Type I and Type II Errors and Power
(topic test)

1(a) A Type I error is when H_0 is rejected when in fact H_0 is true

(b) A Type II error is when H_0 is not rejected when in fact H_0 is false

(c) The power of a hypothesis test is the probability of rejecting H_0 when H_0 is indeed false

(d) One method of decreasing the risk of a Type I error is to decrease the significance level. The disadvantage of this method is that it will increase the risk of a Type II error.

(e) One method of decreasing the risk of a Type II error is to increase the significance level. The disadvantage of this method is that it will increase the risk of a Type I error

either of these

Another method of decreasing the risk of a Type II error is to increase the sample size. The disadvantage of this method is that it may not be practical or possible to collect more data

(f) Power $= 1 - P(\text{Type II error})$

$= 1 - 0.274$

$= \underline{\underline{0.726}}$

Define dependent variable

2 (a) Let X be the number of first day views since the viral post.

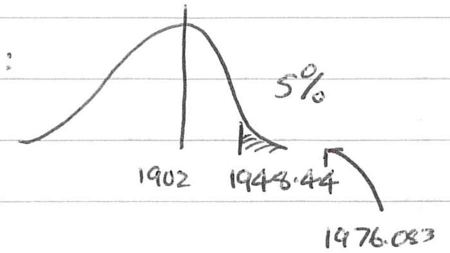
$$H_0: \mu = 1902$$
$$H_1: \mu > 1902$$

Method 1

Use a one-tailed z-test at the 5% level using $\bar{X} \sim N\left(1902, \frac{97.8^2}{12}\right)$

Test statistic: $\bar{x} = 1976.083$

Critical region:



OR p-value
$P(\bar{X} \geq 1976.083)$
$= 0.00434 < 5\%$

(calculator)

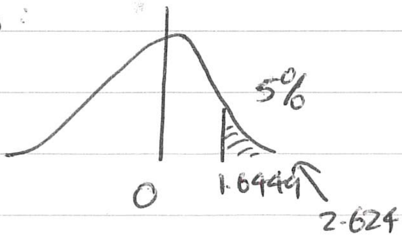
$1976.083 > 1948.44$

Method 2

Use a one-tailed z-test at the 5% level using $Z \sim N(0, 1)$

Test statistic: $\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1976.083 - 1902}{\frac{97.8}{\sqrt{12}}} = 2.624$

Critical region:



OR p-value
$P(Z \geq 2.624)$
$= 0.00434 < 5\%$

(calculator)

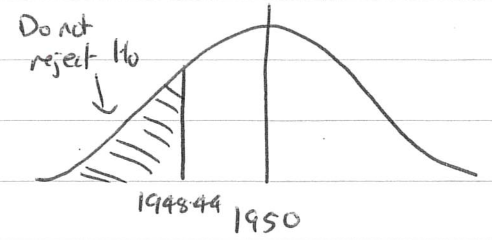
$2.624 > 1.6449$

Result significant.
Reject H_0.
There is significant evidence to suggest the mean number of first day views since the viral post has increased.

conclusion in context, not definite and specific to hypotheses b **Method 1 using $\bar{X}$**

Do not reject H_0 if $\bar{X} \leq 1948.44$

Consider $W \sim N\left(1950, \frac{97.8^2}{12}\right)$ (calculator)



$P(W \leq 1948.44) = 0.478$

so $P(\text{Type II error}) = 0.478$

Hence Power $= 1 - 0.478 = \underline{\underline{0.522}}$

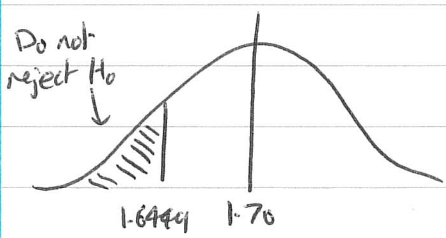
Method 2 using Z

Do not reject H_0 if $Z \leq 1.6449$

Standardised true mean: $\dfrac{1950 - 1902}{\frac{97.8}{\sqrt{12}}} = 1.70$

Consider $W \sim N(1.70, 1)$

(calculator)



$P(W \leq 1.6449) = 0.478$

so $P(\text{Type II error}) = 0.478$

Hence Power $= 1 - 0.478 = \underline{\underline{0.522}}$

(c) If the true mean number of first day views is actually 1950, the probability of correctly concluding the mean number of first day views since the viral post has increased is 0.522.

Must be in context and specific to hypothesis test used

Define dependent variable

3 a) Let X be the number of females appointed.

$$X \sim B(25, 0.48)$$

Lower tail: $\overset{0.0124}{\underset{<2.5\%}{\boxed{0 \quad 6}}}\ 7 \qquad 25$

$P(X \le 6) = 0.0124 < 2.5\%$

$P(X \le 7) = 0.0342 > 2.5\%$

Must show this working

Upper tail: $\overset{0.9868}{\underset{>97.5\%}{0 \quad 17}}\ \overset{0.0132}{\underset{<2.5\%}{\boxed{18 \quad 25}}}$

$P(X \le 16) = 0.965 < 97.5\%$

(calculator) $P(X \le 17) = 0.987 > 97.5\%$

So the critical region is $\underline{X \le 6 \text{ or } X \ge 18}$

(b) $H_0: \pi = 0.48$

$H_1: \pi \neq 0.48$

Use a two-tailed binomial proportion test at the 5% level using $B(25, 0.48)$

Test statistic: $X = 7$

Critical region: $\underset{0.0124}{\boxed{0 \quad 5 \ 6}}\ 7 \quad 17\ \underset{0.0132}{\boxed{18 \ 19 \quad 25}}$

$6 < 7 < 18$

Result not significant.

Do not reject H_0.

There is insufficient evidence to suggest the proportion of females appointed by the company is different to 48%.

conclusion in context, not definite and specific to hypotheses

3 (c) $P(\text{Type I error}) = \text{size of critical region}$

$= 0.0124 + 0.0132 = \underline{0.0256}$

(d) Do not reject H_0 if $7 \leq x \leq 17$

Consider $W \sim B(25, 0.45)$

(calculator)

0 6 $\boxed{7 \; 8 \quad 16 \; 17}$ 18 25

$P(W \leq 6) = 0.0258$ $P(W \leq 17) = 0.9942$

$P(\text{Type II error}) = P(7 \leq W \leq 17) = 0.9942 - 0.0258$

$= \underline{\underline{0.968}}$

(e) Since H_0 was not rejected in part (b), a Type II error could have been made with a 96.8% probability.

Define dependent variables

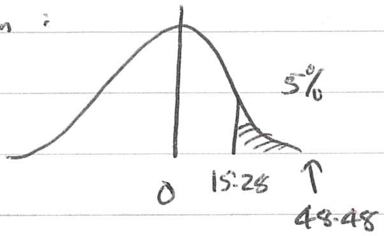
4 (a) Let X be the increase in the number of steps per day for those wearing the gamified device and let Y be that for the standard device.

$$H_0: \mu_x - \mu_y = 0$$
$$H_1: \mu_x - \mu_y > 0$$

Method 1: Use a one-tailed two-sample z-test at the 5% level using
$$\bar{X} - \bar{Y} \sim N\left(0, \frac{4622.25}{143} + \frac{5241.83}{97}\right)$$

Test statistic: $\bar{x} - \bar{y} = 197.73 - 149.25 = 48.48$

Critical region:



OR p-value
$$P(\bar{X} - \bar{Y} \geq 48.48)$$
$$= \underline{0.000000092 < 5\%}$$

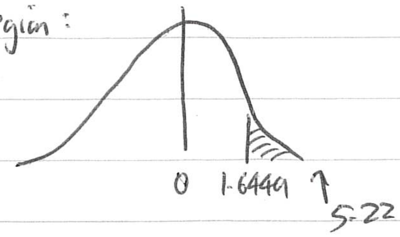
(calculator)

$\underline{48.48 > 15.28}$

Method 2: Use a one-tailed two-sample z-test at the 5% level using $Z \sim N(0,1)$

Test statistic:
$$\frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = \frac{(197.73 - 149.25) - 0}{\sqrt{\frac{4622.25}{143} + \frac{5241.83}{97}}} \approx 5.22$$

Critical region:



OR p-value
$$P(Z \geq 5.22)$$
$$= 0.000000092 < 5\%$$

(calculator)

$5.22 > 1.6449$

conclusion in context, not definite and specific to hypotheses

Result significant. Reject H_0.
There is significant evidence to suggest a higher increase in average number of steps for those wearing the gamified device than the standard device.

4 (b)(i) A Type I error in this case would be to conclude there is significant evidence to suggest a higher mean increase in the number of steps for those wearing the gamified device than those wearing the standard device, when in fact there is no difference.

context and specific to hypotheses

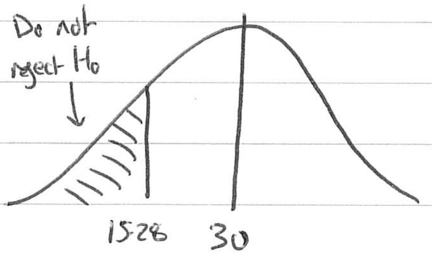
(ii) P(Type I error) = size of critical region = $\underline{0.05}$

(c) A Type II error in this case would be to conclude there is insufficient evidence to suggest a higher mean increase in the number of steps for those wearing the gamified device than those wearing the standard device, when in fact those wearing the gamified device do actually have a higher increase in average number of steps.

(d) <u>Method 1 using $\bar{X} - \bar{Y}$</u>

Do not reject H_0 if $\bar{X} - \bar{Y} \leq 15.28$

Consider $W \sim N\left(30, \frac{4622.25}{143} + \frac{5241.83}{97}\right)$



$P(W \leq 15.28) = \underline{0.0567}$ calculator

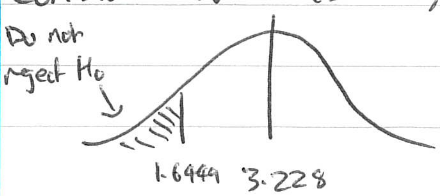
$= P(\text{Type II error})$

<u>Method 2 using Z</u>

Do not reject H_0 if $Z \leq 1.6449$.

standardised difference: $\dfrac{30 - 0}{\sqrt{\dfrac{4622.25}{143} + \dfrac{5241.83}{97}}} = 3.228$

Consider $W \sim N(3.228, 1)$



$P(W \leq 1.6449) = \underline{\underline{0.0567}}$ calculator

Define dependent variable

5(a) Let X be the number of components passing inspection.

$$X \sim B(20000, 0.96)$$

$H_0: \pi = 0.96$

$H_1: \pi < 0.96$

$< 1\%$

$\boxed{0 \quad 19134}\ 19135 \quad 20000$

0.0097

calculator

show this working

$P(X \le 19134) = 0.0097 < 1\%$ Critical region: $\underline{X \le 19134}$

$P(X \le 19135) = 0.011 > 1\%$

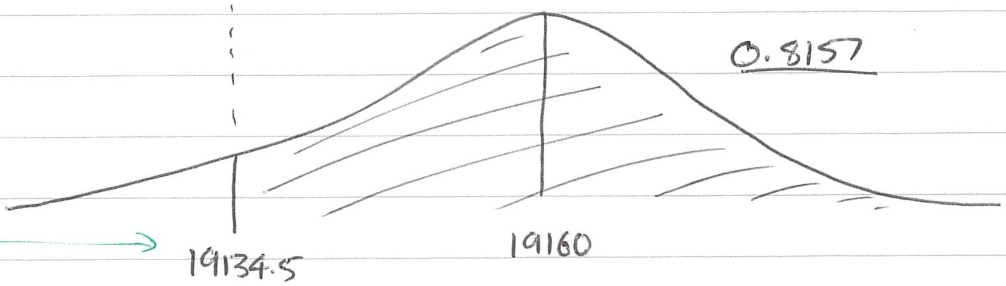
(b) Do not reject H_0 if $X \ge 19135$

Consider $W \sim B(20000, 0.958) \approx N(19160, 804.72)$

Formula book

$$\left[\begin{array}{l} np = 20000 \times 0.958 = 19160 \\ np(1-p) = 20000 \times 0.958 \times 0.042 \\ \quad = 804.72 \end{array}\right]$$

$0 \quad 19134 \ \boxed{19135 \quad 19136 \quad 19160 \quad 20000}$



$\underline{0.8157}$ ← calculate

continuity correction

$P(\text{Type II error}) \approx 0.8157$

so Power $\approx \underline{0.184}$