# HW6

Andrew Shao

2024-11-06

**(a)**

**1.** The model is:

$$brozek = -46.216 + 0.646 \cdot chest$$

Since the p-value is extremely small ($7.373 \cdot 10^{-39}$), we conclude that these two variables are associated.

**2.** Pearson's correlation coefficient is 0.703 between the two variables.
The calculated p-value for $T_{obs}$ is extremely small and close to 0, so we conclude that the two variables are significantly correlated at the $\alpha = 0.05$ level. The value is the same as the value for the slope.
The 95% confidence interval is [0.6344, 0.7604]

**3.**

**(b)**

**(c)**

**(d)**

**(e)**

**(f)**

**(g)**

## Appendix

**(a)**

```r
library(faraway)
```

```
## Warning: package 'faraway' was built under R version 4.3.3
```

```r
head(fat)
```

```
##   brozek siri density age weight height adipos  free neck chest abdom   hip
## 1   12.6 12.3  1.0708  23 154.25  67.75   23.7 134.9 36.2  93.1  85.2  94.5
## 2    6.9  6.1  1.0853  22 173.25  72.25   23.4 161.3 38.5  93.6  83.0  98.7
## 3   24.6 25.3  1.0414  22 154.00  66.25   24.7 116.0 34.0  95.8  87.9  99.2
## 4   10.9 10.4  1.0751  26 184.75  72.25   24.9 164.7 37.4 101.8  86.4 101.2
## 5   27.8 28.7  1.0340  24 184.25  71.25   25.6 133.1 34.4  97.3 100.0 101.9
## 6   20.6 20.9  1.0502  24 210.25  74.75   26.5 167.0 39.0 104.5  94.4 107.8
##   thigh knee ankle biceps forearm wrist
## 1  59.0 37.3  21.9   32.0    27.4  17.1
## 2  58.7 37.3  23.4   30.5    28.9  18.2
## 3  59.6 38.9  24.0   28.8    25.2  16.6
## 4  60.1 37.3  22.8   32.4    29.4  18.2
## 5  63.2 42.2  24.0   32.2    27.7  17.7
## 6  66.0 42.0  25.6   35.7    30.6  18.8
```

```r
## (a)(1) simple linear regression
lma <- lm(brozek ~ chest, data=fat);
summary(lma)
```

```
##
## Call:
## lm(formula = brozek ~ chest, data = fat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.8875  -3.8211  -0.2752   3.4950  13.8989
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -46.21636    4.18460  -11.04   <2e-16 ***
## chest         0.64622    0.04136   15.62   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.524 on 250 degrees of freedom
## Multiple R-squared:  0.494,  Adjusted R-squared:  0.492
## F-statistic: 244.1 on 1 and 250 DF,  p-value: < 2.2e-16
```

```r
## Extract those information for the slope
summary(lma)$coefficients[2,]
```

```
##     Estimate   Std. Error      t value     Pr(>|t|)
## 6.462223e-01 4.136018e-02 1.562426e+01 7.372549e-39
```

```r
##  or the t-statistics and p-value for the slope
##  t-stat=1.562426e+01= 15.62426, p-value= 7.372549e-39
summary(lma)$coefficients[2,3:4]
```

```
##      t value     Pr(>|t|)
## 1.562426e+01 7.372549e-39
```

```r
## (a)(2) Pearson's correlation
# (i) Pearson's correlation
r1 = cor(fat$brozek, fat$chest);
r1
```

```
## [1] 0.7028852
```

```r
# (ii) hypothesis testing via Pearson's correlation
n= dim(fat)[1];
t.obs1 = r1* sqrt((n-2)/ (1-r1^2) );
t.obs1  ### compare with (i)
```

```
## [1] 15.62426
```

```r
# p-value
pvalue1 = 2*(1-pt( abs(t.obs1), df= n-2 ));
pvalue1
```

```
## [1] 0
```

```r
# (iii) 95% CI on Pearson's correlation
alpha = 0.05;
cutoffvalue = qnorm(1- alpha/2);
Zr1 = 0.5*log((1+r1)/(1-r1));
ZCI = Zr1 + c(-1, 1)* cutoffvalue /sqrt(n-3);
rho1.CI = (exp(2*ZCI) -1) / (exp(2*ZCI) +1);
rho1.CI
```

```
## [1] 0.6344161 0.7604106
```

```r
### (a)(3) Spearman's Correlation
## (i) point estimate
rs1= cor(fat$brozek, fat$chest, method= "spearman");
rs1
```

```
## [1] 0.6730803
```

```r
## (ii) hypothesis testing
n= dim(fat)[1];
t.obs2 = rs1* sqrt((n-2)/ (1-rs1^2) );
t.obs2
```

```
## [1] 14.38991
```

```r
# p-value based on Spearman's correlation
pvalue2 = 2*(1-pt( abs(t.obs2), df= n-2 ));
pvalue2
```

```
## [1] 0
```

**(b)**

**(c)**

**(d)**

**(e)**

**(f)**

**(g)**