

## Homework 8 - Aidan Sharpe

### 1

The mean pull-off force of a connector depends on cure time.

#### a)

State the null and alternative hypotheses used to demonstrate that the pull-off force is below 20 newtons.

$$H_0: \mu = 20[\text{N}] \quad H_a: \mu < 20[\text{N}]$$

#### b)

Assume that the previous test does not reject the null hypothesis. Does this result provide strong enough evidence that the pull-off force is greater than or equal to 20 newtons? Please explain.

Failing to reject the null hypothesis does not mean there is enough evidence to confirm it, it simply leaves that possibility open. Additionally, failing to reject the null hypothesis does not automatically mean that we reject the alternative hypothesis.

### 2

According to Chemical Engineering, an important property of fiber is its water absorbency. The average percent absorbency of 27 randomly selected pieces of cotton fiber was found to be 21.2 with a standard deviation of 1.5. A random sample of 23 pieces of acetate yielded an average percent of 12 with a standard deviation of 1.25. Use a Confidence Interval to determine whether there is strong evidence that the population mean percent absorbency is significantly higher for cotton fiber than for acetate? Assume that the percent absorbency is approximately normally distributed and that the population variances in percent absorbency for the two fibers are the same. Let  $\alpha = 0.01$

|           | Sample 1 | Sample 2 |
|-----------|----------|----------|
| $n$       | 27       | 23       |
| $\bar{x}$ | 21.2     | 12       |
| $s$       | 1.5      | 1.25     |

99% confidence interval for sample 1:

```
>>> from scipy.stats import norm
>>> a = 0.01
>>> z_a2 = norm.ppf(1 - (a/2))
>>> x_bar = 21.2
```

```
>>> s = 1.5
>>> n = 27
>>> ci = [x_bar - z_a2*s/n**0.5, x_bar + z_a2*s/n**0.5]
>>> ci
[20.45642212910476, 21.94357787089524]
```

99% confidence interval for sample 2:

```
>>> from scipy.stats import norm
>>> a = 0.01
>>> z_a2 = norm.ppf(1 - (a/2))
>>> x_bar = 12
>>> s = 1.25
>>> n = 23
>>> ci = [x_bar - z_a2*s/n**0.5, x_bar + z_a2*s/n**0.5]
>>> ci
[11.328628077574322, 12.671371922425678]
```

Since there is no overlap between the confidence intervals, there is significantly different absorbency between cotton fiber and acetate.

### 3

A study of asthmatics measured the peak expiratory flow rate (basically, a person's maximum ability to exhale) before and after a walk on a cold winter's day for a random sample of nine asthmatics.

| Subject   | Before | After  | Difference |
|-----------|--------|--------|------------|
| 1         | 312    | 300    | 12         |
| 2         | 242    | 201    | 41         |
| 3         | 340    | 232    | 108        |
| 4         | 388    | 312    | 76         |
| 5         | 296    | 220    | 76         |
| 6         | 254    | 256    | -2         |
| 7         | 31     | 328    | 63         |
| 8         | 402    | 330    | 72         |
| 9         | 290    | 231    | 59         |
| $\bar{x}$ | 323.89 | 267.78 | 56.11      |
| $s$       | 59.83  | 50.01  | 34.17      |

a)

Using a two-sample t-test: Do the data suggest that there is a difference between the peak expiratory flow rate before vs after a walk on a cold winter's day for asthmatics? Use  $\alpha = 0.01$

```

>>> import scipy.stats as st

>>> def p_value(tstat, n1, n2):
...     df = n1 + n2 - 2
...     tstat = abs(tstat)
...     return 1 - st.t.cdf(tstat, df)

>>> def tstar(x1, x2, sp, n1, n2):
...     return (x1 - x2) / (sp * (1/n1 + 1/n2)**0.5)

>>> def sp(x1, x2, s1, s2, n1, n2):
...     return (((n1-1)*(s1**2) + (n2-1)*(s2**2)) / (n1 + n2 - 2))**0.5

>>> x1 = 323.89
>>> x2 = 267.78
>>> s1 = 59.83
>>> s2 = 50.01
>>> n1 = 9
>>> n2 = 9
>>> s_p = sp(x1, x2, s1, s2, n1, n2)
>>> t_star = tstar(x1, x2, s_p, n1, n2)
>>> p_value(t_star, n1, n2)
0.02320508093086404

```

Since the p-value is greater than  $\alpha$ , we do not have enough evidence to reject the notion that going on a walk on a cold winter's day has no effect on the expiratory rate for asthmatics.

b)

Using a paired t-test: Does the data suggest that there is a difference between the peak expiratory flow rate before vs after a walk on a cold winter's data for asthmatics? Use  $\alpha = 0.01$

Using a paired t-test, the degrees of freedom in the `p_value(tstat, n1, n2)` function must be adjusted.

```

>>> def p_value(tstat, n1, n2):
...     df = (n1 + n2)/2 - 2
...     return 1 - st.t.cdf(tstat, df)

>>> p_value(t_star, n1, n2)
0.03387016598071013

```

Using a paired t-test, the p-value increases, but our conclusion does not change.

c)

Are the conclusions for the two-sample t-test and the paired t-test consistent? If not, what could be the reason?

While the conclusions are the same, it seems possible for the conclusions to be different in some scenarios. The paired t-test takes into account the persistent sample population, which changes the degrees of freedom for the p-value.