ID: 0856733
Name: Ming-Han Huang

109-1 Statistics
National Chiao Tung university

Homework #2

Due by Sunday 10/25 23:55

Question #1.A

Please derive the score equations of (γ, β).

Solution:

The Weibull distribution is defined as:

$$f(x; \gamma, \beta) = \frac{\gamma x^{\gamma-1}}{2\beta^2} e^{-x^\gamma / 2\beta^\gamma} \tag{1}$$

And the log-likelihood function is:

$$l(\theta) = logL(\theta) = \sum_{i=1}^{n} log f_\theta(x_i) \tag{2}$$

Therefore, the log-likelihood function of the Weibull distribution is:

$$l(\gamma, \beta) = nlog(\gamma) - nlog(2\beta^\gamma) + (\gamma - 1)\sum_{i=1}^{n} log(x_i) - \frac{\sum_{i=1}^{n}}{2\beta^\gamma} \tag{3}$$

The score equation of (γ, β) is:

$$\frac{\partial l(\gamma, \beta)}{\partial \beta} = \frac{\gamma \sum_{i=1}^{n} x_i^\gamma}{2\beta^{\gamma+1}} - \frac{n\gamma}{\beta} = 0 \tag{4}$$

$$\frac{\partial l(\gamma, \beta)}{\partial \gamma} = \frac{n}{\gamma} + \frac{log(\beta)\sum_{i-1}^{n} x_i^\gamma - \sum_{i-1}^{n} x_i^\gamma log(x_i)}{2\beta^\gamma} - nlog(\beta) + \sum_{i=1}^{n} log(x_i) = 0 \tag{5}$$

Question #1.B

Then choose a test statistics that you learn in the class to test $H_0 : \gamma = 2$. (There are three options) Please state the decision rule based on the test.

Solution:

For convenience, I will choose to use LR Statistic since I've already calculate the MLE and using LR Statistic do not need to calculate the inverse matrix.

Question #1.C

What are the fitted MLE values of (γ, β).

Solution:

From section A, we have

$$\frac{\partial l(\gamma, \beta)}{\partial \beta} = \frac{\gamma \sum_{i=1}^{n} x_i^\gamma}{2\beta^{\gamma+1}} - \frac{n\gamma}{\beta} = 0 \tag{6}$$

And
$$\frac{\partial l(\gamma,\beta)}{\partial\gamma}=\frac{n}{\gamma}+\frac{log(\beta)\sum_{i-1}^{n}x_i^{\gamma}-\sum_{i-1}^{n}x_i^{\gamma}log(x_i)}{2\beta^{\gamma}}-nlog(\beta)+\sum_{i=1}^{n}log(x_i)=0 \tag{7}$$
To find the fitted value of (γ,β), we have to solve the above two equation. Then we have:
$$(\gamma_{MLE},\beta_{MLE})=(2.45,8.692) \tag{8}$$

Question #1.D

Do the data support $H_0:\gamma=2$?

Solution:

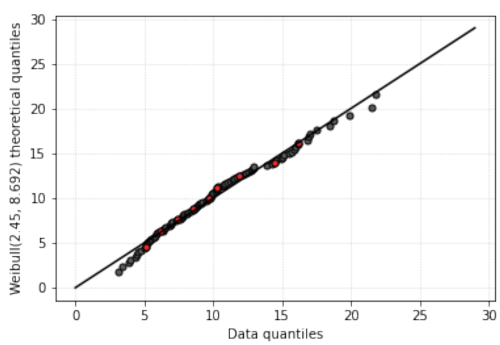
By using LR Statistics, we have $-2log(\frac{L(\tilde{\theta_0})}{L(\hat{\theta})})=-2log(\frac{L(2,7.873)}{L(2.45,8.692)})=9.1468$
Then, by chi-square, we can have $p=0.00249<0.05$.
Therefore, we reject the null hypothesis.

Question #1.E

Please draw a Q-Q plot (X-axis: the data quantiles; Y-axis: the quantiles of the Raleigh distribution).

Solution:

The following is the figure of the Q-Q plot.



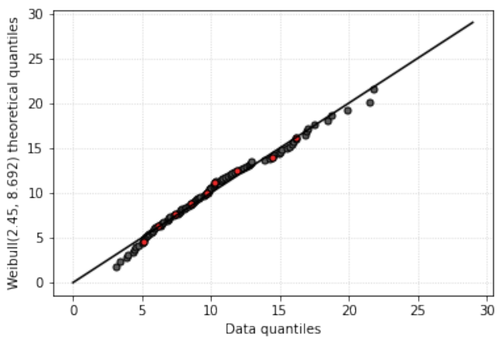
Since the points are located along 45 degree line, the model assumption is plausible.□

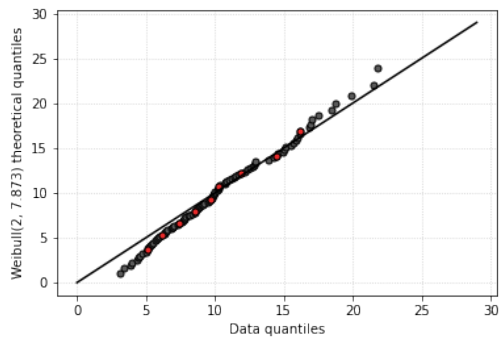
Question #1.F

Please comment on the results of (d) and (e).

Solution:

From (d), we reject the null hypothesis of $\gamma=2$. And from the Q-Q plot in (e) and the Q-Q plot of $\gamma=2$ as following, we can see that the Q-Q plot of $\gamma=2.45$ is more fitted to the 45 degree line comparing to the Q-Q plot of $\gamma=2$.


Weibull(2.45, 8.692) theoretical quantiles
Data quantiles
0
5
10
15
20
25
30


Weibull(2, 7.873) theoretical quantiles
Data quantiles
0
5
10
15
20
25
30

Species	LowAlt	HighAlt	HighAlt-LowAlt	Sign	Sign Rank	
1	5	19	14	+	+9.5	
2	4	10	6	+	+4	
3	12	4	-8	-		-5.5
4	7	10	3	+	+2	
5	17	17	0			
6	4	12	8	+	+5.5	
7	3	17	14	+	+9.5	
8	4	14	10	+	+7	
9	6	3	-3	-		-2
10	15	3	-12	-		-8
11	9	9	0			
12	7	10	3	+	+2	
					+39.5	-15.5

Question #2.A

Perform a sign test based on the paired data (LowAlt,HighAlt). Compute the p-value. Note: the hypothesis is one-sided

Solution:

H_0: HighAlt $\leq$ LowAlt, H_A: HighAlt $>$ LowAlt.
From the figure above, if H_A is true, we expect very few '-' sign. The test statistic is therefore N, the number of '-' sign.
Therefore, p-value $= P[N \leq 3|p = 0.5] = 0.1718 > 0.05$. Since the p value is larger than 0.05, we cannot reject the null hypothesis. That is to say, the tropical plants of a certain genus **do not** tend to produce more flowers at higher altitudes than at lower altitudes.

Question #2.B

Perform a (one-sided) signed-rank test based on the normal approximation. Compute the p-value.

Solution:

H_0: HighAlt $\leq$ LowAlt, H_A: HighAlt $>$ LowAlt.
From the figure above, T = min{39.5,15.5} = 15.5, E(T) = n(n+1)/4 = 27.5, Var(T) = n(n+1)(2n+1)/24 = 96.25.
Then $z = \frac{|T-E(T)|-0.5}{\sqrt{Var(T)}} = 1.172 < 1.65(\alpha = 0.05)$
Therefore, we cannot reject the null hypothesis.

Question #2.C

Perform a signed rank test using a software (based on the exact distribution)

Solution:

```python
from scipy.stats import wilcoxon
y1 = [5,4,12,7,17,4,3,4,6,15,9,7]
y2 = [19,10,4,10,17,12,17,14,3,3,9,10]
wilcoxon(y1,y2,mode='exact')
# WilcoxonResult(statistic=15.5, pvalue=0.21946693040984377)
```

By using python with *scipy.stats.wilcoxon*, we have $p - value = 0.21947 > 0.05$. Therefore, we cannot reject the null hypothesis that $H_0 : HighAlt \leq LowAlt$.

Question #2.D

Even though the data is from 24 different plants, why would it be incorrect to assume that the plants from the lowlands and the plants from higher altitude form two independent samples?

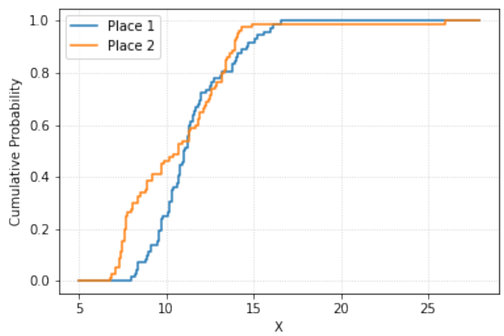
Solution:

The data for two sample needs to be continuous data. Since the number of the flower is a discrete data (there would not be 10.25 flowers), it will be incorrect to assume they form two independent samples.

Question #3.A

Please plot the two empirical distribution functions and compare them.

Solution:



Question #3.B

Please perform a two-sample t test for testing $H_0 : \mu_1 = \mu_2$ under 5% level of significance and state your conclusion. Comment on whether it is appropriate to use the t test.

Solution:

$$t = \frac{differenceofgroupaverages}{standarderrorofdifference} = \frac{11.42 - 10.65}{2.6056 \times \sqrt{1/72 + 1/80}} = 1.82 < 1.976 \tag{9}$$

The t-value with $\alpha = 0.05$ and 150 degrees of freedom is 1.976.
Comparing the value of our statistic (1.82) to the t value. Since $1.82 < 1.976$, we cannot reject the null hypothesis that the means of two place are the same.

Question #3.C

Please perform a Wilcoxon rank sum test for testing $H_0 : F_1 = F_2$ under 5% level of significance. State your conclusion.

Solution:

To answer the question, we pool the two samples, order the observations from the smallest to the largest, retaining their group identity, and rank them from 1 to 152 as the following table.
Since the sample of place 1 (m=72) is smaller than that of place 2 (n=80), the test static W_m is the sum of the ranks associated with the place 1.
Therefore, $W_m = 22.5 + 26 + 27 + 29.5 + ... + 151 = 6055.5$
And, $E(W) = \frac{72(72+80+1)}{2} = 5508$,
$Var(W) = \frac{72*80(72+80+1)}{12} \rightarrow SD = 270.998$
$Z = \frac{6055.5-5508}{270.998} = 2.02$
The z-score with $\alpha = 0.05$ is $1.645 < 2.02$. Thus we reject $H_0 : F_1 = F_2$.

Observation	6.8	6.9	7.1	7.1	7.3	7.3	7.3	7.4	7.4	7.5	7.5	7.5	7.6	7.6	7.6
Place	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Rank	1	2	3.5	3.5	6	6	6	8.5	8.5	11	11	11	14.5	14.5	14.5

Observation	7.6	7.7	7.7	7.7	7.7	7.8	8	8	8.1	8.1	8.2	8.3	8.4	8.4	8.4
Place	2	2	2	2	2	2	1	2	2	2	1	1	1	1	2
Rank	14.5	18.5	18.5	18.5	18.5	21	22.5	22.5	24.5	24.5	26	27	29.5	29.5	29.5

Observation	8.4	8.5	8.8	8.8	8.9	8.9	8.9	8.9	9	9.1	9.1	9.2	9.2	9.5	9.6
Place	2	2	1	2	1	2	2	2	1	1	1	2	2	1	1
Rank	29.5	32	33.5	33.5	36.5	36.5	36.5	36.5	39	40.5	40.5	42.5	42.5	44	46

Observation	9.6	9.6	9.7	9.7	9.7	9.7	9.7	9.7	9.8	9.9	10.1	10.2	10.2	10.2	10.2
Place	1	1	1	1	1	2	2	2	1	2	1	1	1	1	2
Rank	46	46	50.5	50.5	50.5	50.5	50.5	50.5	54	55	56	58.5	58.5	58.5	58.5

Observation	10.3	10.3	10.3	10.4	10.4	10.6	10.7	10.7	10.7	10.7	10.7	10.8	10.8	10.8	10.9
Place	1	1	1	1	2	1	1	1	2	2	2	1	1	1	1
Rank	62	62	62	64.5	64.5	66	69	69	69	69	69	73	73	73	75.5

Observation	10.9	11	11	11	11.1	11.2	11.2	11.2	11.3	11.3	11.3	11.3	11.3	11.3	11.4
Place	2	1	1	1	1	1	1	1	1	1	1	2	2	2	1
Rank	75.5	78	78	78	80	82	82	82	86.5	86.5	86.5	86.5	86.5	86.5	90.5

Observation	11.4	11.5	11.5	11.6	11.7	11.7	11.8	11.8	11.8	11.9	11.9	11.9	12	12	12.1
Place	2	1	1	1	1	2	1	2	2	1	2	2	1	1	2
Rank	90.5	92.5	92.5	94	95.5	95.5	98	98	98	101	101	101	103.5	103.5	105

Observation	12.2	12.3	12.3	12.4	12.5	12.5	12.6	12.6	12.6	12.7	12.8	12.9	13.1	13.2	13.2
Place	2	1	2	2	1	2	1	2	2	1	2	2	1	1	2
Rank	106	107.5	107.5	109	110.5	110.5	113	113	113	115	116	117	118	120.5	120.5

Observation	13.2	13.2	13.3	13.4	13.4	13.4	13.5	13.6	13.8	13.8	13.8	13.9	13.9	13.9	13.9
Place	2	2	2	2	2	2	2	2	1	1	2	1	2	2	2
Rank	120.5	120.5	123	125	125	125	127	128	130	130	130	133.5	133.5	133.5	133.5

Observation	14	14	14.1	14.1	14.2	14.3	14.3	14.6	14.7	14.9	15.1	15.2	15.6	16	16.1
Place	1	2	1	2	2	1	2	1	1	2	1	1	1	1	1
Rank	136.5	136.5	138.5	138.5	140	141.5	141.5	143	144	145	146	147	148	149	150

Observation	16.6	26
Place	1	2
Rank	151	152