MTH5129 Probability & Statistics II
Coursework 7

1. Let X be the number of pounds of butterfat produced by a Holstein cow during the 305 day milking period following the birth of a calf. A random sample of 25 such cows gave the following values of X. Assume the distribution of X is $N(\mu, \sigma^2)$

425	710	661	664	732	714	934	761	744
653	725	657	421	573	535	602	537	405
874	791	721	849	567	568	975		

We wish to test the null hypothesis $H_0 : \sigma^2 = 140^2$ against the alternative hypothesis $H_1 : \sigma^2 > 140^2$.

a) Give the test statistic, its distribution if H_0 is true and a rejection region for a test with significance level $\alpha = 0.05$.

b) Calculate the value of the test statistic and state your conclusion.

Solution:

a) The test statistic is
$$W = \frac{(n-1)S^2}{140^2}.$$
$W \sim \chi^2_{n-1}$ if H_0 is true.

We have $n = 25$ and a one-sided test with $\alpha = 0.05$ so the rejection region is $\{w : w > 36.42\}$. This can be obtained in R as follows:

```
> qchisq(0.95,24)
[1] 36.41503
```

b) $\sum x_i = 16798$ and $\sum x_i^2 = 11828388$.

Thus
$$\begin{aligned}\sum(x_i - \bar{x})^2 &= \sum x_i^2 - \frac{(\sum x_i)^2}{n} \\ &= 11828388 - \frac{16798)^2}{25} \\ &= 541475.84 \\ &= (n-1)s^2.\end{aligned}$$

So the observed value of the test statistic is
$$\frac{541475.84}{140^2} = 27.63.$$

As $27.63 < 36.42$ we cannot reject H_0 at the 5% significance value.

2. Let $X_1, X_2, \ldots, X_{23}$ be a random sample from a normal distribution with variance $\sigma^2 = 100$. Let S^2 be the sample variance. Find the variance of S^2.

   Hint: Remember that if $W \sim \chi^2_\nu$ then $\mathrm{Var[W]} = 2\nu$

> **Solution:** We know $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
>
> Here $n = 23$ and $\sigma^2 = 100$ so
>
> $$\frac{22S^2}{100} \sim \chi^2_{22}.$$
>
> By the hint we know that
>
> $$\mathrm{Var}\left[\frac{22\mathrm{S}^2}{100}\right] = 44.$$
>
> Using the result that $\mathrm{Var[aX]} = \mathrm{a}^2\mathrm{Var[X]}$ for any random variable X, we see that
>
> $$\left(\frac{22}{100}\right)^2 \mathrm{Var[S^2]} = 44.$$
>
> Rearranging we have that
>
> $$\mathrm{Var[S^2]} = \frac{44 \times 100^2}{22^2} = \frac{10000}{11} = 909.09$$

3. Articles produced by a manufacturer are designed to have mean length 5cm and standard deviation 0.06cm. A sample of size 16 from a batch of production has $\bar{x} = 4.96$cm. and sample standard deviation $s = 0.09$cm. Assume the lengths are normally distributed.

   a) Test the hypothesis that the mean length is 5cm against an alternative that it is less than 5cm with significance level $\alpha = 0.025$.
   b) Find the P-value for the test. What is the conclusion?
   c) Find a 95% confidence interval for the population mean.
   d) Test the hypothesis that the population variance is 0.06^2 against a two sided alternative with $\alpha = 0.05$.
   e) Find the P-value for the test. What is the conclusion?
   f) Find a 99% confidence interval for the population variance.

> **Solution:** We have $n = 16$, $\bar{x} = 4.96$ and $s = 0.09$.
>
> a) We test $H_0 : \mu = 5$ against $H_1 : \mu < 5$, so $\mu_0 = 5$.
>
>   The test statistic is $T = \frac{(\bar{X}-\mu_0)\sqrt{n}}{S}$.

$T \sim t_{15}$ if H_0 is true.

The observed value of $t = \frac{(4.96-5)\sqrt{16}}{0.09} = -1.778$.

The rejection region for a 2.5% significance level is $\{t : t < -2.131\}$. This can be obtained in R as follows:

```
> qt(0.025,15)
[1] -2.13145
```

Thus we cannot reject H_0 at the 2.5% significance level.

b) The P value is $P(T < -1.778)$ where $T \sim t_{15}$. This can be obtained in R as follows:

```
> pt(-1.778,15)
[1] 0.04783834
```

The P value shows there is moderate evidence against the null hypothesis. So we have some reason to doubt that the articles have the correct mean value.

c) The form of the confidence interval is

$$\bar{x} \pm t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}.$$

We find from R that $t_{15}(0.025) = 2.131$ as follows

```
> qt(0.975,15)
[1] 2.13145
```

so the confidence interval is

$$4.96 \pm 2.131 \times \frac{0.09}{4} = 4.96 \pm 0.048 = (4.932, 5.008)$$

d) We test $H_0 : \sigma^2 = 0.06^2$ against $H_1 : \sigma^2 \neq 0.06^2$.

The test statistic is $W = \frac{(n-1)S^2}{\sigma_0^2}$.

$W \sim \chi^2_{15}$ if H_0 is true.

The observed value of $w = \frac{15\times 0.09^2}{0.06^2} = 33.75$.

The rejection region for a 5% significance level is $w : w < 6.262 \cup w > 27.49\}$. This can be obtained in R as follows:

```
> qchisq(0.025,15)
[1] 6.262138
> qchisq(0.975,15)
[1] 27.48839
```

Thus we reject H_0 at the 5% significance level.

e) The P value is $2 \times \min\{P(W > 33.75), P(W < 33.75\}$. From R,

```
> 2*(1-pchisq(33.75,15))
[1] 0.007382951
```

Hence we conclude there is strong evidence against H_0 and it is likely that the articles do not have the correct standard deviation.

f) The form of the confidence interval is

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)s^2}{\chi^2_{n-1}(1-\alpha/2)} \right).$$

From R we find $\chi^2_{15}(.995) = 4.601$ and $\chi^2_{15}(.005) = 32.80$ as follows

```
> qchisq(0.005,15)
[1] 4.600916
> qchisq(0.995,15)
[1] 32.80132
```

So the 99% confidence interval for σ^2 is

$$\left(\frac{15 \times 0.09^2}{32.80}, \frac{15 \times 0.09^2}{4.601} \right) = (0.0037, 0.0264).$$

4. Suppose $z_1, z_2, \ldots$ are the observed values of independent standard normal random variables so that $z_1, z_2, \ldots z_n$ would be the observed values of a random sample of size n from a standard normal distribution.

   a) If I calculated $v = \sum_{i=1}^{m} z_i^2$ what distribution would v be an observed value from?

   b) If I then calculated the value of

   $$u_1 = \frac{z_{m+1}}{\sqrt{v/m}}$$

   what distribution would u_1 be an observed value from?

   c) In (b) could I have calculated the value of

   $$u_2 = \frac{z_1}{\sqrt{v/m}}$$

   to achieve the same result?

   Explain your answer.

Solution:

a) A chi-squared distribution with m degrees of freedom.

b) A t distribution with m degrees of freedom

c) No. z_1 is used both in the numerator of u_2 and in calculating v which is in the denominator of u_2. This means the random variables Z_1 and V are not independent which is required in the definition of the t distribution.