Homework 10, Part 3

(46) Does the data indicate that the use of Vitamin C reduces mean time to recover?
Find P value. Do we reject at $\alpha = .05$?

	No Vit C (1)	Vit C (2)
n	35	35
μ	6.9	5.8
σ	2.9	1.2

Null: $H_0: \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$

$H_A: \mu_1 > \mu_2$

Test Statistic: $z = \dfrac{(6.9-5.8) - (0)}{\sqrt{\frac{1}{35}\left((2.9)^2 + (1.2)^2\right)}} = 2.074$

$$\frac{(\mu_1-\mu_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

P-Value: $P(z > 2.074) = .0192$

So at $\alpha = .05$ We reject the null Hypothesis which means the data supports Vitamin C reducing recovery time.

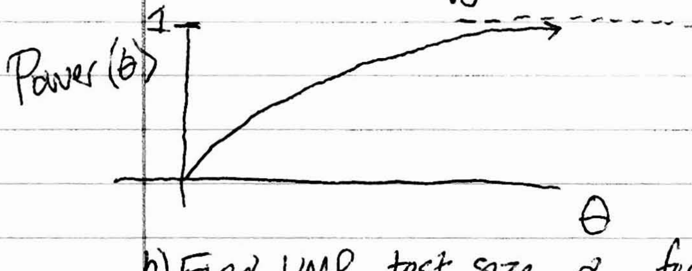
(84) Suppose $f_{Y|\theta} = \begin{cases} \theta y^{\theta-1} & 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$ where $\theta > 0$

a) Sketch the Power function RR $Y > .5$

defn: $\text{Power}(\theta) = \Pr(W \text{ in } RR \mid H_0 = \theta)$ where W is your test statistic $(=Y)$

$$\int_{.5}^{1} \theta y^{\theta-1}\, dy = y^{\theta}\Big|_{.5}^{1}$$

$$= 1 - (.5)^{\theta}$$



b) Find UMP test size α for $H_0: \theta_0 = 1$ vs $H_a: \theta > 1$

Start with a simple test. Let $\theta_a > 1$.

The N-P Lemma Says the RR for a level α test that maximizes power at θ_a is determined by:

$$\frac{L(\theta_0)}{L(\theta_a)} \le K$$

K chosen to give you α

In our case $L(\theta_0 = 1) = 1$

$L(\theta_a) = \theta_a y^{\theta_a - 1}$

so RR has the form $\frac{1}{\theta_a y^{\theta_a - 1}} < K$ or $\frac{1}{K\theta_a} < y^{\theta_a - 1}$

equivalently $\left(\frac{1}{K\theta_a}\right)^{\frac{1}{\theta_a - 1}} < Y$

B/c θ_a is a known constant (we fixed it by assumption), the LHS of the inequality is a constant, call it K^*

So M.P. test of $H: \theta = 1$ vs $A: \theta = \theta_a$ has RR $\{Y > K^*\}$ where the value of K^* is determined by α

$$\alpha = P(Y \text{ in } RR \mid H_0: \theta_0 = 1) = P(Y > K^* \mid \theta_0 = 1)$$

$$= \int_{K^*}^{1} 1\, dy = 1 - K^*$$

so $K^* = 1 - \alpha$

Test statistic Y and RR $\{Y > K^*\}$ of the Level α test do not depend on a particular ~~Level~~ value of θ_a so long as it is larger than 1

i.e. Any value of θ_a Leads to the same RR

(change the original K)

Thus this test is UMP Level α for $H: \theta = 1$ vs $A: \theta > 1$

(86) Let $Y_1 \ldots Y_n \sim f_{Y|\theta} = \begin{cases} \frac{1}{\theta} m y^{m-1} e^{-y^m/\theta} & y > 0 \\ 0 & \text{elsewhere} \end{cases}$

Where m is a known constant

a) Find UMP test $H_0: \theta = \theta_0$ vs $H_A: \theta > \theta_0$

Let $\theta_a > \theta_0$, start with a simple test so N-P applies

$$L(\theta_0) = \prod \frac{1}{\theta_0} m y^{m-1} e^{-y^m/\theta_0}$$

$$= \left(\frac{1}{\theta_0}\right)^n \left(\prod m y^{m-1}\right) e^{-\frac{1}{\theta_0}\sum_i y_i^m}$$

Similarly for $L(\theta_a)$

so

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{\left(\frac{1}{\theta_0}\right)^n e^{-\frac{1}{\theta_0}\sum_i y_i^m}}{\left(\frac{1}{\theta_a}\right)^n e^{-\frac{1}{\theta_a}\sum_i y_i^m}} = \left(\frac{\theta_a}{\theta_0}\right)^n e^{\overbrace{\left(\frac{1}{\theta_a} - \frac{1}{\theta_0}\right)}^{\text{Negative Quantity}}\overbrace{\left(\sum y_i^m\right)}^{\text{sufficient statistic}}} < K$$

Take Logs and solve for $\sum_i y_i^m$ $\quad \left(\frac{1}{\theta_a} - \frac{1}{\theta_0}\right)\left(\sum_i y_i^m\right) < \ln\left(K\left(\frac{\theta_0}{\theta_a}\right)^n\right)$

$$\Rightarrow \sum_i y_i^m > \frac{\ln\left(K \cdot \left(\frac{\theta_0}{\theta_a}\right)^n\right)}{\frac{1}{\theta_a} - \frac{1}{\theta_0}}$$

sign flipped b/c divide by negative number.

Call the RHS K^*, a constant.

<u>Note</u>: I would expect you all could get here... past this point, I am more or less copying the answer booklet. It's a distribution you were not told to know...

Consider the distribution of $Z = Y^m$ since $\frac{dz}{dY} = mY^{m-1}$, if H_0 is true

$g(z) = \frac{1}{\theta_0} e^{-z/\theta_0}$ for $z > 0$

That is Y^m has a gamma distribution $\beta = \theta_0$ $\alpha = 1$

So $\frac{2Y^m}{\theta_0}$ has χ^2 dist. with 2 degrees of freedom

($\alpha = \nu/2$, $\alpha = 1$ so $\nu = 2 =$ d.f.)

Also $\frac{2\sum Y_i^m}{\theta_0}$ is χ^2, $2n$ d.f.

Thus the critical region $\sum Y_i^m > K^* \Rightarrow \frac{2\sum Y_i^m}{\theta_0} > \frac{2K^*}{\theta_0} = K^{**}$

where K^{**} chosen so test has size α

Notice the critical region does not depend on θ_a but only that $\theta_a > \theta_0$

(why? so that $\frac{1}{\theta_a} - \frac{1}{\theta_0}$ is negative, Look back to where we used this)

So the same region holds for all $\theta_a > \theta_0$

Thus this test is UMP Level α

b) $\theta_0 = 100$ $\alpha = \beta = .05$ $\theta_a = 400$ Find appropriate sample size and critical region

if H_0 true $\frac{2\sum Y_i^m}{\theta_0}$ has χ^2 dist $2n$ d.f.

Cutoff point for RR

Statistic to calculate cutoff point for RR

100

so $\alpha = P_r\left(\underbrace{\frac{2\sum Y_i^m}{100} \geq \chi^2_{.05, 2n}}_{\text{Prob. of statistic in RR}} \right) = .05$

$\chi^2_{a,b}$ $0 \le a \le 1$ and $P_r(X > \chi^2_{a,b}) = a$

if H_a true $\frac{2\sum Y_i^m}{400}$ has χ^2 dist $2n$ d.f.

$.05 = \beta = P\left(\frac{2\sum Y_i^m}{400} \le \chi^2_{.95, 2n}\right) = P\left(\frac{1}{4}\frac{2\sum Y_i^m}{400} \le \frac{1}{4}\chi^2_{.95,2n}\right)$

$= P\left(\frac{2\sum Y_i^m}{100} \le 4\chi^2_{.95, 2n}\right)$

Since cutoffs for RR are <u>equal</u> $\Rightarrow 4\chi^2_{.95,2n} = \chi^2_{.05,2n}$

so now find an n that satisfies $P(\chi^2 \le 4\chi^2_{.95,2n}) = .05$
(We defined the cutoff for our RR as $\chi^2_{.05,2n}$, equivalently as $4\chi^2_{.95,2n}$)
or $\frac{1}{4}\chi^2_{.05,2n} = \chi^2_{2n}$

You get $2n = 12$ or $n = 6$
(Use tables in Back of Book)

(94) Regard # voters favoring A in 4 samples as independent Binomial R.V.'s
Construct Likelihood Ratio Test, $\alpha = .05$

$H: p_1 = p_2 = p_3 = p_4$ A: fraction of voters favoring A not the same in all 4 wards

	1	2	3	4	Total
FAVOR	76	53	59	48	236
NOT FAVOR	124	147	141	152	564
	200	200	200	200	800

$$\boxed{\lambda = \frac{L(\hat{\Omega}_H)}{L(\Omega)} = \frac{\max_{\theta \in \Omega_H} L(\theta)}{\max_{\theta \in \Omega} L(\theta)}}$$

In general, the Likelihood is

$$L(\theta) = \prod_{i=1}^{4} \binom{200}{n_i} (p_i)^{n_i} (1-p_i)^{200-n_i}$$

and MLE estimate $\hat{p}_i = \frac{n_i}{200}$ (Besides being obvious, do you know how to derive this?)

In the restricted space Ω_H $L(\theta) = \left(\prod \binom{200}{n_i}\right) \cdot p^{\sum_i n_i} (1-p)^{800-\sum n_i}$

where $p = p_1 = p_2 = p_3 = p_4$

once again MLE of p: $\hat{p} = \frac{\sum_i n_i}{800}$ (easy to verify)

$$\text{so } \lambda = \frac{\max L(\theta_H)}{\max L(\theta)} = \frac{\left(\frac{\sum_i n_i}{800}\right)^{\sum_i n_i} \left(\frac{800-\sum_i n_i}{800}\right)^{800-\sum_i n_i}}{\prod_{i=1}^{4} \left(\frac{n_i}{200}\right)^{n_i} \left(\frac{200-n_i}{200}\right)^{200-n_i}}$$

(put in all your estimates for the p's)

Since the n_i are Large: Thm 10.2 tells us $-2\ln\lambda$ has χ^2 dist. with $r_0 - r$ d.f.

In our case $r_0 = 3$ $r = 0$ so D.f $= 3 - 0 = 3$

(Why?) if $p = p_1 = p_2 = p_3 = p_4$ By picking p_1 we have now fixed the other 3 parameters by specifying ONE of them (free to choose p then other 3 are fixed)

Also, $r = 0$ since we are free to choose all 4 parameters i.e. zero are fixed

$$\text{so } -2\ln\lambda = -2\left\{ \sum_i n_i \ln\left(\frac{\sum_i n_i}{800}\right) + \dots - \left(200 - n_4\right)\ln \frac{200-n_4}{200} \right\}$$

$$= 10.54$$

• n values taken from chart

So the Rejection Region at $\alpha = .05$ contains values of $-2\ln\lambda$ that exceed $\chi^2_{.05,3} = 7.81$

since $10.51 > 7.81$, we reject the hypothesis that the same proportion of voters in each Ward favor candidate A

(97) a) Let $X_1 \dots X_m \sim exp(\theta_1)$ $\quad \frac{1}{\theta_1} e^{-x/\theta_1}$.

$Y_1 \dots Y_N \sim exp(\theta_2)$

Find Likelihood Ratio criterion $H: \theta_1 = \theta_2$ $\quad A: \theta_1 \neq \theta_2$

(Note there would be $\underset{\text{1 variable fixed}}{1} - \underset{\text{zero fixed}}{0} = 1$ d.f for χ^2 test)

$$\lambda = \frac{\max_{\theta \in \Omega_H} L(\theta)}{\max_{\theta \in \Omega} L(\theta)}$$

For the general Likelihood

$$L(\theta, x) = \prod \frac{1}{\theta_1} e^{-x/\theta_1} = \left(\frac{1}{\theta_1}\right)^m e^{-\frac{1}{\theta_1}\sum x_i}$$

$$L(\theta, y) = \qquad = \left(\frac{1}{\theta_2}\right)^N e^{-1/\theta_2 \sum y_i}$$

so full likelihood: $L(\theta, X, Y) = \left(\frac{1}{\theta_1}\right)^m \left(\frac{1}{\theta_2}\right)^N e^{-1/\theta_2 \sum y_i - \frac{1}{\theta_1}\sum x_i}$

To get MLE estimates, take logs, then Partial derivatives setting each to zero

You get $\hat{\theta}_1 = \sum x_i / m$ $\quad \hat{\theta}_2 = \sum y_i / N$

For the Restricted Likelihood we have $\theta_1 = \theta_2$

$$\text{so } L(\theta) = \left(\frac{1}{\theta}\right)^{m+N} e^{-\frac{1}{\theta}\left(\sum_i x_i + \sum y_i\right)}$$

MLE Estimate:

$$\ln(L(\theta)) = -(M+N)\ln\theta - \frac{1}{\theta}\left(\sum_1 X_i + \sum_1 Y_i\right)$$

Derivative

$$\frac{-(M+N)}{\theta} + \frac{1}{\theta^2}\left(\sum_1 X_i + \sum_1 Y_i\right) = 0$$

$$\Rightarrow \hat{\theta} = \frac{\sum_1 X_i + \sum_1 Y_i}{m+N}$$

$\frac{1}{\hat{\theta}}$

$$\text{Thus } \lambda = \frac{\left(\frac{m+N}{\sum_1 X_i + \sum Y_i}\right)^{m+N} e^{-\left(\sum_1 X_i + \sum_1 Y_i\right)\cdot\frac{(M+N)}{\sum_1 X_i + \sum_1 Y_i}}}{\left(\frac{m}{\sum_1 X_i}\right)^{m} e^{-m}\left(\frac{N}{\sum_1 Y_i}\right)^{N} e^{-N}}$$

Notice that $e^{-\left(\sum_1 X_i + \sum_1 Y_i\right)\cdot\frac{(M+N)}{\sum_1 X_i + \sum_1 Y_i}} = e^{-m+N}$ which cancels in denominator

$$\text{so } \lambda = \frac{\left(\frac{m+N}{\sum_1 X_i + \sum Y_i}\right)^{m+N}}{\left(\frac{m}{\sum X_i}\right)^{m}\left(\frac{N}{\sum_1 Y_i}\right)^{N}}$$

If anyone is unclear about Question 86 and cares how to do it...
We can go over it in Section or office hours next week

– Eric