MATH 212

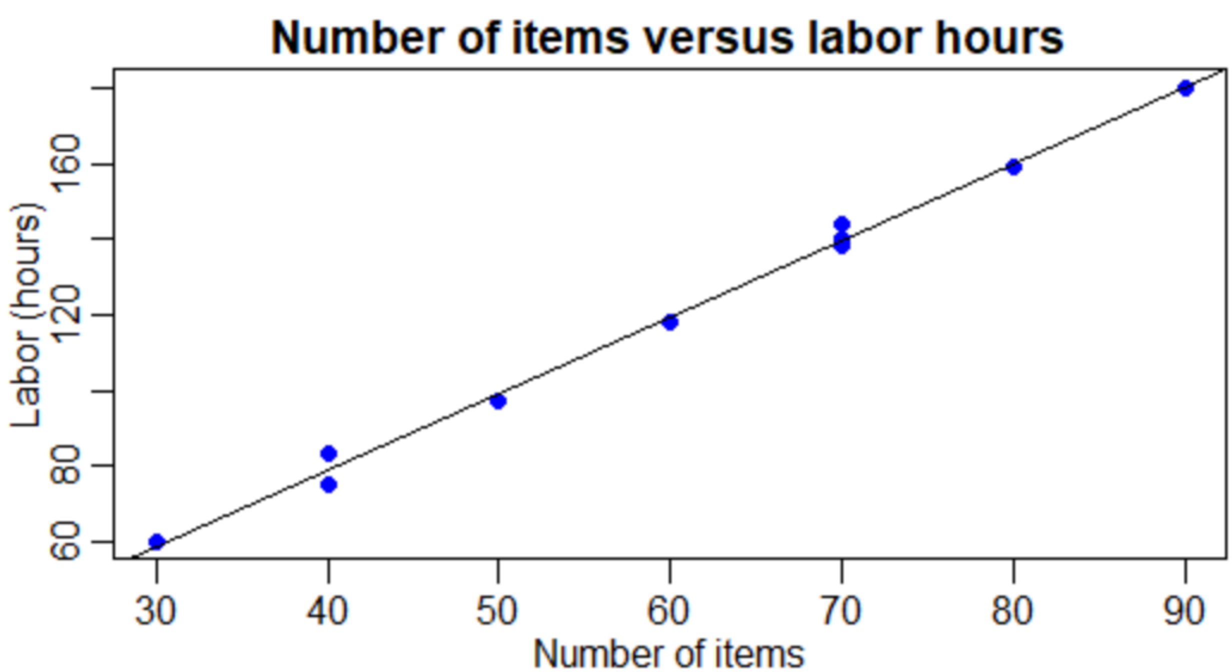
Test I

Name: **Zack Roder**

The Central Company manufactures a certain specialty item once a month in a batch production run. The number of items produced in each run varies from month to month as demand fluctuates. The company is interested in the (x) relationship between the size of the production run (x) and the number of hours of labor (y) required for the run. The company has collected the following data for the ten most recent runs:

Number of items (x)	40	30	70	90	50	60	70	40	80	70
Labor (hours) (y)	83	60	138	180	97	118	140	75	159	144

(1) Construct a scatterplot of y versus x and graph the regression line on the scatter plot. Past your final plot below. Make sure to label the x and y axis using the variable names and add an appropriate tile. (10 points)



(2) Find the least squares regression line relating the hours of labor to number of items produced using R (using the "lm" function). Paste your R output (using the "summary" function in R) below. Make sure that you get a slope of 2.02 and a y-intercept of -1.84 (rounded to two decimals). (8 points)

```
> summary(Model)

Call:
lm(formula = labor ~ items)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.9882 -1.5544 -0.4147  1.0118  4.3941 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.83529    3.01990  -0.608     0.56    
items        2.02059    0.04811  41.999 1.14e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.805 on 8 degrees of freedom
Multiple R-squared:  0.9955,	Adjusted R-squared:  0.9949 
F-statistic:  1764 on 1 and 8 DF,  p-value: 1.138e-10
```

$$\text{LABOR} = 2.02 \cdot \text{ITEMS} - 1.84$$

(3) Tests the hypothesis $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$ at the 5% level of significance. Please show all your work for full credit.

Step 1: State the null and Alternative hypothesis using correct notation. (4 points)

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

Step 2: Calculate the test statistic. (4 points)

$$\begin{aligned} t &= \frac{b_1 - \beta_1^*}{s_{b_1}} \\ &= \frac{2.02 - 0}{0.048} \\ &= 42.08 \end{aligned}$$

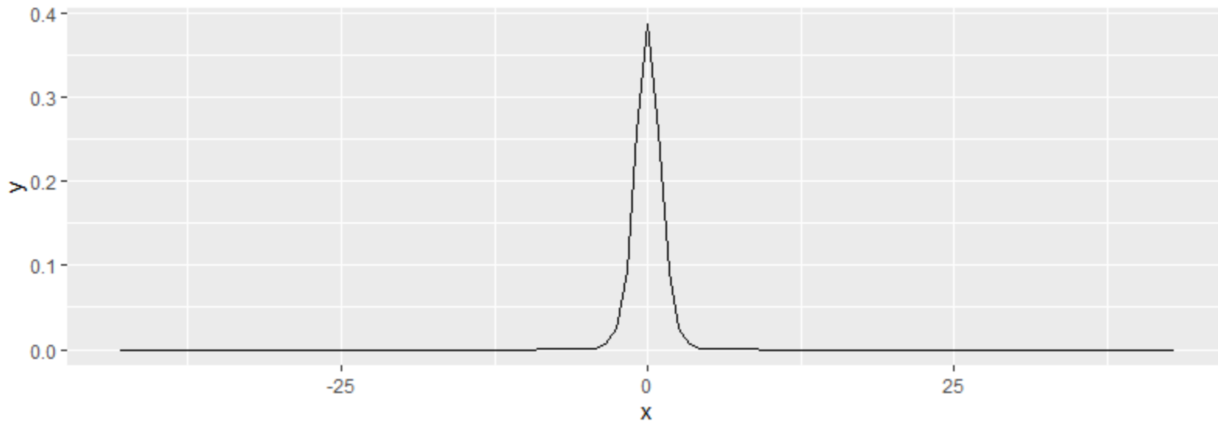
$t = \mathbf{42.08}$

Step 3: Calculate the p-value. Clearly draw a diagram and show complete work. (4 points)

Calculated in R:

```
> 2*(1 - pt(42.08,8))
[1] 1.120894e-10
```

Plotting the t-distribution curve, we clearly see that the p-value is approximately zero:



P-value = **1.12 * 10^-10**

Step 4: Make a decision (4 points)

(a) Reject H_0

(b) Do not reject H_0

Step 5: Write the conclusion is context of the problem (4 points)

Since our p-value is less than 0.05, we reject the null hypothesis $\beta_1 = 0$. Hence, there is sufficient evidence to conclude that there is a linear relationship between labor hours and number of items produced.

(4) Construct a 95% confidence interval for β_1. Please show all your work for full credit.

Step 1: Write the equation (2 points)

$$[b_1 - t_{\alpha/2,n-2} \cdot s_{b_1}, b_1 + t_{\alpha/2,n-2} \cdot s_{b_1}]$$

Step 2: Identify the following values (use R to compute these values)

$\bar{x} = 60$ (2 points)

$s = 2.805$ (2 points)

$n = 10$ (2 points)

$t_{\frac{\alpha}{2},n-2} = 2.306$ (2 points)

Step 3: Substitute to find the upper bound and the lower bound of the confidence interval. Show work

$$[b_1 - t_{\alpha/2,n-2} \cdot s_{b_1}, b_1 + t_{\alpha/2,n-2} \cdot s_{b_1}]$$
$$[2.02 - 2.306 \cdot 0.048, 2.02 + 2.306 \cdot 0.048]$$
$$[1.91, 2.31]$$

Lower bound: $\mathbf{1.91}$ (2 points)

Upper bound: $\mathbf{2.13}$ (2 points)

Step 4: Interpret the confidence interval in context of the problem (6 points):

We are 95% confident that the number of labor hours required to produce each additional item is between 1.909 and 2.131, inclusive.

(5) Tests the hypothesis $H_0: \beta_0 = 0$ vs. $H_a: \beta_0 \neq 0$ at the 5% level of significance. Please show all your work for full credit.

Step 1: State the null and Alternative hypothesis using correct notation. (4 points)

$$H_0 : \beta_0 = 0$$
$$H_a : \beta_0 \neq 0$$

Step 2: Calculate the test statistic. (4 points)

$$\begin{aligned} t &= \frac{b_0 - \beta_0^*}{s_{b_0}} \\ &= \frac{-1.835}{3.020} \\ &= -0.608 \end{aligned}$$

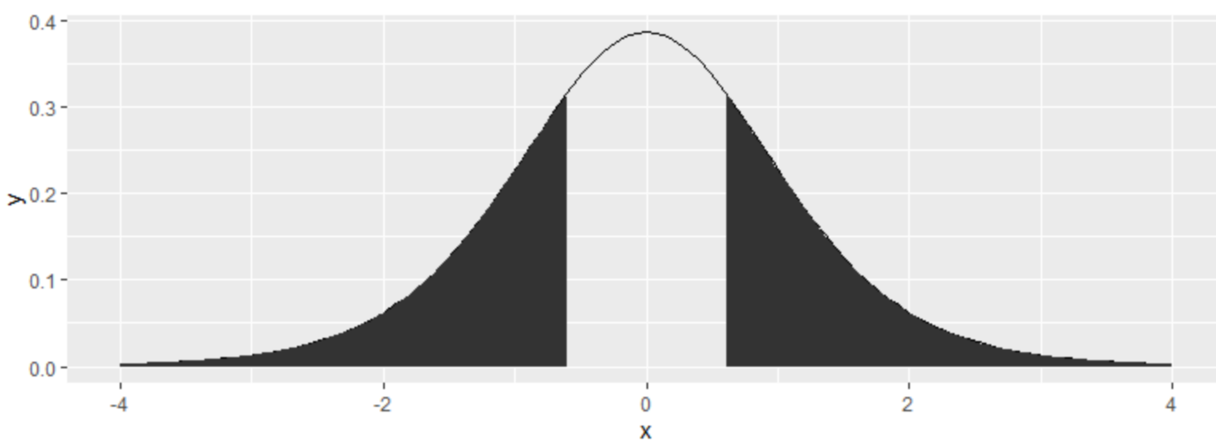
$t = \mathbf{-0.608}$

Step 3: Calculate the p-value. Clearly draw a diagram and show complete work. (4 points)

Calculated in R:

```
> 2*(1 - pt(0.608, 8))
[1] 0.5600478
```

t-distribution plotted in R:



P-value = 0.56

Step 4: Make a decision (4 points)

(a) Reject H_0

(b) Do not reject H_0

Step 5: Write the conclusion is context of the problem (4 points)

Since our p-value 0.56 > 0.05, we fail to reject the null hypothesis. Thus, we do not have sufficient evidence to conclude that the y-intercept is different from 0.

(6) Compute the coefficient of determination R^2 for the regression of the number of labor hours on number of items produced.

Step 1: Write the equation (2 points)

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}}$$

Step 2: Identify the following values (use the "anova" function in R)

$SSE = \mathbf{62.9588}$ (3 points)

$SST = \mathbf{13881.4412}$ (3 points)

```
> anova(Model)
Analysis of Variance Table

Response: labor
          Df Sum Sq Mean Sq F value    Pr(>F)    
items      1  13881 13881.4  1763.9 1.138e-10 ***
Residuals  8     63     7.9                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step 3: Substitute to find R^2. Show work

$$R^2 = 1 - \frac{62.9588}{13881.4412} = 0.9955$$

R^2 = 0.9955 (4 points)

Step 4: Interpret the R^2 value in context of this problem (6 points):

Using the R^2 value, we can show that the model that accounts for 99.55% of the variation in labor hours.

(7) A random sample of 200 men aged between 20 and 60 was selected from a certain city. The linear correlation coefficient between income and blood pressure was found to be r = 0.807. What does this imply? Does this suggest that if a man gets a salary raise his blood pressure is likely to rise? Why or why not? Explain. (4 points)

The correlation coefficient of 0.807 does suggest a direct relationship between income and blood pressure, but we need more information to conclude that there is *statistically significant* relationship between the two variables. Although this model has a fairly well-fitted regression line, it would be necessary to perform a hypothesis test on the slope of the model before concluding that the relationship between income and blood pressure among men is significant.