HW #13 solutions

① First, sort the data:

$$\text{sort}(s) =$$

$$= \{ \overset{1}{51}, \overset{2}{53}, \overset{3}{65}, \overset{4}{68}, \overset{5}{69}, \overset{6}{70}, \overset{7}{72}, \overset{8}{75}, \overset{9}{79}, \overset{10}{82}, \overset{11}{84},$$

$$\overset{12}{87}, \overset{13}{89}, \overset{14}{90}, \overset{15}{94}, \overset{16}{100} \}$$

min: 51 , max: 100

$Q_2 = 50^{th}$ percentile

$$L = \left(\frac{50}{100}\right) 16 = 8 \Rightarrow val_{50} = \frac{\text{sort}(s)_8 + \text{sort}(s)_9}{2}$$

$$\Rightarrow val_{50} = \frac{75 + 79}{2} = 77 = Q_2 = \text{median}$$

$Q_1 = 25^{th}$ percentile

$$L = \left(\frac{25}{100}\right) 16 = 4 \Rightarrow val_{25} = \frac{\text{sort}(s)_4 + \text{sort}(s)_5}{2}$$

$$val_{25} = \frac{68 + 69}{2} = 68.5 = Q_1$$

$Q_3 = 75^{th}$ percentile

$$L = \left(\frac{75}{100}\right) 16 = \frac{3}{4}(16) = 12 \Rightarrow val_{75} = \frac{\text{sort}(s)_{12} + \text{sort}(s)_{13}}{2}$$

$$val_{75} = \frac{87+89}{2} = 88 = Q_3$$

Thus, the 5 # summary is:

$$\{min, Q_1, Q_2, Q_3, max\} = \{51, 68.5, 77, 88, 100\}.$$

$$IQR = Q_3 - Q_1 = 88 - 68.5 = 19.5$$

outliers are values $> Q_3 + 1.5\,IQR$

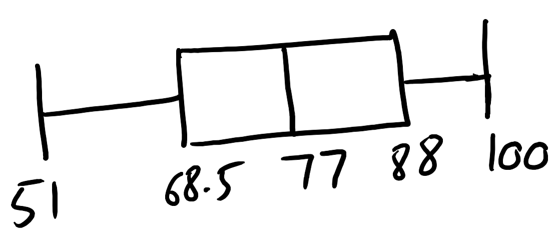
$$= 88 + 1.5(19.5) = 117.25$$

or less than $Q_1 - 1.5\,IQR = 68.5 - 1.5(19.5)$

$$= 39.25$$

Hence, no outliers are present.

The modified boxplot looks like below:



(if outliers were present they would be marked with *)

② Box has 8 red, 3 white, 9 blue balls

We draw 3 out of $8+3+9 = 20$ total.

$$P(\text{all 3 red}) = \frac{{}_8C_3}{{}_{20}C_3} = .049 = \frac{8}{20} \times \frac{7}{19} \times \frac{6}{18}$$

$P(\text{all 3 white}) = \dfrac{{}_3C_3}{{}_{20}C_3} = \dfrac{1}{1140}$

$P(\text{2 are red and 1 is white}) =$

$= \dfrac{({}_8C_2)({}_3C_1)}{{}_{20}C_3} = \dfrac{7}{95}$

$P(\text{at least 1 is white}) =$

$= 1 - P(\text{none are white})$

$= 1 - \dfrac{{}_{17}C_3}{{}_{20}C_3} = 1 - \dfrac{34}{57} = \dfrac{23}{57}$

$P(\text{1 of each color is drawn}) =$

$= \dfrac{({}_8C_1)({}_3C_1)({}_9C_1)}{{}_{20}C_3} = \dfrac{18}{95}$

(3) Two cards drawn from standard 52 card deck.

$P(\text{both cards aces with replacement})$

$= \left(\frac{4}{52}\right)\left(\frac{4}{52}\right) = \frac{1}{169}$ (independent events)

P(both cards aces without replacement) =

$$= \left(\frac{4}{52}\right)\left(\frac{3}{51}\right) = \frac{1}{221} \quad \text{(dependend evends)}$$

After first ace is drawn there are 3 aces and 51 cards left.

(4) Manufacture claims $p = 0.9$ (90% effective)

In a sample of 200 people who had allergy, relief was experienced by 160 people.

$$\hat{p} = \frac{160}{200} = 0.8$$

$H_0 : p = 0.9$ (claim correct)

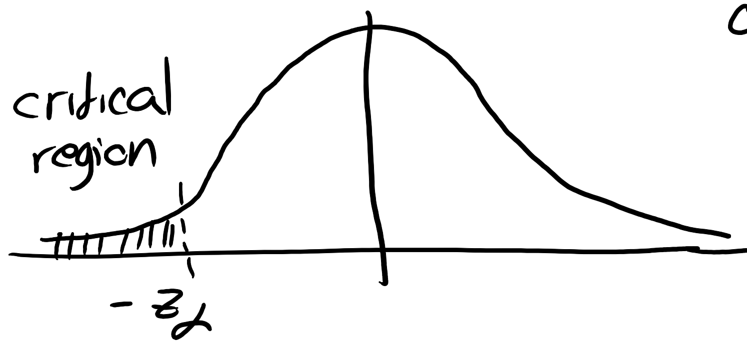
$H_1 : p < 0.9$ (claim is false)

use $\alpha = 0.01$ confidence level

Notice: $H_1 : p \neq 0.9$ does not make sense since we are testing claim that effectiveness is at 90% or higher. If its higher than one would obviously believe $H_0 : p = 0.9$.

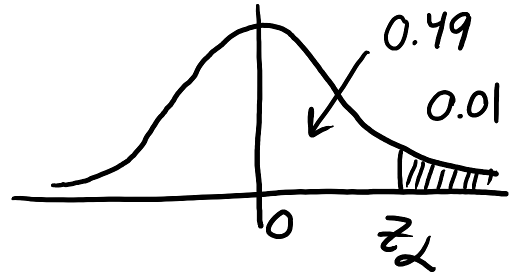
decision rule:

$z < -z_\alpha$



area of shaded region $= \alpha = 0.01$

$\Rightarrow$ same area as this one on right



$P(0 < z < z_\alpha) = 0.49$

use table to deduce that $z_{0.01} = 2.33$

statistic $\quad z = \dfrac{n\hat{p} - np_0}{\sqrt{np_0q_0}} = \dfrac{160 - 180}{\sqrt{200(0.9)(0.1)}}$

$$= \frac{-20}{4.23} \approx -4.73$$

Decision rule: reject H_0 if $z < -z_\alpha = -2.33$.

Hence, since $-4.73 < -2.33$ we can reject H_0 and conclude that at $\alpha = 0.01$ significance level, the manufacturers claim is not legitimate.

(6) $P(|\bar{x} - \mu| < 3) = ?$

$\Rightarrow P(-3 < \bar{x} - \mu < 3) = ?$

by CLT $\quad \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ since $n \geq 30$

$$P\left(\frac{-3}{\sigma/\sqrt{n}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{3}{\sigma/\sqrt{n}}\right) =$$

where we set $\sigma = s$ for this large sample (as specified)

$$= P\left(\frac{-3}{4/\sqrt{50}} < z < \frac{3}{4/\sqrt{50}}\right) = ?$$

$$P(-5.3 < z < 5.3) \approx 1.$$

almost the area under the whole curve

Now what is $P(|\bar{x}-\mu| < 1)$?

$$\Rightarrow P(-1.76 < z < 1.76) = 2P\underbrace{(0 < z < 1.76)}_{\text{use table}}$$

$$= 2 \times 0.4608 \approx 0.92$$

This makes sense, the probability of a tighter estimate is smaller.

(5) (a) confidence interval:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(b) Involves t-distribution (we do not know population std deviations)

$\alpha = 0.05$

$$\left.\begin{array}{l} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{array}\right\}$$ Notice that these are deduced from the problem text.

T statistic

decision rule: reject H_0 if $|T| > t_{\alpha/2, n_1+n_2-2}$

$$T = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ (approximation)

Notice that since $H_1: \mu_1 - \mu_2 \neq 0$, we use a two sided test. Essentially, we reject H_0 if $\bar{X}_1 - \bar{X}_2$ is either far enough to the left or to the right of zero.

For both (a) and (b) we need to find $\bar{X}_1, \bar{X}_2, s_1, s_2$. Notice that both confidence intervals and hypothesis tests use very similar quantities, since the two can be linked mathematically.

$$\bar{X}_1 = \frac{1}{10}\sum_{i=1}^{10} A_i = 4.214 \;; \quad \bar{X}_2 = \frac{1}{10}\sum_{i=1}^{10} B_i = 4.323$$

$$\bar{S}_1 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{10}(A_i - \bar{X}_1)^2}$$

$$\bar{S}_1^2 = \frac{1}{9}\left[(4.21-4.214)^2 + (4.13-4.214)^2 + \ldots + (4.27-4.214)^2\right] = 0.0046 \Rightarrow \bar{S}_1 = 0.068$$

$$\bar{S}_2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{10}(B_i - \bar{X}_2)^2}$$

$$S_2^2 = \frac{1}{9}\left[(4.27-4.323)^2 + (4.38-4.323)^2 + \ldots + (4.41-4.323)^2\right] = 0.0056 \Rightarrow \bar{S}_2 = 0.075$$

$$S = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{0.0046}{10} + \frac{0.0056}{10}} = 0.032$$

For the confidence interval, $\alpha = 0.05$

since $95\% = (1-0.05)\times 100\%$.

We must find:

$$t_{\alpha/2,\, n_1+n_2-2} = t_{0.025,\, 10+10-2} = t_{0.025,18}$$
$$= 2.101 \quad \text{(use t-distribution table)}$$

95% CI:

$$(\bar{X}_1 - \bar{X}_2) - S\times 2.101 < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + S\times 2.101$$

$$\Rightarrow -0.11 - (0.032)(2.101) < \mu_1 - \mu_2 < -0.11 + (0.032)(2.101)$$

95% CI: $-0.18 < \mu_1 - \mu_2 < -0.04$

$\Rightarrow$ This makes sense since $\bar{x}_1 < \bar{x}_2$; hence we suspect that $\mu_1 < \mu_2$. With probability 0.95, $\mu_1 - \mu_2$ lies in the above interval.

For the hypothesis test,

$$T = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{-0.11}{0.032} \simeq -3.44$$

Decision rule: reject H_0 if $|T| > t_{\alpha/2, n_1+n_2-2}$

$\Rightarrow |T| > 2.101$. Since $|-3.44| = 3.44 > 2.101$

we reject H_0 in favor of H_1. We acknowledge that $P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$

$= \alpha = 0.05$ (5%).