M2 - HW1 SOLUTIONS

1 (a) $H_0: \mu = 10$ kg

$H_A: \mu \neq 10$ kg.

(b) actually the ques. wants probability of Type 1 error. (α)

$n=25$ sample size for (25 times)

$x_1, x_2 \ldots x_n$: random sample each $x_i \sim N(\mu, (0.2)^2)$

$\mu \rightarrow$ mean unknown

$\sigma \rightarrow$ standard deviation $= 0.2$

$N(\mu, \frac{\sigma^2}{n}) \rightarrow$ normal dist.

x_i's are independent given the question.

$$\bar{x} \sim N\left(\mu, \frac{(0.2)^2}{25}\right) \sim N(\mu, 0.0016)$$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-\mu}{\frac{0.2}{5}} \Rightarrow 25(\bar{x}-\mu) \sim N(0,1)$$

Given events $\bar{x} \geq 10.1032$ && $\bar{x} \leq 9.8968$ are disjoint so the probability of $P(X_1 \cup X_2) = P(X_1) + P(X_2)$

we assume $\mu = 10$ with null hypothesis is true $\bar{x} \sim N(10, 0.0016)$. we reject if $\bar{x} \geq 10.1032$ or $\bar{x} \leq 9.8968$ is true.

$\alpha = P(\bar{x} \geq 10.1032 \;||\; \bar{x} \leq 9.8968 \mid \mu = 10)$

$\alpha = P(\bar{x} \geq 10.1032 \mid \mu=10) + P(\bar{x} \leq 9.8968 \mid \mu = 10)$

$\alpha = P(25(\bar{x}-10) \geq 25(10.1032-10)) + P(25(\bar{x}-10) \leq 25\cdot(9.8968-10))$

$\alpha = P(z \geq 2.58) + P(z \leq -2.58) = 1 - 2 \cdot \underbrace{P(z \leq -2.58)}_{\text{tablodan aynı } 0.005}$

$\alpha = 1 - 2 \cdot (0.005) = \boxed{0.01}$

(c) question wants to probability of failing to reject the null hypothesis when it's actually false. Type 2 error. (β)

* Firstly assume $\mu = 10.1$

$\beta = P(9.8968 \leq \bar{x} \leq 10.1032 \mid \mu = 10.1)$

$\beta = P(25 \cdot (9.8968 - 10.1) \leq 25(\bar{x} - 10.1) \leq 25(10.1032 - 10.1))$

$\beta = P(-5.08 \leq z \leq 0.08)$

$= P(z \leq 0.08) - P(z \leq -5.08)$ → Tabloda değeri yok 0'a yakınsar

$= 0.5319 - 0 = \boxed{0.5319}$

* Secondly assume $\mu = 9.8$

$\beta = P(9.8968 \leq \bar{x} \leq 10.1032 \mid \mu = 9.8)$

$= P(25 \cdot (9.8968 - 9.8) \leq 25(\bar{x} - 9.8) \leq 25 \cdot (10.1032 - 9.8))$

$= P(2.42 \leq z \leq 7.58) = P(z \leq 7.58) - P(z \leq 2.42)$

$1 - 0.9922 = \boxed{0.0078}$

(d) $z \geq c \Rightarrow \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} : \frac{\bar{x}-10}{\frac{0.2}{\sqrt{25}}} \geq c$

$\bar{x} - 10 \geq \frac{c}{25} \Rightarrow \bar{x} \geq 10 + \frac{c}{25}$.

✓ $10 + \frac{c}{25} \geq 10.1032 \Rightarrow \boxed{c = 2.58}$

$z \leq c \Rightarrow \frac{\bar{x}-10}{\frac{0.2}{5}} \leq -c \Rightarrow \bar{x} - 10 \leq -\frac{c}{25}$

$\bar{x} \leq 10 - \frac{c}{25}$ reject null hypothesis when $\bar{x} \leq 9.8968$

$10 - \frac{c}{25} \leq 9.8968 \Rightarrow \frac{c}{25} \geq 0.1032$ $\boxed{c = 2.58}$

(e) given type 1 error $\alpha = 0.05 \Rightarrow$ sample size $n = 10$

$H_0: \mu = 10$

$H_A: \mu \neq 10$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{x}-10}{\frac{0.2}{\sqrt{10}}} = 15.811(\bar{x}-10)$$

$\alpha =$ type 1 error $= 0.05$

$0.05 = P(z \geq c \;||\; z \leq -c \mid \mu = 10) = P(z \geq c) + P(z \leq -c)$

$0.05 = P(z \geq c) + P(z \geq c) = 2P(z \geq c) = 2 \cdot [1 - P(z \leq c)]$

$0.05 = 2[1 - P(z \leq c)]$

$0.025 = 1 - P(z \leq c) \Rightarrow P(z \leq c) = 0.975$

Tabloda 0.975 yapacak c değeri okunur

$\boxed{c = 1.96}$

Reject Region : $\boxed{z \geq 1.96 \;||\; z \leq -1.96}$

(f) $\mu = 10$, $n = 10$) $z = \frac{\bar{x}-10}{\frac{0.2}{\sqrt{10}}}$

$$\bar{x} = \frac{9.981 + 10.006 + 9.857 + 10.107 + 9.888 + 9.728 + 10.439 + 10.214 + 10.190 + 9.793}{10} = 10.0203$$

$z = \frac{10.0203 - 10}{\frac{0.2}{\sqrt{10}}} = 0.321$

c'nin cevabı $z \geq 1.96 \;||\; z \leq -1.96$

Reject region'a düşmüyor.

(g) $\frac{\bar{x}-10}{\sigma/\sqrt{n}}$ $n = 25$

$z \geq c \Rightarrow \frac{\bar{x}-10}{0.2} \cdot 5 \geq c \Rightarrow \bar{x} \geq 10 + c/25$

reject null hypothesis $\bar{x} \geq 10.1032$

$10 + \frac{c}{25} \geq 10.1032$ $\boxed{c \geq 2.58}$

$z \leq c \Rightarrow \frac{\bar{x}-10}{0.2} \cdot 25 \leq -c \Rightarrow \bar{x} \leq 10 - \frac{c}{25}$

reject null hypothesis $\bar{x} \leq 9.8968$

$10 - \frac{c}{25} \geq 9.8968 \Rightarrow c \geq 2.58$

reject region

$z \geq 2.58$ or $z \leq -2.58$

(2) Null hypothesis (ideal die) ⇒ H_0 : Die is fair
equally likely

H_A = Die is not fair

Binomial Dist.

Score	1	2	3	4	5	6	7	8
freq	7	10	11	9	12	10	14	7
expected	10	10	10	10	10	10	10	10

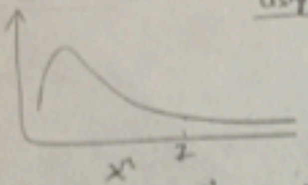
$$X^2 = \frac{(O_i - E_i)^2}{E_i}$$

$$X^2 = \frac{(7-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(7-10)^2}{10}$$

$$X^2 = \frac{9+1+1+4+16+9}{10} = \frac{40}{10} = 4$$

n-1 degree of freedom. $n=8 \Rightarrow X^2_{7,0.05} = 14.07$

using $\alpha = 0.05$ reject Ho



$X^2 = 4 < X^2_{7,0.05} = 14.07$ reject the null.

The die is probably biased. (Ho reject)

(3)

	weight Group1	weight Group2	weight Group3	Totals
bacteria (+)	12	13	31	56
bacteria (-)	35	40	14	89
Totals	47	53	45	145

Cum

Ho: cows weight is independent of mee positive or negative bacteria

H1: cows weight is not independent of - -

$$X_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Reject Ho if : using fixed - significance level test with $\alpha = 0.05$

$r = 2, c = 3 \Rightarrow (r-1)(c-1) = 1 * 2 = 2$ degrees of freedom

reject Ho if $X_0^2 > X^2_{0.05,2}$

$u_1 = \frac{56}{145} = 0.39$; $u_2 = \frac{89}{145} = 0.61$

$v_1 = \frac{47}{145} = 0.32$; $v_2 = \frac{53}{145} = 0.37$; $v_3 = \frac{45}{145} = 0.31$

$E_{11} = n u_1 v_1 = 145(0.39)(0.32) = 18{,}096 \approx 18{,}1$

$E_{12} = n u_1 v_2 = 145(0.39) * (0.37) = 20{,}92 \approx 20{,}92$

$E_{13} = n u_1 v_3 = 145(0.39) * (0.31) = 17{,}530 \approx 17{,}53$

$E_{21} = n u_2 v_1 = 145(0.61)(0.32) = 28{,}3$

$E_{22} = n u_2 v_2 = 145(0.61)(0.37) = 32{,}73$

$E_{23} = n u_2 v_3 = 145(0.61)(0.31) = 27{,}42$

	weight group1	weight group2	weight group3	Total
Bacteria (+)	18.1	20.92	17.53	56.55
Bacteria (-)	28.3	32.73	27.42	88.45
Totals	46.4	53.65	44.95	145

$$X_0^2 = \sum_{i=1}^{2} \sum_{j=1}^{3} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\frac{(12-18.1)^2}{18.1} + \frac{(13-20.92)^2}{20.92} + \frac{(31-17.53)^2}{17.53} + \frac{(35-28.3)^2}{28.3} + \frac{(40-32.73)^2}{32.73} + \frac{(45-27.42)^2}{27.42} \Rightarrow$$

$X_0^2 = 2{,}055 + 2{,}998 + 10{,}35 + 1{,}576 + 1{,}614 + 11{,}27$

$X_0^2 = 29{,}873 \approx 30$

$X^2_{0.05,2} = 5.99$

since 307 5.99 reject the hypothesis of independence!

(4) score-before

$\bar{X}_b = 18.4$ $\sigma_b^2 = 9.93$

after-score

$\bar{X}_a = 2045$ $\sigma_a^2 = 16.47$

paired t-test.

score-before	score-after	difference and fork (after-before)
18	22	4
21	25	4
16	17	1
22	24	2
19	16	-3
24	29	5
17	20	3
21	23	2
23	19	-4
18	20	2
14	15	1
16	15	-1
16	18	2
19	26	7
18	18	0
20	24	4
12	18	6
22	25	3
15	19	4
17	16	-1

$\bar{d} = 2.05$

$S_d = 2.837$

Standard deviation of difference $= \frac{S_d}{\sqrt{n}}$

$SE(\bar{d}) = \frac{2.837}{\sqrt{20}} = 0.634$

$t = \frac{\bar{d}}{SE(\bar{d})} = \frac{2.05}{0.634} = 3.231$

degree of freedom $= n-1 = 19$

$\Rightarrow \boxed{p = 0.004}$

Strong evidence that on average. The certificate program does lead to improvements.

Answer 6 in this "Seç Bakalım" problem we have 3 hypotheses.

H_A = 'The car is behind the door A'
H_B = 'The car is behind the door B'
H_C = 'The car is behind the door C'
Data = D = 'Yazıcıoglu opens door B and reveals a goat'

(a) $P(D|H_A)$ = H_A says the car is behind A. Yazıcıoglu is equally likely to pick B or C and reveal a goat. Thus $P(D|H_A) = 1/2$.
$P(D|H_B)$: H_B says the car is behind B. Yazıcıoglu never choose B so $P(D|H_B) = 0$
$P(D|H_C)$: H_C says the car behind C. Yazıcıoglu doesn't make mistakes he open door B and reveal a goat. So $P(D|H_C) = 1$

H	P(H)	P(D|H)	posterior
H_A	1/3	1/2	1/3
H_B	1/3	0	0
H_C	1/3	1	2/3

equally likely

So Yazıcıoglu should switch, as her chance of winning the car after switching is double that had she stayed with her initial choice.

(b) $P(D|H_A)$: H_A says the car behind A. So Yazıcıoglu is equally likely to show B or C and reveal a goat. So $P(D|H_A) = 1/2$
$P(D|H_B)$: H_B says the car behind B. Yazıcıoglu show B, but if he does we won't reveal a goat so $P(D|H_B) = 0$
$P(D|H_C)$: H_C says the car behind C. Yazıcıoglu is equally likely to B or C so $P(D|H_C) = 1/2$

H	P(H)	P(D|H)	Posterior
H_A	1/3	1/2	1/2
H_B	1/3	0	0
H_C	1/3	1/2	1/2

In this case switching is just good or bad as staying with original choice.

(c) $P(D|H_A) = 1/2$, $P(D|H_B) = 0$
S: be event that Yazıcıoglu is sober $\rightarrow 0.7$
S^c: the event he is drunk $\rightarrow 0.3$
$P(D|H_C, S) = 1$ $\quad P(D|H_C, S^c) = 1/2$
$P(D|H_C) = P(D|H_C,S)P(S) + P(D|H_C,S^c)P(S^c)$
$\quad = 1 \cdot 0.7 + \frac{1}{2} \cdot 0.3 = 17/20$

H	P(H)	P(D|H)	Posterior
H_A	1/3	1/2	10/27
H_B	1/3	0	0
H_C	1/3	17/20	17/27

5 n=200 $\quad \bar{X} = 33{,}01$
σ^2 = varyans = 8,63) genel data

7

predictor	Coeff	SE(coeff)	t	p-value
intercept	-6.72	14.00	-0.48	0.636
Test1	0.256	0.2274	1.13	0.274
Test2	0.3912	0.2198	1.78	0.091
Test3	0.9015	0.2086	4.32	< 0.0001

Analysis of Variance

Source	DF	SS	MS	F	p-value
Regression	3	11961.8	3987.3	12.60	< 0.0001
Error	19	3440.8	181.3		
Total	22	15402.6			

(a) intercept represents the constant term in the model.

$\hat{\beta}_0 = -6.72$, $\hat{\beta}_1 = 0.256$, $\hat{\beta}_2 = 0.3912$, $\hat{\beta}_3 = 0.9015$

$y = -6.72 + 0.256\ Test1 + 0.3912\ Test2 + 0.9015\ Test3$

(b) Ho: coefficient no effect.
in first table p-value column (a low p-value) indicates that we can reject null hypothesis.
* a predictor that has a low p-value is likely to be a meaningful addition our model because changes in the predictor's value are related to changes in the response variable.
changes in the response variable. Larger p-value (insignificant) suggests that changes in the predictor are not associated with changes in the response.
In the first table output Test3 are significant because p-values of Test3 are smaller than 0.0001 when look second table regression model p-value is smaller than 0.0001 so model is statistically significant.

(c) yes if we increase Test1 score our model result increase. (positively correlated)

$y = -6.72 + 0.256\ Test1 + \ldots$

(d) $R^2 = \frac{SSR}{SST} = \frac{11961.8}{15402.6} = 0.7769 \approx \%77.7$

(e) How much of the variation in final exam scores is accounted for by the regression model?
The coefficient of determination measures the proportion of variation in the dependent variable that is explained by the least-squares regression model. $R^2 = \%77.7$ of the variation in final exam scores is explained by the regression model.

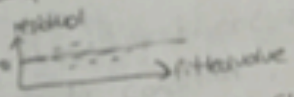
(f) Explain in the context what the coefficient Test3 scores means.

$\hat{\beta}_3 = 0.9015$

The coefficient represents the average increase in the dependent variable per unit of the independent variable.
The final score increases on average by 0.9015 per point on test3.

(g) The p-value Test1 equals 0.274.
Ho: Test1 coefficient has no effect for model.
null hypothesis.
If we select significance level 0.05
our p-value = 0.274 > 0.05 so null hypothesis is rejected.
* There is no sufficient evidence to reject the claim that Test1 has no effect on the final exam score. Actually Test1 scores have some effect on the final score.

(h) First plot:



* Not satisfied. Because there is curvature present in the scatterplot.
As the datapoints lie below zero to the left and right of the graph but lie above 0 in the middle.

Second plot: Satisfied. Because when points increase score is increased positivel correlated.

Third (Histogram Plot): Satisfied. Data pattern in the normal probability plot doesn't contain strong curvature and histogram of the residuals is approximately symmetric. and highest bars are approximately in the middle of the histogram so plot is satisfied.

5.soru için python da veri kümesi plot edilmiştir ve normal dağılım olup olmadığı kontrol edilmiştir. Ödeve .ipynb dosyası da eklenmiştir.