

Vorregistrierung der Abschlussarbeit im Empirischen Projektmodul

Kleingruppe: A

Vor- und Nachnamen der Studierenden: Patrica Hünemörder, Emma Neu, Celina Kuhn, Julia Häußer

Datum: 17.04.2025

1) Hypothese: Welche Hauptfragestellung/-hypothese soll in der Arbeit beantwortet werden?

Gibt es einen Zusammenhang zwischen interindividuellen Unterschieden in Antizipationseffekten beim Lesen und der individuellen Exposition gegenüber Druckmedien (ART-Leistungen)?

2) Abhängige Variable: Beschreiben Sie die abhängige(n) Variable(n). Wie wird/werden diese gemessen? Auf welcher Skala oder in welcher Einheit?

Reaktionszeitdifferenz bei der Antizipation vorhersagbarer und unvorhersagbarer Wörter.
(Reaktionszeitdifferenz = unvorhersagbar – vorhersagbar)

- Gemessen in Trainingsblock (A Training, „Train them to predict a lot) im Testblock mit hoher Validität, in dem die Vorhersage strategisch sinnvoll ist
- Intervallskaliert in Sekunden

3) Unabhängigen Variablen: Beschreiben Sie die unabhängige(n) Variable(n). Wie werden diese gemessen (kategorisch oder kontinuierlich)? Bei faktoriellen Variablen, geben Sie die Namen des Faktors und die Namen der Stufen an. Geben Sie für jede Variable an, ob sie *within-* oder *between-subjects* gemessen wird!

- individuelle Exposition gegenüber Druckmedien (gemessen durch ART-Leistungen)
- kontinuierlich gemessen; Absolutskaliert
- within-subject

4) Analysen: Spezifizieren Sie exakt, welche Analysen bzw. statistischen Tests Sie vornehmen möchten um die Hauptfragestellung/-hypothese zu untersuchen.

Produkt-Moment-Korrelation nach Pearson (zweiseitige Testung)

5) Stichprobengröße: Geben Sie an, wie sie die Stichprobenumfangsplanung durchgeführt haben. Wie viele Probanden benötigen Sie um die Hauptfragestellung zu beantworten?

1. Durchführung einer Poweranalyse (zweiseitiger Test) mit der Funktion `pwr.r.test` für eine Power von 0.8 mit dem Signifikanzlevel 0.05 und einer Effektgröße von $r = 0.25$
2. Durchführung einer Poweranalyse (zweiseitiger Test) mit der Funktion `pwr.r.test` für eine Power von 0.8 mit dem Signifikanzlevel 0.05 und einer Effektgröße von $r = 0.15$
3. Visualisierung des Powerverlaufs in Abhängigkeit der Stichprobengröße bei einer Effektstärke von $r = 0.25$ und $r = 0.15$
-> Erscheint eine weitere Stichprobengröße als sinnvoll?
4. Visualisierung des Powerverlaufs in Abhängigkeit der Effektstärke bei einer maximal zu erwartenden Stichprobengröße von $n = 400$
-> Ist es möglich kleinere Effekte als $r = 0.15$ aufzudecken?

Schlussfolgerung:

Die Festlegung der einzelnen Parameter auf Power = 0.8, Signifikanzlevel = 0.05 und einer Effektgröße von $r = 0.15$ erscheint am sinnvollsten.

Dafür benötigen wir eine Stichprobe, die 346 Probanden umfasst.

Berechnung der Power für eine Stichprobe mit 346 Probanden per Monte-Carlo-Simulation für genauere Schätzung (Signifikanzlevel 0.05 und einer Effektgröße von $r = 0.15$).

Wiederholung der Visualisierung des Powerverlaufs in Abhängigkeit der Stichprobengröße und Wiederholung der Visualisierung des Powerverlaufs in Abhängigkeit der Effektgröße bei einer maximalen Stichprobengröße von $n = 400$.

6) **Anhänge:** Screenshots aus R, wo zutreffend.

1.

```
##{r}
pwr.r.test(power = 0.8, r = 0.25, sig.level = 0.05, alternative = "two.sided")
```

```
approximate correlation power calculation (arctangh transformation)

      n = 122.4466
      r = 0.25
sig.level = 0.05
power = 0.8
alternative = two.sided
```

2.

```
##{r}
pwr.r.test(power = 0.8, r = 0.15, sig.level = 0.05, alternative = "two.sided")
```

```
approximate correlation power calculation (arctangh transformation)

      n = 345.7037
      r = 0.15
sig.level = 0.05
power = 0.8
alternative = two.sided
```

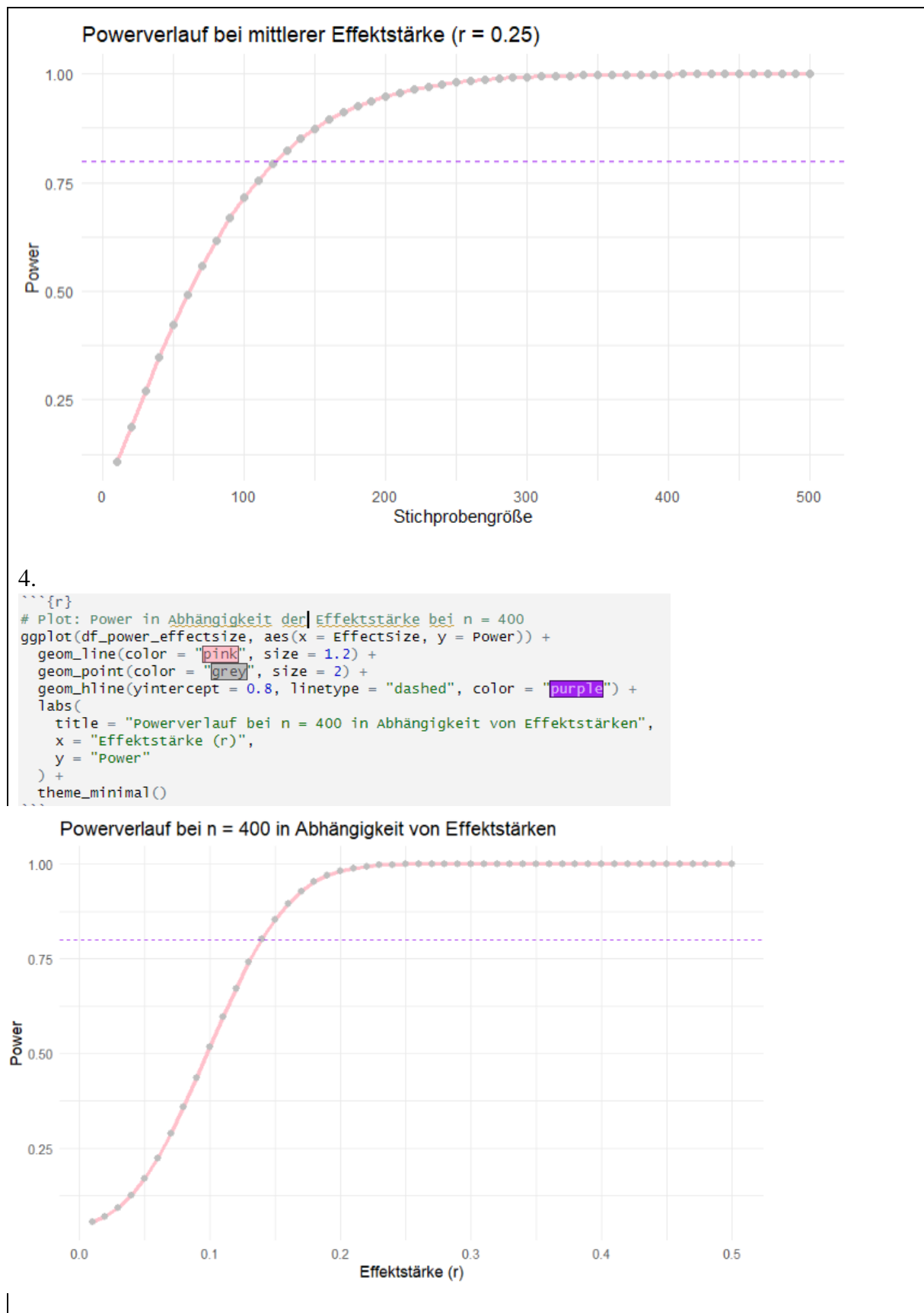
3.

```
# Powerverlauf bei mittleren Effektstärke (r = 0.25) über verschiedene Stichprobengrößen
sample_sizes <- seq(10, 500, by = 10)
effect_size_medium <- 0.25

power_medium_effect <- sapply(sample_sizes, function(n) {
  pwr.r.test(n = n, r = effect_size_medium, sig.level = 0.05, alternative = "two.sided")$power
})

df_power_medium <- data.frame(
  SampleSize = sample_sizes,
  Power = power_medium_effect
)

# Plot: Power in Abhängigkeit der Stichprobengröße (mittlere Effektstärke)
ggplot(df_power_medium, aes(x = SampleSize, y = Power)) +
  geom_line(color = "pink", size = 1.2) +
  geom_point(color = "grey", size = 2) +
  geom_hline(yintercept = 0.8, linetype = "dashed", color = "purple") +
  labs(
    title = "Powerverlauf bei mittlerer Effektstärke (r = 0.25)",
    x = "Stichprobengröße",
    y = "Power"
  ) +
  theme_minimal()
```



Berechnung der Power für eine Stichprobe mit 346 Probanden per Monte-Carlo-Simulation für genauere Schätzung (Signifikanzlevel 0.05 und einer Effektgröße von $r = 0.15$).

```
library(simr)
library(pwr)

# Funktion zur Simulation von Spearman-Korrelation (Monte-Carlo-Simulation)
simulate_power <- function(n = 346, rho = 0.15, alpha = 0.05, reps = 1000) {
  significant_results <- 0

  for (i in 1:reps) {
    # Generiere zwei zufällige, korrelierte Variablen mit Spearman's Rho
    x <- rnorm(n)
    y <- rho * x + sqrt(1 - rho^2) * rnorm(n) # Erzeugt Korrelation
    R_test <- cor.test(x, y, method = "pearson", alternative = "two.sided")

    # Prüfe, ob das Ergebnis signifikant ist; wenn ja erhöhe significant_results um 1
    if (R_test$p.value < alpha) {
      significant_results <- significant_results + 1
    }
  }

  # Berechne die Power (Anteil signifikanter Tests)
  power_estimate <- significant_results / reps
  return(power_estimate)
}

# Gibt Ergebnisse aus
power_result <- simulate_power()
print(power_result)
...
```

Wiederholung der Visualisierung des Powerverlaufs in Abhängigkeit der Stichprobengröße

```
```{r}
Poweranalyse mit variabler Stichprobengröße
simulate_power_vary_n <- function(rho = 0.15, alpha = 0.05, reps = 1000, n_values) {
 # Vektor, um die Power für verschiedene Stichprobengrößen zu speichern
 power_values <- numeric(length(n_values))

 # Für jede Stichprobengröße n
 for (j in 1:length(n_values)) {
 n <- n_values[j] # Aktuelle Stichprobengröße
 significant_results <- 0 # Zähler für signifikante Ergebnisse

 # Durchführung der Simulation für die gegebene Stichprobengröße
 for (i in 1:reps) {
 x <- rnorm(n) # Generiere eine Zufallsvariable x
 y <- rho * x + sqrt(1 - rho^2) * rnorm(n) # Generiere y mit Korrelation rho
 Rn_test <- cor.test(x, y, method = "pearson") # Spearman-Test durchführen

 # Prüfe, ob der p-wert signifikant ist -> wenn ja significant_results +1
 if (Rn_test$p.value < alpha) {
 significant_results <- significant_results + 1
 }
 }

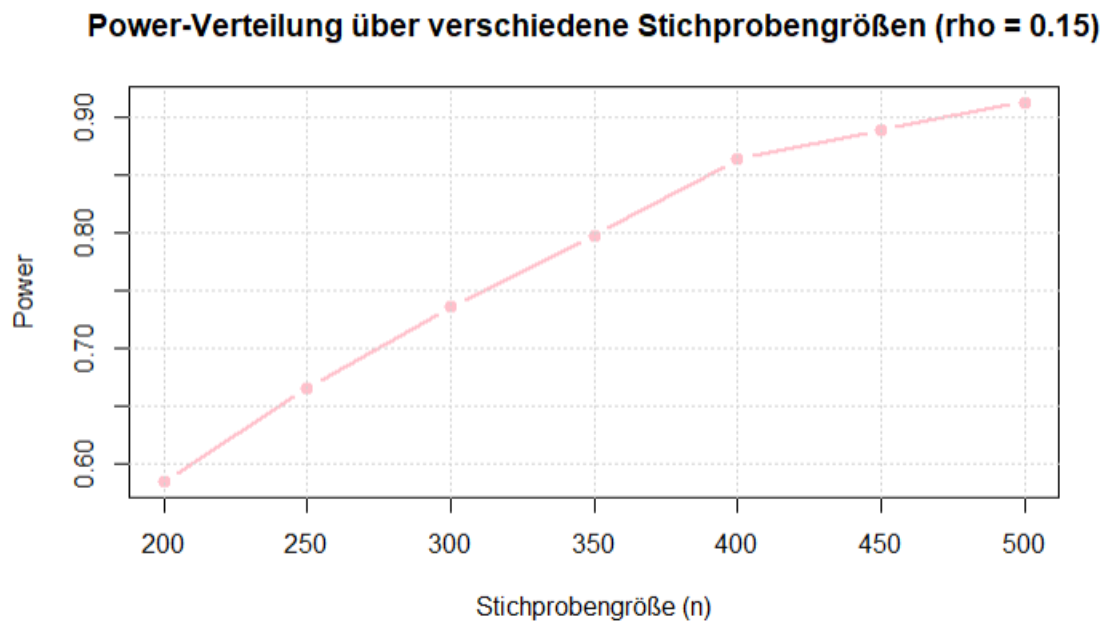
 # Berechne die Power für diese Stichprobengröße
 power_values[j] <- significant_results / reps
 }

 # Power-Werte ausgeben
 return(power_values)
}

Werte für die Stichprobengrößen (n von 200 bis 500)
n_values <- seq(200, 500, by = 50)

Poweranalyse durchführen: rho = 0.25 und n von 200 bis 500
power_distribution <- simulate_power_vary_n(rho = 0.15, n_values = n_values)

Grafik ausgeben
Anmerkung: Diagramm fällt mit jedem Durchlauf leicht anders aus -> Mehr Stichproben (reps) simulieren?
plot(n_values, power_distribution, type = "b",
 xlab = "Stichprobengröße (n)", ylab = "Power",
 main = "Power-verteilung über verschiedene Stichprobengrößen (rho = 0.15)",
 col = "pink", pch = 19, lwd = 2)
grid()
```



Wiederholung der Visualisierung des Powerverlaufs in Abhängigkeit der Effektgröße bei einer maximalen Stichprobengröße von  $n = 400$

```
#Power in Abhängigkeit der Effektgröße

simulate_power_PR_plot <- function(n = 400, alpha = 0.05, reps = 1000, rho_values = seq(0.1, 0.5, by = 0.05)) {
 power_results <- numeric(length(rho_values))

 for (i in seq_along(rho_values)) {
 rho <- rho_values[i]
 significant_results <- 0

 for (j in 1:reps) {
 x <- rnorm(n)
 y <- rho * x + sqrt(1 - rho^2) * rnorm(n)
 PR_test <- cor.test(x, y, method = "pearson")

 if (PR_test$p.value < alpha) {
 significant_results <- significant_results + 1
 }
 }

 power_results[i] <- significant_results / reps
 }

 # Erstellung Dataframe für Visualisierung (Dataframe nochmal durchlesen ums erklären zu können)
 power_data <- data.frame(rho = rho_values, power = power_results)

 # Diagramm mit ggplot2
 ggplot(power_data, aes(x = rho, y = power)) +
 geom_line(color = "pink", size = 1) +
 geom_point(color = "grey", size = 2) +
 labs(title = "Power in Abhängigkeit der Effektgröße",
 x = "Effektgröße (Cohens r)",
 y = "Power",
 caption = paste("n =", n, ", Alpha =", alpha, ", wiederholungen =", reps)) +
 theme_minimal()
}

Funktion ausführen
simulate_power_PR_plot()
```

