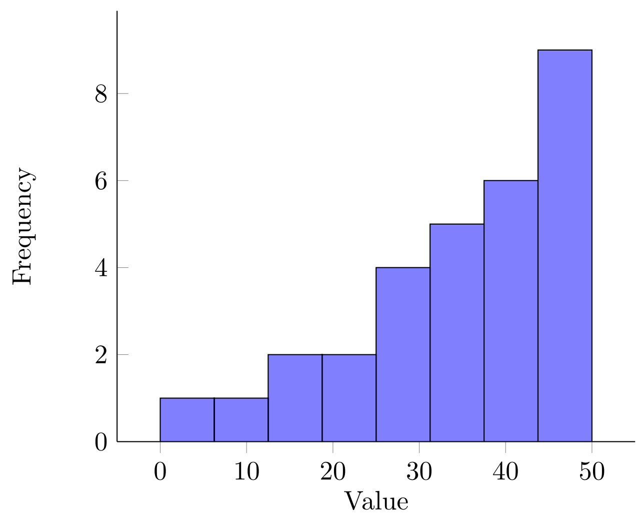
# Math 343 - Homework 2

**Preston Duffield**
Western Washington University
April 14, 2023

## Question 1

Data with a rightward skew would produce a normal probability plot with a positive curvature. Below is an example of a histogram that would produce a positively curved normal probability plot.



## Question 2

$H_0$: $\mu_1 - \mu_2 = 10$
$H_a$: $\mu_1 - \mu_2 > 10$

$$\begin{aligned} Z &= \frac{\bar{y}_1 - \bar{y}_2 - 10}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \\ &= \frac{162.5 - 155 - 10}{\sqrt{\frac{1^2}{10} + \frac{1^2}{12}}} \\ &= -5.838 \end{aligned}$$

By observing that $Z_\alpha = 3.09$ we can conclude the following. There is not enough statistical evidence to support the hypothesis that $\mu_1 - \mu_2 = 10$, ie, the breaking strength of plastic 1 exceeds that of plastic 2 by at least 10 psi. Therefore, based on the sample information, they should not use plastic 1.

A $100(1 - \alpha)$ confidence interval is given by:

$$\begin{aligned} 100(1-\alpha)\ \text{C.I} = \bar{y}_1 - \bar{y}_2 &\pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \\ 7.5 &\pm 3.29\sqrt{\frac{1^2}{10} + \frac{1^2}{12}} \\ 7.5 &\pm 1.40 \end{aligned}$$

We are 99% confident that the true value of $\mu_1 - \mu_2$ is between 6.1 and 8.9. This is consistent with our above test which conlucded that the difference was not greater than 10.

## Question 3

### a)

**Test**

| | |
|---|---|
| Null hypothesis | $H_0: \sigma_1 / \sigma_2 = 1$ |
| Alternative hypothesis | $H_1: \sigma_1 / \sigma_2 \neq 1$ |
| Significance level | $\alpha = 0.05$ |

| Method | Test Statistic | DF1 | DF2 | P-Value |
|---|---|---|---|---|
| Bonett | 0.00 | 1 | | 0.963 |
| Levene | 0.00 | 1 | 18 | 1.000 |

Figure 1: The output of the test for two variances from Minitab.

Since the P-value $>$ $\alpha$ we can conclude the following. There is enough statistical evidence to support the hypothesis that both of the variances are equal.

### b)

**Test**

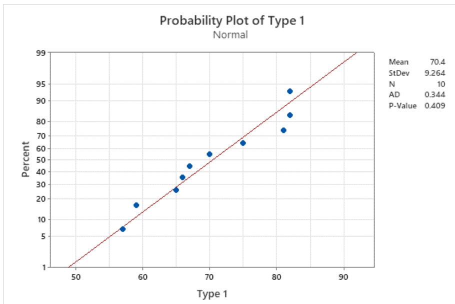
| | |
|---|---|
| Null hypothesis | $H_0: \mu_1 - \mu_2 = 0$ |
| Alternative hypothesis | $H_1: \mu_1 - \mu_2 \neq 0$ |

| T-Value | DF | P-Value |
|---|---|---|
| 0.05 | 18 | 0.962 |

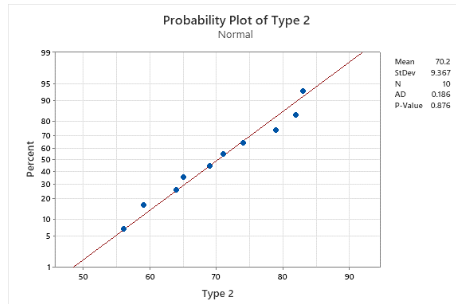
Figure 2: The output of the two sample t test from Minitab. Assuming equal variances.

Since the P-value $= 0.962 > \alpha$ we can conclude the following. There is enough statistical evidence to support the hypothesis that the two means are equal.

## c)




(a) Minitab output showing the probability plot of type 1.




(b) Minitab output showing the probability plot of type 2.

Note that for both Type 1 and Type 2, the hypothesis we will test is as follows,
$H_0$: The data are drawn from a normal distribution.
$H_a$: The data are not drawn from a normal distribution.

**Type 1** Since the P-value $= 0.409 > \alpha$ we can conclude the following. The evidence of the data is consistent with the hypothesis that the data are drawn from a normal distribution.

**Type 2** Similarly, since the P-value $= 0.876 > \alpha$ we can conclude the following. The evidence of the data is consistent with the hypothesis that the data are drawn from a normal distribution.

## Question 4

### a)

The hypothesis to test whether insulation has reduced the average energy consumption is:
$H_0$: $\delta = 0$
$H_a$: $\delta > 0$
This will be a right-tailed paired t-test at $\alpha = 0.1$.

| | |
|---|---|
| Sample Mean of differences | $\bar{d} = 0.54$ |
| Sample Standard Deviation of differences | $S_d = 1.01566$ |
| Sample Size | $n = 10$ |

The test statistic $t$ is:

$$\begin{aligned} t =& \frac{\bar{d}}{S_d/\sqrt{n}} \\ =& \frac{0.54}{1.01566/\sqrt{10}} \\ =& 1.681 \end{aligned}$$

The critical region begins at:

$$t_{\alpha,n-1} = t_{.1,9} = 1.383$$

Since $t = 1.681 > t_{.1,9} = 1.383$, that is, the test statistic is in the right-tailed critical region we can conclude the following. There is enough statistical evidence to support the hypothesis that $\delta > 0$, ie, the mean energy consumption after insulation is less than the mean eneergy consumption before insulation.

### b)

**1.** One factor that could have affected the results is the mean difference in temperature between the two winters. If one winter was significantly colder than another it could have affected the data that was collected.

**2.** Another factor that could have affected the result is that other electronic appliances could have significantly affected the energy consumption. If there was a significant difference in the energy consumption of other appliances between the two winters it could have affected the results.

# Question 5

## a/b)

**Test**

| | |
|---|---|
| Null hypothesis | $H_0$: μ_difference = 0 |
| Alternative hypothesis | $H_1$: μ_difference ≠ 0 |

| T-Value | P-Value |
|---|---|
| 0.43 | 0.674 |

Figure 4: The output of the paired t test from Minitab.

Since the P-value = 0.674 > $\alpha$ we can conclude the following. There is enough statistical evidence to support the hypothesis that the two means are equal.

## c)

**Estimation for Paired Difference**

| Mean | StDev | SE Mean | 95% CI for μ_difference |
|---|---|---|---|
| 0.000250 | 0.002006 | 0.000579 | (-0.001024, 0.001524) |

*μ_difference: population mean of (Caliper 1 - Caliper 2)*

Figure 5: The output of the paired t test containing the confidence interval from Minitab.

From the above confidence interval we can conclude the following. We are 95% confident that the true difference between the population means is between -0.001 and 0.001.

We can also note that the confidence interval contains 0, which is consistent with our hypothesis test.

## Question 6

### a)

Listing 1: R output of Shapiro-Wilk test on Birth Order: 1

```
Shapiro-Wilk normality test

data:  b1
W = 0.84597, p-value = 0.05201
```

Since the P-value = 0.05201 $> \alpha$ we can conclude the following. The evidence of the data is consistent with the hypothesis that the data are drawn from a normal distribution.

Listing 2: R output of Shapiro-Wilk test on Birth Order: 1

```
Shapiro-Wilk normality test

data:  b2
W = 0.92972, p-value = 0.4452
```

Since the P-value = 0.4452 $> \alpha$ we can conclude the following. The evidence of the data is consistent with the hypothesis that the data are drawn from a normal distribution.

### b)

Listing 3: R output of a paired t test

```
Paired t-test

data:  b1 and b2
t = -0.36577, df = 9, p-value = 0.723
alternative hypothesis:
    true mean difference is not equal to 0
95 percent confidence interval:
    -0.3664148  0.2644148
sample estimates:
mean difference
    -0.051
```

The confidence intercal on the difference in mean score leads us to the following conclusion. We are 95% confident that the true difference in the population means is between -0.36 and 0.26. Since the confidence interval contains 0, we can also conclude that the two sample means may be equal.

### c)

$H_0$: $\mu_1 = \mu_2$
$H_a$: $\mu_1 \neq \mu_2$

Since the P-value $= 0.723 > \alpha$ we can conclude the following. There is enough statistical evidence to support the hypothesis that the sample means are equal, ie, $\mu_1 = \mu_2$.

## Question 7

To test the hypothesis that the true population variance of Formulation 1 (new recipe), is greater than the true population variance of Formulation 2 (original recipe) , we will use the following hypothesis:
$H_0$: $\sigma_1^2 = \sigma_2^2$
$H_a$: $\sigma_1^2 > \sigma_2^2$

This will be a right tailed test of two variances at $\alpha = 0.05$. The test statistic for the test is:

$$\begin{aligned} F =& \frac{S_1^2}{S_2^2} \\ =& \frac{0.100}{0.061} \\ =& 1.639 \end{aligned}$$

The critical region begins at:

$$F_{\alpha,n_2-1,n_1-1}, = F_{.05,9,9} = 3.18$$

Since $F = 1.639 < F_{.05,9,9} = 3.18$, the test statistic is not in the right-tailed critical region we can conclude the following. There is enough statistical evidence to support the hypothesis that $\sigma_1^2 > \sigma_2^2$, ie, the true population variance of Formulation 1 (new recipe), is greater than the true population variance of Formulation 2 (original recipe).

# Question 8

To derive the confidence interval for $a$ we can start with the fact that $\frac{\S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$ has an $F$ distribution with $n_1 - 1$ numerator degrees of freedom and $n_2 - 1$ denominator degrees of freedom. This insight leads to:

$$
\begin{aligned}
1 - \alpha &= P\left(F_{1-\frac{\alpha}{2},n_1-1,n_2-1} < \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} < F_{\frac{\alpha}{2},n_1-1,n_2-1}\right) \\
&= P\left(F_{1-\frac{\alpha}{2},n_1-1,n_2-1} \cdot \frac{S_2^2}{\sigma_2^2} < \frac{S_1^2}{\sigma_1^2} < F_{\frac{\alpha}{2},n_1-1,n_2-1} \cdot \frac{S_2^2}{\sigma_2^2}\right) \\
&= P\left(\frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{F_{\frac{\alpha}{2},n_1-1,n_2-1}}{F_{1-\frac{\alpha}{2},n_1-1,n_2-1}} \cdot \frac{S_1^2}{S_2^2}\right)
\end{aligned}
$$

Therefore the $100(1-\alpha)$ confidence interval for the ratio of variances is given by:

$$
\frac{\sigma_1^2}{\sigma_2^2} \in \left(\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2},n_1-1,n_2-1}}, \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\frac{\alpha}{2},n_1-1,n_2-1}}\right)
$$