Charles Hwang
Dr. Whalen
STAT 488-001
2022 February 10

Homework 2

2.1A) $P(-|C)=\frac{1}{4}$, $P(+|\overline{C})=\frac{1}{10}$

2.1B) $P(+|C)+P(-|C)=1$, $P(+|\overline{C})+P(-|\overline{C})=1$
Sensitivity: $\boxed{P(+|C)=\frac{3}{4}}$, Specificity: $\boxed{P(-|\overline{C})=\frac{9}{10}}$

2.1C) $P(C,+)=P(+|C)P(C)=\left(\frac{3}{4}\right)(0.04)=0.03$
$P(C,-)=P(-|C)P(C)=\left(\frac{1}{4}\right)(0.04)=0.01$
$P(\overline{C},+)=P(+|\overline{C})P(\overline{C})=\left(\frac{1}{10}\right)(0.96)=0.096$ $\quad P(C)=0.04$
$P(\overline{C},-)=P(-|\overline{C})P(\overline{C})=\left(\frac{9}{10}\right)(0.96)=0.864$ $\quad P(\overline{C})=1-P(C)=0.96$

	+	−
C	0.03	0.01
$\overline{C}$	0.096	0.864
	0.126	0.874

2.1D) $P(C|+)=\frac{P(+|C)P(C)}{P(+)}=\frac{\left(\frac{3}{4}\right)(0.04)}{(0.126)}=\frac{5}{21}\approx\boxed{0.2381}$

2.3A) $\pi_{US}-\pi_{UK}=\frac{62.4}{1M}-\frac{1.3}{1M}=\frac{61.1}{1M}$, $\frac{\pi_{US}}{\pi_{UK}}=\frac{\frac{62.4}{1M}}{\frac{1.3}{1M}}=48$ Relative risk is more useful because it captures the proportional difference between the two rather than the raw difference.

2.4) Germany: $1-\frac{\frac{9}{2}}{\frac{9}{2}+1}=1-\frac{\frac{9}{2}}{\frac{11}{2}}=1-\frac{9}{11}=\boxed{\frac{2}{11}}$ $\quad \left(1-\frac{\pi}{\pi+1}\right)$
Brazil: $1-\frac{5}{5+1}=1-\frac{5}{6}=\boxed{\frac{1}{6}}$
France: $1-\frac{\frac{11}{2}}{\frac{11}{2}+1}=1-\frac{\frac{11}{2}}{\frac{13}{2}}=1-\frac{11}{13}=\boxed{\frac{2}{13}}$
England: $1-\frac{20}{20+1}=1-\frac{20}{21}=\boxed{\frac{1}{21}}$
Spain: $1-\frac{7}{7+1}=1-\frac{7}{8}=\boxed{\frac{1}{8}}$

2.7A) "The odds of survival for females was 11.4 times the odds of survival for males."

2.7B) Female: $\frac{\pi}{\pi+1}=\frac{2.9}{2.9+1}=\frac{29}{39}\approx\boxed{0.7436}$
Male: $\frac{\pi}{\pi+1}=\frac{\frac{2.9}{4.4}}{\frac{2.9}{4.4}+1}=\frac{2.9}{14.3}\approx\boxed{0.2028}$
$RR=\frac{\frac{2.9}{3.9}}{\frac{2.9}{14.3}}=\frac{11}{3}=\boxed{3.\overline{6}}$

2.10) $\frac{n_{11}n_{22}}{n_{12}n_{21}}=\frac{(433)(554883)}{(570)(8049)}=\frac{8088113}{1527310}\approx 52.369$ The odds of fatality in an auto accident when not wearing a restraint are approximately 52.369 times the odds of fatality when wearing a restraint.

2.11) $\frac{n_{11}n_{22}}{n_{12}n_{21}}=\frac{(802)(494)}{(34)(53)}=\frac{198094}{901}\approx\boxed{219.86}\rightarrow \ln(219.86)\approx 5.393\pm 0.445=(4.948, 5.838)$
$z_{\alpha/2}(SE)=(1.96)\sqrt{\frac{1}{802}+\frac{1}{494}+\frac{1}{34}+\frac{1}{53}}\approx 0.445$ $\quad e^{(4.948, 5.838)}\approx\boxed{(140.89, 343.09)}$

2.17A)

	Democratic	Republican	Independent	Total
White	871 (988.0416)	821 (700.0656)	336 (339.8928)	2028
Black	347 (229.9584)	42 (162.9344)	83 (79.1072)	472
Total	1218	863	419	2500

$X^2=\sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(n_{ij}-\mu_{ij})^2}{\mu_{ij}}\approx\boxed{184.3231}$
$G^2=2\sum_{i=1}^{2}\sum_{j=1}^{3}n_{ij}\ln\left(\frac{n_{ij}}{\mu_{ij}}\right)\approx 213.9010$
$df=(r-1)(c-1)=((2)-1)((3)-1)=\boxed{2}$
$\boxed{p<0.0001}$ — We reject H_0 at the $\alpha=0.05$ level. There is sufficient evidence that there is a relationship between race and party.

$\dfrac{n_{ij}-\hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1-\hat{\pi}_{i+})(1-\hat{\pi}_{+j})}}$

2.17B)	Democratic	Republican	Independent
White	-11.96678	12.99945	-0.53262
Black	11.96678	-12.99945	0.53262

We can see there are more black Democrats ~~than~~ and less white Democrats than expected and more white Republicans and less black Republicans than expected. There are about as many white and black Independents as expected.

2.17C)	Democratic	Republican	Total
White	871 (990.32003)	821 (701.67996)	1692
Black	347 (227.67996)	42 (161.32003)	389
Total	1218	863	2081

$G_1^2 = 2\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}\ln\left(\frac{n_{ij}}{\mu_{ij}}\right) \approx \boxed{213.6202}$

$G_2^2 = G^2 - G_1^2 = (213.9011) - 213.6202 = \boxed{0.2809}$

There is clearly much more difference between observed and expected values for Democrats and Republicans than between observed and expected values for Democrats and Republicans combined and Independents.

2.18) No, it is not valid to apply a χ^2 test of independence to this table. We can clearly see both rows sum to over 100, indicating subjects were allowed to choose more than one answer. However, each of the three variables can have a χ^2 test of independence applied by determining the number of subjects who did not choose the factor. See example table with factor A

Factor A	Yes	No	Total
Men	60	100-60=40	100
Women	75	100-75=25	100
Total	135	65	200

2.28) Clinic	Drug	Success	Failure	%
1	A	18	12	60
	B	12	8	60
2	A	2	8	20
	B	8	32	20
Total	A	20	20	50
	B	20	40	33

$60\% - 60\% = 20\% - 20\% = \underline{0\%}$

$50\% - 33.\overline{3}\% = \underline{16.\overline{6}\%}$

We can see X and Y are conditionally indep. from the difference in percentages, but when Z was ignored, X and Y are marginally associated. It would be misleading to study only the marginal table because it masks the apparent association.

XZ $\hat{\theta}_{X(Y=1)Z} = \frac{18\cdot 8}{12\cdot 2} = 6 \left(\frac{n_{111}\cdot n_{212}}{n_{211}\cdot n_{112}}\right)$

$\hat{\theta}_{X(Y=2)Z} = \frac{12\cdot 32}{8\cdot 8} = 6 \left(\frac{n_{121}\cdot n_{222}}{n_{221}\cdot n_{122}}\right)$

YZ $\hat{\theta}_{(X=1)YZ} = \frac{18\cdot 8}{12\cdot 2} = 6 \left(\frac{n_{111}\cdot n_{122}}{n_{121}\cdot n_{112}}\right)$

$\hat{\theta}_{(X=2)YZ} = \frac{12\cdot 32}{8\cdot 8} = 6 \left(\frac{n_{211}\cdot n_{222}}{n_{221}\cdot n_{212}}\right)$

XY $\hat{\theta}_{XY(Z=1)} = \frac{18\cdot 8}{12\cdot 12} = 1 \left(\frac{n_{111}\cdot n_{221}}{n_{211}\cdot n_{121}}\right)$

$\hat{\theta}_{XY(Z=2)} = \frac{2\cdot 32}{8\cdot 8} = 1 \left(\frac{n_{112}\cdot n_{222}}{n_{212}\cdot n_{122}}\right)$

2.30A) True

2.30B) False

2.30C) True

2.30D) False

2.30E) True