**Individual Data Analysis Exam Part 2**

EPID 600: Principles of Epidemiology for Public Health

1. **Calculate the cohort study measures of comparison/association**

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk Difference, Risk Ratio, Risk Rate and Confidence Intervals from Cohort Study of UNC25 Exposure and Cancer of 1000 Individuals from UNC Family Practice Clinics Between January 1, 2000, and December 31, 2009.** | | | |
| **Measure** | **Estimate** | **95% Confidence Interval** | |
| Risk Difference | *9.502 (per 100 persons)* | 5.247 | 13.76 |
| Risk Ratio | *3.934* | *2.371* | *6.527* |
| Rate Ratio | *.9531* | 0.5634 | 1.612 |

1. Interpret, using complete sentences, each of your estimates:
   1. Risk difference (RD) & 95% CI [*5pts*]

***The risk of cancer was 9.502 cases per 100 persons higher in the group exposed to the virus than in the group unexposed to the virus. We are 95% confident that the risk difference of cancer in those exposed to the virus compared those unexposed is between .5634 and 1.612. Since the 95% CI does not include the null value (0) the finding is statistically significant.***

* 1. Risk ratio (RR) & 95% CI [*5pts*]

***Among those exposed to the virus, the risk of being diagnosed with cancer was 3.934 times as likely compared with those not exposed to the virus. We are 95% confident that the risk ratio of cancer in those exposed to the virus compared those unexposed is between 2.371 and 6.527. Since the 95% CI does not include the null value (1) the finding is statistically significant.***

* 1. Rate ratio (IRR) & 95% CI [*5pts*]

***The rate of cancer among those exposed to the virus was .9531 times the rate of cancer of those not exposed to the virus. We are 95% confident that the rate ratio of cancer in those exposed to the virus compared those unexposed is between .5634 and 1.612. Since the 95% CI includes the null value (1) there is insufficient evidence to conclude that the finding is statistically significantly.***

1. **Calculate the case-control study measure of comparison/association**. Calculate the odds ratio and OR 95% CI for the case-control study of UNC25 exposure and cancer and report in table 2, required format as follows*. [15 total points as indicated in table.]*

|  |  |  |  |
| --- | --- | --- | --- |
| **Odds Ratio and Confidence Interval from Nested Case Control Study of UNC25 Exposure and Cancer of 100 Individuals from UNC Family Practice Clinics Between January 1, 2000, and December 31, 2009.** | | | |
| **Measure** | **Estimate** | **95% Confidence Interval** | |
| **OR from nested case control study** | 4.224 | 1.752 | 10.18 |

1. ***Those exposed to the virus had 4.224 times the odds of cancer diagnosis compared to those who were unexposed. We are 95% confident that the odds ratio of cancer in those exposed to the virus compared those unexposed is between 1.752 and 10.18. Since the 95% CI does not include the null value (1) the finding is statistically significant. We cannot not measure and compare incidence in a case control study since there is no follow up period. This means we cannot directly calculate a risk ratio from the case-control data.***
2. ***The risk ratio was 3.934 and the odds ratio was 4.224. When the odds ratio is interpreted as a relative risk it will always overstate any effect size. Since the odds ratios is greater than one it is expected that it will be larger than the risk ratio. It is also expected to have a larger CI. The risk ratio has a 95% CI of (2.371, 6.527) and the odds ratio has a 95% CI of (1.752, 10.18).***
3. ***Assuming the nested case-control study was sampled accurately, when the disease is uncommon (less than 10% prevalence) the odds ratio is a good estimate of the risk ratio. In the data from the cohort study, the prevalence of cancer was 5.7%, thus the condition was met.***
4. ***A case control study allows you to study multiple exposures and these studies tend to be cheaper and require less time than a cohort study.***
5. ***G: Ecologic (Repeated)***
6. ***B: Cross-sectional***
7. ***G: Ecologic (Repeated)***
8. ***The unit of study is states, which implies ecologic. This was done over time so it is repeated.***
9. Select the study design used by Sivakami et al 2019.

***E: Case Control***

1. Provide an interpretation for the prevalence ratio and 95%CI in Table 5, Sivakami et al 2019, for the “No” response of Menstrual Hygiene Education provided in school.

***A prevalence ratio of 2.41 means that the proportion of missing school during menstruation by adolescent girls in the 3 states in India who report not having menstrual hygiene education in school is 2.41 times the proportion of girls who report having menstrual hygiene education in school. The confidence interval of (1.65,3.53) means that over repeated sampling with 95% confidence the true prevalence ratio is between 1.65 and 3.53. Since the null value (1) is not in the interval, the finding is statistically significant.***

1. Provide an interpretation for the prevalence ratio and 95%CI in Table 5, Sivakami et al 2019, for the “Yes” response to “Pain Medication in school.”

***A prevalence ratio of .39 means that the proportion of missing school during menstruation by adolescent girls in the 3 states in India who report that pain medication is given in their school is .39 times that of girls who report not having pain medication given in their school. The confidence interval of (1.65,3.53) means that over repeated sampling with 95% confidence the true prevalence ratio is between .26 and .59. Since the null value (1) is not in the interval, the finding is statistically significant.***

1. Examine the Tang et al. 2018 Table 2 below. Provide a detailed interpretation for the China “region of birth” ratio measure. The total sample includes patients attending a community health center in NYC, 1997-2017, both those with and without HBV.

***Both Odds Ratios were adjusted for age, sex, region of birth, language preference, insurance, and reported family history of HBV.***

***The Adjusted Odds ratio for China is 40.38. This means that those born in China had 40.38 times the odds of current HBV infection as those born in the United States. We are 95% confident that the adjusted odds ratio of current HBV infection in those born in China compared those born in the United States is between 12.10 and 134.83. Since the 95% CI does not include the null value (1) the finding is statistically significant.***

***The Adjusted Odds ratio for other region of birth is 9.96. This means those born in a region other than China or the United States had 9.96 times the odds of current HBV infection as those born in the United States. We are 95% confident that the adjusted odds ratio of current HBV infection in those born in a region other than China or the United States compared those born in the United States is between 2.76 and 35.92. Since the 95% CI does not include the null value (1) the finding is statistically significant.***

**Appendix**

**R Code**

**etable=function(exposure,disease){**

**t=table(exposure,disease)[2:1,2:1]**

**rownames(t)=c('E+','E-')**

**colnames(t)=c('D+','D-')**

**kable(t)**

**}**

**riskratio=function(a,b,c,d){**

**e=a/(a+b)**

**f=c/(c+d)**

**c(e,f,e/f)**

**}**

**ciriskr=function(ecases,enoncases,uecases,uenoncases){**

**a=ecases**

**b=enoncases**

**c=uecases**

**d=uenoncases**

**e=a/(a+b)**

**f=c/(c+d)**

**rr=e/f**

**z=c(-1.96,1.96)**

**se=sqrt(1/a-1/(a+b)+**

**1/c-1/(c+d))**

**ci=exp(log(rr)+z\*se)**

**c("Risk Ratio"=rr,"95 CI"=ci)**

**}**

**rateratio=function(cases,noncases,pyrcases,pyrnoncases){**

**a=cases**

**b=noncases**

**p=pyrcases**

**q=pyrnoncases**

**(a/p)/(b/q)**

**}**

**cirater=function(cases,noncases,pyrcases,pyrnoncases){**

**a=cases**

**b=noncases**

**p=pyrcases**

**q=pyrnoncases**

**rr=(a/p)/(b/q)**

**se=sqrt(1/a+1/b)**

**z=c(-1.96,1.96)**

**ci=exp(log(rr)+z\*se)**

**c("Rate Ratio"=rr,"95 CI"=ci)**

**}**

**cioddsr=function(ecases,enoncases,uecases,uenoncases){**

**a=ecases**

**b=enoncases**

**c=uecases**

**d=uenoncases**

**or=(a/c)\*(d/b)**

**se=sqrt(1/a+1/b+1/c+1/d)**

**z=c(-1.96,1.96)**

**ci=exp(log(or)+z\*se)**

**c("Odds Ratio"=or,"95 CI"=ci)**

**dat = read\_excel("IDA\_Dataset+8.7.18.xlsx")**

**dat=dat%>%select(-c(X\_\_1,X\_\_2))**

**dat2=dat%>%select(c(CANCER,VIRUS,PERSONDAYS,LTF))**

**dat2=dat2%>%mutate(personyrs=PERSONDAYS/365)**

**taba=table(dat$VIRUS,dat$CANCER)[2:1,2:1]**

**x=c("Exposed","Unexposed")**

**y=c("Cancer","No Cancer")**

**row.names(taba)=x**

**colnames(taba)=y**

**kable(taba)**

**dat2%>%filter(VIRUS==1)%>%summarise(pyrs=sum(personyrs))**

**dat2%>%filter(VIRUS==0)%>%summarise(pyrs=sum(personyrs))**