2018 Prelim

Will Gertsch

8/19/2020

1

a

$P(D|+) = 170/200 = 0.85$. We will use the normal approximation confidence interval for the proportion

$$\hat{p} \pm Z_{.975}\sqrt{Var(\hat{p})}$$

where $Var(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{200}$. We use 200 in the denominator since we are only considering the sample of women who test positive. Therefore the confidence interval is

$$0.85 \pm 1.96\sqrt{.85(.15)/200} \implies (0.8005124, 0.8994876)$$

b Assuming $P(D|-) = 0$.

i

Use the law of total probability

$$P(D) = P(D|+)P(+) + P(D|-)P(-) = P(D|+)P(+) = \frac{170}{200} * \frac{200}{5000} = 0.034$$

ii

$\hat{p} = 0.034$. Using the normal approximation again

$$0.034 \pm 1.96\sqrt{.034(1-.034)/5000} \implies (0.02997658, 0.04002342)$$

c Assuming $P(D|-) = 0.06$. Use the law of total probability again.

$$P(D) = P(D|+)P(+) + P(D|-)P(-) = 0.034 + (0.06)(4800/5000) = 0.0916$$

d

We start with the normal CI and solve for n.

$$1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.02$$

Rearranging, we get

$$n = \frac{\hat{p}(1-\hat{p})}{0.0001041233}$$

Since we do not have an estimate of $\hat{p}$, we could use the critics proposed proportion. However, I would prefer to use $\hat{p} = 1/2$ since that value will maximize n and give us an upper bound for the sample size we need. Therefore $n = 2401$. This is actually less than what the investigators thought they needed.

2

a

$$\hat{\beta}_1 = r_{XY}\frac{SD(Y)}{SD(X)} = 0.366 * \frac{1.09}{1.125} = 0.3546133$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 1.953 - 0.3546133 * 1.108 = 1.560088$$

b Every additional attack is associated with a $10^{0.3546133} = 2.262629$ increase in the level of IgE (no units were given for this).

c

I chose to think of this as an overall F-test first.

$$F^* = \frac{MSR}{MSE} = \frac{\frac{5.73}{1}}{\frac{37.06}{35}} = 5.411495$$

F^* has a $F(1,35)$ distribution under the null. $F_{.95}(1,35) = 4.121338$ so reject the null hypothesis. Therefore is evidence that the number of attacks has a significant linear relationship with log10(IgE).

If we didn't want to use the F distribution, we could note that $F^* = T^2$ and that the equivalent t distribution is $t(35)$ with critical value $t_{.95}(35) = 1.689572$.

d

These models are not nested since the SLR model is not a constrained version of the dummy variable model. This means that we cannot use the F-test to compare the models. However since the problem is obviously set up for a partial F-test, I will do a partial F-test.

$$F^* = \frac{\frac{6.76-5.73}{2}}{\frac{36.02}{33}} = 0.4718212$$

The distribution of F^* is $F(2,33)$ and the critical value is $F_{.95}(2,33) = 3.284918$. Therefore, fail to reject the null hypothesis. The dummy variable model is not significantly better at explaining the variance of log10(IgE).

e

I would recommend a Poisson regression model with link function $g(\lambda) = \log\lambda$. In other words, $Y|X = x_i \sim$ Poisson(λ_i). Thus the model is

$$\log\lambda_i = \beta_0 + \beta_1 X_i$$

Every 10-fold unit increase in IgE is associated with a e^{β_1} increase in the number of attacks.

3

4

5