DailyCheck#11

2025-05-05

DailyCheck #11

1) In Figure 1, why is only (c) of use to us?

Answer: The d subject 1 and subject two are only holding one individual. Which is not enough simple there. For a and b it gives a sense of the variability, but it is not a variance. It is just showing how small was the smallest value, and how big was the biggest value and it is not easy to eyeball it. C is the only one that says mean plus or minus 2 standard deviations which is the one we have to find. From the top to the dot is two standard deviations. From the dot to the bottom is two standard deviations.

2) How do we end up with an estimated variance of 900s?

Answer: From the eyeballing, from the bottom to the dot is about 1 minutes. 1 minutes is equal to two standard deviations. 1 standard deviation is equal to half a minute. Half a minute is equal to 30 seconds. Variance is square of the standard deviation that is the square of 30s which is 900s.

Your Turn Part 1

Take about 5-10 minutes on your own to give it your best shot at

##re-creating the graph on Slide 11. Once you are either finished or ##stuck after giving it an honest effort, then watch the solution video.

```
between.var <- seq(50, 200, by = 10)

n_var1 <- NA
for(i in 1: length(between.var)){
  n_var1[i] <- power.anova.test(groups = 3,
                   between.var = between.var[i],
                   within.var = 900,
                   power = 0.8, sig.level = 0.05, n = NULL)$n
}

n_var2 <- NA
for(i in 1: length(between.var)){
  n_var2[i] <- power.anova.test(groups = 3,
                   between.var = between.var[i],
                   within.var = 225,
                   power = 0.8, sig.level = 0.05, n = NULL)$n
}

n_var3 <- NA
```

```
for(i in 1: length(between.var)){
  n_var3[i] <- power.anova.test(groups = 3,
                    between.var = between.var[i],
                    within.var = 2025,
                    power = 0.8, sig.level = 0.05, n = NULL)$n
}

sample_sizes <- data.frame(
  n = c(n_var1, n_var2, n_var3),
  between.var = rep(between.var, 3),
  within.var = c(rep("900", length(n_var1)),
                rep("225", length(n_var2)),
                rep("2025", length(n_var3)))
)
sample_sizes$within.var <- factor(sample_sizes$within.var,
                                    levels = c("225", "900", "2025"))
library(ggplot2)
ggplot(data = sample_sizes, mapping = aes(x = between.var,
                                          y = n,
                                          group = within.var,
                                          color = within.var)) +
  geom_point() + geom_line()
```
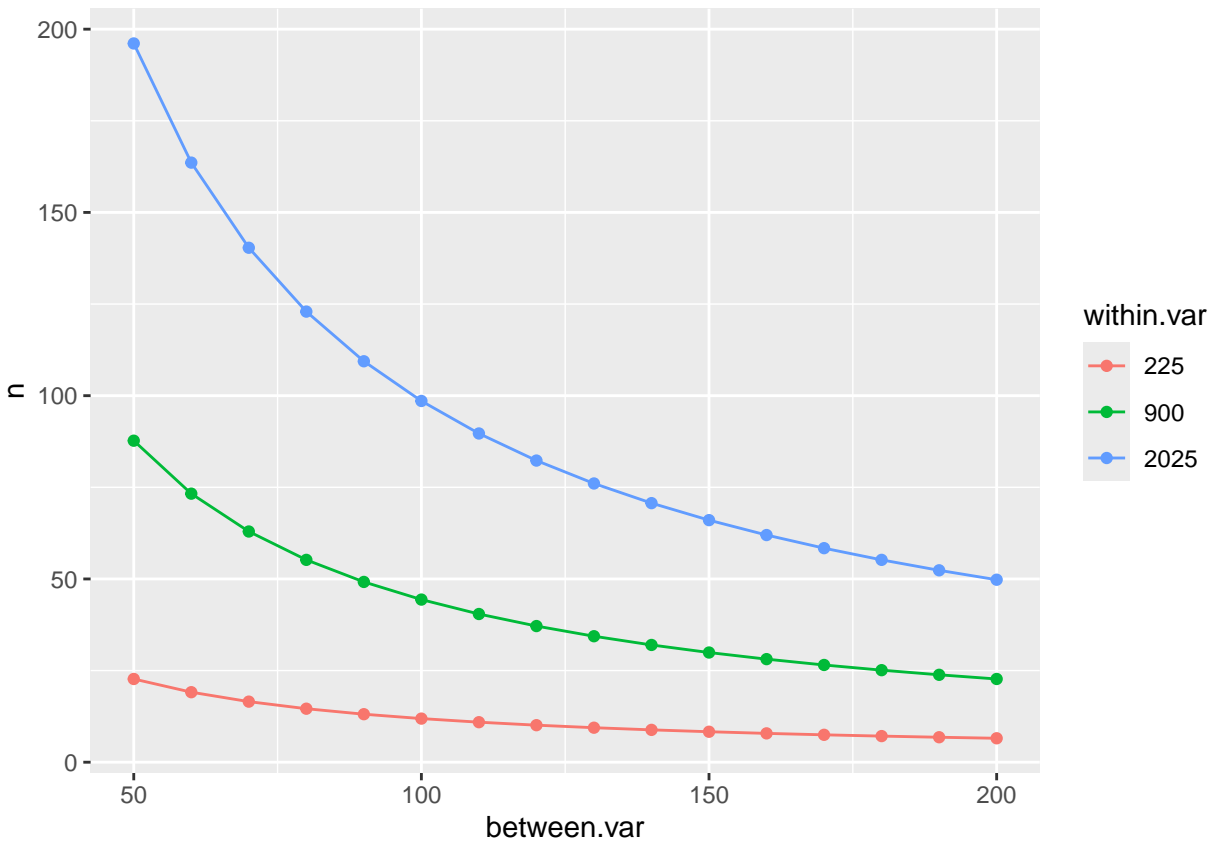

This plot shows the difference between groups increases and the required sample size per group decrees. Bigger differences between groups or less within - group variability both reduce the needed sample sample size for my study

Your Turn Part 2

Run TukeyHSD on the VascGraft.xlsx model with blocking, and report your conclusions

```
library(readxl)
vasc <- read_excel("VascGraft.xlsx")

vasc$PSI <- as.factor(vasc$PSI)
vasc$block <- as.factor(vasc$block)

model1 <- aov(yield ~ PSI + block, data = vasc)
hsd_output <- TukeyHSD(model1)

knitr::kable(hsd_output$PSI)
```

	diff	lwr	upr	p adj
8700-8500	-1.133333	-5.637161	3.370495	0.8854831
8900-8500	-3.900000	-8.403828	0.603828	0.1013084
9100-8500	-7.050000	-11.553828	-2.546172	0.0020883
8900-8700	-2.766667	-7.270495	1.737161	0.3245644
9100-8700	-5.916667	-10.420495	-1.412839	0.0086667
9100-8900	-3.150000	-7.653828	1.353828	0.2257674

Your Turn Part 3

Histogram of residuals

```
model1 <- aov(yield ~ PSI + block, data = vasc)
hist(model1$residuals, xlab = "residuals", main = "Vascular Graft Residuals")
```
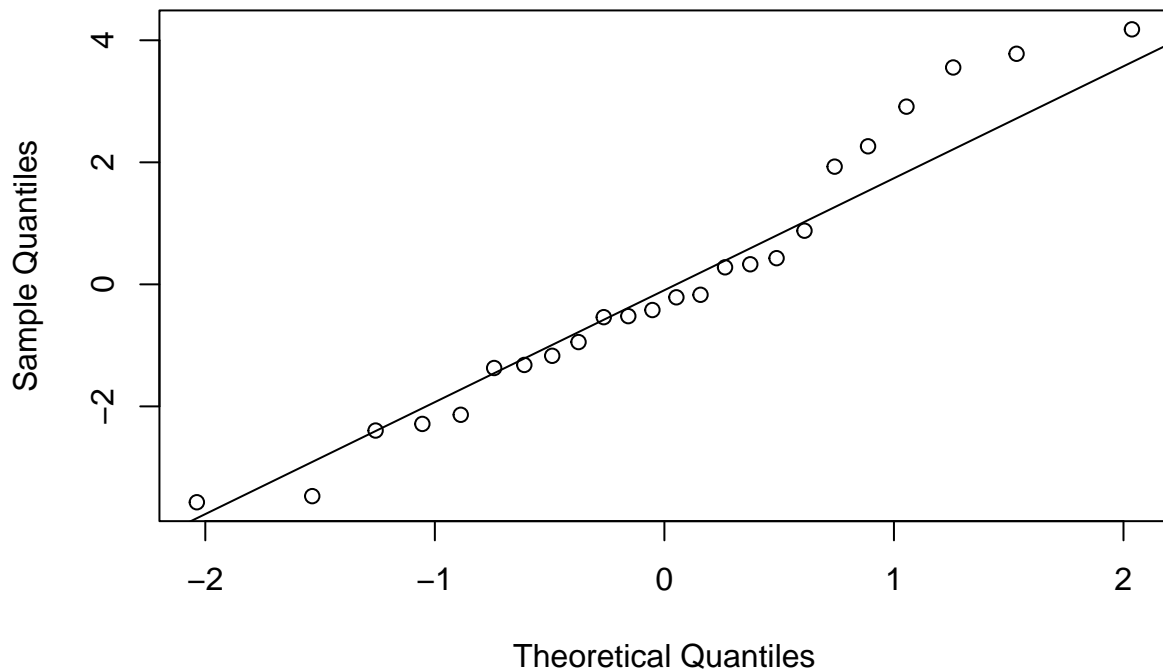
Vascular Graft Residuals
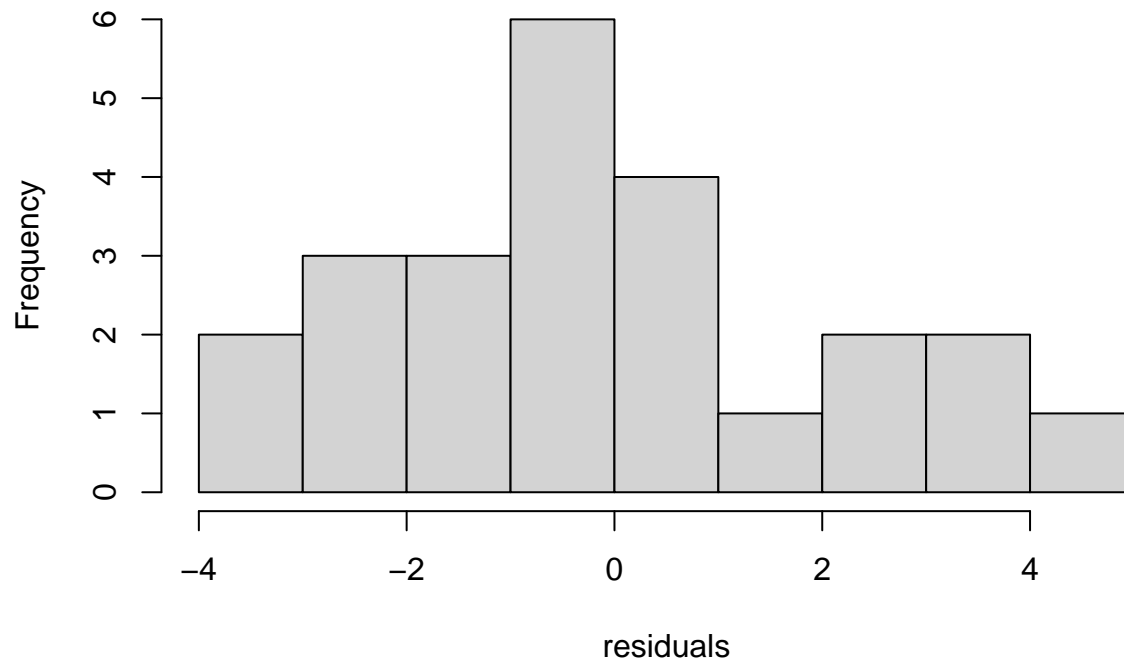
The histogram of residuals appears approximately symmetric and centered around zero, with no extreme outliers. This suggests that the normality assumption of the residuals is reasonably met.

QQ-plot of residuals

```r
qqnorm(model1$residuals)
qqline(model1$residuals)
```

Normal Q–Q Plot

The Q-Q plot shows that most of the residuals fall close to the reference line, with only minor deviations at the tails. This indicates that the residuals are approximately normally distributed, supporting the normality assumption for the ANOVA model.

Shapiro - Wilk Test

```
shapiro.test(model1$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model1$residuals
## W = 0.95631, p-value = 0.3689
```

The Shapiro-Wilk test gave a p-value of 0.3689, which is much greater than 0.05. This means there is no evidence to reject the normality assumption. The residuals can be considered approximately normally distributed.

## Plot of residuals vs. order of data collection

```
x <- 1:length(model1$residuals)
plot(model1$residuals ~ x)
```
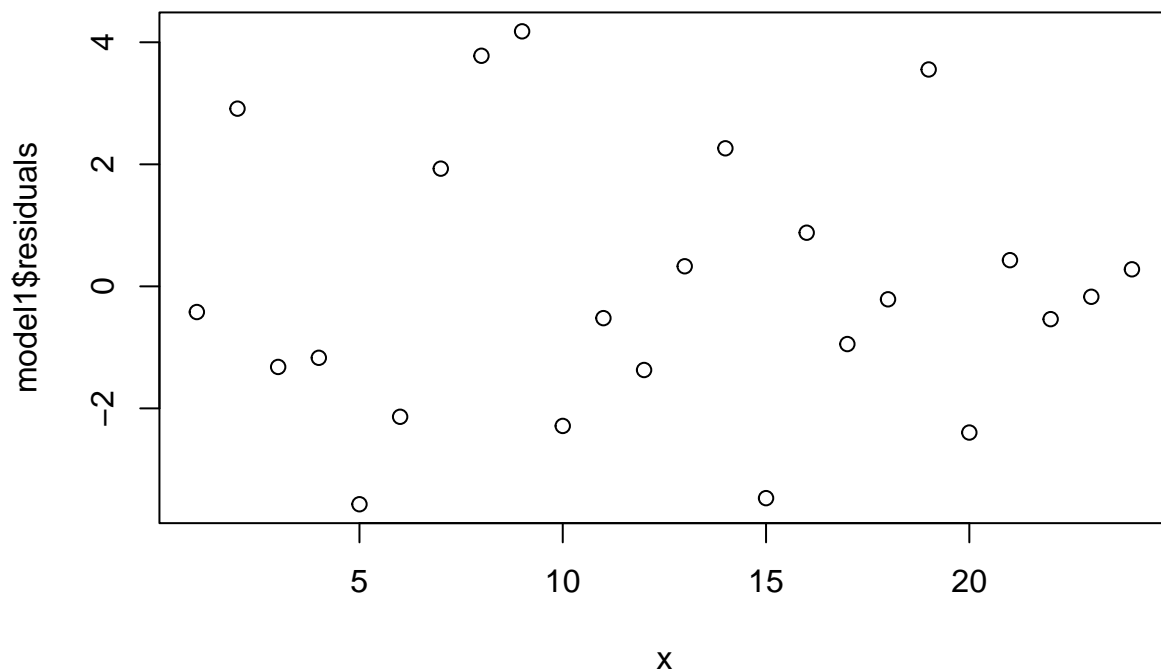
The plot of residuals versus the order of data collection shows no obvious pattern or trend; the points are scattered randomly. This suggests there is no structure to the residuals, supporting the assumption of independence in your model.

Plot of residuals vs fitted values

```
plot(model1$residuals ~ model1$fitted.values)
```

model1$fitted.values
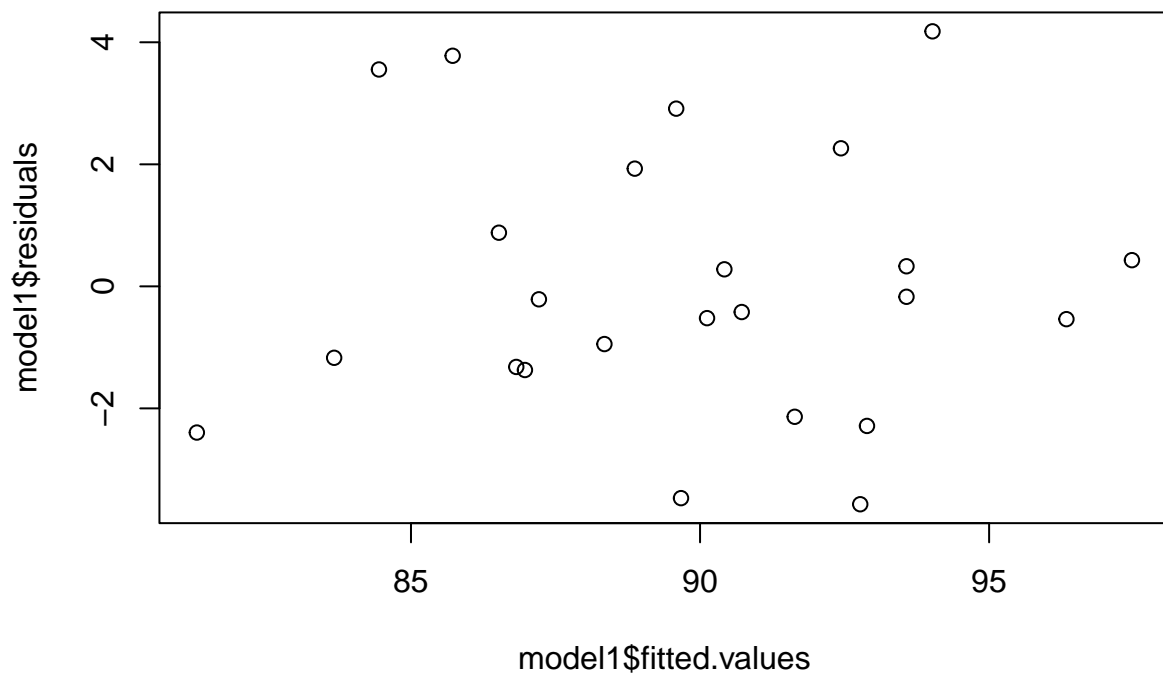
The plot of residuals versus fitted values shows that the spread of residuals is fairly consistent across all fitted values, with no clear pattern or funnel shape. This suggests the assumption of equal variances (homoscedasticity) is satisfied.