NAME: SUCHETA JHUNJHUNWALA

ROLL NO.: MDS202151

PaSWR FINAL EXAM

1) At first we are defining a function ~~for~~ called 'fun' which takes four arguments as parameters x, mu (default value = 0), sigma (default value = 1) and alpha (default value = 0.95).

Inside the function we calculate the value of z s.t.

$P(Z > z) = \frac{1-\text{alpha}}{2}$ by defining the variable

z = qnorm ((1-alpha)/2, lower.tail = FALSE).

qnorm is the quantile function for normal distribution which gives x s.t. $P(X > x) = p$ where p is the parameter given to qnorm. lower.tail = FALSE enables $P(X > x)$ instead of $P(X \le x)$.

sdx = sigma / sqrt (length(x)).

This variable calculates the value of $\frac{\text{sigma}}{\sqrt{\text{length}(x)}}$ where length(x) is the length of the vector x given to fun.

Then, N = pnorm ((mean(x) - mu) / sdx, lower.tail = FALSE) calculate the probability of a normal variable $X \sim \text{Normal}(0,1)$.

$P(X > (\text{mean}(x) - \text{mu}) / \text{sd}x)$ where mean(x) is the mean of the vector x given to fun. lower.tail = FALSE is to calculate $P(X > x)$ instead of $P(X \le x)$.

At last, the function fun returns a vector of length 3 which has three values. The first value is the lower bound of the 95% confidence interval for the expectation of X, second value is the upper bound for the 95% confidence interval and N returns the probability.

Then we apply fun on the given vector x.

Note that, we take y= fun (x, 76, 1.5) so our mu = 76 and sigma = 1.5, though alpha is still 0.95.

y is a vector which has the lower and upper bound of the 95% confidence interval for expectation corresponding the scores given to it.

z is the value s.t. $P(Z > z) = \frac{1-alpha}{2} = 0.025$

where $Z \sim Normal(0,1)$.

$sdx = \frac{sigma}{\sqrt{length(x)}} = \frac{1.5}{\sqrt{9}}$ since x has 9 elements.

N gives the probability that given $Z \sim Normal(0,1)$,

$$P\left(Z > \frac{mean(x) - 76}{1.5/\sqrt{9}}\right)$$

At last we return, $\left(mean(x) - z * \frac{1.5}{\sqrt{9}}, mean(x) + z * \frac{1.5}{\sqrt{9}}, N\right)$

which has values (73.9039069, 75.8688709, 0.9868659)

so the 95% confidence interval for the expected values of scores when the scores (say X) follow Normal distribution s.t. $X \sim Normal(76, (1.5)^2)$ is

(73.9089069, 75.8688709)

N returns $P\left(Z > \dfrac{\text{mean}(x) - 76}{1.5/\sqrt{9}} \right)$

that 0.9868659 is the probability of getting a score more than $\dfrac{\text{mean}(x) - 76}{1.5/\sqrt{9}}$.

2) $X_1, \ldots X_n$ iid $X \sim \text{Rayleigh}(\alpha)$

$$f(x) = \begin{cases} \alpha x \exp\left(-\frac{1}{2}\alpha x^2\right) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The parameter here is α. the maximum likelihood function is given by,

$$L(\alpha \mid X_1, X_2, \ldots X_n) = \prod_{i=1}^{n} \alpha X_i e^{-\frac{1}{2}\alpha X_i^2} \quad \text{when } X_i \geq 0$$
$$\forall i = 1, 2, \ldots n$$

Also note that, if $X_i < 0$ then,

$$L(\alpha \mid X_1, \ldots X_n) = 0.$$

Now, $L(\alpha \mid X_1, \ldots X_n) = \alpha^n \prod_{i=1}^{n} X_i e^{-\frac{1}{2}\alpha X_i^2}$

$$= \alpha^n \prod_{i=1}^{n} X_i \cdot \prod_{i=1}^{n} e^{-\frac{1}{2}\alpha X_i^2}$$

$$= \alpha^n \left(\prod_{i=1}^{n} X_i\right) e^{-\frac{1}{2}\alpha \sum_{i=1}^{n} X_i^2}$$

On taking logarithm on both sides we get,

$$\log L(\alpha) = n \log \alpha + \log\left(\prod_{i=1}^{n} X_i\right) + \left(-\frac{1}{2}\alpha \sum_{i=1}^{n} X_i^2\right) \log e$$

$$= n \log \alpha + \sum_{i=1}^{n} (\log X_i) - \frac{1}{2}\alpha \sum_{i=1}^{n} X_i^2$$

To find the maximum likelihood estimate we need to maximise $L(\alpha)$ that is we need to maximize $\log L(\alpha)$.

In order to do so, we need to find its critical points.

On taking derivative w.r.t. α on both sides we get,

$$(\log L)'(\alpha) = \frac{n}{\alpha} + 0 - \frac{1}{2} \cdot 1 \sum_{i=1}^{n} x_i^2 .$$

(Note that $\log L$ is a function of $\alpha, x_1, \dots x_n$ so here by derivative we mean partial derivative w.r.t. α)

Now any pt. $\hat{\alpha}$ is a critical point if $(\log L)'(\hat{\alpha}) = 0$

$$\Rightarrow \frac{n}{\hat{\alpha}} - \frac{1}{2} \sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow \frac{n}{\hat{\alpha}} = \frac{1}{2} \sum_{i=1}^{n} x_i^2 \Rightarrow \hat{\alpha} = \frac{2n}{\sum_{i=1}^{n} x_i^2} .$$

We shall perform the second derivative test to check if $\hat{\alpha}$ is a point of maximum or not.

$$(\log L)''(\alpha) = -\frac{n}{\alpha^2} + 0 .$$

$$(\log L)''(\hat{\alpha}) = \frac{-n}{\left(\frac{2n}{\sum_{i=1}^{n} x_i^2} \right)^2} = -n \cdot \left(\frac{\sum_{i=1}^{n} x_i^2}{2n} \right)^2 .$$

$$\left(\frac{\sum x_i^2}{2n} \right)^2 \geq 0 \quad \forall n \text{ and } n > 0$$

$\therefore \log D''(\hat{\alpha}) < 0$.

$\Rightarrow \hat{\alpha}$ is the point of maximum for L.

$\Rightarrow$ Maximum likelihood estimate of α is $\dfrac{2n}{\sum_{i=1}^{n} x_i^2}$

4) Brown wrinkled seeds $\rightarrow 219$

Brown round seeds $\rightarrow 81$

White wrinkled seeds $\rightarrow 69$.

White round seeds $\rightarrow 31$.

Ratios should be $9:3:3:1$.

i.e. $\left(\frac{9}{16} \times 100\right)\%$ should be brown wrinkled, $\left(\frac{3}{16} \times 100\right)\%$ should be brown round, $\left(\frac{3}{16} \times 100\right)\%$ should be white wrinkled and $\left(\frac{1}{16} \times 100\right)\%$ should be white round.

The total number of samples is $n = 400$.

This model is for ~~checking~~ is the χ^2-test for goodness of fit.

which is given by $\chi^2 = \sum_{j=1}^{k} \frac{(y_j - np_j)^2}{np_j}$

where y_j is the number of data points and p_j is $P(X = c_j)$. $y_j = \# \{ i : x_i = c_j \}$.

> x = c(219, 81, 69, 31)

Here x represents our Y_i which is the number of data points corresponding to the range of X i.e.

{brown wrinkled, brown round, white wrinkled, white round}

n is the number of samples taken so

> n = 400.

This is χ^2 distribution with (k-1) degrees of freedom. Here since the range of X has 4 ~~xxx~~ elements, k = 4 so we get,

> T = pchisq (X square, df = 4 - 1, lower.tail = FALSE)

∴ the final code becomes,

```
> x = c(219, 81, 69, 31)
+ n = 400
+ prob = c(9/16, 3/16, 3/16, 1/16)
+ Xsquare = sum(((x - n * prob)^2)/n * prob))
> Xsquare
> T = pchisq (Xsquare, df = 4 - 1, lower.tail = FALSE)
> T
```

As output we get two values, Xsquare and T.

Xsquare is the value of the test statistic and it comes out as 2.733333.

whereas T is the value of $P(W \geq \cancel{X}^2)$ where

$W \sim \chi^2(4-1) = \chi^2(3)$ and χ^2 is the test statistic.

$$\chi^2 = \sum_{i=1}^{k} \frac{(y_i - np_i)^2}{np_i}$$

This probability comes out as 0.2549554

i.e. such type of samples are observed for more than approximately 25.49 % times.

The null hypothesis is ~~that~~ the theory of predicting such seeds in the ratio 9 : 3 : 3 : 1.

The level of significance is 0.05.

0.2549554 > 0.05

∴ We can't reject the null hypothesis. Since the value is much larger than the significance level, we can say that there is no evidence to reject the null hypothesis.