HW 5-1 (Uploaded on Apr. 20, 2017) **Name:**

HW 5.1. Inferences of $\mu_1 - \mu_2$ with known σ_1 and σ_2. The concentration of active ingredient in a liquid laundry detergent is thought to be affected by the type of catalyst used in the process. The concentration is known to have normal distribution with the standard deviation 3 grams per liter regardless of the catalyst type. Then observations on concentration are taken with each catalyst, and the data follow:

Catalyst 1: 57.9, 66.2, 65.4, 65.4, 65.2, 62.6, 67.6, 63.7, 67.2, 71.0.
Catalyst 2: 66.4, 71.7, 70.3, 69.3, 64.8, 69.6, 68.6, 69.4, 65.3, 68.8.

(a) Find a 95% confidence interval on the difference in mean active concentrations for the two catalysts.

(b) With significant level $\alpha = 0.05$, is there any evidence to indicate that the mean active concentrations depend on the choice of catalyst (i.e., they are different or not)? What is your conclusion?

Sol

(a) Since population variances are known, we use `2-sampZInt` in calculator or using the equation at page 100 in lecture notes to obtain the confidence interval. A 95% confidence interval for the difference in mean active concentrations for the two catalysts is $(-5.83, -0.5704)$.

Interpretation: We are 95% confident that the population mean difference of concentrations $\mu_1 - \mu_2$ is in the interval $(-5.83, -0.5704)$.

(b) Null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_a : \mu_1 \neq \mu_2.$$

which are equivalent to

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_a : \mu_1 - \mu_2 \neq 0.$$

Set significant level $\alpha = 0.05$.

- Confidence Interval Approach: According to the reject criterion on page 101, since "0" is not in 95% confidence interval, we do reject the null hypothesis.
- P-value approach: Here P-value can be obtaining by `2-SampleZTest` (since population variances are known) in calculator. Since P-value$= 0.0170 < 0.05 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do provide sufficient evidence to conclude that the real difference of concentrations $\mu_1 - \mu_2$ is different from 0, i.e., μ_1 is different from μ_2.

HW 5.2. Inferences of $\mu_1 - \mu_2$ with $\sigma_1 = \sigma_2$. An article in *Nature* (2003, Vol. 48, p. 1013) described an experiment in which subjects consumed different types of chocolate to determine the effect of eating chocolate on a measure of cardiovascular health. We will consider the results for only dark chocolate and milk chocolate. In the experiment, 12 subjects consumed 100 grams of dark chocolate and 200 grams of milk chocolate, one type of chocolate per day, and after one hour, the total antioxidant capacity of their blood plasma was measures in an assay. The subjects consisted of seven women and five men with an average age range of 32.2 ± 1 years, an average weight of 65.8 ± 3.1 kg, and average body mass index of 21.9 ± 0.4 kh/m^2. Data similar to that reported in the article follows.

Dark Chocolate: 118.8, 122.6, 115.6, 113.6, 119.5, 115.9, 115.8, 115.1, 116.9, 115.4, 115.6, 107.9.

Milk Chocolate: 102.1, 105.8, 99.6, 102.7, 98.8, 100.9, 102.8, 98.7, 94.7, 97.8, 99.7, 98.6.

Let μ_1 be the mean blood plasma antioxidant capacity resulting from eating dark chocolate and μ_2 be the mean blood plasma antioxidant capacity resulting from eating milk chocolate.

(a) According to the box plots in Figure 1 or sample standard deviations, do you believe that the population standard deviations are equal? Why?

(b) Calculate a two-sided 95% on the mean difference $\mu_1 - \mu_2$ of blood plasma antioxidant capacities.

(c) Is there evidence to support the claim that consuming dark chocolate produces a higher mean level of total blood plasma antioxidant capacity that consuming milk chocolate? Use significant level $\alpha = 0.05$. What is your conclusion?

(d) Do normal qq plots of blood plasma antioxidant capacity in Figure 2 indicate any violations of the assumptions for the tests and confidence interval that your performed?

Sol.

(a) Yes, since the quartile differences of two samples look similarly so we treat $\sigma_1 = \sigma_2$.

(b) Since population variances are unknown but equal, we use `2-sampTInt` in calculator with `Pooled` option or using the equation at page 103 in lecture notes to obtain the confidence interval. A 95% confidence interval for the difference in mean active concentrations for the two catalysts is $(13.142, 18.608)$.

Interpretation: We are 95% confident that the mean difference $\mu_1 - \mu_2$ of blood plasma antioxidant capacities is in $(13.142, 18.608)$.

(c) Null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_a : \mu_1 > \mu_2.$$

which is equivalent to

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_a : \mu_1 - \mu_2 > 0.$$

Set significant level $\alpha = 0.05$.

- Confidence Interval Approach: According to the reject criterion on page 104 in lecture notes, we construct the confidence lower-bound which is 13.612 (or we can use `2-sampTInt` by setting confidence level to be $2\alpha = 0.1$ and without `Pooled` option, then the smaller number of the output confidence is the 95%lower-bound). Since $\delta_0 = 0 < 13.612$ which does not meet the reject criterion, we do reject the null hypothesis.
- P-value approach: Here P-value can be obtaining by `2-SampTTest` with `Pooled` option (since population variances are unknown but equal) in calculator. Since P-value$= 1.8417 \times 10^{-11} < 0.05 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do provide sufficient evidence to conclude that the real difference of concentrations $\mu_1 - \mu_2$ is larger than 0, i.e., μ_1 is larger than μ_2.

(d) The first sample may not come from normal distribution since there are points departuring the straight line. In the second qq plot, all points are close to the straight so the normal assumption for the second sample is appropriate.

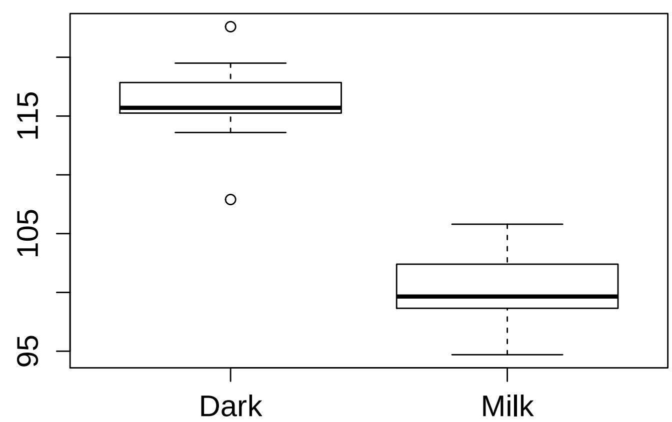

Figure 1: Boxplots for Chocolate data.

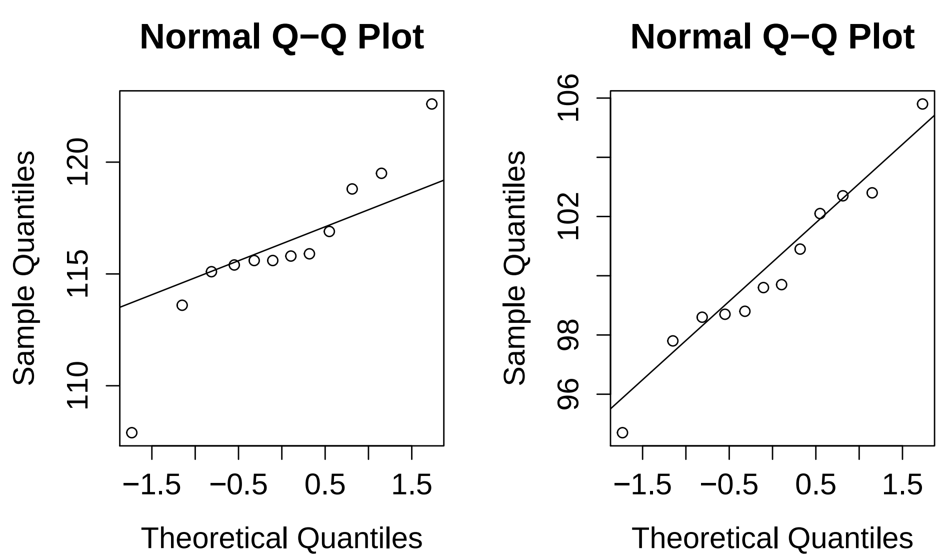

Figure 2: Normal quantile-quantile (qq) plots for Chocolate data.

HW 5.3. Inferences of $\mu_1 - \mu_2$ with $\sigma_1 \neq \sigma_2$. An article in *Polymer Degradation and Stability* (2006, Vol. 91) presented data from a nine-year aging study on S537 foam. Foam samples were compressed to 50% of their original thickness and stored at different temperatures for nine years. At the star of the experiment as well as during each year, sample thickness was measured, and the thicknesses of the eight samples at each storage condition were recorded. The data of two storage conditions follow:

50°C: 0.047, 0.060, 0.061, 0.064, 0.080, 0.090, 0.118, 0.165, 0.183.

60°C: 0.062, 0.105, 0.118, 0.137, 0.153, 0.197, 0.210, 0.250, 0.375.

(a) According to the box plots in Figure 3 or standard deviations, what do you believe: $\sigma_1 = \sigma_2$ or $\sigma_1 \neq \sigma_2$? Why?

(b) Find a 95% confidence interval for the difference in the mean compression for the two temperatures.

(c) Is there evidence to support the claim that mean compression increases with the temperature at the storage condition?

(d) Do normal qq plots of compression in Figure 4 indicate any violations of the assumptions for the tests and confidence interval that your performed?

Sol.

(a) Since the quartile differences of two samples look differently, we treat $\sigma_1 \neq \sigma_2$.

(b) Since population variances are unknown and unequal, we use `2-sampTtest` in calculator without `Pooled` option or using the equation at page 106 in lecture notes to obtain the confidence interval. A 95% confidence interval for the difference in mean active concentrations for the two catalysts is $(-0.1587, -0.0055)$.

We are 95% confident that the difference of population means $\mu_1 - \mu_2$ of thicknesses is in $(-0.1587, -0.0055)$.

(c) Null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_a : \mu_1 < \mu_2,$$

which are equivalent to

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_a : \mu_1 - \mu_2 < 0.$$

Set significant level $\alpha = 0.05$.

- Confidence Interval Approach: According to the reject criterion on page 107 in lecture notes, we construct the confidence upper-bound which is -0.0194 (or we can use `2-sampTInt` by setting confidence level to be $2\alpha = 0.1$ and without `Pooled` option, then the larger number of the output confidence is the 95% upper-bound). Since $\delta_0 > -0.0194$ which does meet the rejection criterion, we do reject the null hypothesis.

- P-value approach: Here P-value can be obtaining by `2-SampTTest` without `Pooled` option (since population variances are unknown and not equal) in calculator. Since P-value$= 0.0189 < 0.05 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do provide sufficient evidence to conclude that the population mean difference of thicknesses $\mu_1 - \mu_2$ is smaller than 0, i.e., μ_1 is smaller than μ_2.

(d) The first qq plot looks fine since all points are not too far away from the straight line. In the second qq plot, the departure on the right-hand side is acceptable. So the normal assumptions for two samples are appropriate.

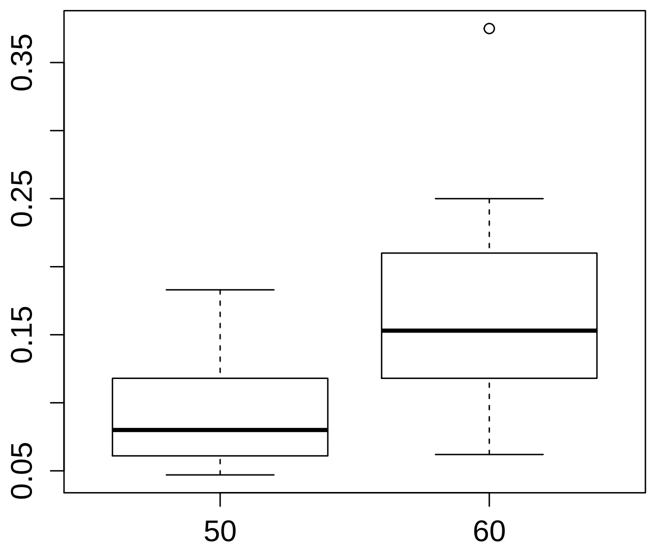

Figure 3: Boxplots for Foam data.

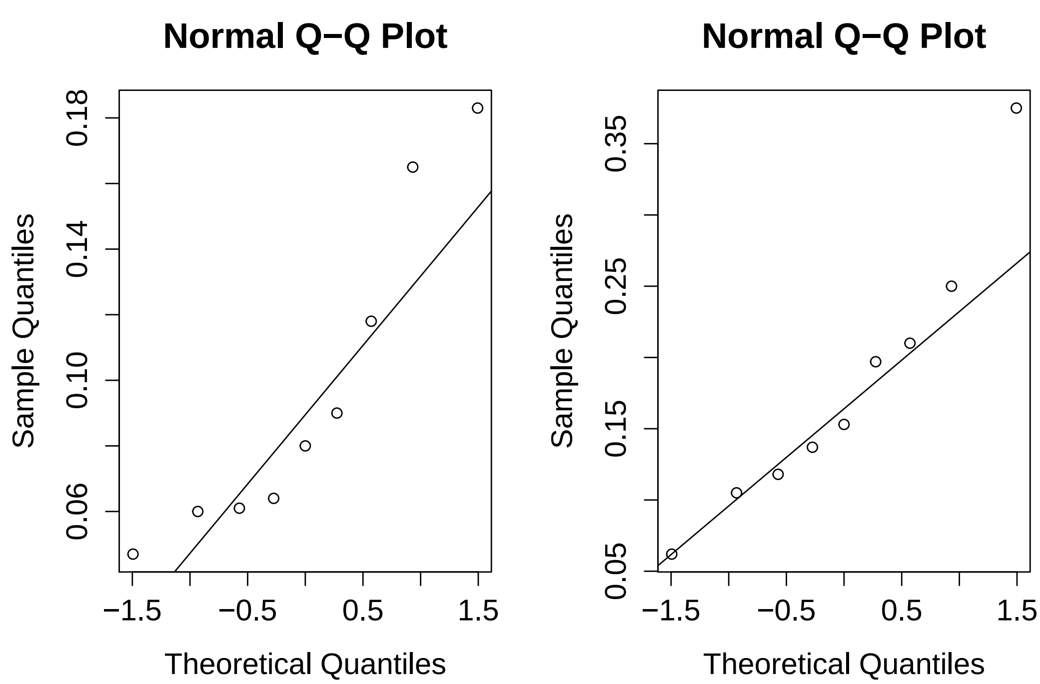

Figure 4: Normal quantile-quantile (qq) plots for Foam data.

HW 5.4. Inference of $p_1 - p_2$. An article in *Knee Surgery, sports traumatology, Arthroscopy* (2005, Vol. 13, pp. 273-279) considered arthroscopic meniscal repair with an absorbable screw. Results showed that for tears greater than 25 millimeters, 14 of 18 (78%) repairs were successful, but for shorter tears, 22 of 30 (73%) repairs were successful.

(a) Calculate a one-sided 95% confidence bound on the difference in proportions.

(b) Is there evidence that the success rate is greater for longer tears? Use significant level $\alpha = 0.05$. What is your conclusion?

Sol.

(a) We use `TwoSamp-PropInt` in calculator or confidence interval stated on page 110 in lecture to obtain the confidence interval. A 95% confidence interval for the the population proportion difference $p_1 - p_2$ of success rates is $(-0.2044, 0.293)$.

Interpretation: We are 95% confident that the population proportion difference $p_1 - p_2$ of success rates is in $(-0.2044, 0.293)$.

(b) Null and alternative hypotheses:

$$H_0 : p_1 - p_2 = 0 \text{ versus } H_a : p_1 - p_2 > 0,$$

which are equivalent to

$$H_0 : p_1 = p_2 \text{ versus } H_a : p_1 > p_2,$$

Set significant level $\alpha = 0.05$.

- P-value approach: Here P-value can be obtaining by `2-PropZTest` in calculator. Since P-value$= 0.3653 > 0.05 = \alpha$, we do not reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do not provide sufficient evidence to conclude that the population proportions difference of success rate $p_1 - p_2$ is larger than 0, i.e., p_1 is not larger than p_2.

HW 5.5. Inferences of σ_1^2/σ_2^2. The thickness of a plastic film (in mils) on a substrate material is thought to be influenced by the temperature at which the coating ins applied. In completely randomized experiment, 11 substracteds are coated at 125°F, resulting in a sample mean coating thickness of sample mean $\overline{x}_1 = 103.5$ and a sample standard deviation of $s_1 = 10.2$ Another 13 substrates are coated at 150°F for which sample mean $\overline{x}_2 = 99.7$ and sample standard deviation $s_2 = 20.1$ are observed. It was originally suspected that raising the process temperature would reduce mean coating thickness.

(a) Construct a 95% confidence interval for σ_1^2/σ_2^2.

(b) Test $H_0 : \sigma_1 = \sigma_2$ against $H_a : \sigma_1 > \sigma_2$ using $\alpha = 0.05$. What is your conclusion?

Sol.

(a) Here we have $\alpha = 0.05$, $s_1^2 = 10.2^2$, and $s_2^2 = 20.1^2$, a 95% confidence interval for σ_1^2/σ_2^2 is given by

$$\begin{aligned} & \left(\frac{s_1^2}{s_2^2} \times \frac{1}{F_{n_1-1,n_2-1,\alpha/2}}, \frac{s_1^2}{s_2^2} \times F_{n_2-1,n_1-1,\alpha/2}\right) \\ = & \left(\frac{(10.2)^2}{(20.1)^2} \times \frac{1}{3.3736}, \frac{(10.2)^2}{(20.1)^2} \times 3.6209\right) \\ = & (0.0763, 0.9325) \end{aligned}$$

Interpretation: We are 95% confident that the ratio σ_1^2/σ_2^2 is in $(0.0763, 0.9325)$.

(b) Null and alternative hypotheses:

$$H_0 : \sigma_1 = \sigma_2 \text{ versus } H_a : \sigma_1 \neq \sigma_2.$$

Set significant level $\alpha = 0.05$.

- Confidence Interval Approach: According to the reject criterion on page 120, since "1" is not in the 95% confidence interval $(0.0763, 0.9325)$, we do reject the null hypothesis.
- P-value approach: Here P-value can be obtaining by `2-SampFTest` in calculator. Since P-value$= 0.0395 > 0.05 = \alpha$, we do not reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do not provide sufficient evidence that the standard deviations are different.

HW 5.6. Inferences of $\mu_1 - \mu_2$ in matched pairs design. The manager of a fleet of automobiles is testing two brands of radial tires and assigns one tire of each brand at random to the two rear wheels of eight cars and runs the cars until the tires wear out. The data (in kilometers) follow:

Car	Brand 1	Brand 2
1	36,925	34,318
2	45,300	42,280
3	36,240	35,500
4	32,100	31,950
5	37,210	38,015
6	48,360	47,800
7	38,200	37,810
8	33,500	33,215

(a) Find a 95% confidence interval on the difference in mean life.

(b) Which brand would you prefer? Test $H_0 : \mu_D = 0$ against $H_a : \mu_D \neq 0$ using $\alpha = 0.05$. What is your conclusion?

Sol. First we calculate the difference, Brand 1 - Brand 2, for each car:

$$2607, 3020, 740, 150, -805, 560, 390, 285.$$

According to those differences, we construct one sample confidence interval for μ_D with unknown population variance and perform one sample $t-$test

(a) We use `T-Interval` in calculator or confidence interval stated on page 122 in lecture to obtain the confidence interval. We are 95% confident that the population mean difference μ_D of distances until tires wearing out is in $(-210.1197, 1946.8697)$.

(b) Null and alternative hypotheses:

$$H_0 : \mu_D = 0 \text{ versus } H_a : \mu_D > 0.$$

Set significant level $\alpha = 0.05$.

- Confidence Approach:According to the reject criterion on page 123 in lecture notes, we construct the confidence upper-bound which is ?? (or we can use `T-Interval` by setting confidence level to be $2\alpha = 0.1$ and without `Pooled` option, then the larger number of the output confidence is the 95% upper-bound).
- P-value approach: Here P-value can be obtaining by `T-test` in calculator. Since P-value$=$ $0.0986 > 0.05 = \alpha$, we do not reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do not provide sufficient evidence to conclude that the population mean difference μ_D of distances until tires wearing out is different.

Even though data do not provide sufficient evidence to tell there are difference between brands, I prefer Brain 1 since most of the differences,283 Brain 1 - Brain 2, are larger than 0.

HW 5.7. The compressive strength of concrete is being studied, and four different mixing techniques are being investigated. The following data have been collected. Test the hypothesis that mixing techniques affect the strength of the concrete. Use $\alpha = 0.05$.

Mixing Technique	Compressive Strength (psi)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

Sol. Null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

versus

$$H_a : \text{Not all population means are equal.}$$

Set significant level $\alpha = 0.05$.

- P-value approach: Here P-value can be obtaining by `ANOVA` in calculator. It can also be obtain by the ANOVA table provided by `R`:

```
Analysis of Variance Table

Response: Compressive Strength
                   Df Sum Sq Mean Sq F value    Pr(>F)
Mixing Technique    3 489740  163247  12.728 0.0004887 ***
Residuals          12 153908   12826
---
Signif. codes:
0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

  Since P-value$= 4.887 \times 10^{-4} < 0.05 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.05$, the data do provide sufficient evidence to conclude that the population means of strength of the concrete are different due to mixing technique.

HW 5.8. An electronics engineer is interested in the effect on tube conductivity of five different types of coating for cathode ray tubes in a telecommunications system display device. The following conductivity data are obtained. Is there any difference in conductivity due to coating type? Use $\alpha = 0.01$.

Coating Type	Conductivity			
1	143	141	150	146
2	152	149	137	143
3	134	133	132	127
4	129	127	132	129
5	147	148	144	142

Sol. Null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

versus

$$H_a : \text{Not all population means are equal.}$$

Set significant level $\alpha = 0.01$.

- P-value approach: Here P-value can be obtaining by `ANOVA` in calculator. It can also be obtain by the ANOVA table provided by `R`:

```
Analysis of Variance Table

Response: Conductivity
             Df  Sum Sq Mean Sq F value    Pr(>F)
Coating Type  4 1060.50 265.125  16.349 2.414e-05 ***
Residuals    15  243.25  16.217
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

  Since P-value$= 2.41 \times 10^{-5} < 0.01 = \alpha$, we do reject the null hypothesis.

Conclusion: At significant level $\alpha = 0.01$, the data do provide sufficient evidence to conclude that the population means of conductivity are different due to coating type.