Midterm 3 Solutions

Q1 a) $n = 1200$ $\quad \hat{p} = \frac{630}{1200} = 0.525$

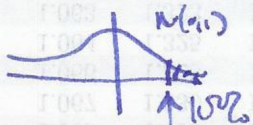
confidence interval: $\hat{p} \pm z_* \sqrt{\hat{p}(1-\hat{p})/n}$ $\qquad z_* = 1.28$

$(0.507, 0.543)$

b) $H_0: p = 0.5$ $\quad$ test statistic: $\dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$

$H_a: p > 0.5$

$$\frac{0.525 - 0.5}{\sqrt{0.5(1-0.5)/1200}} = 1.732$$

critical value 

$z_* = 1.28$ $\qquad 1.732 > 1.28 = z^*$

reject H_0:

conclusion: significant evidence $p > 0.5$

Smith will get more than half the votes

Q2 $H_0: p_1 = p_2$ $\quad$ use difference between two proportions

$H_a: p_1 \neq p_2$ $\quad \hat{p}$ = pooled estimate = $\dfrac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2} = \dfrac{630 + 475}{1200 + 1000} \approx 0.502$

$n_1 = 1200$ $\quad \hat{p}_1 = 0.525$

$n_2 = 1000$ $\quad \hat{p}_2 = 0.475$

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.525 - 0.475}{\sqrt{0.502(1-0.502)\left(\frac{1}{1200} + \frac{1}{1000}\right)}} \simeq 2.336$$

significance level 10% 2-sided $z_* = 1.64 < 2.336$

reject H_0: significant evidence populations are different.

Q3 a) H_0: no correlation between pizza size and topping ②

H_a: some correlation.

b) $\frac{\text{row total} \times \text{col total}}{n} = \frac{33 \times 31}{100} \approx 10$

c) $\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ large veggie: $\frac{(9-10)^2}{10} = \frac{1}{10}$

d) $df = (\#\text{rows} - 1)(\#\text{cols} - 1) = 2 \times 2 = 4$

e) P-value < 0.001

f) reject H_0: some association between rows and cols.

Q4. a) 0

b) 0.5

c) -0.9

Q5 a) F b) T c) F d) T

Q6 a) F b) T c) F d) T

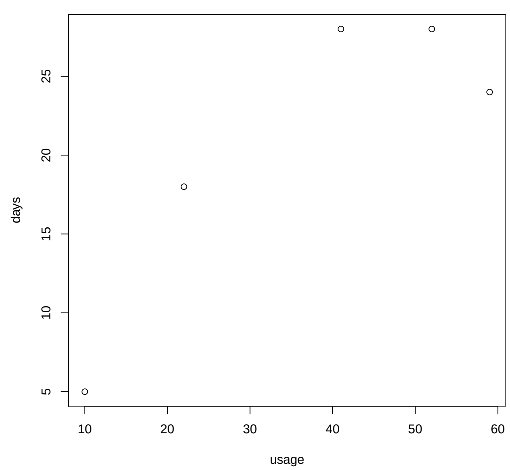
Midterm 3 Lab Solutions

```
> usage<-c(41, 10, 59, 52, 22)
> days<-c(28, 5, 24, 28, 18)
```

(a)
```
> mean(days)
[1] 20.6
```

(b)
```
> sd(usage)
[1] 20.48658
```

(c)
```
> plot(usage, days)
```



(d)
```
> cor(usage, days)
[1] 0.8558274
```

(e)
```
> lm(usage~days)

Call:
lm(formula = usage ~ days)

Coefficients:
(Intercept)         days
    -0.6929       1.8200
```

The least squares regression line is: $y = 1.82x - 0.6929$, where x is days and y is usage.

(f)
```
> summary(lm(usage~days))

Call:
lm(formula = usage ~ days)

Residuals:
      1       2       3       4       5 
 -9.268   1.593  16.012   1.732 -10.068 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.6929    14.1815  -0.049   0.9641
days          1.8200     0.6351   2.866   0.0643 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 12.24 on 3 degrees of freedom
Multiple R-squared: 0.7324,Adjusted R-squared: 0.6433
F-statistic: 8.212 on 1 and 3 DF,  p-value: 0.06427
```

$1.82 \pm t_* \times SE = 1.81 \pm t_* \times 0.6351$

Find t_* :

```
> qt(0.95, df=3)
[1] 2.353363
```

Final answer: $(0.3253792, 3.314621)$

(g) $1.82 \times 25 - 0.6929$

```
> 1.82*25 - 0.6929
[1] 44.8071
```

Predict 45G of internet usage.