Introduction

I will start by presenting my own proof that I came up with when I decided to attack (prove) this theorem. Next I will present a proof of the Factor Theorem that I yoinked from Wikipedia. This proof uses an idea I'll dub 'problem translation'. Then I will present a second proof from the same article that is much less satisfying. And then I will make some connection(s) and get to invoke the idea of problem translation again.

Bonus: Best fit polynomial Taylor Series Given some function $f(x)$, I want to find find a quadratic approximation, $p(x)$, near the point $x = x_0$. That is, I want to say $f(x) \approx p(x) |\ x \approx x_0$ where $p(x) = ax^2 + bx + c$. So $f(x) \approx ax^2 + bx + c |\ x \approx x_0$ and I want to find values for a, b, c for this to be true. And the quadratic approximation, the best fit parabola at the point $x = x_0$ satisfies the three relations: $f(0) = p(0)$, $f'(x_0) = p'(x_0)$, and $f''(x_0) = p''(x_0)$.

Now, similar the proof for the factor theorem, when $x_0 = 0$, it is easy to come up with the values a, b, and c. $p(0) = c$, $p'(0) = b$, and $p''(0) = 2a$. So $c = p(0)$, $b = p'(0)$, and $a = \frac{p''(0)}{2}$, done. I've found values for the coefficients. Verify these for yourself by serially differentiaing $p(x)$ and note how all terms except for the lowest (in power) zero out due to the x's. So let $g(x) = f(x + x_0)$, basically shifting f by x_0 to the left to set it to the origin. So $g(0) = f(x_0)$ and it is easy to find a best fix quadratic approximation for $g(x) |\ x \approx 0$. Namely $g(x) \approx \frac{g''(0)}{2}x^2 + g'(0)x + g(0) |\ x \approx 0$. And given the translation relation of g and f, $g(x) = f(x + x_0)$, $g'(x) = f'(x + x_0)$, and $g''(x) = f''(x + x_0)$ So $g(x) \approx \frac{f''(x_0)}{2}x^2 + f'(x_0)x + f(x_0) |\ x \approx 0$. Again, $g(k) = f(k + x_0)$ or, swapping perspective from g to f, $f(k) = g(k - x_0)$. So to find $f(k) | k \approx x_0$, this is equivalent to finding $g(k - x_0) | k \approx x_0$. As the input to f approaches x_0, the input the g approaches 0 and g's input is f's input minus x_0. So I can replace occurences of x with $x - x_0$ for the approximation of $g(x) | x \approx 0$. To summarize, $f(k) | k \approx x_0 = g(k - x_0) | k \approx x_0 = \frac{f''(x_0)}{2}(x - x_0)^2 + f'(x_0)(x - x_0) + f(x_0)$. The last equality follows from as input, k, approaches x_0, g's input approaches 0, so I can use the quadratic approximation of g near 0 that I just derived.

Also, same underlying mechanism. I state this underlying machinery explicitly as $f + (\Delta f) \approx f(x_0) + f'(x_0)(\Delta x) + \frac{f''(x_0)}{2}(\Delta x)^2 \ |\ x \approx x_0$. And at $x_0 = 0$, $\Delta x = x - x_0 = x$. And this change in x equaling x itself is a nice property.

Let me briefly explain how I reason about this mechanism, inspired by simple kinematics. I personally find thinking in terms of velocity and acceleration helpful for f' and f'' respectively. $f'(x_0)(\Delta x)$ is the contribution toward Δf solely from $f'(x_0)$ initial velocity, if this velocity is held constant so no acceleration. (This is a linear approximation as velocity is constant, no curvature). and $\frac{f''(x_0)}{2}(\Delta x)^2$ is the contribution toward Δf solely from $f''(x_0)$, the initial acceleration and likewise, no initial velocity. Then the velocity would linearly increase so the average velocity over Δx (think of x as t for time) would be $\frac{f''(x_0)(\Delta x)}{2}$ and thus the resulting Δf would be that average velocity times Δx. Alternatively, think of the velocity time graph being a triangle with base Δx and height $(\Delta x) f''(x_0)$.