HW #10 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 $> Q_3 + 1.5\,IQR$

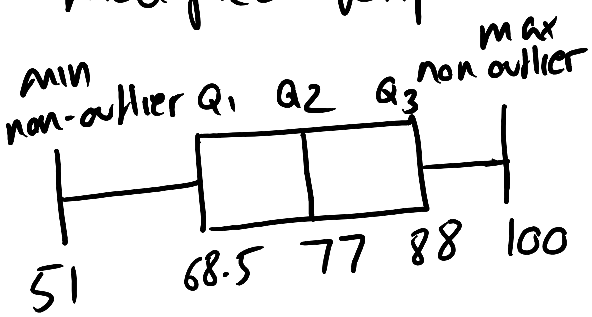
$$= 88 + 1.5(19.5) = 117.25$$

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.

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

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

P(both cards aces without replacement) =

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

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.

$$\hat{p} = \frac{160}{200} = 0.8 \quad \text{(sample proportion)}$$

$H_0 : p = 0.9$ (claim correct)

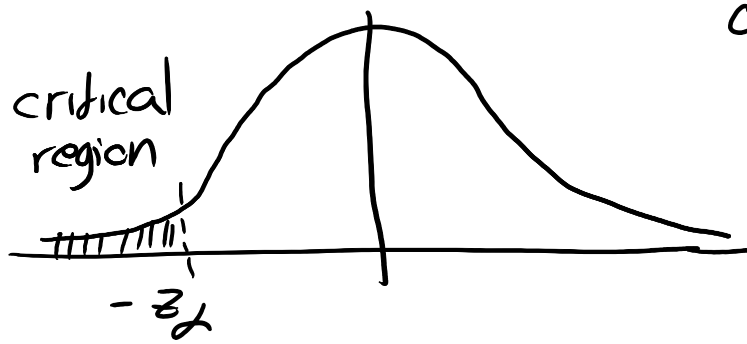
$H_1 : p < 0.9$ (claim is false)

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 $H_0 : p = 0.9$.

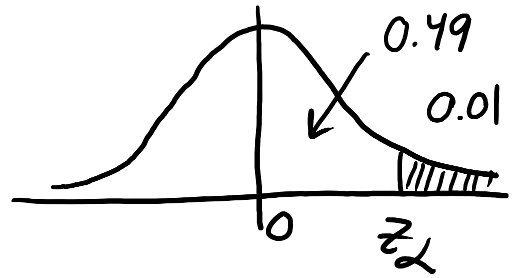
decision rule:

$z < -z_\alpha$



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

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

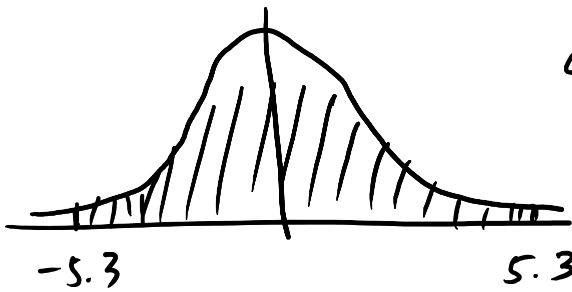
by CLT $\quad \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ since $n \geq 30$

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

where we set $\underline{\sigma = s}$ 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.

(4) p-values

The probability (assuming Ho is true) of obtaining a test statistic that is at least as contradictory to H_0 as one observed.

H_0: $p = 0.9$

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

Let B: binomial r.v. with value equal to # of people experiencing relief out of 200 sampled. Then B is Binomial with parameters $n=200$, $p=0.9$.

$$p\text{-value} = P(B \le 160)$$

$$= \sum_{k=0}^{160} \binom{200}{k} p^k (1-p)^{200-k}$$

$$= \text{pbinom}(160, 200, 0.9) \approx 1.7 \times 10^{-5}$$

The p-value is very small

Since p-value $< \alpha$ (as specified), we reject H_0.

(5) delay time values

sample of 12 contains the measurements:

$\{1.75, 1.89, 2.50, 2.60, 1.10, 0.75, 3.20, 2.30, 1.50, 1.75, 2.15, 2.22\}$

$H_0 : M = 2 = M_0$

$H_1 : M \neq 2$

$S = \max(S_1, S_2)$ where

$S_1 = \#$ measurements $< M_0$

$S_2 = \#$ measurements $> M_0$

$S_1 = 6$, $S_2 = 6 \Rightarrow S = 6$

p-value $= 2P(X \geq 5)$ where

$X \sim \text{Binomial}(n, p)$ with $n=12$, $p=0.5$

X_i measures if value $i = M_0$ or <u>not</u>.

$$P(X \leq 5) = 1 - P(X > 5) = 1 - P(X \geq 6)$$

$$\Rightarrow P(X \geq 6) = 1 - P(X \leq 5)$$

$$= 1 - \sum_{k=0}^{5} \binom{12}{k} (0.5)^k (0.5)^{5-k}$$

$$= 1 - \text{pbinom}(5, 12, 0.5)$$

$$= 1 - 0.38 = 0.613$$

p-value $\to 1$ so essentially, if the true median $M = 2$ then its very possible to get a distribution of values we observe (with half being below the median and half above).

Now, consider instead:

$H_0: M = 1.50 = M_0$

$H_1: M > 1.50$

Let S = # measurements $> M_0$

$= 9$

p-value $= P(B \geq S)$

where $B \sim \text{Binomial}(12, .5)$

$P(B \geq 9) = 1 - P(B < 9)$

$= 1 - P(B \leq 8) \simeq 0.073$

Since p-value is not less than e.g. 0.05 or 0.01, we cannot reject H_0.

(we can reject H_0 if we use $\alpha = 0.1$)

Large scale median test

$H_0: M = M_0$
$H_1: M > M_0$
rejection rule:

$z > z_\alpha$.

$$z = \frac{(S - .5) - .5n}{.5\sqrt{n}}$$

$$\left(\sqrt{npq} = \sqrt{n(.5)(.5)} = .5\sqrt{n}\right)$$

$$z = \frac{(9 - .5) - .5(12)}{.5\sqrt{12}} = \frac{8.5 - 6}{.5\sqrt{2}} = \frac{2.5}{.5\sqrt{12}}$$

$$= \frac{5}{\sqrt{12}} \approx 1.44$$

For above test, choose $\alpha = 0.05$

$z_\alpha = 1.645$

Since $z < z_\alpha$, we cannot reject H_0 (same conclusion as with sign test).

For the original test,

$H_0 : \mu = 2$

$H_1 : \mu \neq 2$

$$z = \frac{(6-.5) - (.5)(12)}{.5\sqrt{12}} \approx -0.29$$

decision rule:

reject H_0 if $|z| > z_{\alpha/2} = 1.96$

since this is not the case, we cannot reject H_0.

Ex) A survey in city 1 found 73% of people considered city tax too high. A random sample of 30 people were asked the same question in city 2. 15 thought the taxes were too high. Is this proportion significantly different from that in city 2? Test at $\alpha = .05$

$H_0: p = .73$ and $H_1: p \neq .73$

$$z_s = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{.50 - .73}{\sqrt{\frac{(.73)(.73)}{30}}}$$

$$\sqrt{\frac{pq}{n}} = \sqrt{\frac{(.73)(.27)}{30}} \approx .081$$

$$z_s \approx -2.84$$

reject H_0 if $|z_s| > z_{\alpha/2} = 1.96$

$\Rightarrow |z_s| = 2.84 > 1.96$ so we reject H_0.

Ex] A box contains 5 red and 4 white balls. Two balls are drawn successively from the box without replacement. If the second ball is white, what is the probability that the first is also white?

W_1: white on 1st draw

W_2: white on 2nd draw

$$P(W_1 | W_2) = \frac{P(W_1 \cap W_2)}{P(W_2)} = \frac{(4/9)(3/8)}{4/9}$$

$$= 3/8$$

since the second is white, there are just 3 out of 8 remaining ways for first to be white.

Ex] Births in a hospital occur randomly at an avg of 1.8 births per hour.

Find probability of observing more than 2 births in 1 hr.

$X \sim \text{Poisson}(1.8)$; $\lambda = E[X]$, the mean rate

$$P(X \geq 2) = 1 - P(X < 2)$$

$$= 1 - [P(X=0) + P(X=1)]$$

$$= 1 - \left(e^{-1.8} \frac{1.8^0}{0!} + e^{-1.8} \frac{1.8^1}{1!}\right)$$

$$\approx 0.537$$