

Stats

Contents

| | |
|-------------------------------------|---|
| Probability Distributions | 2 |
| Binomial | 2 |
| Normal | 2 |
| Hypothesis Testing | 3 |
| Correlation | 3 |
| Hypothesis testing | 3 |

Probability Distributions

Binomial

The binomial distribution is used to model a situation with a fixed number of independent trials each with a constant probability of success.

You can model X as a binomial distribution if:

- There a fixed number of trials, n
- Each trial must succeed or fail
- There is a fixed probability of success, p
- Each trial is independent

If $X \sim B(n, p)$, then

Formula Book

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (0 \leq x \leq n)$$

Normal

The normal distribution $X \sim N(\mu, \sigma^2)$ is symmetrical, meaning the mean and median are equal.

When doing questions that involve the normal distribution, sketching the bell curve on the right is always a good idea.

The standard normal distribution $Z \sim N(0, 1^2)$ is very useful, since it allows you to find values for μ and σ when they're unknown. The normal random vari-

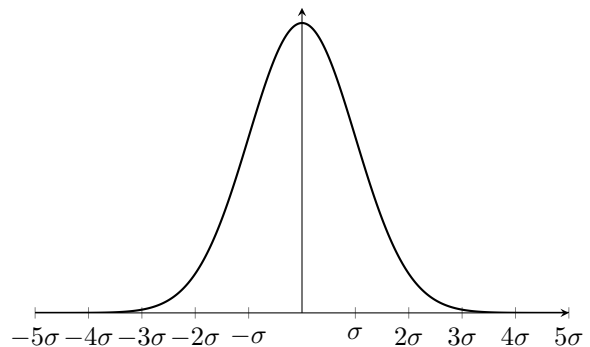
able X can be coded using

Remember

$$Z = \frac{X - \mu}{\sigma}$$

 and you can use this equation to find the unknown parameters of X .

The 'Normal CD' function on your calculator will calculate the area between an upper and lower bound on the bell curve. The 'Inverse Normal' function will find a value for which the area to the left of that value is the area you specify.



Hypothesis Testing

Every hypothesis test has two hypotheses:

H_0 : The null hypothesis - this is what you assume to be true by default

H_1 : The alternative hypothesis

The hypotheses are written in different forms depending on whether the test is one- or two-tailed.

One-tailed:

$$H_0 : p = k$$

$$H_1 : p \leq k$$

Two-tailed:

$$H_0 : p = k$$

$$H_1 : p \neq k$$

If the question says that someone measured and got a value, then you plug that value into a probability calculation with the parameters from the null hypothesis, with the inequality sign in the same direction as the alternative hypothesis. If the probability of the event happening when assuming H_0 is less than the level of significance, then we reject H_0 and accept the alternative hypothesis.

A critical value is the smallest or largest value (depending on the direction of the inequality) obtained by a random variable such that H_0 would be rejected. Finding a critical value is often best done with the tables in the back of the book.

Correlation

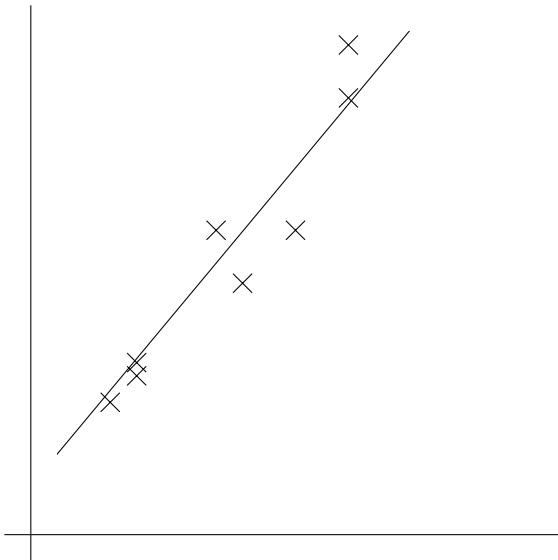


Figure 1: A strong positive correlation

The **product moment correlation coefficient** or “Pearson’s correlation coefficient” is a measure of correlation between two variables. It is often called r and is measured in the range $[-1, 1]$. $r = \pm 1$ means the data perfectly follows a positive or negative correlation respectively.

Hypothesis testing

Use $H_0 : \rho = 0$ and $H_1 : \rho \leq 0$ or $H_1 : \rho \neq 0$. Get r from your calculator (use section 6 and option 4 regression calc) and ρ from the table in the formula book by finding the sample size and significance level. Remember to halve the significance level if the test is two-tailed.