STAT 512 hw 4

1. Let $X \sim \text{Normal}(\mu, 1)$.

   (a) Give an interval (L, U), where U and L are based on X, such that $P(L < \mu < U) = 0.99$.

   We can begin by noting that $P(-z_{\alpha/2} < (X - \mu)/1 < z_{\alpha/2}) = 1 - \alpha$ for any $\alpha \in (0, 1)$, since $(X - \mu)/1 \sim \text{Normal}(0, 1)$. The can be rearranged to $P(X - z_{\alpha/2} < \mu < X + z_{\alpha/2}) = 1 - \alpha$, so a 99% confidence interval for μ is the interval with endpoints

   $$X \pm z_{0.01/2} = X \pm 2.576.$$

   (b) Give an upper bound U based on X such that $P(\mu < U) = 0.99$.

   We can begin by noting that $P(-z_\alpha < (X - \mu)/1) = 1 - \alpha$ for any $\alpha \in (0, 1)$, since $(X - \mu)/1 \sim \text{Normal}(0, 1)$. The can be rearranged to $P(\mu < X + z_\alpha) = 1 - \alpha$, so the 99% upper confidence limit for μ is

   $$X + z_{0.01} = X + 2.326.$$

   (c) Give a lower bound L based on X such that $P(L < \mu) = 0.99$.

   We can begin by noting that $P((X - \mu)/1 < z_\alpha) = 1 - \alpha$ for any $\alpha \in (0, 1)$, since $(X - \mu)/1 \sim \text{Normal}(0, 1)$. The can be rearranged to $P(X - z_\alpha < \mu) = 1 - \alpha$, so the 99% upper confidence limit for μ is

   $$X - z_{0.01} = X - 2.326.$$

2. Let $X \sim \text{Normal}(0, \sigma^2)$.

   (a) Find the distribution of X^2/σ^2. *Hint: It is a pivot quantity.*

   We have $X^2/\sigma^2 = ((X - 0)/\sigma)^2$, which has the same distribution as Z^2, where $Z \sim \text{Normal}(0, 1)$, which is the χ^2_1 distribution. So $X^2/\sigma^2 \sim \chi^2_1$.

   (b) Give an interval (L, U), where U and L are based on X, such that $P(L < \sigma^2 < U) = 0.95$.

   We may write $P(\chi^2_{1,1-\alpha/2} < X^2/\sigma^2 < \chi^2_{1,\alpha/2}) = 1 - \alpha$ for any $\alpha \in (0, 1)$. We may rearrange this to get as $P(X^2/\chi^2_{1,\alpha/2} < \sigma^2 < X^2/\chi^2_{1,1-\alpha/2}) = 1 - \alpha$, so that

   $$(X^2/5.023, X^2/0.00098)$$

is a 95% confidence interval for σ^2.

(c) Give an upper bound U based on X such that $P(\sigma^2 < U) = 0.95$.

We may write $P(\chi^2_{1,1-\alpha} < X^2/\sigma^2) = 1 - \alpha$ for any $\alpha \in (0,1)$. We may rearrange this to get as $P(\sigma^2 < X^2/\chi^2_{1,1-\alpha}) = 1 - \alpha$, so that the upper 95% confidence limit for σ^2 is

$$X^2/\chi^2_{1,1-0.05} = X^2/0.0039.$$

(d) Give a lower bound L based on X such that $P(L < \sigma^2) = 0.95$.

We may write $P(X^2/\sigma^2 < \chi^2_{1,\alpha}) = 1 - \alpha$ for any $\alpha \in (0,1)$. We may rearrange this to get as $P(X^2/\chi^2_{1,\alpha} < \sigma^2) = 1 - \alpha$, so that the upper 95% confidence limit for σ^2 is

$$X^2/\chi^2_{1,0.05} = X^2/3.841.$$

3. The following data from [1] are carapace lengths (mm) of lobsters caught in some region:

$$78 \quad 66 \quad 65 \quad 63 \quad 60 \quad 60 \quad 58 \quad 56 \quad 52 \quad 50$$

Assume that the lengths of lobster carapaces in the region have a Normal distribution.

(a) Give a 95% confidence interval for the mean carapace length of lobsters in the region.

The following R code computes in the bounds of the interval.

```
x <- c(78,66,65,63,60,60,58,56,52,50)
n <- length(x)
S <- sd(x)
alpha <- 0.05
L <- mean(x) - qt(1 - alpha/2,n-1)*S/sqrt(n)
U <- mean(x) + qt(1 - alpha/2,n-1)*S/sqrt(n)
```

The interval is $(55.099, 66.501)$.

(b) Give a 95% confidence interval for the variance of carapace lengths of lobsters in the region.

The following R code computes in the bounds of the interval.

```
L <- (n-1)*S^2/qchisq(1-alpha/2,n-1)
U <- (n-1)*S^2/qchisq(alpha/2,n-1)
```

The interval is $(30.048, 211.673)$.

4. Load the PlantGrowth data set in R by executing the command data(PlantGrowth). Assume that the weights in all the groups come from a Normal distribution.

   (a) Give a 95% confidence interval for σ_2^2/σ_1^2, where σ_1^2 is the variance of weights in the ctrl group and σ_2^2 is the variance of the weights in the trt1 group.

   The following R code computes the lower and upper bounds of the 95% confidence interval.

```
data("PlantGrowth")

ctrl <- PlantGrowth$weight[PlantGrowth$group == "ctrl"]
trt1 <- PlantGrowth$weight[PlantGrowth$group == "trt1"]

n1 <- length(ctrl)
n2 <- length(trt1)

S1 <- sd(ctrl)
S2 <- sd(trt1)

alpha <- .05

L <- S2^2/S1^2 * qf(alpha/2, n1-1,n2-1)
U <- S2^2/S1^2 * qf(1-alpha/2, n1-1,n2-1)
```

   The interval is $(0.460, 7.459)$.

   (b) If $X_1, \ldots, X_{n_1} \stackrel{\text{ind}}{\sim} \text{Normal}(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_{n_2} \stackrel{\text{ind}}{\sim} \text{Normal}(\mu_2, \sigma_2^2)$ and if $\sigma_1^2 = \sigma_2^2$, then we have the pivot quantity result

$$\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - (\mu_1 - \mu_2)}{\sqrt{S^2_{\text{pooled}}(1/n_1 + 1/n_2)}} \sim t_{n_1+n_2-2},$$

   where

$$S^2_{\text{pooled}} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

   Assuming that the variances of the plant growths in the ctrl and the trt1 groups are equal, construct a 95% confidence interval for $\mu_1 - \mu_2$, where μ_1 is the mean of weights in the ctrl group and μ_2 is the mean of the weights in the trt1 group.

Because of the pivot quantity result, we may write

$$1-\alpha = P\left(t_{n_1+n_2-2,1-\alpha/2} < \frac{\bar{X}_{n_1}-\bar{Y}_{n_2}-(\mu_1-\mu_2)}{\sqrt{S^2_{\text{pooled}}(1/n_1+1/n_2)}} < t_{n_1+n_2-2,\alpha/2}\right)$$
$$= P\Big(\bar{X}_{n_1}-\bar{Y}_{n_2}-\sqrt{S^2_{\text{pooled}}(1/n_1+1/n_2)}\,t_{n_1+n_2-2,1-\alpha/2} < \mu_1-\mu_2$$
$$< \bar{X}_{n_1}-\bar{Y}_{n_2}+\sqrt{S^2_{\text{pooled}}(1/n_1+1/n_2)}\,t_{n_1+n_2-2,\alpha/2}\Big),$$

which gives for a $(1-\alpha)100\%$ confidence interval the bounds

$$\bar{X}_{n_1}-\bar{Y}_{n_2} \pm \sqrt{S^2_{\text{pooled}}(1/n_1+1/n_2)}\,t_{n_1+n_2-2,\alpha/2}.$$

The following R code computes this on the data.

```
Sp <- sqrt(  ((n1-1)*S1^2  + (n2-1)*S2^2 ) / (n1 + n2 - 2)  )

L <- mean(ctrl) - mean(trt1) - sqrt( Sp^2*(1/n1 + 1/n2) ) * qt(1-alpha/2,n1+n2-2)
U <- mean(ctrl) - mean(trt1) + sqrt( Sp^2*(1/n1 + 1/n2) ) * qt(1-alpha/2,n1+n2-2)
```

The interval is $(-0.283, 1.0253)$.

5. Let Z_1, Z_2, Z_3, and Z_4 be independent random variables having the $\text{Normal}(0,1)$ distribution and let $\bar{Z} = (Z_1+Z_2+Z_3+Z_4)/4$. Give the distributions of the following quantities.

(a) $\bar{Z}$

This has the $\text{Normal}(0,1/4)$ distribution.

(b) $Z_1^2+Z_2^2+Z_3^2$

This has the χ^2_3 distribution.

(c) $Z_1/\sqrt{(Z_2^2+Z_3^2)/2}$

This has the t_2 distribution.

(d) $2\bar{Z}[(1/3)\sum_{i=1}^4(Z_i-\bar{Z}_4)^2]^{-1/2}$

This has the t_3 distribution.

(e) $\sum_{i=1}^4(Z_i-\bar{Z}_4)^2$

This has the χ^2_3 distribution.

(f) $(Z_1^2 + Z_2^2)/(Z_3^2 + Z_4^2)$

This has the $F_{2,2}$ distribution.

6. Let Z and W be independent random variables such that $Z \sim \text{Normal}(0,1)$ and $W \sim \chi^2_\nu$, for some $\nu > 0$. Recall that the random variable

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t_\nu.$$

(a) Find the mean of a random variable having the t_ν distribution.

We have $\mathbb{E}\frac{Z}{\sqrt{W/\nu}} = \sqrt{\nu}\mathbb{E}X\mathbb{E}W^{-1/2} = 0$, where we have used that fact that Z and W are independent and $\mathbb{E}Z = 0$.

(b) Show that $\mathbb{E}W^{-1} = 1/(\nu - 2)$, assuming $\nu > 2$.

We have

$$\begin{aligned}
\mathbb{E}W^{-1} &= \int_0^\infty \frac{1}{w}\frac{1}{\Gamma(\nu/2)2^{\nu/2}} w^{\nu/2-1}e^{-w/2}dw \\
&= \frac{\Gamma(\nu/2-1)2^{\nu/2-1}}{\Gamma(\nu/2)2^{\nu/2}} \underbrace{\int_0^\infty \frac{1}{\Gamma(\nu/2-1)2^{\nu/2-1}} w^{(\nu/2-1)-1}e^{-w/2}dw}_{=1} \\
&= \frac{\Gamma(\nu/2-1)}{(\nu/2-1)\Gamma(\nu/2-1)2} \quad \text{(property of Gamma function)} \\
&= 1/(\nu-2).
\end{aligned}$$

(c) Find the variance of a random variable having the t_ν distribution.

Since $\mathbb{E}T = 0$, we have $\operatorname{Var} T = \mathbb{E}T^2 = \nu\mathbb{E}Z^2\mathbb{E}W^{-1} = \nu/(\nu-2)$, since $\mathbb{E}Z^2 = 1$.

7. Two javelin throwers will compete. Suppose the distances thrown by the first follow the $\text{Normal}(70, 25)$ distribution (units in meters) and those of the second follow the $\text{Normal}(68, 36)$ distribution. Give the probability that the second javelin thrower throws further than the first, assuming that their throws are independent.

Let T_1 and T_2 be the distances thrown by the two javelin throwers. Then we wish to find $P(T_1 < T_2)$. Finding this probability directly would require taking a double integral over the product of two Normal pdfs, which is very complicated. Instead, we may formulate the event $T_1 < T_2$ in terms of the difference between T_1 and T_2 by defining the new random variable $D = T_1 - T_2$, so that $P(T_1 < T_2) = P(T_1 - T_2 < 0) = P(D < 0)$. From the Lecture 2 notes, we find that the distribution of D, which is a linear combination of independent Normal random variables, is the $\text{Normal}(-2, 25)$ distribution. So we have

$$P(D < 0) = P((D - 2)/5 < -2/5) = P(Z < -2/5), \quad Z \sim \text{Normal}(0, 1),$$

and the answer is $\Phi(-2/5) =$ `pnorm(-2/5)` $= 0.3445783$.

References

[1] William B Jeffries, Chang Man Yang, and Harold K Voris. Diversity and distribution of the pedunculate barnacle octolasmis gray, 1825 epizoic on the scyllarid lobster, thenus orientalis (lund, 1793). *Crustaceana*, 46(3):300–308, 1984.