MATH20802: STATISTICAL METHODS
SEMESTER 2
SOLUTIONS TO PROBLEM SHEET 6

1. The hypotheses are: $H_0 : \mu_X \leq \mu_Y$ and $H_1 : \mu_X > \mu_Y$. The p-value is:

$$\begin{aligned}
p\text{-value} &= \Pr\left(Z \geq \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}}\right) \\
&= \Pr\left(Z \geq \frac{137.667-127.778}{\sqrt{\frac{10^2}{9}+\frac{\sigma_Y^2}{9}}}\right) \\
&= \Pr\left(Z \geq \frac{3 \times 9.889}{\sqrt{100+\sigma_Y^2}}\right).
\end{aligned}$$

This reduces to $\Pr(Z \geq 2.653) = 0.004$ if $\sigma_Y = 5$, $\Pr(Z \geq 2.098) = 0.018$ if $\sigma_Y = 10$ and $\Pr(Z \geq 1.327) = 0.092$ if $\sigma_Y = 20$.

2. Let μ_X = mean of values relating to message 1 and let μ_Y = mean of values relating to message 2. The hypotheses are: $H_0 : \mu_X \leq \mu_Y$ and $H_1 : \mu_X > \mu_Y$. The test statistic is:

$$\begin{aligned}
\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_X^2}{m}+\frac{\sigma_Y^2}{n}}} &= \frac{5.6-4.1}{\sqrt{\frac{2^2}{9}+\frac{2^2}{9}}} \\
&= \frac{3 \times 1.5}{2\sqrt{2}} \\
&= 1.591
\end{aligned}$$

and the table value $z_{0.01} = 2.326$. So there is no evidence against the claim.

3. Let μ_X = mean IQ of urban students and let μ_Y = mean IQ of rural students. The hypotheses are: $H_0 : \mu_X = \mu_Y$ and $H_1 : \mu_X \neq \mu_Y$. In the case of equal variances, the test statistic is:

$$\begin{aligned}
\frac{\mid \bar{x}-\bar{y} \mid}{\sqrt{s_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}} &= \frac{\mid 102.2-105.3 \mid}{\sqrt{\frac{(100-1)(11.8)^2+(60-1)(10.6)^2}{100+60-2}\left(\frac{1}{100}+\frac{1}{60}\right)}} \\
&= \frac{3.1}{1.856} \\
&= 1.670
\end{aligned}$$

and the table value $t_{m+n-2,0.025} = t_{158,0.025} = 1.975$. So there is no evidence to suggest that there is a difference between mean IQs. In the case of unequal variances, the test statistic is:

$$\begin{aligned}
\frac{\mid \bar{x}-\bar{y} \mid}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}} &= \frac{\mid 102.2-105.3 \mid}{\sqrt{\frac{(11.8)^2}{100}+\frac{(10.6)^2}{60}}} \\
&= \frac{3.1}{1.807} \\
&= 1.716,
\end{aligned}$$

$$\begin{aligned}
\nu &= \frac{\left(\frac{s_X^2}{m}+\frac{s_Y^2}{n}\right)^2}{\frac{s_X^2}{m(m-1)}+\frac{s_Y^2}{n(n-1)}} \\
&= \frac{\left(\frac{(11.8)^2}{100}+\frac{(10.6)^2}{60}\right)^2}{\frac{1}{100-1}\left(\frac{(11.8)^2}{100}\right)^2+\frac{1}{60-1}\left(\frac{(10.6)^2}{60}\right)^2} \\
&= 134.9 \\
&\approx 135
\end{aligned}$$

and the table value $t_{\nu,0.025} = t_{135,0.025} = 1.978$. So again there is no evidence to suggest that there is a difference between mean IQs (consistent with the conclusion above).

4. Let μ_X = mean IQ of urban students and let μ_Y = mean IQ of rural students. The hypotheses are: $H_0 : \mu_X \geq \mu_Y$ and $H_1 : \mu_X < \mu_Y$. In the case of equal variances, the test statistic is:

$$\begin{aligned}
\frac{\bar{x}-\bar{y}}{\sqrt{s_p^2\left(\frac{1}{m}+\frac{1}{n}\right)}} &= \frac{102.2-105.3}{\sqrt{\frac{(100-1)(11.8)^2+(60-1)(10.6)^2}{100+60-2}\left(\frac{1}{100}+\frac{1}{60}\right)}} \\
&= -\frac{3.1}{1.856} \\
&= -1.670
\end{aligned}$$

and the table value $-t_{m+n-2,0.05} = -t_{158,0.05} = -1.655$. So there is evidence to support the claim that $\mu_X < \mu_Y$. In the case of unequal variances, the test statistic is:

$$\begin{aligned}
\frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_X^2}{m}+\frac{s_Y^2}{n}}} &= \frac{102.2-105.3}{\sqrt{\frac{(11.8)^2}{100}+\frac{(10.6)^2}{60}}} \\
&= -\frac{3.1}{1.807} \\
&= -1.716,
\end{aligned}$$

$$\begin{aligned}
\nu &= \frac{\left(\frac{s_X^2}{m}+\frac{s_Y^2}{n}\right)^2}{\frac{s_X^2}{m(m-1)}+\frac{s_Y^2}{n(n-1)}} \\
&= \frac{\left(\frac{(11.8)^2}{100}+\frac{(10.6)^2}{60}\right)^2}{\frac{1}{100-1}\left(\frac{(11.8)^2}{100}\right)^2+\frac{1}{60-1}\left(\frac{(10.6)^2}{60}\right)^2} \\
&= 134.9 \\
&\approx 135
\end{aligned}$$

and the table value $-t_{\nu,0.05} = -t_{135,0.05} = -1.656$. So again there is evidence to support the claim that $\mu_X < \mu_Y$ (consistent with the conclusion above).

5. Let μ_X = mean starting salary of women and let μ_Y = mean starting salary of men.

Company	1	2	3	4	5	6	7	8
Woman's salary (x_i)	52	53.2	78	75	62.5	72	39	49
Man's salary (y_i)	54	55.5	78	81	64.5	70	42	51
d_i	-2	-2.3	0	-6	-2	2	-3	-2

The hypotheses are: $H_0 : \mu_X = \mu_Y$ and $H_1 : \mu_X \neq \mu_Y$. The test statistic is:

$$\begin{aligned}\frac{\sqrt{n} \mid \bar{d} \mid}{S_d} &= \frac{\sqrt{8} \times \mid -1.913 \mid}{2.299} \\ &= 2.352.\end{aligned}$$

and the table value $t_{n-1,0.05} = t_{7,0.05} = 1.895$. So there is evidence to suggest that there are differences in starting salaries between women and men.

6. Let μ_X = mean marriage rate in 1987 and let μ_Y = mean marriage rate in 1989.

Country	1987 Rate (x_i)	1989 Rate (y_i)	d_i
Belgium	5.8	6.4	-0.6
Finland	5.4	5.1	0.3
Greece	6.3	6.0	0.3
Israel	6.9	7.0	-0.1
New Zealand	6.0	6.1	-0.1
Norway	5.0	4.9	0.1
Switzerland	6.6	6.8	-0.2
United States	9.9	9.7	0.2
Yugoslavia	7.0	6.7	0.3

The hypotheses are: $H_0 : \mu_X \leq \mu_Y$ and $H_1 : \mu_X > \mu_Y$. The test statistic is:

$$\begin{aligned}\frac{\sqrt{n} \mid \bar{d} \mid}{S_d} &= \frac{\sqrt{9} \times 0.022}{0.303} \\ &= 0.220.\end{aligned}$$

and the table value $t_{n-1,0.05} = t_{8,0.05} = 1.860$. So there is no evidence against the hypothesis that 1987 rates are no greater than that for 1989.

7. Let p_X = smoking proportion in 1986 and let p_Y = smoking proportion at the time of ACS poll. The hypotheses are: $H_0 : p_X \leq p_Y$ and $H_1 : p_X > p_Y$. The test statistic is:

$$\begin{aligned}\frac{\bar{x} - \bar{y}}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m} + \frac{\bar{y}(1-\bar{y})}{n}}} &= \frac{\frac{640}{2000} - \frac{738}{2500}}{\sqrt{\frac{\frac{640}{2000}\left(1-\frac{640}{2000}\right)}{2000} + \frac{\frac{738}{2500}\left(1-\frac{738}{2500}\right)}{2500}}} \\ &= \frac{0.0248}{0.014} \\ &= 1.790.\end{aligned}$$

and the table value $z_{0.05} = 1.645$. So there is evidence to suggest that the smoking rate has decreased since 1986.

8. Let p_X = survival probability for hospital A and let p_Y = survival probability for hospital B. The hypotheses are: $H_0 : p_X = p_Y$ and $H_1 : p_X \neq p_Y$. The p-value is:

$$p\text{-value} = Pr\left(|Z| \geq \frac{\mid \bar{x} - \bar{y} \mid}{\sqrt{\frac{\bar{x}(1-\bar{x})}{m} + \frac{\bar{y}(1-\bar{y})}{n}}}\right)$$

$$
\begin{aligned}
&= Pr\left(|Z| \geq \frac{\mid \frac{72}{480} - \frac{30}{360} \mid}{\sqrt{\frac{\frac{72}{480}\left(1-\frac{72}{480}\right)}{480} + \frac{\frac{30}{360}\left(1-\frac{30}{360}\right)}{360}}}\right) \\
&= \Pr\left(|Z| > \frac{0.067}{0.022}\right) \\
&= \Pr\left(|Z| > 3.05\right) \\
&= 2\left\{1 - \Phi(3.05)\right\} \\
&= 0.002.
\end{aligned}
$$

So there is evidence against the hypothesis that the survival probabilities are the same.

9. Let σ_X^2 = variance for caliper A and let p_Y = variance for caliper B. The hypotheses are: $H_0 : \sigma_X = \sigma_Y$ and $H_1 : \sigma_X \neq \sigma_Y$. The test statistic is:

$$
\begin{aligned}
\frac{s_X^2}{s_Y^2} &= \frac{831}{592} \\
&= 1.404
\end{aligned}
$$

and the table values are $F_{m-1,n-1,0.025} = F_{9,9,0.025} = 4.026$ and $F_{m-1,n-1,0.975} = F_{9,9,0.975} = 0.248$. So there is no evidence to suggest that the calpers have different variability.

10. The hypotheses are: $H_0 : \sigma_X = \sigma_Y$ and $H_1 : \sigma_X > \sigma_Y$. The test statistic is:

$$
\begin{aligned}
\frac{s_X^2}{s_Y^2} &= \frac{(2.3)^2}{(1.1)^2} \\
&= 4.372
\end{aligned}
$$

and the table value is $F_{m-1,n-1,0.05} = F_{8,8,0.05} = 3.438$. So there is no evidence to suggest that the new oven has lesser variability than the current one.