BIOST/EPI 536 Homework 1

Ivy Zhang

10/5/2021

Question 1

```
#Question 1 Table
leuk$cr = factor(leuk$cr, levels = c("Y","N"), labels = c("Yes","No"))
leuk$tx = factor(leuk$tx, levels = c("D","I"), labels = c("Daunorubicin","Idarubicin"))
leuk$sex = factor(leuk$sex, levels=c("M","F"),labels=c("Male", "Female"))
leuk$fab = as.factor(leuk$fab)
label(leuk$sex) = "Sex"
label(leuk$age) = "Age"
label(leuk$karn) = "Karnofsky score (0-100)"
label(leuk$wbc) = "Baseline white blood cells"
units(leuk$wbc) = "1000/mm^3"
label(leuk$plt) = "Baseline platelets"
units(leuk$plt) = "1000/mm^3"
label(leuk$hgb) = "Baseline hemoglobin"
units(leuk$hgb) = "g/dl"
label(leuk$fab) = "FAB classification of AML subtype  (1 - 6)"
table1(~sex+age+karn+wbc+plt+hgb+fab|tx, data = leuk,caption = "Baseline characteristics by treatment g
```

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Daunorubicin	Idarubicin	Overall
	(N=65)	(N=65)	(N=130)
Sex			
Male	35.0 (53.8%)	30.0 (46.2%)	65.0 (50.0%)
Female	30.0 (46.2%)	35.0 (53.8%)	65.0 (50.0%)
Age			
Mean (SD)	39.8 (13.4)	38.0 (12.5)	38.9 (12.9)
Median [Min, Max]	40.0 [19.0, 60.0]	36.0 [17.0, 61.0]	36.5 [17.0, 61.0]
Karnofsky score (0-100)			
Mean (SD)	79.5 (12.6)	79.5 (11.6)	79.5 (12.1)
Median [Min, Max]	80.0 [40.0, 100]	80.0 [30.0, 100]	80.0 [30.0, 100]
Baseline white blood cells (1000/mm^3)			
Mean (SD)	43.3 (55.0)	29.0 (36.3)	36.1 (46.9)
Median [Min, Max]	16.7 [0.700, 215]	11.8 [0.400, 154]	13.8 [0.400, 215]
Missing	1.00 (1.5%)	0 (0%)	1.00 (0.8%)
Baseline platelets (1000/mm^3)			
Mean (SD)	93.6 (92.4)	66.6 (57.8)	80.0 (77.8)
Median [Min, Max]	62.0 [11.0, 457]	50.0 [11.0, 370]	57.0 [11.0, 457]
Missing	1.00 (1.5%)	0 (0%)	1.00 (0.8%)
Baseline hemoglobin (g/dl)			
Mean (SD)	9.64 (1.49)	9.22 (1.82)	9.43 (1.67)
Median [Min, Max]	9.45 [6.40, 13.9]	9.20 [2.80, 13.7]	9.30 [2.80, 13.9]
Missing	1.00 (1.5%)	0 (0%)	1.00 (0.8%)

	Daunorubicin	Idarubicin	Overall
FAB classification of AML subtype (1 - 6)			
0	1.00 (1.5%)	0 (0%)	1.00 (0.8%)
1	6.00 (9.2%)	13.0 (20.0%)	19.0 (14.6%)
2	15.0 (23.1%)	15.0 (23.1%)	30.0 (23.1%)
3	9.00 (13.8%)	11.0 (16.9%)	20.0 (15.4%)
4	12.0 (18.5%)	8.00 (12.3%)	20.0 (15.4%)
5	13.0 (20.0%)	12.0 (18.5%)	25.0 (19.2%)
6	1.00 (1.5%)	2.00 (3.1%)	3.00 (2.3%)
Missing	8.00 (12.3%)	4.00 (6.2%)	12.0 (9.2%)

Question 2

```
#Question 2 Table
label(leuk$cr) = "Complete Remission"
table1(~cr|sex*tx, data =leuk)
```

```
## Get nicer `table1` LaTeX output by simply installing the `kableExtra` package
```

	Daunorubicin	Idarubicin	Daunorubicin	Idarubicin	Daunorubicin	Idarubicin
	(N=35)	(N=30)	(N=30)	(N=35)	(N=65)	(N=65)
Complete Remission						
Yes	17.0 (48.6%)	21.0 (70.0%)	21.0 (70.0%)	30.0 (85.7%)	38.0 (58.5%)	51.0 (78.5%)
No	18.0 (51.4%)	9.00 (30.0%)	9.00 (30.0%)	5.00 (14.3%)	27.0 (41.5%)	14.0 (21.5%)

Question 3

In order to do the unadjusted analysis, we will calculate the Odds Ratio, Relative Risk, Risk Difference: Based on the previous table from Question 2, we have Odds Ratio (calculated by logistic regression):

```
#Question 3 Logistic Regression
leuk$cr01 = 1
leuk$cr01[which(leuk$cr == "No")] = 0
log = glm(cr01~tx , family = binomial, data = leuk)
kable(summary(log)$coefficients)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.3417493	0.2516999	1.357765	0.1745382
txIdarubicin	0.9510190	0.3929238	2.420365	0.0155049

OR $= \frac{51/14}{38/27} = 2.588 = e^{0.951}$ According to the calculation, we estimated that the odds of individuals who take idarubicin will have complete remission is 1.588 times higher than the odds of people who take daunorubicin.Treatment is statistically significant at the level of 0.05.

Relative Risk:

```
#Question 3 Relative Risk
rr = glm(cr01~tx, family = poisson(link = "log"), data = leuk)
kable(summary(rr)$coefficients)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.5368011	0.1622213	-3.309066	0.0009361
txIdarubicin	0.2942395	0.2142979	1.373039	0.1697402

RR = $\frac{51/65}{38/65} = 1.342 = e^{0.294}$ According to the calculation, we estimated that individuals who take idarubicin are 34.2% more likely to have complete remission than the odds of people who take daunorubicin. Treatment is not statistically significant at the level of 0.05.

Risk Difference:

```
#Question 3 Risk Difference
rd = glm(cr01~tx, family = binomial(link = "identity"), data = leuk)
kable(summary(rd)$coefficients)
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.5846154	0.0611229	9.564595	0.0000000
txIdarubicin	0.2000000	0.0795984	2.512612	0.0119841

RD = $51/65 - 38/65 $ = 0.2

According to the calculation, we estimated that individuals who take idarubicin will have 20 additional cases of complete remission per 100 people compared to people who take daunorubicin.Treatment is statistically significant at the level of 0.05.

I think AML patients should be most interested in Odds Ratio. Relative Risk could vary greatly due to the variation among baseline risks and have some limitation, and Odds Ratio does not have this limitation.

Question 4

I don't think sex is a confounder in this study. Based on the definition, confounder must be associated with the treatment group(E). However, in this study, participants are randomly assigned to the different treatment groups and have no association with sex. Therefore, I think the analysis of treatment effect is not potentially confounded by sex.

Question 5-7

Question 8

(a) $PAR = \frac{P[E]\times(RR-1)}{P[E]RR+1-P[E]}$

(b) $PAR = \frac{0.35\times(22-1)}{0.35\times22+1-0.35} = 0.880$

In the population, 88.0% of the overall risk of lung cancer death is due to smoking.

(c) $PAR = \frac{0.05\times(22-1)}{0.05\times22+1-0.05} = 0.512$

In the population, 51.2% of the overall risk of lung cancer death is due to smoking.