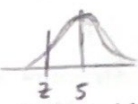
STAT 516 sp 2026 exam 01

75 minutes, one page of handwritten notes allowed, no calculat

1. Inference on the mean of a normal population

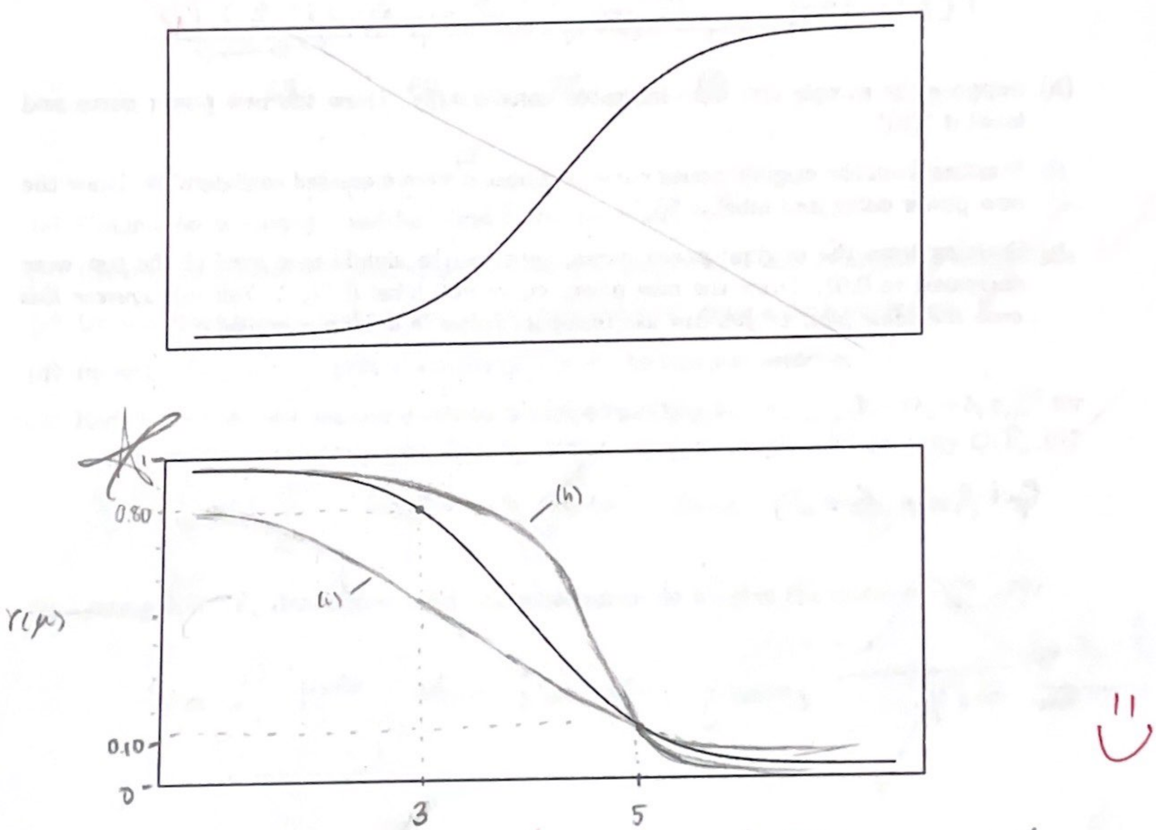
Suppose $X_1, \dots, X_n$ is a random sample from the $\mathcal{N}(\mu, \sigma^2)$ distribution, where σ^2 is known, and that we plan to test H_0: $\mu \geq 5$ versus H_1: $\mu < 5$ with a test that rejects H_0 when

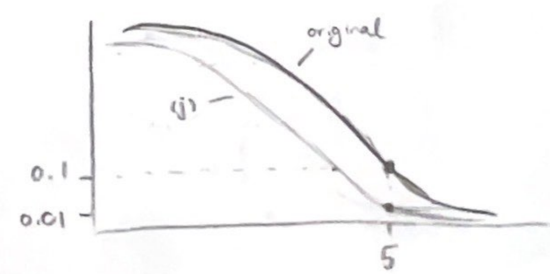
$$Z_{\text{test}} = \frac{\bar{X} - 5}{\sigma/\sqrt{n}} < -z_{0.10},$$



where $z_{0.10}$ is the value such that $P(Z > z_{0.10}) = 0.10$, $Z \sim \mathcal{N}(0, 1)$. Suppose this test would have power 0.80 if μ were equal to 3.

(a) Indicate the panel in which the curve could be the power curve for this test.





(b) Put a tick mark at $y = 0.10$ on the vertical axis and at $x = 5$ on the horizontal axis.

(c) Draw a horizontal line from the $y = 0.10$ tick mark and a vertical line from the $x = 5$ tick mark.

(d) What is the maximum height of the power curve?

1

(e) What is the minimum height of the power curve?

0

(f) Draw a horizontal line and having height $y = 0.80$ and, where this crosses the power curve, drop a vertical line down to the horizontal axis; label the position where this vertical line touches the horizontal axis.

(g) What is the probability that this test would make a Type II error if μ were equal to 3?

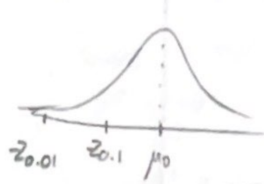
$\gamma(3) = 0.80$ $\quad$ $P(\text{Type II}) = 1 - 0.80 = \boxed{0.20}$

(h) Suppose the sample size were increased considerably. Draw the new power curve and label it "(h)".

(i) Starting from the original power curve, suppose σ were increased considerably. Draw the new power curve and label it "(i)".

(j) Starting from the original power curve, suppose the significance level of the test were decreased to 0.01. Draw the new power curve and label it "(j)". *You can squeeze this onto the same plot, or you can use the space below to draw a new plot.*

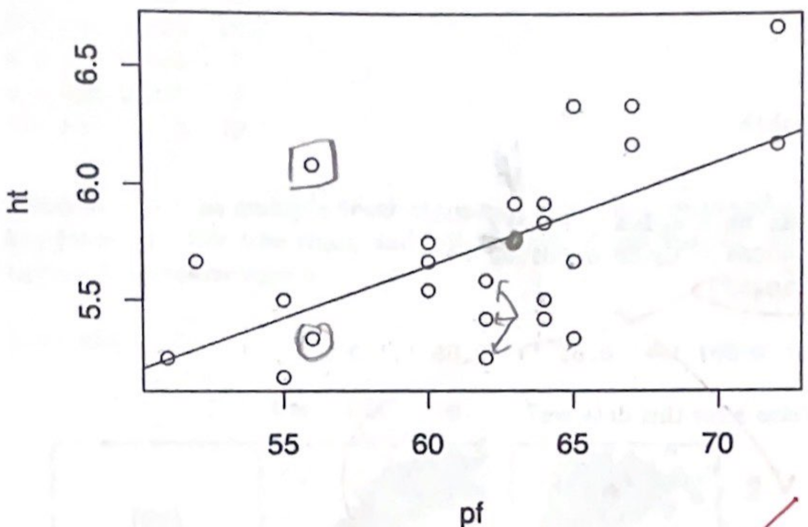
$P_\mu(z_{test} < z_\alpha)$



2. Simple linear regression

The scatterplot shows the heights (in feet) and pinky finger lengths (in millimeters) of several students with least-squares line overlaid.

```
plot(ht ~ pf)
abline(lm(ht ~ pf))
```



(a) Circle a point having a residual that is less than zero.

(b) Draw a box around the point appearing to have the residual that is largest in magnitude.

(c) Do your best to locate the point $(\bar{x}_n, \bar{Y}_n)$ on the plot; draw a dark circle here.

(d) Indicate the point (or points) on the plot with the smallest leverage. (arrows)

(e) Recall that the least-squares criterion is defined as $Q(b_0, b_1) = \sum_{i=1}^{n}(Y_i - (b_0 + b_1 x_i))^2$ for any candidate intercept b_0 and slope b_1. What name do we give the quantity $Q(\bar{Y}_n, 0)$?

Total sum of squares

(f) Letting $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ denote the residuals, what name do we give the quantity $\sum_{i=1}^{n} \hat{\varepsilon}_i^2$?

Error sum of squares

(g) Give an expression in terms of quantities appearing in the output below for the coefficient of determination R^2 (you do not need to evaluate your expression).

$$R^2 = \frac{1.4005}{1.4005 + 2.2556} = \frac{SS_{Reg}}{SS_{Tot}} = \frac{SS_{Reg}}{SS_{Reg} + SS_{Err}}$$

```
anova(lm(ht ~ pf))

Analysis of Variance Table

Response: ht
          Df Sum Sq Mean Sq F value   Pr(>F)
pf         1 1.4005 1.40054   13.66 0.001262 **
Residuals 22 2.2556 0.10253
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reg
Err

(h) What is the sample size n for this data set?

$n = 24$ ∴ $1 + 22 = n - 1$

$23 = n - 1$

(i) Give the value of $\hat{\sigma}^2$.

0.10253

(j) Suppose you find hand prints left by two individuals—one with pinky finger length 65mm and one with pinky finger length 70mm. Of which individual can you more accurately predict the height based on the model fitted to this data set? Carefully explain your answer.

The individual with the 65 mm finger is easier to predict height of due to the way standard deviation is structured in the PI equation:

variance term: $t_{n-2,\alpha/2}\, \sigma \sqrt{1 + \frac{1}{n} + \frac{(x_{new} - \bar{x}_n)^2}{S_{xx}}}$

A larger distance from $\bar{x}_n$ will result in a higher st. dev. and larger prediction interval, so the 70 mm finger individual will have a less accurate prediction since 70 is farther from $\bar{x}_n$

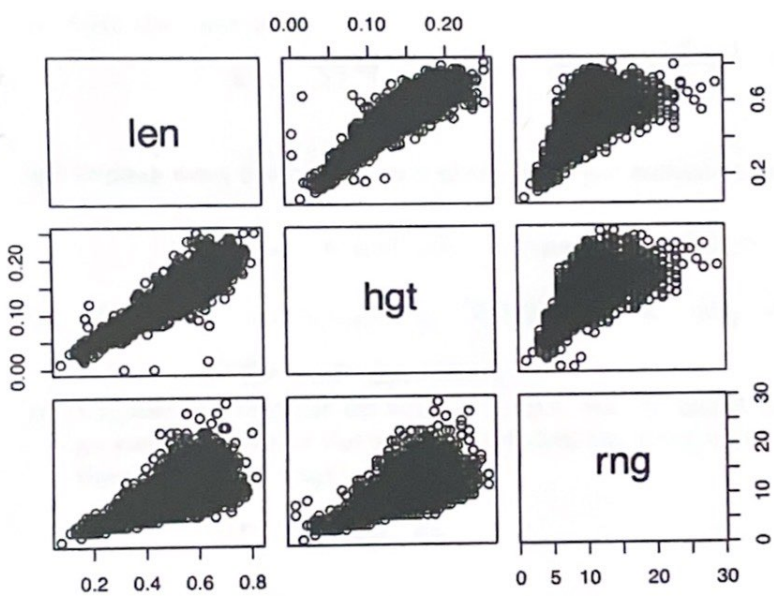
3. Multiple linear regression

A data set of which the first few records are shown below contains length and height measurements as well as the number of rings of a sample of abalones.

```
    len   hgt rng
1 0.350 0.090   7
2 0.530 0.135   9
3 0.440 0.125  10
4 0.330 0.080   7
5 0.425 0.095   8
6 0.530 0.150  20
```

Consider using the multiple linear regression model to predict the number of rings an abalone has (these are like tree rings, and tell the age of the abalone, but counting them requires cutting the abalone open).

```
plot(ab)
```




```
lm1 <- lm(rng ~ len + hgt, data = ab)
summary(lm1)
```

```
Call:
lm(formula = rng ~ len + hgt, data = ab)

Residuals:
    Min      1Q  Median      3Q     Max
-6.0433 -1.6385 -0.5679  0.8442 16.6717

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.6651     0.1792  14.874   <2e-16 ***
len           1.1304     0.7547   1.498    0.134
hgt          47.9564     2.3550  20.364   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.554 on 4171 degrees of freedom
Multiple R-squared:  0.3729,    Adjusted R-squared:  0.3726
F-statistic:  1240 on 2 and 4171 DF,  p-value: < 2.2e-16
```

(a) Give the value of $\hat{\sigma}$.

$\hat{\sigma} = 2.554$

(b) Explain what the parameter σ describes in the multiple linear regression model.

The parameter σ describes the standard deviation of the error term. The error term is represented by the ϵ_i value and is the noise of the ~~predicted values~~ vs. the actual values.

(c) Suppose two abalones are equal in length, but the height measurement of one is 1 unit greater than that of the other. What does the fitted model say about the difference in their numbers of rings?

If the lengths are the same, the height measurement being one unit greater will cause a 47.9564 unit increase in the number of rings

(d) Give the value of the ratio SS_{Error} / SS_{Tot}.

$.3729 = 1 - \frac{SS_{err}}{SS_{tot}}$

$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}$ $R^2 = 0.3729$

$\frac{SS_{err}}{SS_{tot}} = 1 - .3729 = .6271$

(e) How many abalones are in the data set?

$4171 = n - (2+1)$

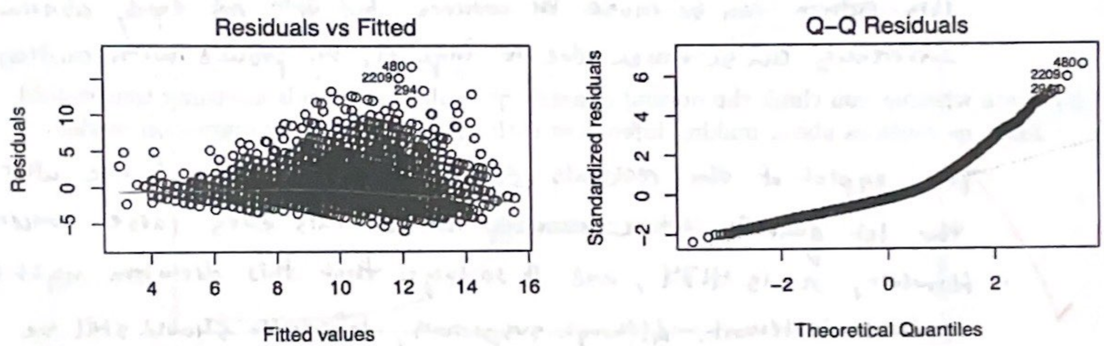
$4171 = n - 3$

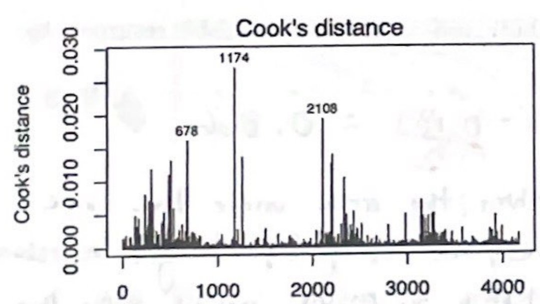
4174 abalones in data set

(f) Give the p value for testing whether the slope coefficient for the length covariate is equal to zero. Interpret this value.

$p = 0.134$ This is greater than significance levels meaning we fail to reject $H_0: \beta_{len} = 0$. Therefore, length has no substantial impact on predicting # of ~~age~~ rings in abalones.

```
par(mfrow=c(2,2))
plot(lm1,which=1)
plot(lm1,which=2)
plot(lm1,which=4)
```





(g) Comment on whether you think there are any outliers in this data set.

3 outliers are reasonably marked in the plots

But maybe only the plots in Cookie's Distance plots marked with 1174 and 2108 could be assumed as obvious outliers.

The data marked with 678 may not be a outlier with strong influence of the MLR.

(h) State whether you think the residuals versus fitted values plot reveals anything that should make us cautious about making inferences with the multiple linear regression model.

I believe. this residuals vs. fitted values plot. slightly show a linear relationship. because the vertical line is flat, not very curved. and the residuals are randomly scattered around 0.

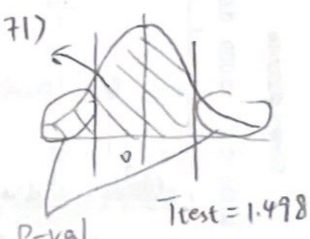
~~But~~ It is not very strong to express linear relationship because it maybe may have a funnel-shaped pattern like ~~[illegible]~~ But overall shows a reasonable linearity.

(i) State whether you think the normal quantile-quantile plot reveals anything that should make us cautious about making inferences with the multiple linear regression model.

Most points fall roughly along the inference line. however, almost all the points on the right tail do not fall on the line.

Residuals are approximately normal distributed but not give a clear. linear regression because of the strong right tail deviations.

(j) Use numbers given in the output to compute the value which will be returned by pt(1.498,4171)



$$pt(1.498, 4171) = 1 - \frac{0.134}{2}$$
$$= 1 - 0.067$$
$$= 0.933$$