

Yield Trial Data Analysis

Contents

Install and load packages	1
Data Preparation	2
Data Overview	2
Dataset Structure	2
Missing Values	2
Statistical Analysis	3
Overall Statistics	3
By Genotype	3
By Environment	4
Data Visualization	5
Yield Distribution	5
Genotype Performance	5
Quality Control	6
Normality Test	6
Outlier Detection	7
Replicate Correlation	7

Install and load packages

```
suppressMessages({  
  if (!require("readxl")) install.packages("readxl")  
  if (!require("tidyverse")) install.packages("tidyverse")  
  if (!require("ggpubr")) install.packages("ggpubr")  
  if (!require("knitr")) install.packages("knitr")  
  if (!require("kableExtra")) install.packages("kableExtra")  
  
  library(readxl)  
  library(tidyverse)  
  library(ggpubr)
```

```
library(knitr)
library(kableExtra)
})
```

Data Preparation

```
field_data <- read_excel("field_data.xlsx") %>%
  mutate(
    Genotype = as.factor(Genotype),
    Environment = as.factor(Environment),
    Rep = as.factor(Rep)
  )
```

Data Overview

Dataset Structure

```
cat("\\textbf{Dataset dimensions:} ", paste(dim(field_data), collapse = " x "), "\\n\\n")
```

Dataset dimensions: 900 x 4

```
# Custom function to format glimpse output for PDF
format_glimpse <- function(data) {
  glimpse_out <- capture.output(glimpse(data))
  cat(paste(glimpse_out, collapse = "\\n"))
}

format_glimpse(field_data)
```

Rows: 900 Columns: 4 \$ Genotype G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1, G1~ \$ Environment E1, E1, E1, E2, E2, E2, E3, E3, E3, E4, E4, E4, E5, E5, E5~ \$ Rep 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2~ \$ Yield 7.721363, 8.098298, 8.735440, 10.665111, 10.135101, 10.532~

Missing Values

```
cat("\\textbf{Missing values:}\\n\\n")
```

Missing values:

```
colSums(is.na(field_data))
```

Genotype Environment Rep Yield 0 0 0 0

Statistical Analysis

Overall Statistics

```
field_data %>%
  summarise(
    Observations = n(),
    Mean = mean(Yield),
    SD = sd(Yield),
    Min = min(Yield),
    Max = max(Yield),
    CV = (SD/Mean)*100
  ) %>%
  kable(caption = "Overall Statistics", digits = 3, booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 1: Overall Statistics

Observations	Mean	SD	Min	Max	CV
900	10.114	1.679	5.453	15.691	16.597

By Genotype

```
field_data %>%
  group_by(Genotype) %>%
  summarise(
    Mean = mean(Yield),
    SD = sd(Yield),
    CV = (SD/Mean)*100,
    .groups = 'drop'
  ) %>%
  arrange(desc(Mean)) %>%
  kable(caption = "By Genotype", digits = 3, booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 2: By Genotype

Genotype	Mean	SD	CV
G16	12.620	1.180	9.348
G6	11.847	0.939	7.928
G3	11.758	1.027	8.732
G11	11.442	1.310	11.452
G17	10.596	0.951	8.979
G19	10.586	1.183	11.171
G7	10.422	1.208	11.589
G14	10.406	1.189	11.425
G13	10.378	1.181	11.376

G12	10.328	1.202	11.643
G4	10.164	1.095	10.771
G5	9.791	1.299	13.269
G2	9.722	1.209	12.438
G10	9.438	1.132	11.990
G1	9.426	1.174	12.453
G15	9.318	1.249	13.405
G20	9.257	1.214	13.115
G9	9.248	1.492	16.137
G8	8.011	1.245	15.536
G18	7.523	1.044	13.881

By Environment

```
field_data %>%
  group_by(Environment) %>%
  summarise(
    Mean = mean(Yield),
    SD = sd(Yield),
    CV = (SD/Mean)*100,
    .groups = 'drop'
  ) %>%
  arrange(desc(Mean)) %>%
  kable(caption = "By Environment", digits = 3, booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

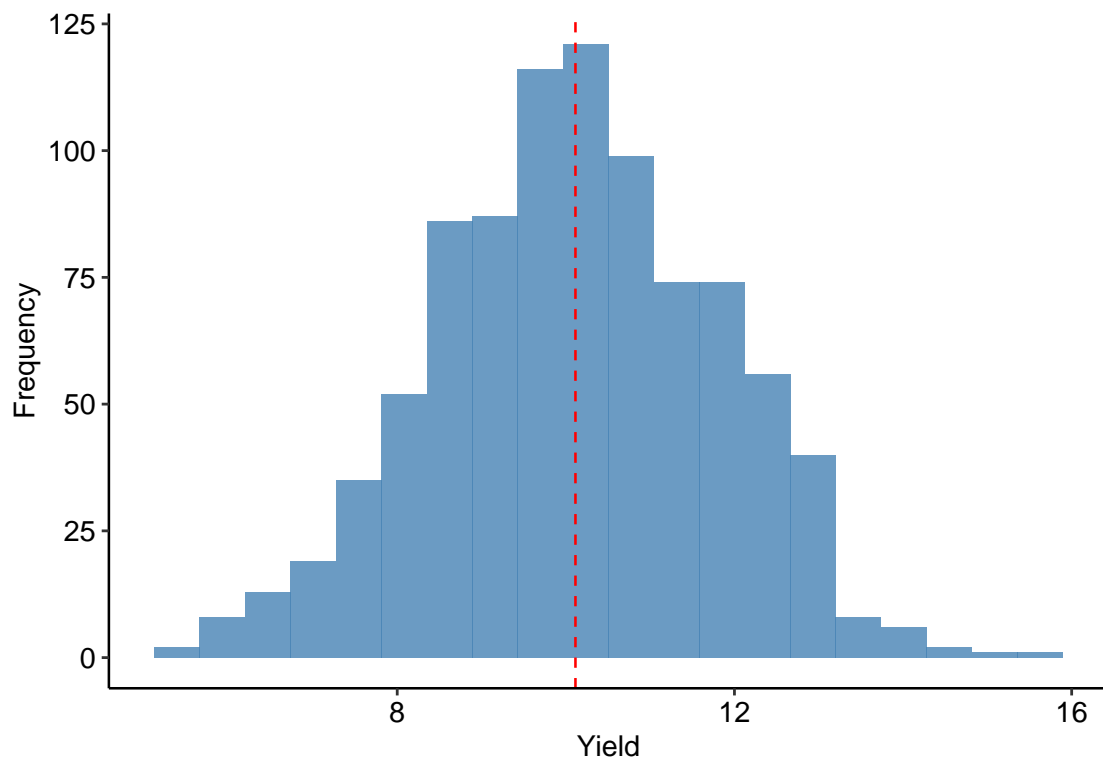
Table 3: By Environment

Environment	Mean	SD	CV
E10	11.268	1.493	13.247
E13	11.138	1.259	11.305
E7	11.128	1.500	13.481
E15	11.094	1.561	14.070
E14	11.047	1.661	15.036
E11	10.421	1.379	13.236
E8	10.124	1.432	14.141
E12	9.981	1.395	13.975
E2	9.932	1.453	14.626
E5	9.507	1.427	15.012
E3	9.459	1.296	13.705
E4	9.430	1.473	15.616
E1	9.357	1.474	15.747
E9	9.217	1.470	15.950
E6	8.606	1.724	20.037

Data Visualization

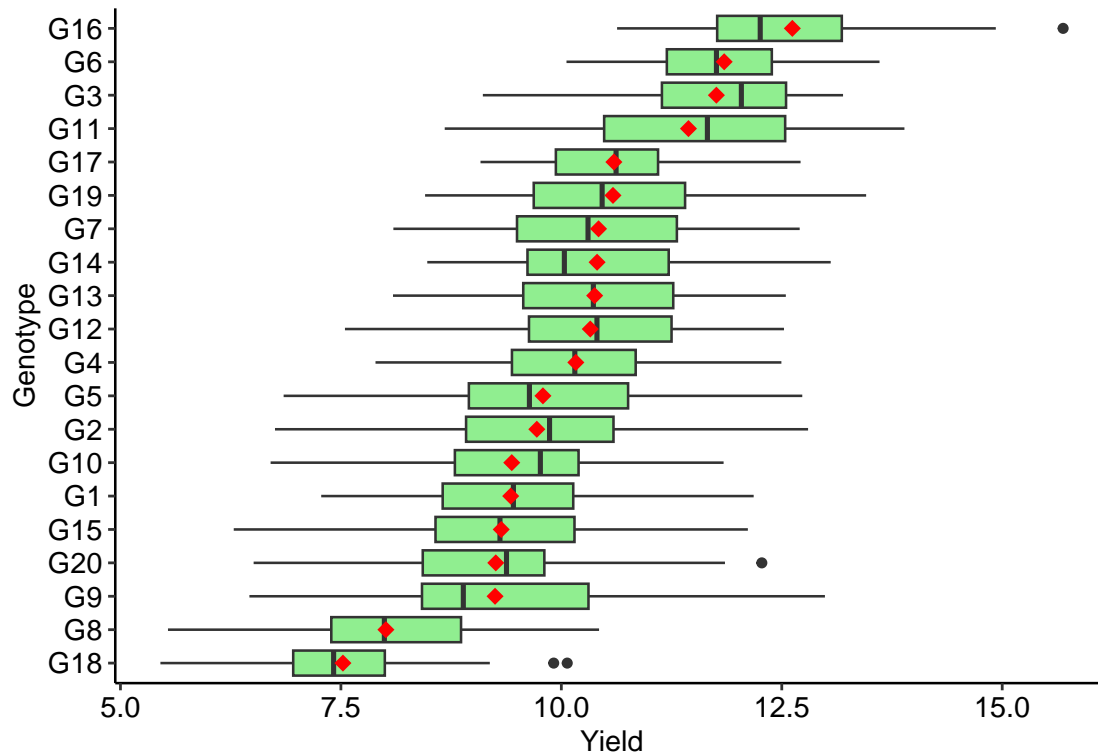
Yield Distribution

```
ggplot(field_data, aes(Yield)) +  
  geom_histogram(bins = 20, fill = "steelblue", alpha = 0.8) +  
  geom_vline(aes(xintercept = mean(Yield)), color = "red", linetype = "dashed") +  
  labs(x = "Yield", y = "Frequency") +  
  theme_pubr()
```



Genotype Performance

```
ggplot(field_data, aes(reorder(Genotype, Yield, mean), Yield)) +  
  geom_boxplot(fill = "lightgreen") +  
  stat_summary(fun = mean, geom = "point", shape = 18, size = 3, color = "red") +  
  labs(x = "Genotype", y = "Yield") +  
  coord_flip() +  
  theme_pubr()
```



Quality Control

Normality Test

```
shapiro_result <- shapiro.test(field_data$Yield) %>%
  broom::tidy()

shapiro_result %>%
  kable(caption = "Normality Test (Shapiro-Wilk)", digits = 4, booktabs = TRUE) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 4: Normality Test (Shapiro-Wilk)

statistic	p.value	method
0.9979	0.3257	Shapiro-Wilk normality test

```
if(shapiro_result$p.value > 0.05) {
  cat("Interpretation: The data follows a normal distribution (p > 0.05).")
} else {
  cat("Interpretation: The data does NOT follow a normal distribution (p > 0.05).")
}
```

Interpretation: The data follows a normal distribution (p > 0.05).

Outlier Detection

```
field_data %>%
  group_by(Genotype, Environment) %>%
  mutate(z_score = scale(Yield)) %>%
  filter(abs(z_score) > 3) %>%
  {
    if(nrow(.) > 0)
      kable(., caption = "Potential Outliers", booktabs = TRUE) %>%
      kable_styling(latex_options = c("striped", "hold_position"))
    else
      cat("No extreme outliers detected (z-score > 3).")
  }
```

No extreme outliers detected (z-score > 3).

Replicate Correlation

```
if(length(unique(field_data$Rep)) > 1) {
  field_data %>%
    pivot_wider(names_from = Rep, values_from = Yield, names_prefix = "Rep_") %>%
    select(starts_with("Rep_")) %>%
    cor(use = "pairwise.complete.obs") %>%
    kable(caption = "Replicate Correlation", digits = 3, booktabs = TRUE) %>%
    kable_styling(latex_options = c("striped", "hold_position"))
}
```

Table 5: Replicate Correlation

	Rep_1	Rep_2	Rep_3
Rep_1	1.000	0.883	0.875
Rep_2	0.883	1.000	0.901
Rep_3	0.875	0.901	1.000