STA2002 - Homework 4

Xue Zhongkai 122090636

PROBLEM 1.

(a) Given $X \sim N(\mu_1, \frac{\sigma^2}{n})$, the type-II error is

$$\beta = P(\bar{x} > \mu_0 - \mathbf{z}_\alpha \frac{\sigma}{\sqrt{n}} \,|\, \mu = \mu_1) = P(\bar{x} - \mu_1 + \mu_1 - \mu_0 > -\mathbf{z}_\alpha \frac{\sigma}{\sqrt{n}}) = P(\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} > -\mathbf{z}_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}})$$

By definition of the normal cdf, it is equivalent to

$$\beta = 1 - \Phi(-\mathbf{z}_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}) = \Phi(\mathbf{z}_\alpha + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}})$$

With the symmetry of the bell-like cdf,

$$\mathbf{z}_\beta = 2 \cdot \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} - (\mathbf{z}_\alpha + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}) = -\mathbf{z}_\alpha + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}$$

That is equivalent to,

$$n = (\mathbf{z}_\alpha + \mathbf{z}_\beta)^2 \frac{\sigma^2}{(\mu_1 - \mu_0)^2}$$

(b)With the requirements, while $\mathbf{z}_{0.05} = -1.6449,\ \mathbf{z}_{0.025} = -1.9600$

$$n \geq (-1.6449 - 1.9600)^2 \times \frac{0.12^2}{(-0.02)^2} = 467.86 \approx 468$$

The least value for n should be 468.

PROBLEM 2.

Set the test statistic to be

$$\mathbf{z} = \frac{cX + dY - \sigma}{\sqrt{\frac{\sigma^2}{m} + \frac{\sigma^2}{n}}} \sim N(0, 1)$$

For two-sided test, the reject region is

$$|\mathbf{z}| > \mathbf{z}_{\alpha/2}$$

PROBLEM 3.

(a) Since X_i are independent to each other, the joint pdf of ith and jth order statistics is

$$f_{X_{(i)}, X_{(j)}}(x, y) = \frac{n![F(x)]^{i-1}}{(i-1)!(j-i-1)!(n-j)!} \cdot [F(y) - F(x)]^{j-i-1} \cdot [1 - F(y)]^{n-j} \cdot f(x) \cdot f(y)$$

Then joint pdf of $(X_{(1)}, X_{(n)})$ is

$$f_{X_{(1)},X_{(n)}}(u_1,u_2) = \frac{n!}{(n-2)!} \cdot [F(u_2) - F(u_1)]^{n-2} \cdot f(u_1) \cdot f(u_2),\ u_1 < u_2$$

Given the sample range stastistics $R = x_{(n)} - x_{(1)}$,

$$F(r) = P(R \leq r) = P(x_{(n)} - x_{(1)} \leq r) = \iint_{u_2 - u_1 \leq r} f_{X_{(1)},X_{(n)}}(u_1,u_2)\, du_1 du_2$$

To compute this double integral,

$$F(r) = \int_{-\infty}^{\infty} \int_{u_1}^{u_1+r} \frac{n!}{(n-2)!} \cdot [F(u_2) - F(u_1)]^{n-2} \cdot p(u_1) \cdot p(u_2)\, du_2 du_1$$

Set $y = F(u_2) - F(u_1)$, then $\partial y = f(u_2) \partial u_2$. Plug it in,

$$\int_{u_1}^{u_1+r} y^{n-2} \cdot p(u_1)\, dy = \frac{1}{n-1} [F(u_1 + r) - F(u_1)]^{n-1} p(u_1)$$

That is why

$$F(r) = \int_{-\infty}^{\infty} \frac{n!}{(n-2)!} \cdot \frac{1}{n-1} \cdot [F(u_1+r) - F(u_1)]^{n-1} p(u_1)\, du_1 = n \int_{-\infty}^{\infty} [F(u_1+r) - F(u_1)]^{n-1} p(u_1)\, du_1,\ r > 0$$

which is exactly the same as given.

(b) Set $v = x_{(1)}$, then $x_{(n)} = R + v$, we have pdf

$$f_{R,v}(r,v) = n(n-1) \cdot [F(r+v) - F(v)]^{n-2}$$

Given the fact it follows $U(0,1)$,

$$f_{R,v}(r,v) = n(n-1) \cdot [r + v - v]^{n-2} = n(n-1) \cdot r^{n-2}$$

As $r + v < 1$, further we have

$$f_R(r) = \int_0^{1-r} n(n-1) \cdot r^{n-2},\, dv = n(n-1) \cdot r^{n-2},\ 0 < r < 1$$

To make it clear,

$$f_R(r) = \frac{r^{(n-1)-1} \cdot (1-r)^{2-1}}{\frac{(n-2)! \cdot 1!}{n!}} = \frac{r^{(n-1)-1} \cdot (1-r)^{2-1}}{\frac{\Gamma(n-1) \cdot \Gamma(2)}{\Gamma(n+1)}} = \frac{r^{(n-1)-1} \cdot (1-r)^{2-1}}{\beta(n-1,2)}$$

Thus, it is true that

$$R \sim \mathbf{Beta}(n-1, 2)$$

The expectation for the Beta distribution is

$$\mathbf{E}(R) = \frac{\alpha}{\alpha + \beta} = \frac{n-1}{n+1}$$

Problem 4.

The mean for the word "also" of the assumed Poisson distribution in author's 200 passages is

$$\lambda = \frac{0 \times 22 + 1 \times 53 + 2 \times 58 + 3 \times 39 + 4 \times 20 + 5 \times 5 + 6 \times 2 + 7 \times 1}{200} = \frac{41}{20} = 2.05$$

From the Poisson distribution with parametre 2.05, we may compute p_i, the hypothesized probability associated with ith interval:

$$p_1 = P(X = 0) = \frac{e^{-2.05}(2.05)^0}{0!} = 0.1287$$
$$p_2 = P(X = 1) = \frac{e^{-2.05}(2.05)^1}{1!} = 0.2639$$
$$p_3 = P(X = 2) = \frac{e^{-2.05}(2.05)^2}{2!} = 0.2705$$
$$p_4 = P(X = 3) = \frac{e^{-2.05}(2.05)^3}{3!} = 0.1848$$
$$p_5 = P(X = 4) = \frac{e^{-2.05}(2.05)^4}{4!} = 0.0947$$
$$p_6 = P(X \geq 5) = 1 - \sum_{i=1}^{6} p_i = 0.0574$$

The last two are combined for too small E_i. By calculating $E_i = np_i$, the expected frequencies follow:

Number of "also"	Observed frequeny	Expected frequency
0	22	25.74
1	53	52.78
2	58	54.10
3	39	36.96
4	20	18.94
≥5	8	11.48

Set the null hypothesis H_0: The distribution is Poisson, with H_1: The distribution is NOT Poisson.

Assume the test stastic to be

$$\chi^2 = \sum_{i=0}^{5} \frac{(O_i - E_i)^2}{E_i}$$

With stastics listed above,

$$\chi^2 = \frac{(22-25.74)^2}{25.74} + \frac{(53-52.78)^2}{52.78} + \frac{(58-54.10)^2}{54.10} + \frac{(39-36.96)^2}{36.96} + \frac{(20-18.94)^2}{18.94} + \frac{(8-11.48)^2}{11.48} = 2.0523$$

With rejection region

$$\chi^2 < \chi^2_{0.05}(6-1) = 11.070,$$

we fail to reject H_0.

As a result, it is reasonable to assume it to be a Poisson distribution.

Problem 5.

(a) Use 30 scores as a sample, we have the sample mean and the sample variable

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = 64.9$$

$$\hat{\sigma^2} = \mathbf{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) = 230.56$$

Set the null hypothesis H_0: The score is adequately described by a normal distribution.

For $X \sim N(64.9, 230.56)$ with $k = 5$ cells,

Class interval	Observed frequeny	Expected frequency
$-\infty < x < 52.12$	8	6
$52.12 \leq x < 61.05$	6	6
$61.05 \leq x < 68.75$	4	6
$68.75 \leq x < 77.70$	4	6
$77.70 \leq x < \infty$	8	6

Check the test stastic to be

$$\chi^2 = \sum_{i=0}^{5} \frac{(O_i - E_i)^2}{E_i} = 2.667$$

Since two parametres have been estimated,

$$\chi^2_{0.05}(5-2-1) = \chi^2_{0.05}(2) = 5.991$$

With the fact that

$$\chi^2 < \chi^2_{0.05}(2),$$

we fail to reject H_0.

As a result, we can reasonably assume it is a normal distribution.

(b) Set the null hypothesis H_0: Class I and Class II are independent, *i.e.* $p_{i,j} = p_{i,\cdot} \times p_{\cdot,j}$

Under H_0,estimate the probabilities like

$$\hat{p_{1\cdot}} = \frac{8+5+2}{30} = 0.5 \qquad \hat{p_{2\cdot}} = \frac{4+7+4}{30} = 0.5$$

and

$$\hat{p_{\cdot 1}} = \frac{8+4}{30} = 0.4 \qquad \hat{p_{\cdot 2}} = \frac{5+7}{30} = 0.4 \qquad \hat{p_{\cdot 3}} = \frac{2+4}{30} = 0.2$$

we have the test statistic

$$\sum_{i=1}^{2}\sum_{j=1}^{3} \frac{(y_{i,j} - n\hat{p_{i,\cdot}}\hat{p_{\cdot,j}})^2}{n\hat{p_{i,\cdot}}\hat{p_{\cdot,j}}} = 2.33$$

Under significance level $\alpha = 0.05$,

$$\chi^2_{0.05}(2 \times 1) = \chi^2_{0.05}(2) = 5.991$$

With the fact that

$$\chi^2 < \chi^2_{0.05}(2),$$

fail to reject H_0. That is, they are independent.

As a result, we can reasonably assume they are independent.

Problem 6.

Set H_0: Working life of these four batches have no significant differences, *i.e.* $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu$.

By analysing the illustrative data, we find it is a single-factor experiment with 4 treatments of the factor (A, B, C, D), each has 7, 5, 8, 6 observations respectively. That is,

$$m = 4 \qquad n = n_1 + n_2 + n_3 + n_4 = 7 + 5 + 8 + 6 = 26$$

With a simple numerical process, we have

Treatments	$\sum_{j=1}^{7} x_{ij}$	$\sum_{j=1}^{7} x_{ij}^2$
Treatment A	11,760	19,785,400
Treatment B	8,310	13,828,100
Treatment C	13,290	22,191,700
Treatment D	9,410	14,778,700
Totals	42,770	70,583,900

Further we have

$$SS(TO) = 70,583,900 - \frac{42770^2}{26} = 227,250$$
$$SS(T) = \frac{1}{7} \times 11760^2 + \frac{1}{5} \times 8310^2 + \frac{1}{8} \times 13290^2 + \frac{1}{6} \times 9410^2 - \frac{42770^2}{26} = 47,399.167$$
$$SS(E) = 227,250 - 47,399.167 = 179,850.833$$

Given the significance level $\alpha = 0.05$,

$$F = \frac{SS(T)/(m-1)}{SS(E)/(n-m)} = \frac{47,399.167/(4-1)}{179,850.833/(26-4)} = 1.93 < 3.0491 = F_{0.05}(3, 22)$$

Thus we fail to reject H_0.

As a result, we can reasonably assume these four batches have NO significant differences.

******************************** **End of Homework 4** ********************************