Problem 1

a) Distribution of responses:
$y_i \sim Bernoulli(p_i), 0 \leq p_i \leq 1$; where $p_i = P(y_i = 1)$

Linear Predictor:
$\eta_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 = \boldsymbol{x_i^T \beta}$; where $x_{i1}$is the person's height and x_{i2} is the number of cigarettes smoked per day

Link Function
$$p_i = \frac{\exp{(\eta_i)}}{1 + \exp{(\eta_i)}} = \frac{\exp{(x_i^T \beta)}}{1 + \exp{(x_i^T \beta)}}$$

b) $\log\left[\frac{p}{1-p}\right] = x^T\beta = -4.50161 + 0.02521x_1 + 0.02313x_2$

c) The log-odds of getting CHD are expected to increase by 0.02521 for each additional inch of height a person has (if number of cigarettes is constant).

The odds of getting CHD are expected to increase by 2.55% for each additional inch of height (if number of cigarettes is constant).

d) The log-odds of getting CHD are expected to increase by 0.2313 for each additional 10 cigarettes smoked (if height stays constant).

The odds of getting CHD are expected to increase by 26% for each additional 10 cigarettes smoked (if height stays constant).

e) The LRT returned a p-value of 0.3374, which is greater than any reasonable significance level. This means that we will fail to reject the null hypothesis that $\beta_1 = 0$.

The 95% confidence interval for β_1 is (-0.02619902, 0.07702835). Because 0 is within our interval, that means that 0 is a possible value for β_1.

f) and (g) Below is a table for the estimated log-odds, odds, and probability for a person with 70 inches in height and 12 cigarettes smoked, estimated at 95% confidence.

Log – Odds	**Estimate**	**Lower**	**Upper**
	-2.45954	-2.591835	-2.327245
Odds	**Estimate**	**Lower**	**Upper**
	0.08547426	0.07488251	0.09756415
Probability	**Estimate**	**Lower**	**Upper**
	0.0787437	0.6966576	0.8889152

Problem 2

a) <u>Distribution of responses:</u>
$y_i \sim Poisson(\lambda_i), \lambda_i > 0,$
Where y is the number of discoveries made within the year.

<u>Linear Predictor:</u>
$\log(\lambda_i) = \eta_i = \beta_0 = \boldsymbol{\beta}$

<u>Link Function</u>
$\lambda = e^{\beta_0}$

$\beta_0 = 1.131,$
$e^{\beta_0} = 3.098754$

```
> disc_glm

Call:  glm(formula = discoveries ~ 1, family = poisson(), data = disc_df)

Coefficients:
(Intercept)
      1.131

Degrees of Freedom: 99 Total (i.e. Null);  99 Residual
Null Deviance:      164.7
Residual Deviance: 164.7        AIC: 435.7
> exp(1.131)
[1] 3.098754
```

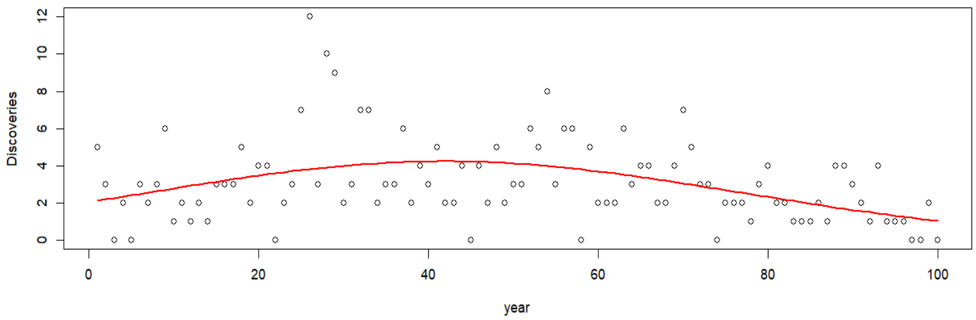
b) Mean(discoveries) = 3.1
The average discoveries per year is approximately equal to the value of λ determined by our model. The mean (expected value) of a Poisson distribution is λ.

c) Based on the model in part (a), we would predict 3.1 discoveries made in the next year.

d) Used the model: $\log(\lambda_i) = \beta_0 + x_i\beta_1 + x_i^2\beta_2$

Using the LRT, testing the null hypothesis ($\beta_2 = 0$) versus the alternative ($\beta_2 \neq 0$) determined a p-value approximately 0, leading us to reject the null hypothesis in favor of the alternative.

e) The shape of the graph does seem to have some curvature to it, but I do not believe that I would say that this is a good fit.



f) I did not have time to get to parts (f) and (g) before I needed to wrap up the homework paper and submit it. My apologies. (I have been sick with the stomach bug for over half the week and lost time!)

Appendix (R Code)

```
library(faraway)

library(lmtest)


#Getting Data

wcgs_small <- data.frame(height = wcgs$height,
                         cigs = wcgs$cigs,
                         chd = ifelse(wcgs$chd == "yes", 1, 0))


#Plot

plot(wcgs_small$cigs, wcgs_small$height,
     col = wcgs_small$chd+1, pch = wcgs_small$chd+10,
     xlab = "cigs", ylab = "height")


#Linear Model

out <- lm(chd ~ ., data = wcgs_small)

head(out$fitted.values)


#GLM Full

wcgs_glm <- glm(chd ~ ., data = wcgs_small, family = binomial())

wcgs_glm


#Reduced Model

wcgs_red <- glm(chd ~ cigs, data = wcgs_small, family = binomial())


#LRT

lrtest(wcgs_red, wcgs_glm)


#confidence interval

confint(wcgs_glm)
```

```
#log odds
alpha = 0.05
zcrit <- qnorm(alpha/2, lower.tail = FALSE)
newx <- data.frame(height = 70, cigs = 12)
lds <- predict(wcgs_glm, newdata = newx,
        type = "link", se.fit = TRUE)
logodds <- data.frame(estimate = lds$fit,
              Lower = lds$fit - zcrit*lds$se.fit,
              Upper = lds$fit + zcrit*lds$se.fit)
logodds

## Odds
odds <- exp(logodds)
odds

## P(y = 1)
prob <- odds/(1 + odds)
prob

#PROBLEM 2
#making the data numeric and putting it in a dataframe
disc_df <- data.frame(year = as.numeric(time(discoveries)-1859),
              discoveries = as.numeric(discoveries))

#poisson with no predictors but the intercept
disc_glm <- glm(discoveries ~ 1, family = poisson(), data = disc_df)
disc_glm

#trying to make the quadratic poisson
```

```
disc_glm_quad <- glm(discoveries ~ year + I(year^2), family = poisson, data = disc_df)
disc_glm_line <- glm(discoveries ~ year, family = poisson, data = disc_df)

#Checking the fitted values
disc_glm_quad$fitted.values

#Graphing the data with the fitted curve
plot(disc_df$year, disc_df$discoveries,
    xlab = "year", ylab = "Discoveries")
lines(disc_df$year, disc_glm_quad$fitted.values, col = "red", lwd = 2)
```