Contingency Tables

Wald Test $\hat{\pi} = \frac{Y}{n} \dot{\sim} N\left(\pi, \frac{\hat{\pi}(1-\hat{\pi})}{n}\right)$

$se(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$

$Z_W = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} \dot{\sim} N(0,1)$ $\quad H_0: \pi = \pi_0$

$H_a: \pi \neq \pi_0$

- When H_0 is true $W = Z_W^2 \dot{\sim} \chi_1^2$

Score Test $\hat{\pi} = \frac{Y}{n} \dot{\sim} N\left(\pi, \frac{\pi_0(1-\pi_0)}{n}\right)$ $\quad H_0: \pi = \pi_0$

$Z_S = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \dot{\sim} N(0,1)$ $\quad H_a: \pi \neq \pi_0$

- When H_0 is true $S = Z_S^2 \sim \chi_1^2$

LRT When $H_0: \pi = \pi_0$ is true

$$L = 2\left(Y \log\left(\frac{\hat{\pi}}{\pi_0}\right) + (n-Y)\log\left(\frac{1-\hat{\pi}}{1-\pi_0}\right)\right) \sim \chi_1^2$$

Exact Inference
- Under $H_0: \pi = \pi_0$ $\quad Y \sim \text{binomial}(n, \pi_0)$
- Say we reject if $Y < a$ or $Y > b$
- Then set $P(\text{Type I error}) = \alpha$
  - ie: $P(Y < a \mid \pi = \pi_0) < \alpha/2$ and $P(Y > b \mid \pi = \pi_0) < \alpha/2$
  - & choose largest a and b
- We can invert this to get an exact confidence Interval

R CODE

```
out1 <- prop.test(x=0, n=25, conf.level=.95, correct=F)
out1$conf.int

out2 <- binom.test(x=0, n=25, conf.level=.95)
out2$conf.int
```

n_{ij}	$Y=1$	$Y=2$	$\cdots$	$Y=J$	Totas
$X=1$	n_{11}	n_{12}	$\cdots$	n_{1J}	n_{1+}
$X=2$	n_{21}	n_{22}	$\cdots$	n_{2J}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$X=I$	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	n_{I+}
totals	n_{+1}	n_{+2}	$\cdots$	n_{+J}	n_{++}

Counts

n_{ij}	$Y=1$	$Y=2$	$\cdots$	$Y=J$	totals
$X=1$	π_{11}	π_{12}	$\cdots$	π_{1J}	π_{1+}
$X=2$	π_{21}	π_{22}	$\cdots$	π_{2J}	π_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$X=I$	π_{I1}	π_{I2}	$\cdots$	π_{IJ}	π_{I+}
totals	π_{+1}	π_{+2}	$\cdots$	π_{+J}	1

Probabilities

Independent if $\pi_{ij} = \pi_{i+} \cdot \pi_{+j}$

Difference in Probability $\pi_1 - \pi_2$

Relative Risk π_1/π_2

Odds Ratio $\theta = \dfrac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \dfrac{n_{11}n_{22}}{n_{12}n_{21}}$

Confidence Interval ① $\log(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\dfrac{1}{n_{11}} + \dfrac{1}{n_{12}} + \dfrac{1}{n_{21}} + \dfrac{1}{n_{22}}} \rightarrow (L, U)$

CI for $\theta \rightarrow (e^L, e^U)$ Odds Ratio

② $\hat{\pi}_1 - \hat{\pi}_2 \sim N\left(\pi_1 - \pi_2, \dfrac{\pi_1(1-\pi_1)}{n_{1+}} + \dfrac{\pi_2(1-\pi_2)}{n_{2+}}\right)$ difference

$(\hat{\pi}_1 - \hat{\pi}_2) \pm z_{\alpha/2}\,\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2)$

③ $\log\left(\dfrac{\pi_1}{\pi_2}\right) \pm z_{\alpha/2}\sqrt{\dfrac{1-\pi_1}{\pi_1 n_{1+}} + \dfrac{1-\pi_2}{\pi_2 n_{2+}}} \rightarrow (L, U)$ Relative Risk

CI for $\dfrac{\pi_1}{\pi_2} \rightarrow (e^L, e^U)$

Test for independence $H_0: \pi_{ij} = \pi_{i+} \cdot \pi_{+j} \equiv \mu_{ij} = n\pi_{i+} \cdot \pi_{+j}$

Pearson Stat $X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$ where $\hat{\mu}_{ij} = n\left(\frac{n_{i+}}{n}\right)\left(\frac{n_{+j}}{n}\right)$

- when H_0 is true $X^2 \sim \chi^2_{(I-1)(J-1)}$
- better in smaller samples

LRT Stat $G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right)$

when H_0 is true $G^2 \sim \chi^2_{(I-1)(J-1)}$

Pearson Residual $e_{ij} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$

standardized $\rightarrow$ $r_{ij} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1-p_{i+})(1-p_{+j})}}$

very rare to have $|r_{ij}| > 3$ when $X \perp Y$

$\boxed{\text{SAS}}$

```
Proc Freq order=data;                          /* 2*2 */
    weight = count;
    tables X Y / chisq expected;
    exact or chisq;
    Pearson's test;
run;
Proc freq order=data;                          /* single proportion */
    weight = count;
    tables X / binomial testp = (.75, .25);  /* H0 p1 = .75 */
    exact binomial chisq;
run;
```

```
proc freq order=data;                    /* 3x3 */
     weight= count;
     tables X*Y / chisq expected;
proc genmod   order=data;
     weight = count;
     class X Y;
     model count = X Y / dist = poi link=log residuals;
     run;
```


woot!