Pearson’s Meta-Analysis

http://arxiv.org/abs/0911.3531

1. - Regresion lineal

http://www.stat.cmu.edu/~hseltman/309/Book/chapter9.pdf

Dado un par de vectores $\vec{X}$ y $\vec{Y}$ quiero establecer un modelo lineal, $y = a + bx$ que relacione a ambos. Tengo n datos para cada par (x,y) por lo que los grados de libertad del modelo serán $df = n - 2$. Los coeficientes de la correlación se calculan según, $$ b = \frac{\sum_{i=1}^{n}{\left(x_i - \bar{x} \right)\left(y_i - \bar{y} \right) }}{\sum_{i=1}^{n} \left(x_i - \bar{x} \right) ^2} $$ y $$ a = \bar{y} - b\bar{x}$$

El mejor estimado de la varianza de los datos respecto al modelo, $\sigma ^2$, es, $$ s^2 = \frac{\sum_{i=1}^{n} \left(y_i - a - b x_i \right) ^2}{n -2}$$ y los errores estandard de cada coeficiente del modelo son, $$ SE(a) = s \sqrt{ \frac{\sum_{i=1}^{n}{x_i}^2}{c\sum_{i=1}^{n}{x_i}^2 - \left(\sum_{i=1}^{n}{x_i} \right)^2}} $$ y $$ SE(b) = s \sqrt{ \frac{n}{c\sum_{i=1}^{n}{x_i}^2 - \left(\sum_{i=1}^{n}{x_i} \right)^2}} $$

El coeficiente de correlación de Pearson, r, se calcula según, $$ r = \frac{n \sum_{i=1}^{n}{x_i y_i} - \sum_{i=1}^{n}{x_i} \sum_{i=1}^{n}{y_i}}{\sqrt{\left[n \sum_{i=1}^{n}{{x_i}^2} - \left(\sum_{i=1}^{n}{x_i} \right) ^2 \right] \left[n \sum_{i=1}^{n}{{y_i}^2} - \left(\sum_{i=1}^{n}{y_i} \right) ^2 \right] }}$$ y es lo que da una idea de que tan bueno es el ajuste (o la regresión de los datos) a una línea recta.

Asumiendo que hayamos hecho la hipotesis nula ($a=0, b=0$), los t-estadísticos de cada coeficiente se calculan como, $$ t_a = \frac{a}{SE(a)}, t_b = \frac{b}{SE(b)}$$ que corresponden a una distribución t-Student con $n -2$ grados de libertad. De aquí pueden calcularse los p-values (p_a, p_b) correspondientes invirtiendo la distribución de Student, $$ p_j = \Theta \left(t_j, n-2 \right) $$

http://search.cpan.org/~mikek/Statistics-Distributions-1.02/Distributions.pm

Lo que no entiendo aqui es porque no se calculan los p-values directamente a partir del inverso de la CDF que ha quedado caraterizada al calcular la desviacion estandard del ajuste. Tal y como yo lo veo $p = 1 - \phi(1.96 \sigma) $, siendo ϕ la *normal cumulative distribution function*.

2.- Meta analisis de las regresiones

2.1.- Método que encontre por ahí

http://hmg.oxfordjournals.org/content/17/R2/R122.long

2.2.- Pearson's Method (en desarrollo)

Si se han hecho varias regresiones es posible hacer un meta-analisis de las mismas. Supongamos que la hipotesis nula es, $$ H_0 : (b_0, b_1, \ldots , b_k) = 0 $$ siendo k el numero de regresiones totales que se han hecho. Supongamos tambien que de cada regresión disponemos de los valores de b, n y p. (Notese que b indica la dirección del efecto por lo que el *odd ratio*, OR, sería de la misma utilidad aquí).

Para calcular el meta-análisis utilizando el *Stouffer*'s *Z-score method* (podría usarse cualquier otro) se calcula (Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach, M. C. WHITLOCK, doi: 10.1111/j.1420-9101.2005.00917.x), $$ Z_{-} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \phi^{-1}(1 - p_i)$$ es el *z-score* indicando que $b<0$ y $$ Z_{+} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \phi^{-1}(p_i)$$ indicando que0 $b_i >0$.

Ahora escogemos el *z-score* resultante como $Z = \max (Z_-, Z_+)$ y ~~nuestro *p-value* final será, $p = \phi \left(Z \right)$. Esto además nos proporciona la "dirección del efecto", o sea, el signo de la pendiente del ajuste (b).~~