Response and adverse events to chemotherapy: A mock study

# Basic sanity checking/EDA here

glimpse(mockdata)

## Observations: 1,499  
## Variables: 25  
## $ case <dbl> 110754, 99706, 105271, 105001, 112263, 86205, 99508,…  
## $ age <dbl> 67, 74, 50, 71, 69, 56, 50, 57, 51, 63, 61, 59, 61, …  
## $ arm <chr> "F: FOLFOX", "A: IFL", "A: IFL", "G: IROX", "F: FOLF…  
## $ sex <chr> "Male", "Female", "Female", "Female", "Female", "Mal…  
## $ race <chr> "Caucasian", "Caucasian", "Caucasian", "Caucasian", …  
## $ fu\_time <dbl> 922, 270, 175, 128, 233, 120, 369, 421, 387, 363, 16…  
## $ fu\_stat <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 2, 2…  
## $ ps <dbl> 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1…  
## $ hgb <dbl> 11.5, 10.7, 11.1, 12.6, 13.0, 10.2, 13.3, 12.1, 13.8…  
## $ bmi <dbl> 25.09861, 19.49786, NA, 29.42922, 26.35352, 19.03673…  
## $ alk\_phos <dbl> 160, 290, 700, 771, 350, 569, 162, 152, 231, 492, 74…  
## $ ast <dbl> 35, 52, 100, 68, 35, 27, 16, 12, 25, 18, 45, 16, 50,…  
## $ mdquality\_s <dbl> NA, 1, 1, 1, NA, 1, 1, 1, 1, 1, NA, NA, 1, 0, 1, 1, …  
## $ age\_ord <chr> "60-69", "70-79", "40-49", "70-79", "60-69", "50-59"…  
## $ ethnicity <chr> "White (not Hispanic)", "White (not Hispanic)", "Whi…  
## $ name <chr> "Daniel Robinson", "Delaney Quinby", "Lauren Ginn", …  
## $ first\_name <chr> "Daniel", "Delaney", "Lauren", "Julia", "Iffat", "Tr…  
## $ last\_name <chr> "Robinson", "Quinby", "Ginn", "Carney", "al-Sahli", …  
## $ low\_wbc <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ neuropathy <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ diarrhea <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ vomiting <dbl> 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ blood\_clot <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ site <chr> "Portland", "Portland", "Portland", "Portland", "Por…  
## $ country <chr> "USA", "USA", "USA", "USA", "USA", "USA", "USA", "US…

# skim(mockdata)  
  
mockdata %>%   
 group\_by(arm) %>%   
 skim()

## Skim summary statistics  
## n obs: 1499   
## n variables: 25   
## group variables: arm   
##   
## ── Variable type:character ─────────────────────────────────────────────────  
## arm variable missing complete n min max empty n\_unique  
## A: IFL age\_ord 0 428 428 5 5 0 7  
## A: IFL country 0 428 428 3 5 0 2  
## A: IFL ethnicity 0 428 428 8 33 0 6  
## A: IFL first\_name 0 428 428 2 11 0 279  
## A: IFL last\_name 0 428 428 3 16 0 398  
## A: IFL name 0 428 428 8 27 0 428  
## A: IFL race 0 428 428 5 16 0 7  
## A: IFL sex 0 428 428 4 6 0 2  
## A: IFL site 0 428 428 6 11 0 13  
## F: FOLFOX age\_ord 0 691 691 5 5 0 8  
## F: FOLFOX country 0 691 691 3 10 0 5  
## F: FOLFOX ethnicity 0 691 691 8 33 0 6  
## F: FOLFOX first\_name 0 691 691 2 11 0 397  
## F: FOLFOX last\_name 0 691 691 2 16 0 617  
## F: FOLFOX name 0 691 691 8 27 0 691  
## F: FOLFOX race 0 691 691 5 16 0 7  
## F: FOLFOX sex 0 691 691 4 6 0 2  
## F: FOLFOX site 0 691 691 6 14 0 18  
## G: IROX age\_ord 0 380 380 5 5 0 7  
## G: IROX country 0 380 380 3 3 0 1  
## G: IROX ethnicity 0 380 380 8 33 0 6  
## G: IROX first\_name 0 380 380 2 11 0 262  
## G: IROX last\_name 0 380 380 3 16 0 357  
## G: IROX name 0 380 380 9 23 0 380  
## G: IROX race 0 380 380 5 16 0 7  
## G: IROX sex 0 380 380 4 6 0 2  
## G: IROX site 0 380 380 6 11 0 12  
##   
## ── Variable type:numeric ───────────────────────────────────────────────────  
## arm variable missing complete n mean sd p0  
## A: IFL age 0 428 428 59.67 11.36 27   
## A: IFL alk\_phos 69 359 428 175.58 128.61 11   
## A: IFL ast 69 359 428 37.29 28.04 10   
## A: IFL blood\_clot 0 428 428 0.063 0.24 0   
## A: IFL bmi 9 419 428 27.29 5.55 14.05  
## A: IFL case 0 428 428 95252.21 8200.84 76170   
## A: IFL diarrhea 0 428 428 0.18 0.39 0   
## A: IFL fu\_stat 0 428 428 1.96 0.2 1   
## A: IFL fu\_time 0 428 428 553.58 419.61 9   
## A: IFL hgb 69 359 428 12.28 1.69 9.06  
## A: IFL low\_wbc 0 428 428 0.16 0.37 0   
## A: IFL mdquality\_s 55 373 428 0.89 0.31 0   
## A: IFL neuropathy 0 428 428 0.16 0.37 0   
## A: IFL ps 69 359 428 0.53 0.6 0   
## A: IFL vomiting 0 428 428 0.18 0.39 0   
## F: FOLFOX age 0 691 691 60.3 11.63 19   
## F: FOLFOX alk\_phos 141 550 691 161.98 121.98 10   
## F: FOLFOX ast 141 550 691 35.2 26.66 7   
## F: FOLFOX blood\_clot 0 691 691 0.046 0.21 0   
## F: FOLFOX bmi 20 671 691 27.21 5.17 16.65  
## F: FOLFOX case 0 691 691 1e+05 9617.76 78845   
## F: FOLFOX diarrhea 0 691 691 0.29 0.45 0   
## F: FOLFOX fu\_stat 0 691 691 1.86 0.35 1   
## F: FOLFOX fu\_time 0 691 691 731.25 487.74 0   
## F: FOLFOX hgb 141 550 691 12.38 1.76 9   
## F: FOLFOX low\_wbc 0 691 691 0.26 0.44 0   
## F: FOLFOX mdquality\_s 156 535 691 0.9 0.3 0   
## F: FOLFOX neuropathy 0 691 691 0.21 0.41 0   
## F: FOLFOX ps 141 550 691 0.55 0.59 0   
## F: FOLFOX vomiting 0 691 691 0.25 0.43 0   
## G: IROX age 0 380 380 59.76 11.5 26   
## G: IROX alk\_phos 56 324 380 173.51 138.56 7   
## G: IROX ast 56 324 380 35.67 25.81 5   
## G: IROX blood\_clot 0 380 380 0.039 0.19 0   
## G: IROX bmi 4 376 380 27.11 5.75 15.43  
## G: IROX case 0 380 380 94869.23 6960.2 78841   
## G: IROX diarrhea 0 380 380 0.092 0.29 0   
## G: IROX fu\_stat 0 380 380 1.93 0.25 1   
## G: IROX fu\_time 0 380 380 607.24 435.51 17   
## G: IROX hgb 56 324 380 12.37 1.68 9   
## G: IROX low\_wbc 0 380 380 0.068 0.25 0   
## G: IROX mdquality\_s 41 339 380 0.91 0.29 0   
## G: IROX neuropathy 0 380 380 0.068 0.25 0   
## G: IROX ps 56 324 380 0.54 0.61 0   
## G: IROX vomiting 0 380 380 0.092 0.29 0   
## p25 p50 p75 p100 hist  
## 53 61 68 88 ▁▂▃▆▇▆▃▁  
## 89 133 217 858 ▇▆▂▁▁▁▁▁  
## 21 29 42 205 ▇▃▁▁▁▁▁▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▁  
## 23.57 26.23 30.59 53.01 ▁▅▇▅▂▁▁▁  
## 90563 93317 1e+05 108884 ▂▁▂▇▆▁▃▆  
## 0 0 0 1 ▇▁▁▁▁▁▁▂  
## 2 2 2 2 ▁▁▁▁▁▁▁▇  
## 255.5 446.5 724.25 2170 ▇▇▅▂▂▁▁▁  
## 11 12.1 13.45 17.3 ▃▇▇▇▆▃▁▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▂  
## 1 1 1 1 ▁▁▁▁▁▁▁▇  
## 0 0 0 1 ▇▁▁▁▁▁▁▂  
## 0 0 1 2 ▇▁▁▆▁▁▁▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▂  
## 52 61 69 88 ▁▁▂▆▇▇▅▁  
## 85 116 194.75 1014 ▇▃▁▁▁▁▁▁  
## 19 25.5 40 174 ▇▃▁▁▁▁▁▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▁  
## 23.75 26.52 30.12 49.13 ▂▆▇▅▂▁▁▁  
## 92512.5 105126 111018.5 112488 ▁▁▂▃▁▂▂▇  
## 0 0 1 1 ▇▁▁▁▁▁▁▃  
## 2 2 2 2 ▂▁▁▁▁▁▁▇  
## 345 601 1046 2472 ▅▇▅▃▃▁▁▁  
## 11.1 12.2 13.6 18.2 ▃▆▇▆▅▂▁▁  
## 0 0 1 1 ▇▁▁▁▁▁▁▃  
## 1 1 1 1 ▁▁▁▁▁▁▁▇  
## 0 0 0 1 ▇▁▁▁▁▁▁▂  
## 0 0 1 2 ▇▁▁▇▁▁▁▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▂  
## 52 61 68 85 ▁▂▂▅▆▇▅▁  
## 87.75 122 210.25 982 ▇▅▂▁▁▁▁▁  
## 20 27 41 176 ▇▆▂▁▁▁▁▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▁  
## 23.17 25.98 29.61 60.24 ▂▇▅▂▁▁▁▁  
## 90638.75 93126 1e+05 107746 ▁▁▂▇▃▁▃▃  
## 0 0 0 1 ▇▁▁▁▁▁▁▁  
## 2 2 2 2 ▁▁▁▁▁▁▁▇  
## 306.5 515.5 807 2118 ▆▇▆▃▁▁▁▁  
## 11.17 12.4 13.62 17 ▃▅▆▇▇▃▂▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▁  
## 1 1 1 1 ▁▁▁▁▁▁▁▇  
## 0 0 0 1 ▇▁▁▁▁▁▁▁  
## 0 0 1 2 ▇▁▁▆▁▁▁▁  
## 0 0 0 1 ▇▁▁▁▁▁▁▁

# Make table one here

Let’s make a basic Table 1 grouped by arm with details on sex and age in each group.

#summary by groups  
tab1 <- tableby(arm ~ sex + age, data = mockdata)  
summary(tab1, text=TRUE)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A: IFL (N=428) | F: FOLFOX (N=691) | G: IROX (N=380) | Total (N=1499) | p value |
| sex |  |  |  |  | 0.190 |
| - Female | 151 (35.3%) | 280 (40.5%) | 152 (40.0%) | 583 (38.9%) |  |
| - Male | 277 (64.7%) | 411 (59.5%) | 228 (60.0%) | 916 (61.1%) |  |
| age |  |  |  |  | 0.614 |
| - Mean (SD) | 59.673 (11.365) | 60.301 (11.632) | 59.763 (11.499) | 59.985 (11.519) |  |
| - Range | 27.000 - 88.000 | 19.000 - 88.000 | 26.000 - 85.000 | 19.000 - 88.000 |  |

Let’s make a Table 1 - but ungrouped, with stats on BMI, sex, Age in each group.

#summary without groups  
tab.noby <- tableby(~ bmi + sex + age, data = mockdata)  
summary(tab.noby)

|  |  |
| --- | --- |
|  | Overall (N=1499) |
| **bmi** |  |
| N-Miss | 33 |
| Mean (SD) | 27.206 (5.432) |
| Range | 14.053 - 60.243 |
| **sex** |  |
| Female | 583 (38.9%) |
| Male | 916 (61.1%) |
| **age** |  |
| Mean (SD) | 59.985 (11.519) |
| Range | 19.000 - 88.000 |

Let’s make a Table 1 but now control # of digits

summary(tableby(arm ~ sex + fu\_time, data = mockdata),   
 digits = 4, digits.p = 2, digits.pct = 1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A: IFL (N=428) | F: FOLFOX (N=691) | G: IROX (N=380) | Total (N=1499) | p value |
| **sex** |  |  |  |  | 0.19 |
| Female | 151 (35.3%) | 280 (40.5%) | 152 (40.0%) | 583 (38.9%) |  |
| Male | 277 (64.7%) | 411 (59.5%) | 228 (60.0%) | 916 (61.1%) |  |
| **fu\_time** |  |  |  |  | < 0.01 |
| Mean (SD) | 553.5841 (419.6065) | 731.2460 (487.7443) | 607.2421 (435.5092) | 649.0841 (462.5109) |  |
| Range | 9.0000 - 2170.0000 | 0.0000 - 2472.0000 | 17.0000 - 2118.0000 | 0.0000 - 2472.0000 |  |

# Run some stats and make a table here

Arm x fu\_stat chi square table (1 = lived, 2 = died)

mockdata %>%   
 tabyl(arm, fu\_stat) %>%   
 adorn\_totals("row") %>% # can also do "col", or c("row", "col")  
 adorn\_percentages() %>%   
 adorn\_pct\_formatting() %>%   
 adorn\_ns() %>%   
 knitr::kable()

|  |  |  |
| --- | --- | --- |
| arm | 1 | 2 |
| A: IFL | 4.2% (18) | 95.8% (410) |
| F: FOLFOX | 14.3% (99) | 85.7% (592) |
| G: IROX | 6.8% (26) | 93.2% (354) |
| Total | 9.5% (143) | 90.5% (1356) |

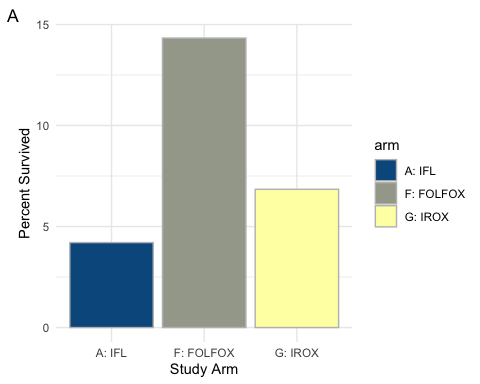
mockdata %>%   
 specify(fu\_stat ~ arm) %>% # alt: response = fu\_stat, explanatory = arm  
 calculate(stat = "Chisq")

mockdata %>%   
 specify(formula = fu\_stat ~ arm, success = "1") %>%   
 hypothesize(null = "independence") %>%   
 generate(reps = 1000, type = "permute") %>%   
 calculate(stat = "diff in props", order = c("1", "2"))

make bar/lollipop chart of proportions here (% survived)

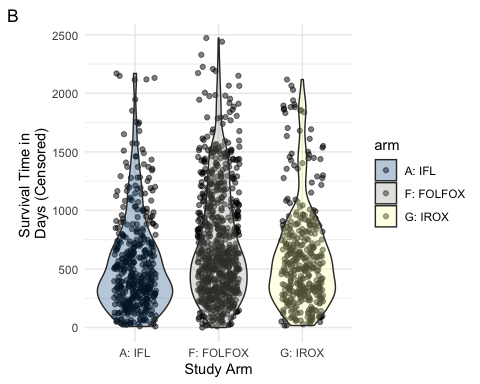
Start with a barplot for percent survival tag it as panel A for a multipanel plot

mockdata %>%   
 group\_by(arm) %>%   
 summarize(surv = length(which(fu\_stat==1)),  
 died = length(which(fu\_stat==2)),  
 pct\_surv = surv\*100/(died+surv)) %>%   
 select(arm, surv, died, pct\_surv) %>%   
 ggplot() +  
 aes(x=arm, y = pct\_surv, fill=arm) +  
 geom\_col(colour = "gray") +  
 labs(y= "Percent Survived", x= "Study Arm", tag ="A") +  
 scale\_fill\_scico\_d(palette = 'nuuk') ->  
p1  
  
p1



Then distribution plot of survival time (in days censored)

ggplot(mockdata) +  
 aes(x=arm, y = fu\_time, fill=arm) +  
 geom\_jitter(width =0.25, alpha=0.5) +  
 geom\_violin(alpha =0.3) +  
 labs(y= "Survival Time in \nDays (Censored)", x= "Study Arm", tag = "B") +  
 theme\_minimal() +  
 scale\_fill\_scico\_d(palette = 'nuuk') ->  
p2  
  
p2



(for colors, maybe show scale\_fill\_gray + scico)

# Make a multi-panel plot here

## Acknowledgments

This is a place to recognize people and institutions. It may also be a good place to acknowledge and cite software that makes your work possible.

## Author Contributions

We strongly encourage you to include an author contributions statement briefly describing what each author did.