STA 250/MTH 342 Intro to Mathematical Statistics

Homework 11

Solution 1: Exercise 10.1.8 (page 633)

When H_0 holds true, we denote X the height of a man selected randomly from the city, so $Z = X - 68 \sim \mathcal{N}(0, 1)$. We have

$$p_1^0 = \Pr(X < 66) = \Pr(Z < -2) = 0.02275,$$
$$p_2^0 = \Pr(66 \leq X < 67.5) = \Pr(-2 \leq Z < -0.5) = 0.2858,$$
$$p_3^0 = \Pr(67.5 \leq X < 68.5) = \Pr(-0.5 \leq Z < 0.5) = 0.3829,$$
$$p_4^0 = \Pr(68.5 \leq X < 70) = \Pr(0.5 \leq Z < 2) = 0.2858,$$
$$p_5^0 = \Pr(X \geq 70) = \Pr(Z \geq 2) = 0.02275.$$

Since the sample size is $n = 500$, we compare the frequency and its expectation under H_0 in the following table.

	N_i	np_i^0
$X < 66$	18	11.35
$66 \leq X < 67.5$	177	142.9
$67.5 \leq X < 68.5$	198	191.5
$68.5 \leq X < 70$	102	142.9
$X \geq 70$	5	11.35

We do the χ^2-test, and the below testing statistic has approximately a χ_4^2 distribution.

$$Q = \sum_{i=1}^{5} \frac{(N_i - np_i^0)^2}{np_i^0} = 27.50.$$

Therefore the p-value is $1 - F_{\chi_4^2}(27.5) = 1.575 \times 10^{-5}$, and one rejects H_0 so long as the level α is greater than 1.575×10^{-5}.

Solution 2: Exercise 10.3.4 (page 645)

We carry out the test of independence. The contingency table is given below

	Wears a moustache	Does not wear a moustache	Total
Between 18 and 30	12	28	40
Over 30	8	52	60
Total	20	80	100

The MEL's $\hat{m}_{ij} = \frac{X_{i+}X_{+j}}{n}$ are listed below.

$\hat{m}_{ij}$	Wear a moustache	Does not wear a moustache
Between 18 and 30	8	32
Over 30	12	48

So the observed χ^2 testing statistic is $Q = 25/6$. The degree of freedom is 1, and the p-value is $1 - F_{\chi^2_1}(25/6) = 0.04123$. Therefore when the level α is greater than 0.04123 we reject H_0.

Solution 3: Exercise 10.4.5 (page 652)
The correct table to be analyzed is as follows:

Supplier	Defectives	Nondefectives
1	1	14
2	7	8
3	7	8

The value of Q found from this table is 7.2. If Q has the χ^2 distribution with $(3-1)(2-1) = 2$ degrees of freedom, then $\Pr(Q \geq 7.2) = 0.027 < 0.05$.

Solution 4:
For Problem 3 in HW10, we have the rejection region $\mathscr{R}(K) = \{\Lambda > K\}$, where

$$\Lambda = \frac{\max_{\Theta} L}{\max_{\Theta_0} L} = \prod_{i=1}^{m} \left(\frac{X_i \sum n_j}{n_i \sum X_j}\right)^{X_i} \left(\frac{(n_i - X_i)\sum n_j}{n_i \sum (n_j - X_j)}\right)^{n_i - X_i}. \tag{1}$$

We know that under "smoothness" conditions, when the sample size is large, the LR statistic $2\log\Lambda \sim_{approx} \chi^2_h$. The degrees of freedom is $h = p - q$, where p is the number of free parameters estimated in computing the global MLE, and q is the number of free parameters estimated in computing the restricted MLE. In this case,

$$\Theta_0 = \{(p_1, \cdots, p_m) | 0 \leq p_1 = \cdots = p_m \leq 1\},$$
$$\Theta = \{(p_1, \cdots, p_m) | 0 \leq p_1, \cdots, p_m \leq 1\}.$$

So we have $h = m-1$, thus $2\log\Lambda \sim_{approx} \chi^2_{m-1}$. The corresponding p-value of this generalized likelihood ratio test is $1 - F_{\chi^2_{m-1}}(2\log\Lambda)$, here Λ is calculated based on (1) from the observed data.

For Problem 4 in HW10, we have the rejection region $\mathscr{R}(K) = \{\Lambda > K\}$, where

$$\Lambda = \frac{\max_{\theta\in\Theta} L}{\max_{\theta\in\Theta_0} L} = \frac{L(\bar{X}, \bar{Y}, \bar{W}, \hat{\sigma}_X, \hat{\sigma}_Y, \hat{\sigma}_W)}{L(\bar{X}, \bar{Y}, \bar{W}, \hat{\sigma}, \hat{\sigma}, \hat{\sigma})}. \tag{2}$$

where $\hat{\sigma}$ is defined in

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2 + \sum_{i=1}^{n_3}(W_i - \bar{W})^2}{n_1 + n_2 + n_3}.$$

$\hat{\sigma}_X$ is defined in

$$\hat{\sigma}_X^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(X_i - \bar{X}\right)^2.$$

and $\hat{\sigma}_Y$ and $\hat{\sigma}_W$ are defined similarly as $\hat{\sigma}_X$. In this case,

$$\Theta_0 = \{\theta = (\mu_X, \mu_Y, \mu_W, \sigma_X, \sigma_Y, \sigma_W) | 0 < \sigma_X = \sigma_Y = \sigma_W < \infty\},$$
$$\Theta = \{\theta = (\mu_X, \mu_Y, \mu_W, \sigma_X, \sigma_Y, \sigma_W) | 0 < \sigma_X, \sigma_Y, \sigma_W < \infty\}.$$

So we have $h = 3 - 1 = 2$, thus $2\log\Lambda \sim_{approx} \chi_2^2$. The corresponding p-value of this generalized likelihood ratio test is $1 - F_{\chi_2^2}(2\log\Lambda)$, here Λ is calculated based on (2) from the observed data.

Solution 5:

(a). When $\mu_X = \mu_Y$, T has a t-distribution with freedom $n + m - 2$.

(b). Denote σ^2 the variance of both X_i's and Y_j's. Let

$$Z = \frac{\bar{X} - \bar{Y} - \mu_X + \mu_Y}{\sigma \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}}, \quad W = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{\sigma^2}.$$

We have learned that $Z \sim \mathcal{N}(0,1)$, $W \sim \chi^2_{n+m-2}$ and that Z and W are independent, therefore

$$\frac{\bar{X} - \bar{Y} - \mu_X + \mu_Y}{s_{pooled}\sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{Z}{\sqrt{W/(m+n-2)}} \sim t_{n+m-2}.$$

(c). Since with confidence $1 - \alpha$,

$$F^{-1}_{t_{n+m-2}}\left(\frac{\alpha}{2}\right) \leq \frac{\bar{X} - \bar{Y} - \mu_X + \mu_Y}{s_{pooled}\sqrt{\frac{1}{m} + \frac{1}{n}}} \leq F^{-1}_{t_{n+m-2}}\left(1 - \frac{\alpha}{2}\right),$$

we solve the inequalities to give the confidence interval,

$$\bar{X} - \bar{Y} - F^{-1}_{t_{n+m-2}}\left(1 - \frac{\alpha}{2}\right) s_{pooled}\sqrt{\frac{1}{m} + \frac{1}{n}} \leq \mu_X - \mu_Y \leq \bar{X} - \bar{Y} - F^{-1}_{t_{n+m-2}}\left(\frac{\alpha}{2}\right) s_{pooled}\sqrt{\frac{1}{m} + \frac{1}{n}}.$$

(d). For testing the hypothesis $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X \neq \mu_Y$, with level α the rejection region could be chosen as

$$\mathscr{R}(\alpha) = \left\{|T| > F^{-1}_{t_{n+m-2}}\left(1 - \frac{\alpha}{2}\right)\right\}.$$

One has

$$
\begin{aligned}
T \in \mathscr{R}(\alpha) \Longleftrightarrow & T > F^{-1}_{t_{m+n-2}}\left(1-\frac{\alpha}{2}\right) \text{ or } T < -F^{-1}_{t_{m+n-2}}\left(1-\frac{\alpha}{2}\right) = F^{-1}_{t_{m+n-2}}\left(\frac{\alpha}{2}\right) \\
\Longleftrightarrow & 0 < \bar{X} - \bar{Y} - F^{-1}_{t_{n+m-2}}\left(1-\frac{\alpha}{2}\right) s_{pooled}\sqrt{\frac{1}{m}+\frac{1}{n}}, \text{ or} \\
& \bar{X} - \bar{Y} - F^{-1}_{t_{n+m-2}}\left(\frac{\alpha}{2}\right) s_{pooled}\sqrt{\frac{1}{m}+\frac{1}{n}} < 0 \\
\Longleftrightarrow & 0 \text{ is not in the CI found in (c).}
\end{aligned}
$$

Solution 6:

H_0: death rate (death/day) is expected to be equal for each month H_A: death rate (death/day) is not equal for each month expected death/day $= \frac{totalnumberofdeath}{365} = \frac{1668+1407+...+1526}{365} = 46.3452$

Month	Day Number	Exp Death	$x_i - m_i$
Jan	31	1436.70137	231.2986
Feb	28	1297.6658	109.34247
Mar	31	1436.70137	-66.7014
Apr	30	1390.35616	-81.3562
May	31	1436.70137	-95.7014
Jun	30	1390.35616	-52.35616
Jul	31	1436.70137	-30.70137
Aug	31	1436.70137	9.2986
Sep	30	1390.35616	-58.35616
Oct	31	1436.70137	-73.70137
Nov	30	1390.35616	19.648356
Dec	31	1436.70137	89.29863

$Q = \sum_{i=1}^{m} \frac{(x_i-m_i)^2}{m_i} = 75.4273$ where d.f = 12-1=11 and $Q \sim \chi^2(11)$. Note that here $p = 12$ (12 different incidence rates) and $q = 1$ (a common incidence rate), and so $p - q = 11$. Therefore, $P(Q \geq 75.4273) = 1.122 * 10^{-11}$, since the p-value is extremely small, we will reject the null hypothesis. We can see a seasoned pattern in the death rate by looking at $x_i - m_i$, which shows that more deaths occur in the death winter.

Solution 7:

H_0: probability of breaking at each point p_i is equal: $p_1 = \ldots = p_5 = p$.

$x_i = P(\text{bar breaks in } i \text{ places}) = \binom{5}{i}p^i(1-p)^{5-i}, 0 \leq i \leq 5$. N_i=number of bar breaks at i.

$L(p) = \frac{280!}{N_0!N_1!N_2!N_3!N_4!N_5!} X_0^{N_0} X_1^{N_1} X_2^{N_2} X_3^{N_3} X_4^{N_4} X_5^{N_5}$

$log(L(p)) = log280! - \sum_{i=0}^{5} logN_i! + \sum_{i=0}^{5} N_i log X_i =$
$log280! - \sum_{i=0}^{5} logN_i! + \sum_{i=0}^{5} N_i(log\binom{5}{i} + ilogp + (5-i)log(1-p))$
$\frac{d}{dp}logL(p) = \sum_{i=0}^{5} N_i[\frac{i}{p} - \frac{5-i}{1-p}]$
$\frac{d^2}{dp^2}logL(p) = \sum_{i=0}^{5} N_i(-\frac{i}{p^2} - \frac{5-i}{(1-p)^2}) < 0$
Then given $\frac{d}{dp}logL(p) = \sum_{i=0}^{5} N_i[\frac{i}{p} - \frac{5-i}{1-p}] = 0$, the MLE is:

$$\hat{p} = \frac{\sum_{i=0}^{5} iN_i}{5 * 280} = 0.142$$

Then the table of the MLE of the expected counts is:

Breaks/Bar	Expected Frequency
0	130.2
1	107.7
2	35.7
3	5.9
4	0.49
5	0.02

We need to pool 3,4,5 to have at least 5 expected counts in each category and the table of the observed count with pooling is:

Breaks/Bar	Observed	Expected
0	157	130.2
1	69	107.7
2	35	35.7
3,4 or 5	19	6.4

$Q = \frac{(157-130.1)^2}{130.1} + \frac{(69-107.8)^2}{107.8} + \frac{(35-35.7)^2}{35.7} + \frac{(19-6.4)^2}{6.4} = 44.3$
$Q \sim \chi^2$ distribution with 3-1=2 d.f. Note that after the pooling there are a total of 4 categories and so the number of parameters in the unrestricted case is $p = 4 - 1 = 3$, while under the null, the number of free parameters is $q = 1$.
$P(Q \geq 44.3) = 1 - F(44.3)$ where F is the cdf of χ^2 with 2 d.f.

~~END~~