Both contrasts sum to zero, fulfilling the requirement of the contrast. Additionally, the weighted sum of both contrasts is 0, so we know that our contrasts are orthogonal, easing our analysis. Recall that since Region is unbalanced we use Type III Sum of Squares, and since we are using FDR at level 0.1, we must adjust the p-values using Benjamini & Hochberg.

Code Appendix

```r
# Setting Document Options ----
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center"
)

# Load Packages ----
packages <- c("tidyverse", "knitr", "kableExtra",
```

```r
              "parameters", "car", "hasseDiagram",
              "psych", "DescTools", "emmeans", "openxlsx")
invisible(lapply(X = packages, FUN = library, character.only = TRUE, quietly = TRUE))

# Setting Global Options ----
options(knitr.kable.NA = "")
options("contrasts" = c("contr.sum", "contr.poly"))

# Load Additional Tools ----
source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")
source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")

# Read in Data
data <- read_csv("data.csv")
data_mod <- data %>%
  dplyr::select(-Player) %>%
  mutate(Region=as.factor(Region),
         Role=factor(Role, levels=c("Top", "Jungle","Mid", "Bot", "Support")),
         Team=as.factor(Team))

data_num <- data_mod %>%
  mutate(Region=as.numeric(Region),
         Role=as.numeric(Role),
         Team=as.numeric(Team))

data_major <- data_mod %>%
  filter(Region %in% c("China", "EU", "Korea", "North America")) %>%
  mutate(Region=factor(Region, levels=c("China", "EU", "Korea", "North America")),
         ChampsPlayed=ChampsPerGame*6)

# Create side-by-side box plots of the distances ----
ggplot(
  data = data_major,
  mapping = aes(
    x = Region,
    y = ChampsPerGame,
    fill = Role,
  )
) +
  geom_boxplot() +
  theme_bw() +
  xlab("Region") +
  ylab("Unique Champions per Game") +
  labs(
    fill = "Role"
  ) +
  theme(
    legend.position = "bottom",
    text = element_text(size = 12)
  )
  # scale_fill_manual(
  #   values = boastUtils::boastPalette
  # )
```

```r
# Create an interaction plot ----
ggplot(
  data = data_major,
  mapping = aes(
    x = Region,
    y = ChampsPlayed,
    shape = Role,
    color = Role,
    linetype = Role,
    group = Role
  )
) +
  stat_summary(fun = "mean", geom = "point", size = 3) +
  stat_summary(fun = "mean", geom = "line", linewidth = 1) +
  geom_jitter(width = 0.1, height = 0.1, alpha = 0.5, size = 1) +
  ggplot2::theme_bw() +
  xlab("Player Region") +
  ylab("Played Champions") +
  labs(
    color = "Player Role",
    shape = "Player Role",
    linetype = "Player Role"
  ) +
  theme(
    legend.position = "bottom",
    text = element_text(size = 12)
  )

# Create an interaction plot ----
ggplot(
  data = data_major,
  mapping = aes(
    x = Role,
    y = ChampsPlayed,
    shape = Team,
    color = Team,
    linetype = Team,
    group = Team
  )
) +
  stat_summary(fun = "mean", geom = "point", size = 3) +
  stat_summary(fun = "mean", geom = "line", linewidth = 1) +
  geom_jitter(width = 0.1, height = 0.1, alpha = 0.5, size = 1) +
  ggplot2::theme_bw() +
  xlab("Player Role") +
  ylab("Played Champions") +
  labs(
    color = "Player Team",
    shape = "Player Team",
    linetype = "Player Team"
  ) +
  theme(
    legend.position = "bottom",
```

```
    text = element_text(size = 12)
  )


modelLabels <- c("1 Champion 1", "5 Role 4", "4 Region 3", "20 Role × Region 12", "70 (Player) 50")
modelMatrix <- matrix(
  data = c(FALSE, FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALS
  nrow = 5,
  ncol = 5,
  byrow = FALSE
)
hasseDiagram::hasse(
 data = modelMatrix,
 labels = modelLabels
)
# Fit model
model <- aov(
  formula = ChampsPlayed ~ Region*Role,
  data = data_major
)
parameters::model_parameters(
  model = model,
  effectsize_type = c("eta", "omega", "epsilon"),
  type = 3, # We used type 3
  drop = "(Intercept)",
  verbose = FALSE
  ) %>%
  dplyr::mutate(
  p = ifelse(
  test = is.na(p),
  yes = NA,
  no = pvalRound(p)
  )
) %>%
  knitr::kable(
  digits = 4,
  col.names = c("Source", "SS", "df", "MS", "F", "p-value",
               "Partial Omega Sq.", "Partial Eta Sq.", "Partial Epsilon Sq."),
  caption = "ANOVA Table for Unique Champions Played Study",
  align = c('l',rep('c',8)),
  booktab = TRUE,
) %>%
  kableExtra::kable_styling(
  bootstrap_options = c("striped", "condensed"),
  font_size = 12,
  latex_options = c("scale_down", "HOLD_position")
)
champsRole <- emmeans::emmeans(
  object = model,
  specs = pairwise ~ Role,
  adjust = "bh", # Where you specify your chosen method
  level = 0.9 # 1--Type I Risk
)
```

```r
as.data.frame(champsRole$contrasts) %>%
  knitr::kable(
  digits = 4,
  col.names = c("Comparison", "Estimate","SE", "DF",
              "t Statistic", "p-value"),
  caption = "Pairwise Comparisons of Role - Benjamini and Hochberg 90\\% Adjustment",
  align = rep("c", 7),
  booktabs = TRUE
) %>%
  kableExtra::kable_styling(
  bootstrap_options = c("striped", "condensed"),
  font_size = 12,
  latex_options = c("HOLD_position")
)
as.data.frame(
  eff_size(
    object = champsRole,
    sigma = sigma(model),
    edf = df.residual(model)
  )
) %>%
  dplyr::mutate(
    ps = probSup(effect.size),
    .after = effect.size
  ) %>%
  dplyr::select(contrast, effect.size, ps) %>%
  knitr::kable(
    digits = 3,
    col.names = c("Comparison",
                  "Cohen's d", "Probability of Superiority"),
    align = "lccc",
    caption = "Effect Sizes for Player Role Comparisons",
    booktab = TRUE
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("striped", "condensed"),
    font_size = 12,
    latex_options = "HOLD_position"
  )
c1 <- c(-1/2, 0, -1/2, 1/2, 1/2)
c2 <- c(1/4, -1, 1/4, 1/4, 1/4)

contrasts(data_major$Role) <- cbind(c1, c2)

### Redefine model ###
contrasts_car_model <- car::Anova(
  lm(ChampsPlayed ~ Region*(C(Role, c1, 1)+C(Role, c2, 1)), data=data_major),
  type=3  # Type 3 SS
)

# Adjust row names
rownames(contrasts_car_model) = c("(Intercept)",
                                  "Region",
```

```
                                   "T and M vs. B and S",
                                   "J vs. T, M, B, and S",
                                   "Region * Role: T and M vs. B and S",
                                   "Region * Role: J vs. T, M, B, and S",
                                   "Residuals")

# Manually set adjusted p values
contrasts_car_model$adj_p <- contrasts_car_model$`Pr(>F)`
contrasts_car_model[3, "adj_p"] <- p.adjust(contrasts_car_model[3, "adj_p"], method="fdr", 4)
contrasts_car_model[4, "adj_p"] <- p.adjust(contrasts_car_model[4, "adj_p"], method="fdr", 4)
contrasts_car_model[5, "adj_p"] <- p.adjust(contrasts_car_model[5, "adj_p"], method="fdr", 4)
contrasts_car_model[6, "adj_p"] <- p.adjust(contrasts_car_model[6, "adj_p"], method="fdr", 4)


contrasts_car_model[-1, ] %>%  # Remove intercept row
  knitr::kable(
    digits=4,
    col.names=c("SS", "DF", "F", "Raw p-value", "Adj. p-value"),
    caption="ANOVA Table for Unique Champions Played with Role Contrasts",
    booktabs=T,
    align=rep("c", 4)
  ) %>%
  kableExtra::kable_styling(
    font_size=12,
    latex_options=c("HOLD_position")
  )
# Produce values of descriptive statistics
region_stats <- psych::describeBy(
  data_major$ChampsPlayed,
  group=data_major$Region,
  # ChampsPerGame ~ Region,
  # data = data_mod,
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = TRUE,
  mat = TRUE,
  digits = 4
)

# Create a table of values ----
region_stats %>%
  tibble::remove_rownames() %>%
  dplyr::select(
    group1, n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
  ) %>%
  knitr::kable(
    caption = "Summary Statistics for Unique Champions Played by Region",
    digits = 2,
    format.args = list(big.mark = ","),
    align = rep(c("l", "c"), times = c(2, 11)),
    col.names = c("Region", "n", "Min", "Q1", "Median", "Q3", "Max", "MAD",
```

```
                  "SAM", "SASD", "Sample Skew", "Sample Ex. Kurtosis"),
    booktabs = TRUE
  )  %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("HOLD_position", "scale_down")
  )
# Produce values of descriptive statistics
role_stats <- psych::describeBy(
  data_major$ChampsPlayed,
  group=data_major$Role,
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = TRUE,
  mat = TRUE,
  digits = 4
)

# Create a table of values ----
role_stats %>%
  tibble::remove_rownames() %>%
  dplyr::select(
    group1, n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
  ) %>%
  knitr::kable(
    caption = "Summary Statistics for Unique Champions Played by Role",
    digits = 2,
    format.args = list(big.mark = ","),
    align = rep(c("l", "c"), times = c(2, 11)),
    col.names = c("Role", "n", "Min", "Q1", "Median", "Q3", "Max", "MAD",
                  "SAM", "SASD", "Sample Skew", "Sample Ex. Kurtosis"),
    booktabs = TRUE
  )  %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("HOLD_position", "scale_down")
  )
# Produce values of descriptive statistics
two_descp_stats <- psych::describeBy(
  data_major$ChampsPlayed,
  group=list(data_major$Region, data_major$Role),
  # ChampsPerGame ~ Region + Role,
  # data = data_mod,
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = TRUE,
  mat = TRUE,
  digits = 4
)
```

```r
# Create a table of values ----
two_descp_stats %>%
  tibble::remove_rownames() %>%
  dplyr::select(
    group1, group2, n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
    # group1, group2, n., min., median., max., mad., mean., sd., skew., kurtosis.
  ) %>%
  as.data.frame() %>%
  knitr::kable(
    caption = "Summary Statistics for Champions per Game by Region and Role",
    digits = 2,
    format.args = list(big.mark = ","),
    align = rep(c("l", "c"), times = c(2, 11)),
    col.names = c(# "Region", "Team", "n", "Min", "Median", "Max", "MAD",
                  "Region", "Role", "n", "Min", "Q1", "Median", "Q3", "Max", "MAD",
                  "SAM", "SASD", "Sample Skew", "Sample Ex. Kurtosis"),
    booktabs = TRUE
  )  %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("HOLD_position", "scale_down")
  )

# Generate the qq plot ----
car::qqPlot(
  x = residuals(model),
  distribution = "norm",
  envelope = 0.90,
  id = FALSE,
  pch = 20,
  ylab = "Residuals (Champions)"
)

# Generate the Tukey-Anscombe plot ----
ggplot(
  data = data.frame(
    residuals = residuals(model),
    fitted = fitted.values(model)
  ),
  mapping = aes(x = fitted, y = residuals)
) +
  geom_point(size = 2) +
  geom_hline(
    yintercept = 0,
    linetype = "dashed",
    color = "grey50"
  ) +
  geom_smooth(
    formula = y ~ x,
    method = stats::loess,
    method.args = list(degree = 1),
    se = FALSE,
    linewidth = 0.5
```

```r
  ) +
  theme_bw() +
  xlab("Fitted values (Champions)") +
  ylab("Residuals (Champions)")
```