Hypothesis Test Conclusion

This is also a fixed answer and you are usually required to state in SIMPLE LANGUAGE.

"Since p-value is small/large, there is ________ evidence against the null hypothesis that true IQ = 115."

To know how much evidence there is against the null hypothesis, use lecture slide 8.45

Evidence	Very Strong	Strong	Moderate	Weak	None
P value	<0.001	0.01	0.05	0.1	0

The key thing to note is if p value is small, there is STRONG evidence against the null hypothesis.

If p value is large, there is NO evidence against the null hypothesis.

Linear Regression Analysis (R Output)

```
> summary(eruption.lm)

Call:
lm(formula = eruptions ~ waiting, data = faithful)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2992 -0.3769  0.0351  0.3491  1.1933

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.87402    0.16014   -11.7   <2e-16 ***
waiting      0.07563    0.00222    34.1   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.497 on 270 degrees of freedom
Multiple R-squared: 0.811,     Adjusted R-squared: 0.811
F-statistic: 1.16e+03 on 1 and 270 DF,  p-value: <2e-16
```

Y-Intercept, b_0

Slope, b_1

Standard Error of Slope, $SE(b_1)$

Test statistic, t:

$$t = \frac{b_1}{\mathrm{SE}(b_1)}$$

$$\frac{0.07563}{0.00222} = 34.1$$

IF ASKED IS THERE A LINEAR RELATIONSHIP?

P value for hypothesis test where Ho: $\beta_1 = 0$ (no association). In this case, p-value is much less than 0.05 so we reject the null hypothesis, i.e. there is a significant relationship

CONFIDENCE INTERVAL:

$$(b_1 - t^* \mathrm{SE}(b_1)\,,\, b_1 + t^* \mathrm{SE}(b_1))$$

t^* is the value from that $t(n-2)$ distribution

HOW STRONG IS THE RELATIONSHIP

$$r^2 = \frac{\text{variance of } \hat{y} \text{ values}}{\text{variance of } y \text{ values}}$$

So r^2 is the % of variation in y that is explained by the linear regression.

- **Understand what is a lurking variable 2016** A lurking variable is one that is not considered but also has an association with a variable being tested. For example, there is an "increase in temperature" would lead to an increase in both "ice-cream sales" & "heat stroke cases". Hence, there would be an association between ice-cream sales and heat stroke cases. However, it is silly to say link ice-cream sales with heat stroke cases. In this case, temperature was the lurking variable. To test the association between ice-cream sales and heat stroke cases, temperature should be kept constant.
- **Know the 4 different sample types.** Simple Random Sample (SRS), Voluntary sample, Stratified Random Sample (population divided into groups of similar individuals first, eg gender or age group) and Multi-stage sample (sample successively smaller groups from the population in stages, eg state, then suburb, then postal)

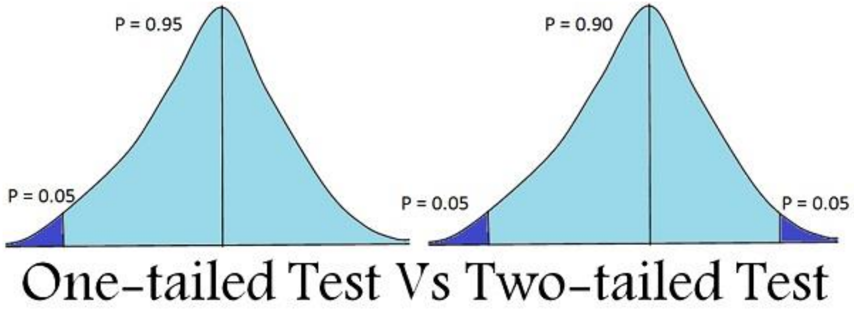
<u>Hypothesis Tests</u>

- **(IMPORTANT) Assumptions & How to check for them (Memorise for all hypothesis tests)**
    - All 1-sample tests
        - Independence – All observations are independent of each other. CHECK: Usually unable to ascertain with given information. Ensured by taking a SRS.
        - Normality – Normal quantile plot (Straight line?), box plot (Symmetrical?) or Large enough sample size (CLT ensures normality)
    - Matched Pairs / Paired T-test (Looks like a 2-sample but more like a 1-sample)
        - Independence – All the PAIRED DIFFERENCES are independent – CHECK – Usually unable to ascertain from given information.
        - Normality – Distribution of PAIRED DIFFERENCES are approximately normal. CHECK: Formula works as long as distribution of x is approximately normal (Distribution of differences need not be approximately normal).
    - 2 Sample t- test **(IINE)**
        - **I**ndependence – 2 samples are independent of each other. CHECK: Should be independent if both variables are unrelated (eg Male/Female) and from different populations. Use given question details
        - **I**ndependence – The measurements within each sample are independent of each other. CHECK: Usually unable to ascertain from given information. Do a SRS.
        - **N**ormality – The distribution of each variation is approximately Normal. CHECK: Check from Normal Quantile Plot if given or Box Plot (Symmetrical shape?). If sample size is big, Central Limit Theorem also ensures Normality.
        - **E**qual standard deviation – Both Samples have the same Standard Deviation. CHECK: It is unlikely that both sample standard deviations are identical. But as long as n1/n2 AND SD1/SD2 are not very different from 1, its fine (from lecture notes)
    - Least Squared Regression **(LINE)**
        1. **L**inear relationship – CHECK: From R output (Use P value for test of no association) and also state R^2 value (Strength of linear relationship – see 2.73)
        2. The observations are **I**ndependent – CHECK: Usually unable to check from given information. If it's an SRS, it'll be independent

Possible Final Exam R Formulas

- Finding p-value of a hypothesis test using given r output
  - To find the P value in one-sided hypothesis test (< or >)
  - P(Z>z) = 1 - P(Z<z) = 1 – pnorm(z)
  - *eg. if test statistic, z = 1.5, and we want to find P(Z > 1.5),*
  - ***P value = 1 – pnorm(1.5)***

  - To find the P value in a two-sided hypothesis test (≠)
  - Simply multiply the one-sided p-value by 2
  - *eg.if test statistic, z = 1.5, and we want o find P(Z ≠* 1.5)
  - ***P value = 2 x (1 – pnorm(1.5))***



the diagram is just meant to show why we multiple p value by 2 for a 2-sided/tailed hypothesis test.

Conditional Probability

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

Recall the whitehouse/roundhouse question (2014 S2 / Qn 2)

- P(eating at whitehouse) = $\frac{1}{8}$
- P(eating at roundhouse) = $\frac{1}{8}$
- P (eating at quad) = $\frac{6}{8}$
- Question: What is the probability of eating at whitehouse if roundhouse is closed
- i.e. Find: P(whitehouse|roundhouse CLOSED)
- *So when roundhouse is closed, what is the reduced probability space? (not what is probability of eating at roundhouse!) therefore* ***P(roundhouse CLOSED)*** *= 1 -* $\frac{1}{8} = \frac{7}{8}$
- *What is P(whitehouse & roundhouse CLOSED)? Remember it is the* *overlapping* *area of P(whitehouse) and P(roundhouse CLOSED), therefore* ***P(whitehouse & roundhouse CLOSED) = P(whitehouse) =*** $\mathbf{\frac{1}{8}}$

$$P(whitehouse|roundhouse\ CLOSED) = \frac{P(Whitehouse\ \&\ Roundhouse\ CLOSED)}{P(roundhouse\ CLOSED)}$$

- *Ans: P(whitehouse|roundhouse CLOSED) =* $\frac{\frac{1}{8}}{\frac{7}{8}} = \frac{1}{7}$