Proof of the "trace-formula"
$E(X^T A X) = tr(A\Sigma) + \mu^T A \mu$

Background:

$\underset{p\times 1}{X}$ is a random vector with mean $E(X) = \mu$
and covariance matrix $Cov(X) = \underset{p\times p}{\Sigma}$.

A is a constant matrix.

Result:

$$E(X^T A X) = tr(A\Sigma) + \mu^T A \mu$$

Proof:

i) $E(\underbrace{X^T A X}_{1\times 1}) = tr(E(X^T A X))$

the trace of a 1x1 scalar
is the value itself

$= E(tr(X^T A X))$

since $X^T A X = tr(X^T A X)$ then also
$E(X^T A X) = E(tr(X^T A X))$

ii) $E(tr(\mathbb{X}^T A \mathbb{X})) = E(tr(A \mathbb{X}\mathbb{X}^T))$

$tr(AB) = tr(BA)$

$= tr(E(A\mathbb{X}\mathbb{X}^T))$

the mean of a sum is the sum of the means

"trace

iii) $tr(E(A\mathbb{X}\mathbb{X}^T)) = tr(A\, E(\mathbb{X}\mathbb{X}^T))$

$E(AY) = A\,E(Y)$

iv) We know that $Cov(\mathbb{X}) = E(\mathbb{X}\mathbb{X}^T) - \mu\mu^T$

so $E(\mathbb{X}\mathbb{X}^T) = \underbrace{Cov(\mathbb{X})}_{\Sigma} + \mu\mu^T = \Sigma + \mu\mu^T$

v) $tr(A\, E(\mathbb{X}\mathbb{X}^T)) = tr(A(\Sigma + \mu\mu^T))$

$= tr(A\Sigma + A\mu\mu^T)$

$= tr(A\Sigma) + tr(A\mu\mu^T)$

since $tr(C+D) = tr(C) + tr(D)$

iv) $\mathrm{tr}(A\Sigma) + \mathrm{tr}(A\mu\mu^T)$

$= \mathrm{tr}(A\Sigma) + \mathrm{tr}(\mu^T A \mu)$

$\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$

$$\underline{\underline{= \mathrm{tr}(A\Sigma) + \mu^T A \mu}}$$ q.e.d.

since $\underset{1\times 1}{\mu^T A \mu}$ is a sceler and tr(scaler) = sceler