Stats HW 8

1 We want to test $H_0: \mu_x = \mu_y$ vs $H_1: \mu_x < \mu_y$

Let $W = \dfrac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$ $\quad \mu_x - \mu_y = 0$

$\bar{x} = -4.7 \quad \bar{y} = -1.6$

$S_x = 7.05 \quad S_y = 5.36$

$$\bar{w} = \frac{-4.7 + 1.6}{\sqrt{\frac{7.05^2}{77} + \frac{5.36^2}{79}}} = -3.08$$

To check to use Z / t test we can use

$\hat{\theta} = \frac{7.03^2}{5.36^2}$; $df = \dfrac{\left(1.72 + \frac{77}{79}\right)^2}{\frac{1}{76}(1.72)^2 + \frac{1}{78}\left(\frac{77}{79}\right)^2} = 142.7100$, use Z test

So $\alpha = 0.05$, have Z test $Z_{0.05} = -1.64$

$-3.08 < -1.64$, so we reject H_0 so this is statisically signifigant.

2 We have $H_0: \mu_x = \mu_y$ and $H_a: \mu_x < \mu_y$

Since told $\sigma_x = \sigma_y$ in problem, $S_p = \dfrac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$

$$S_p = \frac{108(2017)^2 + 493(1897)^2}{109 + 494 - 2} = 3,683,010.73$$

We can use Student t distribution with 601 dfs.

$$T_{601} = \frac{(10732 - 10970)}{\sqrt{3,683,010.73}\sqrt{\frac{1}{109} + \frac{1}{494}}} = -1.1762$$

Since $\mu_x < \mu_y$ $\quad -1.1762 < -t_{0.05, 602}$

At df = 602, we can say it's approximately normal, so $-Z_{0.05} = 1.64$, Since $-1.172 < 1.64$, the result is statistically signifigant and we can reject H_0.

3) GLRT: $\lambda = \dfrac{L(\omega_e)}{L(\Omega_e)} =$

When $\sigma_x^2 = \sigma_y^2 = \sigma^2$, the MLE for σ^2 max $L(\sigma^2)$ for $\forall$ x and y,

let $X_i' = X_i - \mu$, $y_i' = y_i - \mu$, so $\mu_{x'} = \mu_{y'} = 0$.

The $x_1', \ldots x_n', y_1', \ldots y_m'$ is normal w/ $\mu = 0$ and STD $= \sigma$

For a normal distribution, we have shown that $L(\sigma^2) = \frac{1}{n}\sum(x_i - \bar{x})^2$

So for our distribution, $L(\sigma^2) = \frac{1}{m+n}[x_1'^2 + \ldots + x_n'^2 + y_1'^2 + \ldots + y_m'^2]$

an simplify for one ... er by following above formula.

So $\hat{\sigma}^2 = \frac{1}{n+m}\left(\sum_{i=1}^{n} x_i' + \sum_{j=1}^{m} y_j'\right)$

So $L(\hat{\omega}) = \prod_{i=1}^{n} f(x_i; \sigma_x^2) \prod_{i=1}^{m} f_y(y, \sigma^2)$

Therefore, from $\hat{\sigma}^2$, as $f(x; \sigma_x^2) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)}$

$L(\hat{\omega})$ then becomes $\left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n+m}{2}} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{m}(y_i \ldots\right)}$

$= \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{\frac{n+m}{2}} e^{-(n+m)/2}$

$L(\hat{\Omega}) = \left(\frac{1}{2\pi\hat{\sigma}_x^2}\right)^{n/2} e^{\frac{1}{2\sigma_2}\sum(x_i - \bar{x})^2} \cdot \left(\frac{1}{2\pi\hat{\sigma}_y^2}\right)^{m/2} e^{\frac{1}{2\sigma_y^2}\sum_{i=1}^{m}(\ldots}$

Follows from product definition

$\left(\frac{1}{2\pi\hat{\sigma}_x}\right)^{n/2}$, $-m/2$

This happens as it cancels out with $\hat{\sigma}^2$ estimator as $\hat{\sigma}_x^2 = \frac{1}{n}\sum(x - \bar{x})^2$ for X alone, so it becomes $\frac{1}{2\sigma}$ ...

So $L(\hat{\Omega}) = \left(\frac{1}{2\pi\hat{\sigma}_x^2}\right)^{n/2} e^{-m/2} \left(\frac{1}{2\pi\sigma_y^2}\right)^{m/2} e^{-n/2}$

3) $$GLRT = \frac{L(\omega)}{L(\Omega)} = \frac{\left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{n+m/2} e^{-\left(\frac{n+m}{2}\right)}}{\left(\frac{1}{2\pi\hat{\sigma}_x^2}\right)^{n/2}\left(\frac{1}{2\pi\hat{\sigma}_y^2}\right)^{m/2} e^{-\left(\frac{m+n}{2}\right)}}$$

$$GLRT = \frac{(\hat{\sigma}_x^2)^{n/2}(\hat{\sigma}_y^2)^{m/2}}{(\hat{\sigma}^2)^{\left(\frac{n+m}{2}\right)}}$$

$$= \frac{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^{n/2}\left(\frac{1}{m}\sum_{j=1}^{m}(y_j-\bar{y})^2\right)^{m/2}}{\left[\frac{1}{m+n}\left(\sum(x_i-\bar{x})^2+\sum(y_j-\bar{y})^2\right)\right]^{\frac{n+m}{2}}}$$

$$\lambda = \frac{(m+n)^{\frac{m+n}{2}}}{n^{n/2}m^{m/2}} \cdot \frac{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{n/2}\left[\sum_{j=1}^{m}(y_j-\bar{y})^2\right]^{m/2}}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{j=1}^{m}(y_j-\bar{y})^2\right]^{(m+n)/2}}$$

$\downarrow$

This matches the book!

4) Find 95% confidence interval for $\mu_x - \mu_y$

From definition, formula for this confidence interval is:

$$\left[\bar{x}-\bar{y}-t_{\alpha/2,n+m-2}S_p\sqrt{\tfrac{1}{n}+\tfrac{1}{m}},\ \bar{x}-\bar{y}+t_{\alpha/2,n+m-2}S_p\sqrt{\tfrac{1}{n}+\tfrac{1}{m}}\right]$$

$\bar{x} = 83.96$, $\bar{y} = 84.84$, $\alpha = 0.05$, and $n = 5$, $m = 7$,

So the CI becomes:

$$\left[83.96-84.84-t_{0.025,10}\,11.2\sqrt{\tfrac{1}{5}+\tfrac{1}{7}},\ 83.96-84.84+t_{0.025,10}\,11.2\sqrt{\tfrac{1}{5}+\tfrac{1}{7}}\right]$$

$= [-15.492, 13.732]$, I don't think the data suggests that waving is influenced by a dome, as the 95% confidence interval is pretty symmetrically distributed close to 0, and it isn't reasonable to say results mean that w/ a dome means more burying

5) Have 2 groups, born 12AM-4, born elsewhere. $E(12 \text{ to } 4) = \frac{1}{6}(2650) = 441.\overline{66}$

$H_0: p_1 = p_{1_0}, p_2 = p_{2_0}$, $H_1: p_1 \neq p_{1_0}$ ... so ...

$E(\text{anywhere else}) = \frac{5}{6}(2650) = 2208.\overline{3}$

However, actually was 494 between 12-4, so 2156 rest of year

We can find $d = \left(\frac{2156 - 2208\frac{1}{3}}{2208\frac{1}{3}}\right)^2 + \left(\frac{494 - 441.\frac{2}{3}}{441\frac{2}{3}}\right)^2$

$d = 7.44$

Can compare to $X^2_{0.95,1}$ as have 2 buckets

$X^2_{0.95,1} = 3.841 < 7.44$, so we reject H_0 and the data is not what we would expect if the data were uniformly distributed in all time periods

6 If the data was Poisson distributed,

$\lambda = \frac{\#\text{turnovers}}{\#\text{games}} = \frac{800}{440} = 1.82$

Under Poisson w/ r.v. X, and $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$P(X=0) = 0.162$ $\quad P(X=4) = 0.074$
$P(X=1) = 0.295$ $\quad P(X=5) = 0.027$
$P(X=2) = 0.268$ $\quad P(X=6) = 0.008$ } used 3 sig-figs
$P(X=3) = 0.163$

We then multiply each by 440, and ~~tag~~ the ~~follow~~

$d = \sum_{i=0}^{6} \frac{(X_i - 440p_i)^2}{440p_i}$ where $p_i : p(X=i)$ and X_i is num observed

→ evaluated via excel

$d = 6.45$, $X^2_{0.95,6} = 12.592$.

Since $H_0: p_0 = p_{1_0} \ldots p_6 = p_{6_0}$ and $H_1: p_0 \neq p_{1_0} \ldots p_6 \neq p_{1_6}$

And $d = 6.45 < X^2_{0.95,6}$ we can fail to reject H_0 and state the data fits the Poisson distribution

Scanned with CamScanner